NBER WORKING PAPER SERIES

WHO MARRIED, (TO) WHOM, AND WHERE? TRENDS IN MARRIAGE IN THE UNITED STATES, 1850-1940

Claudia Olivetti
M. Daniele Paserman
Laura Salisbury
E. Anna Weber

Working Paper 28033
http://www.nber.org/papers/w28033

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2020

We would like to thank Fadi Hassan, Jeanne Lafortune, Bob Pollak, and conference participants at the “Intergenerational Transmission of Economic Status: Exposure, Heritability and Opportunity” at Universidad Carlos III, the Sixth EH-Clio Lab Conference at Pontificia Universidad Catolica de Chile, the ColumbiaNYU Joint Conference on Spatial Economics Through the Lens of Economic History, the NBER Summer Institute, the Society of the Economics of the Household; and seminar participants at the Bank of Italy, Bocconi University, Dartmouth College, INSEAD, Nanyang Technical University, National University of Singapore, University of British Columbia, University of Michigan, University of Tennessee Knoxville, Uppsala University, and Washington University in St. Louis. Juan Conde Carvajal provided excellent research assistance. All errors are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Who Married, (to) Whom, and Where? Trends in Marriage in the United States, 1850-1940
Claudia Olivetti, M. Daniele Paserman, Laura Salisbury, and E. Anna Weber
NBER Working Paper No. 28033
October 2020
JEL No. J12,J62,N31,N32,N91,N92

ABSTRACT

We present new findings about the relationship between marriage and socioeconomic background in the United States in the late 19th and early 20th Centuries. Imputing socioeconomic status of family of origin from first names, we document a socioeconomic gradient for women in the probability of marriage and the socioeconomic status of husbands. This socioeconomic gradient becomes steeper over time. We investigate the degree to which it can be explained by occupational income divergence across geographic regions. Regional divergence explains about one half of the socioeconomic divergence in the probability of marriage, and almost all of the increase in marital sorting. Differences in urbanization rates and the share of foreign-born across states drive most of these differences, while other factors (the scholarization rate, the sex ratio and the share in manufacturing) play a smaller role.

Claudia Olivetti
Department of Economics
Dartmouth College
6106 Rockefeller Hall
Hanover, NH 03755
and NBER
claudia.olivetti@dartmouth.edu

M. Daniele Paserman
Department of Economics
Boston University
270 Bay State Road
Boston, MA 02215
and IZA
and also NBER
paserman@bu.edu

Laura Salisbury
Department of Economics
York University
Vari Hall 1092
4700 Keele Street
Toronto, ON M3J 1P3
CANADA
and NBER
lsalisbu@yorku.ca

E. Anna Weber
Boston University
270 Bay State Road
Boston, MA 02215
eaweber@bu.edu

1 Introduction

Marriage patterns by socioeconomic status are of interest to social scientists for a variety of reasons. Marriage has historically been a mechanism for social mobility, especially for women. Mothers, as well as fathers, transmit traits and preferences to their children and invest in their upbringing. By determining who has children and what resources these children can access, socioeconomic selection into marriage and systematic marital sorting have implications for the intergenerational transmission of inequality and the extent of cross-group interactions within a society (Kalmijn, 1998).

In this paper, we present one of the first accounts of marital patterns by family background in the United States in the late 19$^{\text{th}}$ and early 20$^{\text{th}}$ centuries. Our approach allows us to focus on the role of *family of origin* in determining who marries and to whom. Was marriage an outcome equally prevalent across all classes? If not, did higher social status facilitate marriage for women, or did it enable them to remain unmarried? How much did marriage contribute to social stratification? Did these patterns evolve as the U.S. underwent the transition from a rural and agricultural to an urban and industrial society? How did these patterns vary across geographic regions?

We answer these questions by studying the socioeconomic determinants of marriage and the degree of marital sorting in the United States for cohorts of women born roughly between 1840 and 1910. Using decennial census data, we impute the socioeconomic status (SES) of women's and men's family of origin from first names. This methodology works as long as first names have socioeconomic content (Olivetti and Paserman, 2015). An advantage of this approach is that we can calculate a marker for the SES of the family of origin without the need to create links (exact or probabilistic) across censuses. This is particularly important for women, who cannot be easily linked across censuses, because they changed their surnames upon marriage.

We show that marriage market outcomes became more stratified by SES during our period of study. In the mid-19$^{\text{th}}$ century, marriage was negatively correlated with SES. Women born to families in the bottom quartile of the occupational earnings distribution were more likely to marry than those in the top quartile. This socioeconomic gradient became steeper over time. The marriage gap between women from the bottom and top quartiles of the parental occupational earnings distribution was less than 2 percentage points for those born in the 1840s but grew to more than 5 percentage points for those born in the 1900s. Marital sorting also increased during this period. The difference in the mean parental occupational earnings of the husbands of women from the bottom and top quartiles of the occupational earnings distribution more than tripled, and the correlation in occupational earnings of husbands' and wives' families of origin more than doubled.

In aggregate, marriage outcomes in the U.S. diverged by SES during this period. But marriage markets are local rather than national. As such, our aggregate results may conflate divergence in marriage outcomes by SES *within* and *between* marriage markets.

Divergence *within* marriage markets may occur if, within a given geographic area, high SES women experienced a relative decline in the gains from marriage, and if the gains from positive sorting increased. Divergence *between* marriage markets may occur if average SES diverged across geographic areas. By the early 20th century, the Northeast had completed its transition to an industrial economy, while the South was predominantly agrarian. So, workers in the top quartile of the national occupational income distribution became more concentrated in the Northeast, while those in the bottom quartile become more concentrated in the South. Therefore, men and women from the same quartile became more likely to marry one another, by virtue of being more likely to be born in the same place. Moreover, because the marriage rate was lower in the Northeast than in the South, being born into the bottom occupational income quartile more strongly predicts marrying, as women from this quartile became increasingly concentrated in the South.

To understand the relative importance of these two components, we estimate changes in the SES gradient in marriage outcomes conditional on state of birth, which roughly captures divergence within marriage markets. The difference between the aggregate and the within-state change captures divergence in the SES gradient in marriage outcomes between marriage markets.

We find that state of birth fixed effects explain about half of the change in the relationship between SES and the marriage rate, and almost all of the change in the correlation between husbands' and wives' SES. This seems to be driven by differing rates of urbanization across states. Controlling for states' urban concentration has the same effect on the SES gradient in marriage outcomes as a full set of state fixed effects. Our findings indicate that the increase in marital sorting is driven primarily by increased sorting *between* marriage markets. However, even within marriage markets, low SES women became increasingly likely to marry compared with high SES women. This indicates that the well-documented SES gradient in the probability of marriage for women born in the late 19$^{\text{th}}$ and early 20$^{\text{th}}$ centuries was not a permanent historical feature of the American marriage market, but rather emerged over time.

Literature Review. Scholars have devoted considerable attention to understanding marriage outcomes by SES. This literature focuses on acquired characteristics like education, occupation, and earnings. Highly educated women born in the first half of the 20th century were substantially less likely to marry (Goldin, 2004, 2017; Bailey, Guldi and Hershbein, 2014). This pattern has reversed in recent years: the marital gap now favors the college educated (Goldin, 2017). We complement these studies by focusing on marital outcomes by socioeconomic background of the family of origin, and by documenting the link between marriage and SES for earlier cohorts of women.

There is evidence of increased positive sorting by spouses' education since 1950 (Mare, 1991, 2016; Schwartz and Mare 2005; Greenwood et al., 2014; Eika et al., 2019; Gihleb and Lang, 2020). Information about marital sorting prior to 1940 is limited, but there is evidence that the educational attainment gap between spouses rose during the early 20$^{\text{th}}$ century (Iyigun and Lafortune, 2016). Other studies in sociology and economics have emphasized sorting on other *acquired* traits such as occupation and labor income (Schwartz, 2010).

There is less work investigating the correlation in spousal *ascribed* characteristics, like parental occupation or wealth. Charles et al. (2013) document positive sorting into marriages based on parental wealth for a cohort of US couples in the late 1990s, above and beyond sorting on the basis of own education. Fremeaux (2014) measures marital sorting by expected inherited wealth in France for 1990-2010. He finds that the correlation in spouses' expected inherited wealth is larger than the correlation in current income and declines only slightly after controlling for education. Wagner et al. (2020) study assortative mating on parental wealth using high quality Danish administrative data, and find parental wealth correlations that are lower than those found in the U.S. and France, even though parental wealth homogamy is substantial for those at the very top of the distribution. Our contribution to this literature is to document the extent of marital sorting in a different time period and economic setting. The nature of our data also allows us to analyze long-run trends and investigate how these patterns are affected by geographic heterogeneity in economic development.

Finally, our work is tied to the recent literature emphasizing the importance of place of birth and childhood neighborhood characteristics on economic mobility (Chetty et al., 2014; Chetty and Hendren, 2018a, 2018b; Derenoncourt, 2019). Tan (2019) also documents substantial heterogeneity in economic mobility in the United States in the early 20^{th} century. Our analysis shows that there were important geographic differences in patterns of marriage and marital sorting, which contributed to the overall stratification of U.S. society during this period.

2 Methodology and Data

We seek to estimate women's marriage outcomes by the SES of their *family of origin*. This is, in part, by necessity. In a context in which married women overwhelmingly did not work, and in the absence of information on educational attainment, there are no good measures of acquired socioeconomic traits. However, there are reasons to favor this measure regardless of data constraints. Women had few opportunities to invest in education or marketable skills. As such, inherited class position may have been the salient feature in determining their marriage prospects.

We face the challenge of inferring inherited class position without intergenerationally linked data. Because women changed their surnames upon marriage, it is impossible to generate panels that follow women from childhood into adulthood using data from decennial censuses.

Following the methodology introduced by Olivetti and Paserman (2015, henceforth OP2015), we use an individual's *first name* to impute father's economic status. This methodology, which has been used to study intergenerational mobility for men and women across two and three generations (OP2015; Olivetti, Paserman and Salisbury, 2018), rests on the assumption that names carry information about SES. There is evidence for this in both modern (Bertrand and Mullainathan, 2004; Fryer and Levitt, 2004) and historical data: OP2015 document that in our sample period, 10 to 17 % of total variation in fathers' socioeconomic status can be explained by the variation *between* names given to their children.

