2024-08-22

The emergence of the socioeconomic gradient in women's marriage outcomes

The Emergence of the Socioeconomic Gradient in Women's Marriage Outcomes*

Claudia Olivetti
Dartmouth College and NBER

M. Daniele Paserman
Boston University and NBER

Laura Salisbury
York University and NBER

E. Anna Weber
U.S. Military Academy

May 2024

Abstract

We present new findings about the relationship between marriage and socioeconomic background in the United States in the late 19^{th} and early 20^{th} 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, which widens over this period. Regional divergence in occupational structure explains half of the divergence in the probability of marriage, and most of the increase in marital sorting. Urbanization and the associated improvement in women's labor market opportunities drive most of these differences.

Keywords: Marriage, Assortative Mating, Gender, Intergenerational Mobility, Regional Convergence.
JEL codes: J12, J62, N31, N32, N91, N92

*This paper previously circulated under the title "Who married, (to) whom, and where? Trends in Marriage in the United States, 1850-1940." We would like to thank Fadi Hassan, Jeanne Lafortune, Bob Pollak, and multiple conference and seminar participants. Juan Conde Carvajal provided excellent research assistance. All errors are our own. All opinions expressed in this paper are those of the authors and do not represent the opinions of the United States Military Academy (USMA), the United States Army, or the Department of Defense.

1 Introduction

Mothers and 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 marital sorting may affect the intergenerational transmission of inequality and the extent of cross-group interactions within a society (Kalmijn, 1998).

This paper presents one of the first accounts of marriage patterns by socioeconomic status in the United States for cohorts of women born between 1840 and 1910. Was marriage an outcome equally prevalent across social classes? How much did marriage contribute to social stratification? In answering these questions, we shed light on the relative roles of economic opportunities for women, demographic factors, and housing market constraints in shaping marital decisions.

We focus on the role of *family of origin* in determining who marries, and to whom. Using decennial census data, we impute the socioeconomic status (SES) of women's and men's family of origin from first names. This method works if first names have socioeconomic content (Olivetti and Paserman, 2015, henceforth OP2015). An advantage of this approach is that we can calculate a marker for familial SES without creating 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.[1]

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, but the differences in marriage rates between high and low SES women were small. The 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

[1]Recent work uses digitized marriage certificates (Withrow, 2020; Eriksson et al., 2023), vital records (Bailey and Lin, 2022) or data from online genealogy platforms (Buckles et al., 2023) to study historical intergenerational mobility of women and men. We see our work as complementary since our method allows to measure national trends for a large, representative sample.

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.

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 is 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 largely reflects increased sorting *between* marriage markets. However, even within marriage markets, high SES women became relatively less likely to marry.

The negative gradient in the probability of marriage by educational attainment (a strong correlate of SES) for women born in the early part of the 20^{th} century is well documented. Our findings indicate that this gradient was not an intrinsic historical feature of the American marriage market. Rather, it emerged during the second half of the 19^{th} century. We argue that increasing economic opportunities, especially for high SES women in urban areas, contributed to the divergence in marriage rates between low- and high-SES women. The expansion of urban manufacturing and clerical jobs rendered marriage a less attractive option. This mirrors the patterns observed more recently in countries undergoing rapid structural transformation, where marriage rates of highly educated women have plummeted (Bertrand et al., 2021; Hwang, 2016). Our results suggest that a part of the growing socioeconomic divergence in the probability of marriage is compositional. Daughters of high-income fathers are more likely to be located in urban areas, where job opportunities for women are more abundant.

What are the implications of these patterns for intergenerational persistence? Intuitively, increased sorting leads to more intergenerational persistence and, potentially, economic in-

equality (Kremer, 1997; Fernandez and Rogerson, 2001). Intergenerational persistence in the United States increased substantially around the turn of the 20th century, even within regions (OP2015). As we find no evidence of increased marital sorting *within* marriage markets, our findings suggest that other factors (increase in economic inequality, higher returns to human capital) may have been the driving force behind the increase in persistence. In fact, the increasing negative socioeconomic gradient in marriage that we document may have the opposite effect. Withdrawal of high-SES women from the marriage market implies better marriage opportunities for low-SES women. The children of low SES women will therefore have access to more resources, which can lead to an increase in mobility.