A second challenge is the lack of information on earnings prior to 1940 in the decennial censuses. Therefore, we measure the SES of women's family of origin using their fathers' occupational status. Specifically, we use the OCCSCORE variable in IPUMS, which assigns to each occupation the median income among persons employed in that occupation in the 1950 Census.[1] One advantage of OCCSCORE is that it is available throughout the period. More than 200 distinct occupations appear in our data; however, certain occupations appear at a high frequency. Farmers comprise approximately 40% of the workforce in 1850, though this declines to 10% by 1940. We perform a number of sensitivity analyses in which we use the occupational distribution from other years, or alternative measures of occupational standing. In all cases, we pay particular attention to the imputation of farmers' income (see Online Appendix for details).

As in OP2015, we impute fathers' OCCSCORE using an individual's *first name*. We calculate the mean log occupational income of fathers of children ages 0-15 in year t with a given first name, and then assign that value to all adults with that first name in year $t + 30$. In much of our analysis we assign each name an SES quartile based on the rank of that name's mean parental log occupational income.[2]

[1]The OCCSCORE variable has been used in numerous articles published in top journals in economics and other social sciences. See Saavedra and Twinam (2020) for references.

[2]Each name is weighted by its frequency in the overall distribution of children's names in year t. This yields quartiles of unequal size, as the children's name distribution in year t is different from the adult name distribution in year $t + 30$. This is due to immigration, attrition of unique names that are not matched across the child and adult samples, and highly popular names (e.g. John, Mary) that straddle quartile boundaries. Our results are robust to assigning names to SES quartiles using weights that reflect each name's frequency in the distribution of adult names in year $t + 30$, which yields quartiles of equal size (see Online Appendix).

This methodology generates estimates of selection and sorting on fathers' income that are different from what we would obtain with individually linked data. First names are an imperfect proxy for father's SES, so that our estimates will suffer from attenuation bias. Our estimates may also be biased if there is a causal effect of first names on marriage market outcomes. For example, if high-status names are attractive on the marriage market, having such a name may increase the likelihood of marrying a high SES spouse, independent of the effect of being born into a high SES family. This would generate upward bias in our estimates.[3]

OP2015 evaluate the relative magnitude of these biases in pseudo-estimates of father-son intergenerational elasticity. Estimates using the name-based methodology are roughly 30% smaller than those obtained using the linked IPUMS 1850-1880 and 1880-1910 samples. So, the attenuation bias in the name-based methodology seems to dominate. Santavirta and Stuhler (2020) further explore the properties of grouping estimators based on names. They emphasize that the probability limit of the estimator depends crucially on the overlap between the samples from the child and parent generation. In the case of perfect overlap (i.e., the parent and child samples are taken from exactly the same families), a name-based grouping estimator will be upward biased.[4] On the other hand, if samples are non-overlapping, the bias may be positive or negative. Using three different data sets with direct links between parents and children, they find that the name-based estimator in non-overlapping samples is 0 to 50% smaller than the direct estimator.

We stress that, even if this methodology yields estimates that differ from those obtained from true intergenerationally linked samples, it will not present a misleading picture of trends if the bias in the pseudo-estimator is roughly constant over time. A time-varying bias would occur if the information content of names changes over time. If names become more informative

[3]This is an important consideration when deciding how to group members of the child generation in order to impute parental SES. First name is only one permanent individual characteristic that is correlated with parental SES. Race and place of birth are others. However, both of these alternative variables have a strong independent effect on outcomes in the child generation. This can introduce a great deal of upward bias in our estimator. Our conjecture is that first names strike the optimal balance: they are correlated with parental SES, but have a smaller causal effect on individual outcomes than the alternatives.

[4]Feigenbaum (2018) estimates intergenerational elasticities by linking the 1915 Iowa State Census and the 1940 Federal Census. Since the former has information on earnings, occupational scores and names, he can compare the estimates using traditional approaches with the OP2015 name-based method, and finds that the two methods yield remarkably similar results in a fully overlapping sample.

about SES, our estimates will show the marriage outcomes of high and low SES women diverging, even if the underlying process has not changed. OP2015 find only limited evidence of increased informativeness of names, so this is unlikely to have a first-order impact on our findings. Time-varying bias may also occur if names follow a geographic pattern and the process of economic development varies across regions. We address this point explicitly in Section 4.

Data. We use U.S. census microdata from 1850 to 1940, which contain information on first names. For 1850 to 1930, we use 1% IPUMS samples (Ruggles et al., 2010). For 1940, we create a 1% extract of the IPUMS Restricted Complete Count Data (Minnesota Population Center and Ancestry.com, 2013).[5] We restrict the sample to whites to avoid issues associated with the near absence of Blacks in the pre-Civil War censuses. Individual level data are available from IPUMS for every decade from 1850 to 1940, with the exception of 1890. This means that we can calculate marriage rates and assortative mating for five cohorts of men and women observed between age 30 and 45 (in 1880, 1900, 1910, 1930 and 1940).

3 Trends in Marriage Outcomes

In this section, we document trends in marriage market outcomes by SES and region of birth for women ages 30-45. These women are old enough to have reasonably complete marriage histories, but young enough to have lived with their parents in year $t-30$. We focus on the marriage rate and the SES of women's spouses.

3.1 Marriage Outcomes by SES

The top panel of Figure 1 shows the evolution of the fraction of women ages 30-45 ever married by SES. Throughout the period, the marriage rate was quite high, ranging from 85% to 92%. There is a clear socioeconomic gradient in the marriage rate: women born in the bottom quartile of the SES distribution are always more likely to marry than those born in

[5] Using the 1% samples implies that our samples are essentially non-overlapping, so that our estimates will likely understate differences in marriage outcomes by socioeconomic status. We return to this point in Section 3.1.

the top quartile. This socioeconomic gradient grows over time. The difference in marriage rates between the upper and lower quartiles was less than 2 percentage points for women born in the mid-19th century, and more than 5 percentage points for those born in the early 20th century.

The bottom panel of Figure 1 shows the average log parental OCCSCORE of the husbands of women from each parental SES quartile. Women in the highest quartile consistently have husbands with the highest parental SES, and women in the lowest quartile have the lowest-status husbands, indicating positive assortative mating. This gradient increases for later cohorts, suggesting an increase in assortativeness over time. The change in the gradient in spousal SES is small in comparison to the overall increase in occupational status (a result of the transformation of the U.S. economy from an agricultural one to an industrial one), but not negligible: the 2.5 log point increase in the top-versus-bottom quartile gap is about one sixth the size of the overall increase in spousal SES.

Both figures point to a divergence in marriage outcomes between women in the top and bottom quartiles of the SES distribution. Over time, women from the top quartile became less likely to be married, and, conditional on marriage, more likely to marry men from the top of the SES distribution.

3.2 Correlation in Spousal Parental Status

To characterize marital sorting, we also calculate the correlation between husbands' and wives' parental SES, both imputed using our names-based methodology. Our sample consists of couples in which both spouses are 30-45 years old. The thick black line in Figure 2 plots the correlation in the imputed log occupational earnings of husbands' and wives' fathers at the national level. There is clear evidence of an increase in assortativeness throughout the period, except for a small dip for cohorts born between 1860 and 1880.

The correlation coefficient increased from around 0.04 to 0.09. These coefficients are much smaller than recent estimates of the correlation between spouses' educational attainment or parental wealth in the United States, which are closer to 0.6 and 0.4, respectively (Ghileb and

Lang 2016; Charles et al 2016). Our results are also smaller than those obtained by Fremeaux (2014) for France (between 0.12 and 0.25). They are more in line with the estimates obtained for couples formed in Denmark between the late 1980s and early 2010s, which range between 0.1 and 0.15 (Wagner et al., 2020).

There are a number of plausible explanations for why our estimates are lower than others in the literature, especially those based on spouses' own characteristics. First, because we have an imperfect proxy for parental socioeconomic background, our estimates may be attenuated by measurement error. OP2015 and Santavirta and Stuhler (2020) find that intergenerational elasticity estimates based on a name-based methodology can be biased downwards by between 0 and 50%. In our case, the attenuation bias can be more severe because we calculate the correlation between two imputed values. In Online Appendix A we discuss this point in more detail, and, using the IPUMS Linked 1850-1880 and 1880-1910 census samples, we estimate that the true correlation in parents' SES is roughly 3 times larger than the correlation coefficients reported in Figure 2. This implies that the true correlation in spouses' parental log occupational incomes falls roughly between 0.1 and 0.3 during our period of study. These values are quite consistent with the literature.

Second, the correlation in couples' *parental* economic status and *own* economic status need not be the same. The relationship between the two depends on whether couples sort on their own or on their parents' characteristics. For example, assume that couples are perfectly matched on their own characteristics, but parents' characteristics are only imperfectly transmitted to their children. Then, the correlation in own characteristics would be extremely high, but the correlation in parental characteristics could be much lower. The difference between the two will depend on the strength of the transmission between parents and children. Finally, it is also possible that the marriage market in the United States around the turn of the 20^{th} century was less stratified than it is today.

3.3 Marriage Outcomes by Region

The divergence in marriage outcomes highlighted in the previous section may be a mechanical result of differences in marriage patterns across regions and geographic divergence in economic development. To assess this hypothesis, we start by documenting the evolution of marriage outcomes by region.

In the top panel of Figure 3, we plot the marriage rate for women by region of birth. Clearly, there are large differences in marriage rates across regions: marriage is generally most common in the South and least common in the Northeast. Moreover, marriage rates diverged strikingly across regions. The fraction of women ever married in the Northeast and the Midwest fell over the second half of the 19$^{\text{th}}$ century, then stabilized. The marriage rate for Southern women increased throughout the period. As a consequence, the marriage rate gap between the South and the Northeast grew from about 2 percentage points at the beginning of the period to almost 10 percentage points by the end. Even if we discount the first data point (as southern women born between 1835 and 1850 may have been disproportionately affected by the Civil War), there is clear evidence of geographic divergence.

The bottom panel of Figure 3 shows the average log parental OCCSCORE of husbands by the wife's region of birth. Consistent with regional patterns of development over this period, women in the Northeast have the highest-status husbands and women in the South have the lowest, and this gap increases over time. We also calculate the correlation in spouses' parental SES *within* regions (Figure 2, thin colored lines). Sorting declined in the Northeast and increased in the South.

Regional correlations are everywhere smaller than the national correlation coefficient. This reflects the locality of marriage markets and the fact that average occupational incomes differ systematically by region. These findings suggest that part of the overall divergence can be attributed to geographic divergence. We explore this point further in the following section.

4 Interpreting Trends in the Relationship between SES and Marriage

The previous section highlighted the divergence in marriage outcomes by parental socioeconomic status. We now focus on distinguishing between divergence *within* and *between* marriage markets. As we have shown, there are large level differences in key marriage market outcomes across regions. For example, marriage rates are generally higher in the South than in the Northeast. We also know that the Northeast industrialized over the period under investigation, while the South remained predominantly agricultural (Kim and Margo, 2004; Lindert and Williamson, 2016). This regional divergence in occupational structure meant that people from the top of the national occupational income distribution were increasingly concentrated in the Northeast, while those from the bottom were increasingly concentrated in the South. This between-marriage market divergence may have contributed to the increase in assortative mating at the national level, as well as to the increasing socioeconomic gradient in the probability of marriage.

4.1 Empirical Approach

We use a regression model to assess the extent to which the aggregate increase in the SES gradient in marriage outcomes is driven by divergence between marriage markets. We estimate the following three regression equations at the individual level, using a sample of women ages 30-45.

$$Y_{iqst} = \sum_{q=2}^{4} \sum_{t=1880}^{1940} \gamma_{qt} \left(Q_{iq} \times T_{it} \right) + \tau_t + \delta' Z_{iqst} + u_{iqst} \tag{1}$$

$$Y_{iqst} = \sum_{q=2}^{4} \sum_{t=1880}^{1940} \overline{\gamma}_{qt} \left(Q_{iq} \times T_{it} \right) + \tau_t + \zeta_{st} + \delta' Z_{iqst} + u_{iqst} \tag{2}$$

$$\begin{aligned} Y_{iqst} = &\sum_{q=2}^{4} \sum_{t=1880}^{1940} \overline{\overline{\gamma}}_{qt,X} \left(Q_{iq} \times T_{it} \right) + \tau_t \\ &+ \sum_{t=1880}^{1940} \beta'_{st} \left(X_{st} \times T_{it} \right) + \delta' Z_{iqst} + u_{iqst} \end{aligned} \tag{3}$$

Here, Y_{iqst} is a marriage outcome (an indicator for ever being married or the log occupational score of the husband's father) for individual i, of parental income quartile q, born in state s, observed at time t; τ_t is a calendar time fixed effect ; Q_{iq} and T_{it} are parental income quartile and time dummy variables; ζ_{st} is a state of birth-by-year fixed effect; X_{st} is a vector of characteristics that vary at the state-year level; and Z_{iqst} is vector of individual-level controls, including age, urban status, and foreign born status.

Equation (1) estimates aggregate divergence in marriage outcomes by SES. The estimated coefficients γ_{qt} reflect the mean difference in marriage outcome Y between a woman from quartile q of the parental occupational income distribution (assigned by first name) and a woman in the lowest quartile, conditional on Z_{iqst}. For instance, suppose Y is an indicator for ever having married. If $\gamma_{4,1940}$ exceeds $\gamma_{4,1880}$ in magnitude, this indicates that the socioeconomic gradient in the probability of marriage has grown over time. This method essentialy replicates the trends described in Figure 1, and has the advantage of allowing us to test whether the change in the socioeconomic gradient in marriage outcomes over time is statistically significant.

Equation (2) estimates divergence in marriage outcomes by SES within marriage markets. Because we condition on state-year fixed effects, $\overline{\gamma}_{4,1940}$ will only exceed $\overline{\gamma}_{4,1880}$ in magnitude if the socioeconomic gradient in Y grew within states. Define $\Delta\gamma_4 \equiv \gamma_{4,1940} - \gamma_{4,1880}$ (from equation 1) and $\Delta\overline{\gamma}_4 \equiv \overline{\gamma}_{4,1940} - \overline{\gamma}_{4,1880}$ (from equation 2). If $\Delta\overline{\gamma}_4$ captures within-marriage market divergence in marriage outcomes by SES, then $\Delta\gamma_4 - \Delta\overline{\gamma}_4$ captures between-marriage market divergence.

Equation (3) seeks to uncover the drivers of divergence in marriage outcomes by SES between marriage markets. Here, we replace ζ_{st} in turn with one of several state-year characteristics: the urban population share, the manufacturing employment share, the foreign born population share, the male share of the white population ages 15-30, the share of children ages 5-20 attending school, and the share of the state's area located within 15 miles of a railroad. State characteristics are assigned according the woman's state of birth. Characteristics are measured at year $t-10$ to reflect the conditions in the woman's presumed marriage market, with the exception of the scholarization rate, which is measured at $t-20$ to reflect childhood conditions.[6]

We estimate Equation 3 separately for each one of these variables, and define, for each variable X, $\Delta\overline{\overline{\gamma}}_{4,X} \equiv \overline{\overline{\gamma}}_{4,1940,X} - \overline{\overline{\gamma}}_{4,1880,X}$. If $\Delta\overline{\overline{\gamma}}_{4,X} \approx \Delta\overline{\gamma}_4$, this would suggest that characteristic X can explain much of the between-marriage market divergence in Y by SES.

4.2 Results

Results from equations (1) and (2) are shown in Figure 4. We plot estimates of γ_{4t} and $\overline{\gamma}_{4t}$ against t. The dependent variable in Panel A is an indicator equal to 1 if the woman has ever been married, and the dependent variable in Panel B is the woman's husband's log parental occupational income (assigned by the husband's first name). The progression of γ_{4t} over time confirms our previous findings, which did not include individual level controls. There is a negative socioeconomic gradient in the probability of marriage, which grows over time. And, the expected log parental occupational income of husbands of women from higher SES quartiles is greater than that of women from lower SES quartiles. This gap also increases over time.

For both outcomes, $\overline{\gamma}_{4t}$ is everywhere smaller in magnitude than γ_{4t}. This indicates that socioeconomic differences in marriage outcomes are partly driven by the fact that people from different economic strata are concentrated in particular parts of the country, and marriage

[6] All characteristics except railroad access are derived from decennial census data (Ruggles et al. 2010; Haines and ICPSR 2010). Railroad data are from Atack et al. (2010).

market outcomes vary across space. It is also clear that $\Delta\overline{\gamma}_4$ is smaller in magnitude than $\Delta\gamma_4$, indicating that a substantial portion of the aggregate divergence in marriage outcomes by SES occurred between marriage markets. Our results suggest that between-marriage market divergence can explain almost all of the nationwide increase in assortative mating (Panel B); however, it only explains about half of the increase in the negative socioeconomic gradient in marriage for women (Panel A).

Which geographic characteristics caused marriage outcomes to diverge by SES between marriage markets? To shed light on this question, we estimate equation (3), in which we replace birth state-by-year fixed effects with time-varying controls for state characteristics interacted with year fixed effects.[7] In Figure 5, we plot $\Delta\gamma_4$, six estimates of $\Delta\overline{\overline{\gamma}}_{4,X}$ (each controlling for a different state characteristic), and $\Delta\overline{\gamma}_4$ (obtained by including a full set of state-year fixed effects).

The scholarization rate and the sex ratio explain very little of the between-marriage market divergence in marriage outcomes by SES. This is not to say that sex ratios do not affect marriage outcomes: it is well established in the literature that they do (Grossbard-Schechtman, 1984; Angrist, 2002; Abramitzky et al., 2011). Rather, this finding suggests that changes in sex ratios over time were not strongly correlated with changes in SES, so sex ratios cannot explain the changing relationship between SES and marriage.

The largest drivers of between-marriage market divergence are the urban population share and the foreign born population share. Because these two characteristics are highly correlated, it is difficult to tell which has the stronger causal effect. We focus our discussion on the urban population share. Urbanization may have affected the SES gradient in marriage through two channels. One is regional development: the Northeast and Midwest urbanized at a faster rate than the South, and both typically had lower marriage rates for women. So, as lower SES families became more concentrated in the South, the link between marriage and low SES would have strengthened. The other possibility is that urbanization has a causal (negative) effect on marriage rates (perhaps because cities offered more attractive labor market options

[7]Interacting these controls with year fixed effects allows their effect on marriage outcomes to change over time. Our results look very similar if we do not allow state characteristics to influence marriage outcomes differently in different years.

to single women), and higher SES regions urbanized faster than lower SES regions.

4.3 Robustness

We test the sensitivity of our results in several ways. The results are described in detail in Appendix B.

We experiment with alternative measures of occupational status: (1) We use the 1900 occupational wage distribution (Preston and Haines 1991) with a wage for farmers calculated from the 1900 census of agriculture (Abramitzky et al 2012; Olivetti and Paserman 2015); we assign farm income both nationally and at the state level. (2) We assign occupational status using the occupational personal wealth distribution from 1870 (Long and Ferrie 2018; Olivetti et al 2018). (3) We assign occupational status according to the occupational real estate wealth distribution from the full count 1850 decennial census, adjusting the wealth of farmers by subtracting the mean value of a farm in 1850 from the census of agriculture. (4) We use LIDO scores, which assign income by occupation-industry-state-sex-race-age using the 1950 income distribution (Saavedra and Twinam, 2018). Our results are robust to all of these alternative definitions of occupational income.

A potential concern is that our methodology may spuriously inflate the degree of assortativeness during periods of high immigration, if foreign-born individuals have distinctive names and also marry spouses of their own nationality. Our results are robust to restricting the sample to the native born.

In our baseline sorting estimates, we restrict the sample to couples in which the husband and wife are both ages 30-45. Because the average husband is roughly 4.5 years older than his wife, this restriction skews the age distribution relative to the sample used to study selection into marriage. Generally, men aged 30-35 and women aged 40-45 are underrepresented. Our results are robust to several adjustments to the age distribution, such as limiting to a sample of couples with women 30-40 and men 35-45, and to including husbands aged 45-55.

Finally, we experiment with different ways of grouping individuals in order to impute parental SES. One important alternative is to group members of the child generation by name-region of birth, controlling for region of birth fixed effects in the specification, to reduce bias coming from the strong causal effect of region of birth on the outcomes we study. We also use Soundex codes for first names and assign quartiles according to the distribution of names in the adult sample population, rather than the sample of children. Our results are robust to these alternatives.

5 Conclusion

We have documented two key changes in the relationship between SES of the family of origin and marriage for American women born between the early 19^{th} and early 20^{th} centuries. High SES women became increasingly less likely to marry, relative to their low SES counterparts. And, women became increasingly likely to marry men from similar backgrounds to themselves.