Literature Review. The literature on marriage outcomes has focused on acquired characteristics like education, occupation, and earnings as markers of SES. Highly educated women born in the first half of the 20th century were substantially less likely to marry (Goldin, 2004; Bailey, Guldi and Hershbein, 2014). We complement these studies by focusing on marital outcomes by socioeconomic status of the family of origin, and by documenting the link between marriage and SES for earlier cohorts of women. Our results indicate that the negative SES gradient in marriage observed in the early 20th century was not as pronounced for these earlier cohorts.

Studies in economics and sociology have analyzed sorting on *acquired* traits such as education, occupation and labor income.[2] There is a smaller literature investigating the correlation in spousal *ascribed* characteristics, like parental occupation or wealth, which uniformly finds evidence of positive assortative matching on these traits (Charles et al. 2013; Fremeaux 2014; Wagner et al. 2020; Eriksson et al. 2023) 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

[2]Mare (2016); Greenwood et al. (2014); Eika et al. (2019); Gihleb and Lang, (2020).

of birth and childhood neighborhood characteristics on economic mobility (Chetty et al., 2014; Chetty and Hendren, 2018a, 2018b; Derenoncourt, 2022). Tan (2023) also documents substantial heterogeneity in economic mobility in the United States in the early 20$^{\text{th}}$ 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 Methods and Data

We estimate women's marriage outcomes by the SES of their *family of origin*. This is, in part, by necessity. As married women overwhelmingly did not work, and information on educational attainment is scarce, there are no good measures of acquired socioeconomic traits for our period. However, there are reasons to favor this measure regardless of data constraints. Women had few opportunities to invest in education or marketable skills at the turn of the 20th Century. 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 method introduced by OP2015, we use an individual's *first name* to impute father's economic status. This method, 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. Among names occurring at least 20 times in a U.S. census year, there is a high degree of continuity in the names that are given to children with high- or low-income parents, which suggests that SES affected naming patterns.[3] OP2015 document

[3]For example, OP2015 show that variants of Anna, Kathryn, Carolyn, and Elizabeth are among names

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. Online Appendix A contains further details.

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.[4] 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 Appendix C 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.[5]

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

with the highest average parental SES in multiple census years between 1850-1930, while Sally, Nancy, Lucinda, Martha, and Rachel are repeatedly among the lowest.

[4]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.

[5]Each name is weighted by its frequency in the overall distribution of children's names in year t. 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 $t+30$, and to using finer partitions of the parental income distribution (see Appendix Table C6).

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.[6]

OP2015 evaluate the magnitude of these biases in pseudo-estimates of the 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, attenuation bias seems to dominate. Santavirta and Stuhler (2024) explore the properties of grouping estimators based on names. They emphasize that the probability limit of the estimator depends on the overlap between the child and parent samples. In the case of perfect overlap (i.e., the parent and child samples are taken from exactly the same families), a name-based estimator will be upward biased. On the other hand, if samples are non-overlapping, the bias may be positive or negative. Our samples are non-overlapping since we use 1% census samples to measure fathers' economic status and children's outcomes.[7] In Appendix A, we provide evidence that the pseudo-estimators used in this paper are biased toward zero.

Given the potential for bias, the name-based methodology is best used for uncovering trends, provided the bias in the pseudo-estimator is roughly constant over time. Time-varying bias will occur if the information content of names changes over time. If names become more informative about SES, our estimates will show the marriage outcomes of high and low SES women diverging, even if the underlying process has not changed. However, we argue that changes in the informativeness of first names is unlikely to have a first-order

[6]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. First names are likely to suffer less from this problem: they are correlated with parental SES, but have a smaller causal effect on individual outcomes than the alternatives.