The bulk of the paper is devoted to understanding why the relationship between women's SES and marriage changed over time. In particular, whether this change is driven by divergence in marriage outcomes by SES within or between marriage markets. We find that more than half of the change in the SES gradient in the probability of marriage occurred within marriage markets. However, almost all of the increase in marital sorting occurred between marriage markets.

There are three takeaways from this paper. First, the large negative SES gradient in the marriage rate for cohorts born in the late 19th to early 20th centuries emerged over time. To our knowledge, we are the first to document this feature of the historical American marriage market. Second, differences in regional development substantially increased the tendency for high SES men and women to form families together, even though there is no evidence that positive sorting become more desirable within marriage markets. If positive sorting tends to increase the intergenerational persistence of economic status, this may be another channel through which regional inequality persists in the United States. Third, urbanization had a first order effect on the marriage market. Thus, cities not only contribute to innovation and economic growth (Glaeser et al., 1992), but can also have important effects on the social and demographic landscape.

References

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

[2] Abramitzky, Ran; Delavande, Adeline and Vasconcelos, Luis (2011). "Marrying Up: The Role of Sex Ratio in Assortative Matching." *American Economic Journal: Applied Economics*, 3(3): 124-157.

[3] Angrist, Joshua D. (2002) "How Do Sex Ratios Affect Marriage and Labor Markets? Evidence from America's Second Generation." *Quarterly Journal of Economics*, 117(3): 997-1038.

[4] Atack, Jeremy; Bateman, Fred; Haines, Michael and Margo, Robert (2010). "Did Railroad Induce or Follow Economic Growth? Urbanization and Population Growth in the American Midwest, 1850-1860." *Social Science History.* 34(2): 171-197.

[5] Bailey, Martha J.; Guldi, Melanie and Hershbein, Brad J. (2014). "Is There a Case for a 'Second Demographic Transition'? Three Distinctive Features of the Post-1960 U.S. Fertility Decline," in *Human Capital in History: The American Record*, Leah Platt Boustan, Carola Frydman and Robert A. Margo eds., University of Chicago Press, pp. 273-312.

[6] Bertrand, Marianne and Sendil Mullainathan (2004). "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review.* 94(4): 991-1013.

[7] Black, Sandra E. and Paul J. Devereux (2011). "Recent Developments in Intergenerational Mobility," in (O. Ashenfelter and D. Card, eds.), *Handbook of Labor Economics*, Volume 4B: 1487-1541, Amsterdam: North-Holland.

[8] Breen, Richard and Leire Salazar (2011). "Educational Assortative Mating and Earnings Inequality in the United States." *American Journal of Sociology*, 117(3): 808-843 (November).

[9] Charles, Kerwin Kofi; Hurst, Eric and Killewald, Alexandra (2013). "Marital Sorting and Parental Wealth." *Demography*, 50(1): 51-70.

[10] Chetty, Raj and Hendren, Nathaniel (2018a). "The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects", *Quarterly Journal of Economics*, 133(3): 1107-1162.

[11] Chetty, Raj and Hendren, Nathaniel (2018b). "The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates", *Quarterly Journal of Economics*, 133(3): 1163-1228.

[12] Chetty, Raj; Hendren, Nathaniel; Kline, Patrick and Saez, Emmanuel (2014). "Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States", *Quarterly Journal of Economics*, 129(4): 1553-1623.

[13] Clark, Gregory (2013) "What is the True Rate of Social Mobility? Evidence from the Information Content of Surnames," manuscript, UC Davis.

[14] Corak, Miles (2013). "Inequality from Generation to Generation," in (R. Rycroft, ed.) *The Economics of Inequality, Poverty, and Discrimination in the 21st Century*, ABCCLIO.

[15] Cvrcek, Tomas. 2012. "America's Settling Down: How Better Jobs and Falling Immigration Led to a Rise in Marriage, 1880-1930." *Explorations in Economic History* 49(3): 335-51.

[16] Derenoncourt, Ellora (2019). "Can You Move to Opportunity? Evidence from the Great Migration." Mimeo., UC Berkeley.

[17] Eika, Lasse; Mogstad, Magne and Zafar, Basit (2018). "Educational Assortative Mating and Household Income Inequality." Forthcoming *Journal of Political Economy.*

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

[19] Fryer, Roland G. and Steven D. Levitt (2004). "The Causes and Consequences of Distinctively Black Names." *Quarterly Journal of Economics.* 119(3): 767-805.

[20] Gihleb, Rania and Lang, Kevin (2017). "Educational Homogamy and Assortative Mating Have Not Increased." NBER Working Paper 22927.

[21] Glaeser, Edward L.; Kallal, Hedi D.; Scheinkman, José A. and Shleifer, Andrei (1992). "Growth in Cities." *Journal of Political Economy*, 100(6): 1126-1552.

[22] Goldin, Claudia (2004). "The Long Road to the Fast Track: Career and Family." *The ANNALS of the American Academy of Political and Social Science*, 1: 20-35.

[23] Goldin, Claudia (2017). "A Long Road: Women's Quest for Career and Family." Lindahl Lecture, Uppsala University.

[24] Greenwood, Jeremy; Guner, Nezih; Kocharkov, Georgi and Santos, Cezar (2014). "Marry Your Like: Assortative Mating and Income Inequality." *American Economic Review*, 104(5): 348-353.

[25] Grossbard-Schechtman, Amyra (1984). "A Theory of Allocation of Time in Markets for Labour and Marriage." *Economic Journal*, 94(376): 863-882.

[26] Haines, Michael R., and Inter-university Consortium for Political and Social Research (2010). Historical, Demographic, Economic, and Social Data: The United States, 1790-2002. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2010-05-21. https://doi.org/10.3886/ICPSR02896.v3.

[27] Iyigun, Murat and Lafortune, Jeanne (2016). "Why Wait? A Century of Education, Marriage Timing and Gender Roles." IZA Discussion Paper No. 9671.

[28] Jones, Larry E., and Michele Tertilt (2008). "An Economic History of Fertility in the United States: 1826-1960." In *Frontiers of Family Economics*, Vol. 1, edited by Peter Rupert, 165-230. Bingley, UK: Emerald Group Publishing.

[29] Katz, Lawrence F., and Robert A. Margo (2014). "Technical Change and the Relative Demand for Skilled Labor: The United States in Historical Perspective." In *Human Capital in History: The American Record*, edited by Leah Platt Boustan, Carola Frydman, and Robert A. Margo, 15-57. Chicago: University of Chicago Press.

[30] Kim, Sukkoo and Robert A. Margo (2004). "Historical Perspectives on U.S. Economic Geography." In *Handbook of Regional and Urban Economics*, volume 4, chapter 66, pages 2981-3019, edited by J. V. Henderson and J. F. Thisse. Amsterdam: North Holland.

[31] Lindert, Peter H. and Jeffrey G. Williamson (2016). *Unequal Gains: American Growth and Inequality since 1700.* Princeton: Princeton University Press.

[32] Long, Jason and Ferrie, Joseph (2018). "Grandfathers Matter(ed): Occupational Mobility across Three Generations in the US and Britain, 1850-1911." *The Economic Journal*, 128(612): F422-F445.

[33] Mare, Robert D. (2016). "Educational Homogamy in Two Gilded Ages: Evidence from Inter-generational Social Mobility Data." *Annals of the American Academy of Political and Social Science.*

[34] Minnesota Population Center and Ancestry.com (2013). IPUMS Restricted Complete Count Data: Version 1.0 [Machine-readable database]. Minneapolis: University of Minnesota.

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

[36] Olivetti, Claudia; M. Daniele Paserman and Laura Salisbury (2018). "Three-generation Mobility in the United States, 1850-1940: The Role of Maternal and Paternal Grandparents." *Explorations in Economic History.*

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

[38] Ruggles, Steven J.; Trent Alexander; Katie Genadek; Ronald Goeken; Matthew B. Schroeder, and Matthew Sobek (2010). Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]. Minneapolis: University of Minnesota.

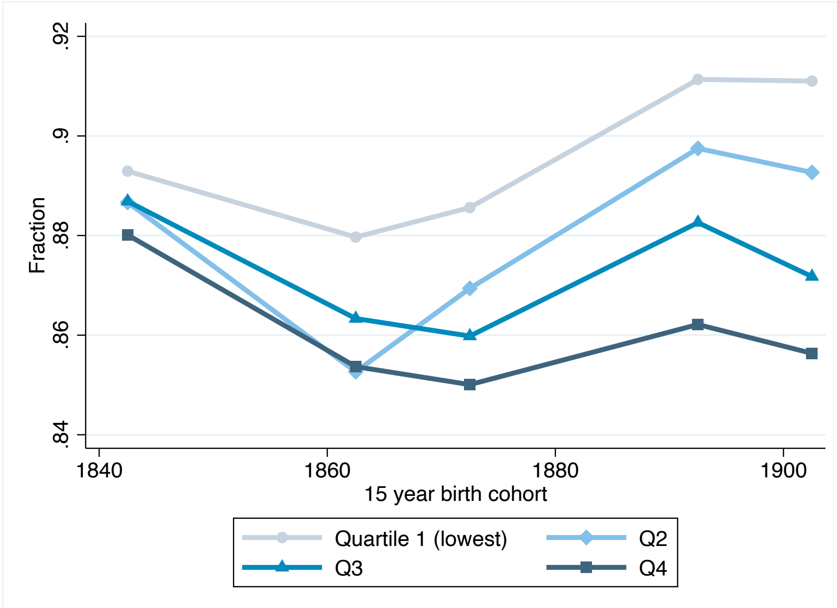
[39] Saavedra, Martin Hugo and Twinam, Tate (2020). "A Machine Learning Approach to Improving Occupational Income Scores." *Explorations in Economic History*, 75, 101304.

[40] Santavirta, Torsten and Stuhler, Jan (2020). "Name-Based Estimators of Intergenerational Mobility." Mimeo., Stockholm University.

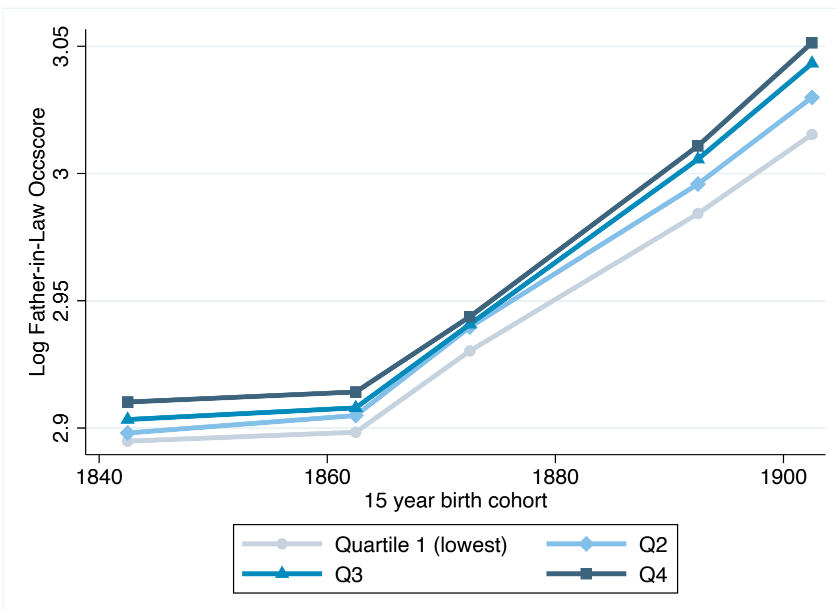
[41] Solon, Gary R. (1999). "Intergenerational Mobility in the Labour Market," in (O. Ashenfelter and D. Card, eds.) *Handbook of Labor Economics*, Volume 3A: 1761-800, Amsterdam: North-Holland.

[42] Tan, Hui Ren (2019). "A Different Land of Opportunity: the Geography of Intergenerational Mobility in the Early 20^{th} US."' Mimeo., National University of Singapore.

Figure 1: Trends by Woman's Parental SES



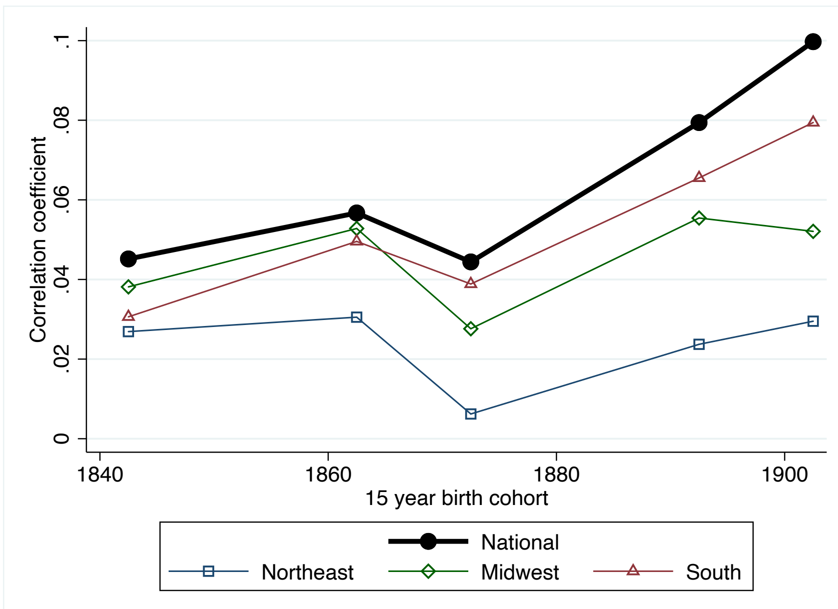
(a) Fraction of Women 30-45 Ever Married



(b) Father-in-Law's SES

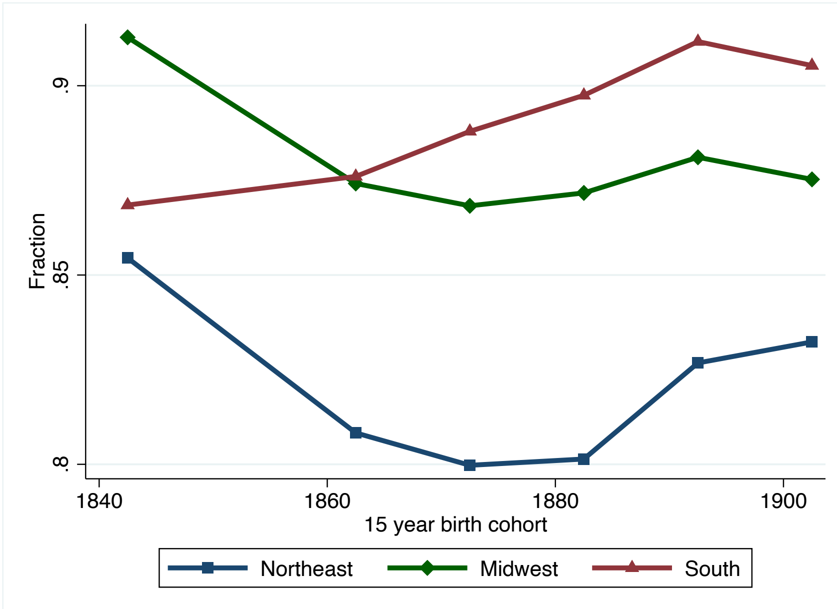
Note: The top panel shows the fraction ever married by birth cohort and parents' quartile of occupational income score. The bottom panel shows the average log occupational score of fathers of husbands aged 30-45 married to wives 30-45 by the woman's parental quartile of occupational income score. The occupational income score is imputed based on first names. Source: Authors' calculations based on the 1% IPUMS samples (1850-1930) and a 1% extract of the 1940 Restricted Complete Count Census Data.

Figure 2: Correlation in Parental SES

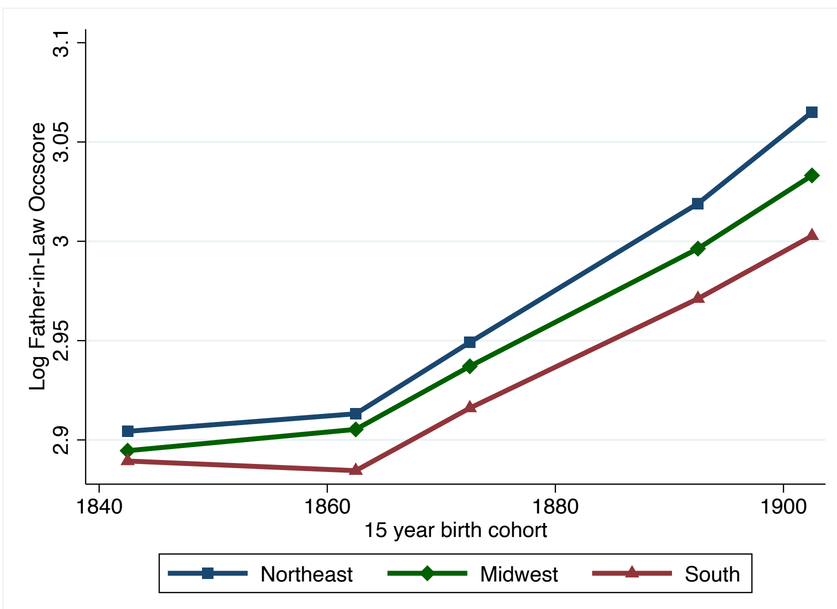


Note: This figure shows the correlation in parental occupational income score between husband and wife. Region assignment is by the wife's region of birth. The occupational income score is imputed based on first names. Source: Authors' calculations based on the 1% IPUMS samples (1850-1930) and a 1% extract of the 1940 Restricted Complete Count Census Data.

Figure 3: Trends by Region of Woman's Birth



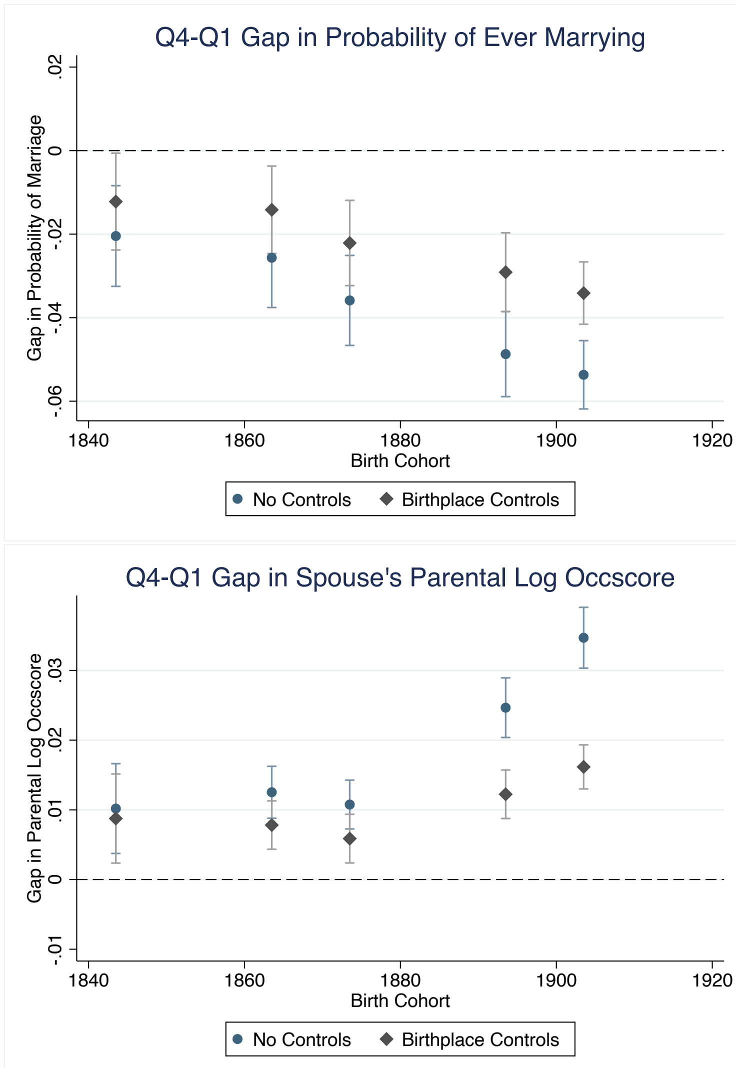
(a) Fraction of Women 30-45 Ever Married



(b) Father-in-Law's SES

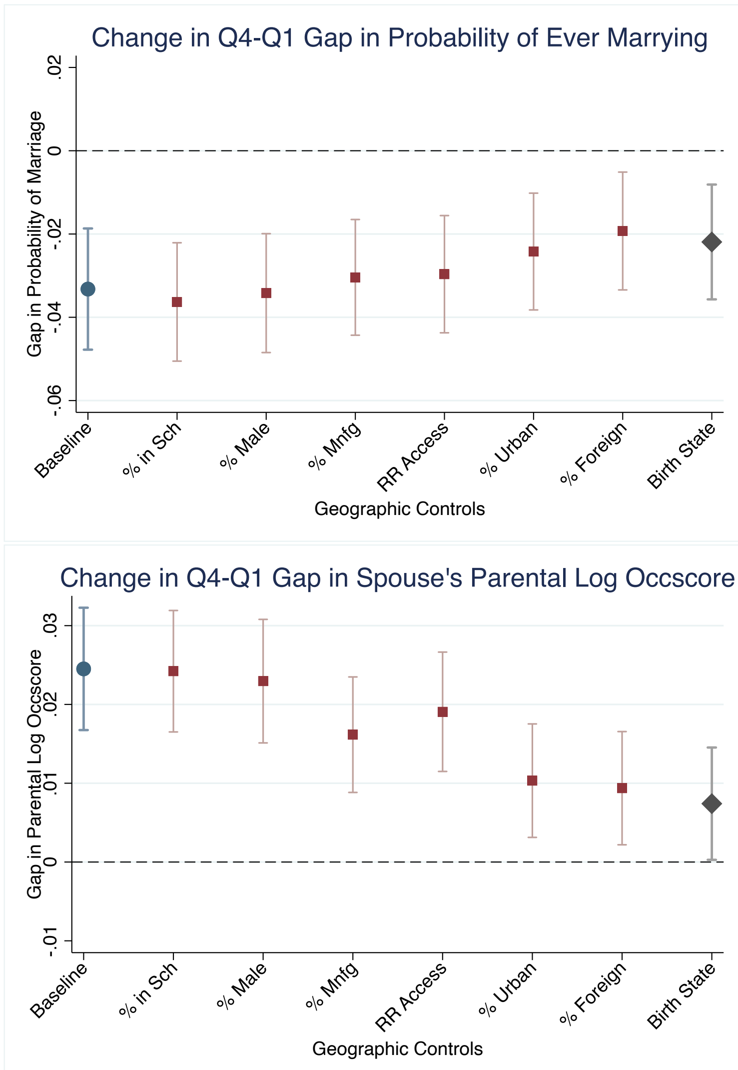
Note: The top panel shows the fraction ever married by birth cohort and woman's region of birth. The bottom panel shows the average log occupational score of fathers of husbands aged 30-45 married to wives 30-45 by woman's region of birth. The occupational income score is imputed based on first names. Source: Authors' calculations based on the 1% IPUMS samples (1850-1930) and a 1% extract of the 1940 Restricted Complete Count Census Data.

Figure 4: Effect of Parental SES and Birthplace on Probability of Ever Marrying and Spouse's Parental Log Occupational Score, Women 30-45



Note: 95% confidence intervals are shown. Plotted coefficients represent the coefficient on the indicator for the top SES quartile, as compared to the bottom SES quartile (the omitted SES category). The Birthplace Controls specification includes a set of dummies for the woman's state of birth interacted with cohort dummies. Source: Authors' calculations based on the 1% IPUMS samples (1850-1930) and a 1% extract of the 1940 Restricted Complete Count Census Data.

Figure 5: Which social and economic characteristics explain the gap between baseline divergence and divergence conditional on state of birth fixed effects?



Note: Figure plots the change in the Q1-Q4 parental SES gradient in the probability of marriage (top panel) or spousal's parental SES (bottom panel) between the cohorts observed as adults in 1880 and 1940 estimated using different controls. All control variables are interacted with a full set of cohort dummies. The Birthplace Controls specification includes a set of dummies for the woman's state of birth interacted with cohort dummies. Source: Authors' calculations based on the 1% IPUMS samples (1850-1930) and a 1% extract of the 1940 Restricted Complete Count Census Data., 1850-1940; IPUMS 100% population databases from the Decennial Censuses, 1850, 1880, and 1900-1940; and ICPSR summary census data, 1850-1940.

Appendix

A Magnitude of Correlation Coefficients

Let y_{ij} denote the parental log occupational income of husband i with first name j, and let $\overline{y}_j$ denote the mean parental log occupational income of men with first name j. Similarly, let y_{kl} denote the parental log occupational income of wife k with name l, and $\overline{y}_l$ denote the mean parental log occupational income of women with first name l. We estimate correlation coefficients $\rho(\overline{y}_j, \overline{y}_l)$ that range from 0.04 to 0.10. What do these estimates imply about the magnitude of the true object of interest, $\rho(y_{ij}, y_{kl})$?

We present here a simple back-of-the-envelope calculation. This should be considered a ballpark estimate of $\rho(y_{ij}, y_{kl})$, given our estimates of $\rho(\overline{y}_j, \overline{y}_l)$.

Following OP2015, suppose y_{ij} and y_{kl} are determined in the following way:

$$y_{ij} = \mu_j + z_{ij} \tag{A1}$$

$$y_{kl} = \mu_l + z_{kl} \tag{A2}$$

where μ_j and μ_l are first name fixed effects and z_{ij} and z_{kl} are random shocks. Then,

$$\overline{y}_j = \mu_j + \frac{1}{N_j} \sum z'_{ij} \tag{A3}$$

where N_j is the number of boys with first name j, and the z'_{ij}s are drawn from a different sample than z_{ij}. Similarly,

$$\overline{y}_l = \mu_l + \frac{1}{N_l} \sum z'_{lk} \tag{A4}$$

where N_l is the number of girls with first name l, and the z'_{kl}s are drawn from a different sample than z_{kl}.

We first compare the correlation coefficients $\rho(y_{ij}, y_{kl})$ and $\rho(y_{ij}, \overline{y}_l)$. If couples sort on both the systematic and random component of parental income (which we expect to be the case), then $\rho(y_{ij}, y_{kl})$ embeds a positive correlation between *both* μ_j and z_{ij} and μ_l and z_{kl}.

Because $\overline{y}_l$ is computed from a sample that doesn't contain person k, $\rho(y_{ij}, y_{kl})$ only embeds a positive correlation between μ_l and both μ_j and z_{ij}. As such, we expect $\rho(y_{ij}, \overline{y}_l)$ to be attenuated relative to $\rho(y_{ij}, y_{kl})$. For simplicity, we write that

$$\rho(y_{ij}, \overline{y}_l) = \lambda \rho(y_{ij}, y_{kl}) \tag{A5}$$

where λ is an attenuation factor.

Now, we compare the correlation coefficients $\rho(y_{ij}, \overline{y}_l)$ and $\rho(\overline{y}_j, \overline{y}_l)$. Again, $\rho(\overline{y}_j, \overline{y}_l)$ embeds a positive correlation between μ_j and μ_l only, while $\rho(y_{ij}, \overline{y}_l)$ embeds a positive correlation between μ_l and *both* μ_j and z_{ij}. As such, we expect $\rho(\overline{y}_j, \overline{y}_l)$ to be attenuated relative to $\rho(y_{ij}, \overline{y}_l)$, As above, we write that

$$\rho(\overline{y}_j, \overline{y}_l) = \chi \rho(y_{ij}, \overline{y}_l) \tag{A6}$$

where χ is another attenuation factor. Combining Equations A5 and A6,

$$\rho(y_{ij}, y_{kl}) = \frac{1}{\lambda \chi} \rho(\overline{y}_j, \overline{y}_l) \tag{A7}$$

We can estimate χ using intergenerationally linked data. Re-arranging Equation A6,

$$\chi = \frac{\rho(\overline{y}_j, \overline{y}_l)}{\rho(y_{ij}, \overline{y}_l)} \tag{A8}$$

Everything on the right hand side of this equation can be estimated using linked data in which married couples *and* husbands' fathers are observed. Linked data in which wives' fathers are observed are never available.

To obtain a ballpark estimate of $\rho(y_{ij}, y_{kl})$, we need an estimate of the attenuation factor λ. We believe that it is reasonable to assume that λ is approximately on the same order of magnitude as χ. Therefore, we calculate the implied correlation assuming that $\lambda = \delta\chi$, for values of $\delta \in \{0.5, 0.75, 1, 1.25, 1.75\}$. We can then estimate $\rho(y_{ij}, y_{kl})$ by

$$\rho(y_{ij}, y_{kl}) = \frac{1}{\delta\chi^2}\rho(\overline{y}_j, \overline{y}_l) \tag{A9}$$

We have samples of men linked between the 1850-1880 censuses and the 1880-1910 censuses. We present the adjustment from Equation A9 for men ages 30-45 in 1880 and 1910 below.

	(1)	(2)	(3)	(4)	(5)	(6)
Years	$\rho(\overline{y}_j, \overline{y}_l)$	χ	δ	λ	$\rho(y_{ij}, y_{kl})$	Ratio
	Full sample	Linked sample		(2) × (3)	(1)/[(2) × (4)]	(5)/(1)
1850-1880	0.045	0.536	0.5	0.268	0.315	6.96
1850-1880	0.045	0.536	0.75	0.402	0.210	4.64
1850-1880	0.045	0.536	1	0.536	0.157	3.48
1850-1880	0.045	0.536	1.25	0.670	0.126	2.79
1850-1880	0.045	0.536	1.5	0.804	0.105	2.32
1880-1910	0.044	0. 586	0.5	0.293	0.260	5.84
1880-1910	0.044	0. 586	0.75	0.439	0.173	3.89
1880-1910	0.044	0. 586	1	0.586	0.130	2.92
1880-1910	0.044	0. 586	1.25	0.732	0.104	2.33
1880-1910	0.044	0. 586	1.5	0.878	0.086	1.95

We estimate that the true correlation in parental log occupational income is anywhere from 2 to 7 times larger than the correlation coefficients we present in the paper. If we take $\delta = 1$ as our preferred estimate, this would mean that the true correlation in parental log occupational income is roughly 3 times larger than the correlations presented in the paper. Extrapolating forward, this would suggest that the correlation in parental log occupational income among couples born in the early 20th century was roughly 0.30.

B Robustness

Measurement of Occupational Status

Our main measure of socioeconomic status is the logarithm of occupational income (the OCC-SCORE variable in IPUMS). We also experiment with alternative measures of occupational status: (1) We use the 1900 occupational wage distribution (Preston and Haines 1991) with a wage for farmers calculated from the 1900 census of agriculture (Abramitzky et al, 2012; Olivetti and Paserman, 2015); we assign farm income both nationally and at the state level. (2) We assign occupational status using the occupational personal wealth distribution from 1870 (Long and Ferrie 2018; Olivetti et al 2018). (3) We assign occupational status according to the occupational real estate wealth distribution from the full count 1850 decennial census, adjusting the wealth of farmers by subtracting the mean value of a farm in 1850 from the census of agriculture. (4) We use LIDO scores, which assign income by occupation-industry-state-sex-race-age using the 1950 income distribution (Saavedra and Twinam, 2018).

The analysis, summarized in Table B-1, shows that, with the exception of personal wealth, these alternative definitions of occupational status yield results that are not qualitatively different from our baseline. When we use personal wealth our results show smaller SES differences in marriage outcomes in the cross section and a smaller change in this gradient over time. However, we note that this is unsurprising since this measure of occupational status is the least precise. Personal wealth is measured with noise and it is only available in the 1% sample of the 1850 decennial census. Therefore the attenuation problem is likely to be most severe.

Foreign Born Status

Our results are essentially unchanged when restricting the sample to the native-born or the children of native-born parents. These results are shown in Table B-2.

Alternative Assignment of Names to Quartiles

We consider three changes to our methodology of assigning parental SES quartiles to names. The results of these alternative assignments are shown in Table B-3.

To consider robustness to the grouping of names, we group names by Soundex instead of exact spelling. This check produces qualitatively similar results to our main specification, with the exception of a particularly low relationship between spousal SES for the 1900-1930 cohort.

We also impute parental SES by name-region of birth instead of by first name alone. In general, while the trend over time in this model is similar to the baseline model, the magnitude of the coefficients are noticeably larger. This level change arises because imputation by name-region mechanically includes more of the regional divergence in the estimated effect of SES than imputation by name alone. Not shown, when state of birth fixed effects are included in the regressions, the levels of the coefficients on Q4 are similar to the baseline model.

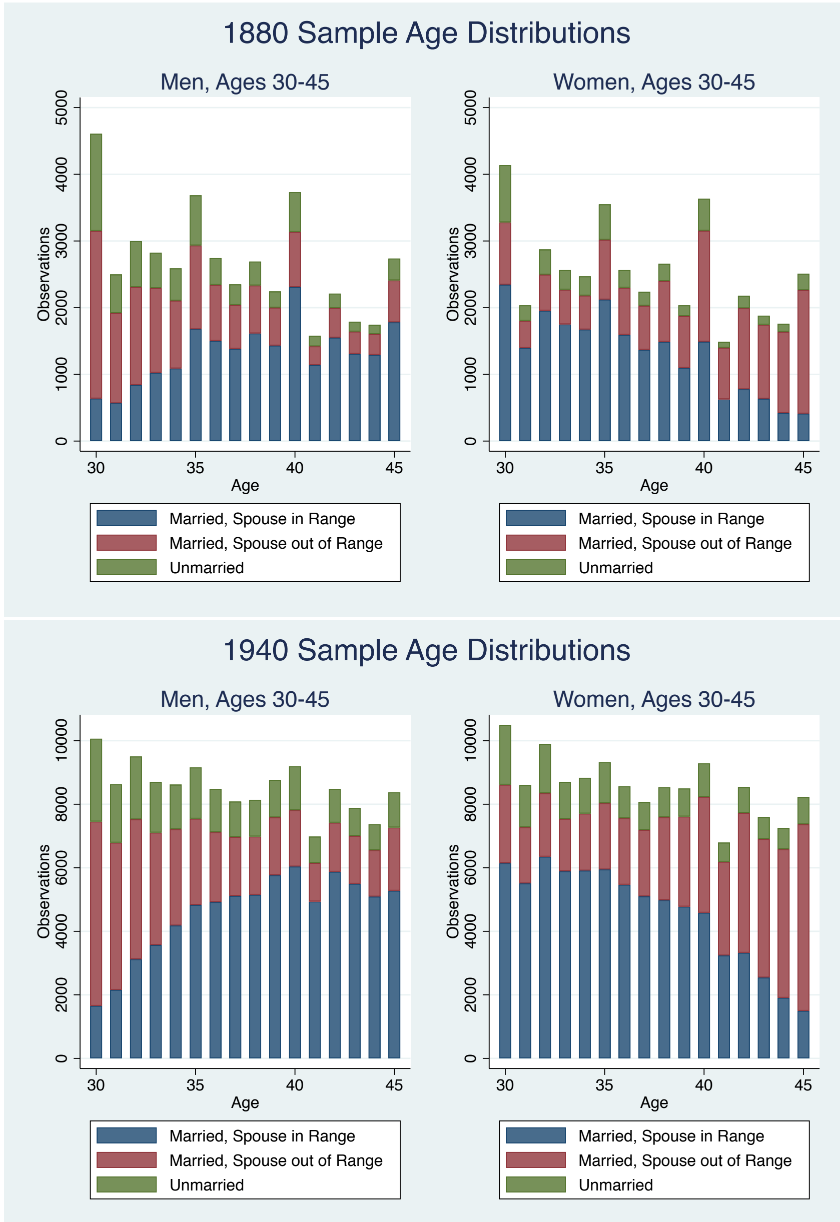
In our main specification, we assign each name an SES quartile according to the rank of that name's mean parental occupational income among the full set of names, weighted by the name distribution when the cohort is observed as children. This results in quartiles of uneven size, as the distribution of the cohort's names as children names differs from the cohort's name distribution as adults due to immigration between the sample years, the attrition of unique names which are not matched across the two samples, and the presence of highly popular names (e.g. John, Mary) which often straddle quartile boundaries. To ensure quartiles of equal size, we can instead assign SES quartiles among the population of adults for whom we can infer the father's income. Our results are qualitatively similar if we use this specification.

Age Range

For our measure of assortative mating by SES, we need to impute parental income for both the husband and a wife. Therefore, our sample in these analyses consists of couples where both spouses are ages 30-45. Because husbands are an average of approximately 4.5 years older than their wives, this restriction results in a skewed age distribution relative to the prior analyses. In Figure B-1 we show how the age distribution of the remaining sample is affected. Generally, men aged 30-35 and women aged 40-45 are underrepresented. This sample restriction could affect our estimates of the level of assortative mating if couples with a smaller age gap (who are more likely to be included in the tails of the age distribution) are differentially assortatively matched by SES, and could affect our interpretation of the trends if the extent of this differential sorting changes over time.

We therefore confirm that our results are robust to several adjustments to the age distribution. First, we limit to a sample of couples with women 30-40 and men 35-45. Next, we include women ages 30-45 with husbands ages 30-55, imputing the parental income of all husbands using the income of fathers of sons aged 0-15 30 years prior. Finally, for the cohorts for which there is available data, we use the same sample of couples but impute the income of husbands using both 30- and 40-year lags of fathers' income. Husbands aged 30-40 are imputed only based on the 30-year lag, husbands aged 45-55 are based only on the 40-year lag, and husbands aged 40-45 are assigned the average value across the two lagged samples. The results of all three of these specifications are shown in Table B-4.

Figure B-1: Age and Marital Status Distribution in Census Samples



Source: Authors' calculations based on the 1% IPUMS samples (1850-1930) and a 1% extract of the 1940 Restricted Complete Count Census Data.

Table B-1: Robustness of Ever-Married Regressions to Alternative Occupational Rankings

	1850-1880	1870-1900	1880-1910	1900-1930	1910-1940
Dependent Variable: Ever Married					
Occscore	-0.02036	-0.02576	-0.03665	-0.04868	-0.05366
	(0.00603)	(0.00604)	(0.00545)	(0.00520)	(0.00417)
	[34990; 1207]	[60007; 1832]	[76284; 2305]	[114391; 3412]	[125014; 3409]
1900 Occupational Income	-0.01608	-0.02469	-0.03011	-0.04531	-0.05344
	(0.00618)	(0.00575)	(0.00533)	(0.00546)	(0.00401)
	[34988; 1206]	[60005; 1830]	[76283; 2304]	[114383; 3406]	[125026; 3414]
LIDO Score	-0.01201	-0.01581	-0.01265	-0.04371	-0.05344
	(0.00621)	(0.00813)	(0.01144)	(0.00532)	(0.00400)
	[34987; 1204]	[59987; 1818]	[76264; 2291]	[114116; 3292]	[124934; 3372]
Occupational Personal Wealth	-0.00173	0.01767	0.02538	0.00555	-0.00030
	(0.00623)	(0.00521)	(0.00747)	(0.00643)	(0.00510)
	[34989; 1206]	[60007; 1832]	[76284; 2305]	[114360; 3405]	[124961; 3379]
Occupational Real Estate Wealth	-0.01935	-0.02190	-0.03560	-0.04616	-0.05295
	(0.00612)	(0.00601)	(0.00500)	(0.00538)	(0.00417)
	[34989; 1206]	[60007; 1832]	[76284; 2305]	[114363; 3405]	[124979; 3391]
Dependent Variable: Husband's Parental Log Occupational Status					
Occscore	0.00968	0.01296	0.01094	0.02465	0.03469
	(0.00323)	(0.00191)	(0.00179)	(0.00218)	(0.00223)
	[16659; 906]	[29063; 1401]	[37642; 1735]	[60763; 2720]	[61978; 2542]
1900 Occupational Income	0.00900	0.01386	0.01062	0.02633	0.03405
	(0.00327)	(0.00205)	(0.00186)	(0.00185)	(0.00228)
	[16658; 905]	[29062; 1400]	[37641; 1734]	[60758; 2716]	[61984; 2546]
LIDO Score	0.01188	0.01377	0.01143	0.02456	0.03472
	(0.00269)	(0.00229)	(0.00209)	(0.00222)	(0.00213)
	[16657; 904]	[29051; 1391]	[37634; 1729]	[60637; 2643]	[61942; 2516]
Occupational Personal Wealth	0.00216	0.00067	-0.00206	-0.00526	-0.00686
	(0.00214)	(0.00215)	(0.00195)	(0.00281)	(0.00310)
	[16659; 906]	[29063; 1401]	[37642; 1735]	[60749; 2714]	[61958; 2527]
Occupational Real Estate Wealth	0.00959	0.01165	0.01022	0.02445	0.03424
	(0.00326)	(0.00200)	(0.00238)	(0.00174)	(0.00209)
	[16659; 906]	[29063; 1401]	[37642; 1735]	[60750; 2713]	[61967; 2535]

Notes: Entries in this table represent OLS coefficients from the estimation of equation (3). Coefficients reflect the difference in the outcome variable between the highest- and lowest-status quartiles of names for the cohort specified in the column. Standard errors in parentheses, clustered by the woman's first name. The number of observations and the number of women's names used to impute father's income for each decade for each regression are specified under each estimate. Different rows use different measures of occupational status 1900 Occupational Income is calculated from the 1900 occupational wage distribution (Preston and Haines 1991) with a wage for farmers calculated from the 1900 census of agriculture (Abramitzky et al 2012; Olivetti and Paserman 2015). LIDO Score assigns income by occupation-industry-state-sex-race-age using the 1950 income distribution (Saavedra and Twinam, 2018). Occupational Personal Wealth using the occupational personal wealth distribution from 1870 (Long and Ferrie 2018; Olivetti et al 2018). Occupational Real Estate Wealth uses the occupational real estate wealth distribution from the full count 1850 decennial census, adjusting the wealth of farmers by subtracting the wealth of farmers by subtracting the mean value of a farm in 1850 from the census of agriculture.
Source: 1% IPUMS samples (1850-1930), 1% extract of the 1940 Restricted Complete Count Census Data.

Table B-2: Robustness of Ever-Married Regressions to Restricting to Native-Born Sample

	1850-1880	1870-1900	1880-1910	1900-1930	1910-1940
Dependent Variable: Ever Married					
All Women	-0.02036	-0.02576	-0.03665	-0.04868	-0.05366
	(0.00603)	(0.00604)	(0.00545)	(0.00520)	(0.00417)
	[34990; 1207]	[60007; 1832]	[76284; 2305]	[114391; 3412]	[125014; 3409]
Native-Born Women	-0.02108	-0.02937	-0.04275	-0.05771	-0.05836
	(0.00586)	(0.00689)	(0.00594)	(0.00493)	(0.00431)
	[25376; 1105]	[46616; 1662]	[59516; 2077]	[93525; 3144]	[110287; 3221]
Daughters of Native-Born Parents	-0.01952	-0.02117	-0.03433	-0.05350	-0.04615
	(0.00572)	(0.00703)	(0.00616)	(0.00476)	(0.00405)
	[22259; 1052]	[32474; 1487]	[40794; 1839]	[65502; 2843]	[104505; 3172]
Dependent Variable: Husband's Parental Log Occupational Status					
All Couples	0.00968	0.01296	0.01094	0.02465	0.03469
	(0.00323)	(0.00191)	(0.00179)	(0.00218)	(0.00223)
	[16659; 906]	[29063; 1401]	[37642; 1735]	[60763; 2720]	[61978; 2542]
Native-Born Wives	0.01223	0.01621	0.01310	0.02746	0.03584
	(0.00373)	(0.00214)	(0.00202)	(0.00196)	(0.00226)
	[11876; 815]	[22090; 1243]	[28839; 1532]	[49430; 2489]	[55409; 2400]
Native-Born Husbands	0.01066	0.01385	0.01257	0.02599	0.03458
	(0.00361)	(0.00218)	(0.00209)	(0.00192)	(0.00220)
	[11397; 798]	[21473; 1231]	[28239; 1526]	[48086; 2473]	[54039; 2375]
At Least One Spouse Native Born	0.01109	0.01511	0.01219	0.02674	0.03521
	(0.00355)	(0.00208)	(0.00200)	(0.00193)	(0.00225)
	[12491; 824]	[23523; 1274]	[30784; 1572]	[51502; 2529]	[57561; 2437]
Both Spouses Native Born	0.01193	0.01498	0.01362	0.02675	0.03521
	(0.00379)	(0.00225)	(0.00211)	(0.00196)	(0.00222)
	[10782; 789]	[20040; 1199]	[26294; 1483]	[46014; 2427]	[51887; 2335]

Notes: Entries in this table represent OLS coefficients from the estimation of equation (3). Coefficients reflect the difference in the outcome variable between the highest- and lowest-status quartiles of names for the cohort specified in the column. Standard errors in parentheses, clustered by the woman's first name. The number of observations and the number of women's names used to impute father's income for each decade for each regresson are specified under each estimate. Different rows represent restrictions the estimation sample related to national origin.
Source: 1% IPUMS samples (1850-1930), 1% extract of the 1940 Restricted Complete Count Census Data.

Table B-3: Robustness of Ever-Married Regressions to Alternative Quartile Assignment Methods

	1850-1880	1870-1900	1880-1910	1900-1930	1910-1940
Dependent Variable: Ever Married					
Baseline	-0.02036	-0.02576	-0.03665	-0.04868	-0.05366
	0.00603	0.00604	0.00545	0.00520	0.00417
	[34990; 1207]	[60007; 1832]	[76284; 2305]	[114391; 3412]	[125014; 3409]
Soundex	-0.01525	-0.02039	-0.03540	-0.04586	-0.04884
	0.00586	0.00637	0.00550	0.00562	0.00420
	[39684; 674]	[64993; 846]	[81491; 937]	[122168; 1214]	[135948; 1339]
Name-Region	-0.03514	-0.05174	-0.07025	-0.07488	-0.06734
	0.00698	0.00561	0.00597	0.00479	0.00374
	[31590; 859]	[56630; 1297]	[71593; 1642]	[108792; 2489]	[120370; 2473]
Adult Name Distribution	-0.02054	-0.02317	-0.03318	-0.04858	-0.05576
	0.00628	0.00579	0.00529	0.00574	0.00427
	[34990; 1207]	[60007; 1832]	[76284; 2305]	[114391; 3412]	[125014; 3409]
Dependent Variable: Husband's Parental Log Occupational Status					
Baseline	0.00968	0.01296	0.01094	0.02465	0.03469
	0.00323	0.00191	0.00179	0.00218	0.00223
	[16659; 906]	[29063; 1401]	[37642; 1735]	[60763; 2720]	[61978; 2542]
Soundex	0.00950	0.01488	0.00930	0.00402	0.01735
	0.00176	0.00202	0.00144	0.00138	0.00117
	[20271; 582]	[33724; 729]	[42455; 787]	[93771; 1158]	[100603; 1266]
Name-Region	0.09608	0.11847	0.17497	0.21785	0.19055
	0.00849	0.00792	0.00744	0.00939	0.00684
	[14245; 634]	[26359; 992]	[33942; 1228]	[55715; 1965]	[57997; 1851]
Adult Name Distribution	0.00956	0.01209	0.00967	0.02409	0.03250
	0.00297	0.00192	0.00206	0.00194	0.00224
	[16659; 906]	[29063; 1401]	[37642; 1735]	[60763; 2720]	[61978; 2542]

Notes: Entries in this table represent OLS coefficients from the estimation of equation (3). Coefficients reflect the difference in the outcome variable between the highest- and lowest-status quartiles of names for the cohort specified in the column. Standard errors in parentheses, clustered by the woman's first name. The number of observations and the number of women's names used to impute father's income for each decade for each regression are specified under each estimate. Different rows represent changes to the way quartiles are imputed. The Soundex specification uses the Soundex of men and women's first names to assign father's income. The Adult Name Distribution assigns father's income to men and women as adults and then defines the quartile indicators, instead of assigning quartiles among children's names.

Source: 1% IPUMS samples (1850-1930), 1% extract of the 1940 Restricted Complete Count Census Data.

Table B-4: Robustness of Ever-Married Regressions to Alternative Age Restrictions

	1850-1880	1870-1900	1880-1910	1900-1930	1910-1940
Dependent Variable: Ever Married					
Women 30-45	-0.01947	-0.02535	-0.03632	-0.04839	-0.05399
	(0.00599)	(0.00602)	(0.00545)	(0.00520)	(0.00462)
	[34990; 1207]	[60007; 1832]	[76284; 2305]	[114391; 3412]	[126002; 3456]
Women 30-40	-0.02015	-0.02253	-0.03826	-0.04923	-0.05551
	(0.00708)	(0.00715)	(0.00628)	(0.00529)	(0.00492)
	[26366; 1105]	[44668; 1652]	[57150; 2042]	[83331; 3086]	[90597; 3085]
Dependent Variable: Husband's Parental Log Occupational Status					
Wives and Husbands 30-45	0.00977	0.01300	0.01096	0.02464	0.03029
	(0.00324)	(0.00191)	(0.00179)	(0.00219)	(0.00238)
	[16659; 906]	[29063; 1401]	[37642; 1735]	[60763; 2720]	[60699; 2562]
Wives 30-40 and Husbands 35-45	0.01385	0.01222	0.01084	0.02537	0.03134
	(0.00409)	(0.00210)	(0.00213)	(0.00258)	(0.00247)
	[11207; 767]	[19041; 1198]	[24934; 1442]	[39362; 2309]	[38442; 2078]
Wives 30-45 and Husbands 30-55	0.00853	0.01215	0.01136	0.02369	0.02875
Using 30-year lags only	(0.00303)	(0.00171)	(0.00158)	(0.00207)	(0.00222)
	[22311; 1017]	[37862; 1555]	[49807; 1956]	[78475; 2995]	[79580; 2872]
Wives 30-45 and Husbands 30-55		0.01014	0.01247		0.02875
Using 30- and 40-year lags		(0.00188)	(0.00185)		(0.00193)
		[37879; 1555]	[49779; 1954]		[79778; 2878]

Notes: Entries in this table represent OLS coefficients from the estimation of equation (3). Coefficients reflect the difference in the outcome variable between the highest- and lowest-status quartiles of names for the cohort specified in the column. Standard errors in parentheses, clustered by the woman's first name. The number of observations and the number of women's names used to impute father's income for each decade for each regresson are specified under each estimate. Different rows restrictions the estimation sample related to age. In the specification where husbands aged 30-55 are included and only 30-year lags are used, the father's income of husbands aged 45-55 are assigned the mean income by first name of 0-15 year olds at a 30-year lag. In the specification where husbands aged 30-55 are included and both 30-year and 40-year lags are included, husbands aged 45-55 are assigned the mean income of children 0-15 at a 40-year lag and husbands 40-45 are assigned the average of the income of children at 30- and 40-year lags. Source: 1% IPUMS samples (1850-1930), 1% extract of the 1940 Restricted Complete Count Census Data.