[7]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.

impact on our findings.[8]

We can roughly measure the socioeconomic content of names by regressing parental occupational income on a full set of fixed effects for children's names and computing the R^2. OP2015 find that this measure of the socioeconomic content of names was stable for boys but increased slightly for girls between 1850 and 1910, from 0.136 to 0.166. We explore the extent to which this affects our results in Appendix A-1. We exploit the fact that, in the 1940 census, we can observe first names and a measure of SES (educational attainment). We draw two different 1% samples from the full count 1940 census, and we use these to estimate the relationship between women's SES and marriage outcomes using: (i) actual educational attainment, and (ii) mean educational attainment by first name from a different (non-overlapping) sample. This allows us to gauge the degree of attenuation bias in our estimates. Then, we run a simulation in which we assign first names to the individuals in our samples, using a method that allows us to vary the socioeconomic content of names. We are able to replicate the results that use *true* first names very closely. We show that, in order to fully explain our findings, the socioeconomic content of names would have to increase such that the R^2 from a regression of SES on name fixed effects *doubles*. A 20% increase in R^2 (what we observe for girls' names over our sample period) has a negligible effect on the relationship between marriage and women's (pseudo) SES.

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). 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 observed between age 30 and 45 (in 1880, 1900, 1910, 1930 and 1940).

[8]See Appendix A for further details.

Our married sample consists of couples in which both spouses are 30-45 years old. This restriction skews the age distribution of the married sample relative to the sample used to study selection into marriage, because during our sample period the average husband was roughly 4.5 years older than his wife. Our descriptives and the results in Section 3.1 are robust to several adjustments to the age distribution (see Appendix C).

3 Trends in Marriage Outcomes

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 high, between 85% and 92%. There is a clear socioeconomic gradient in the marriage rate: women from the bottom quartile of the SES distribution are always more likely to marry than those from the top quartile. This socioeconomic gradient grew 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. In Appendix A.1, we use the 1940 census to show that the *true* difference in the probability of marriage between women in the top and bottom quartiles is likely 3-4 times larger than what we estimate using pseudo quartiles.

The bottom panel of Figure 1 shows the average log parental OCCSCORE of husbands of women from each SES quartile. Women in the highest quartile consistently have husbands with the highest 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-bottom quartile gap is about one sixth the size of the overall

increase in average spousal SES. Our estimates are also severely attenuated, as we show in Appendix Table A-2.

How does the change in the SES gradient affect fertility? In Appendix E, we show that high-SES women have lower fertility than low-SES women, and that fertility declined during this period for all socioeconomic groups. The SES-fertility gradient increased, but this was almost entirely driven by the increase in the SES-marriage gradient – for married women, there was almost no increase in the gradient. This implies that there were no significant changes in the SES gradient of available resources *per child*.

Marital sorting is typically measured as the correlation between spousal traits reflecting SES. To compare our findings to the literature, we calculate the correlation between husbands' and wives' parental SES, both imputed using our names-based methodology. 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 over the period.

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 2013). 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).

Because we have an imperfect proxy for parental SES, our estimates will be attenuated by measurement error. OP2015 and Santavirta and Stuhler (2024) find that intergenerational elasticity estimates using 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 Appendix A, we discuss this point in more detail. We estimate the degree of attenuation in a pseudo-estimate of the correlation between spousal

educational attainment using the 1940 census. Our results suggest that the true correlation in parents' SES is likely 3-6 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 0.6. These values are consistent with the literature.

4 Interpreting Trends in the Relationship between SES and Marriage

We now focus on distinguishing between divergence *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. There are large differences in key marriage market outcomes across regions, with marriage rates higher in the South than in the Northeast (see Appendix B). 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.

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

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

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

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 replicates the trends described in Figure 1, but allows 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 average log OCCSCORE of never married women, the share of children ages 5-20 attending school, the average farm value, and the average number of people per dwelling, and the share of the state's area located within 15 miles of a railroad. State characteristics are assigned by 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 school enrollment rate, which is measured at $t-20$ to reflect childhood conditions.[9]

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 3. 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

[9] 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).

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 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 the between-marriage market divergence can explain seventy percent of the nationwide increase in assortative mating between 1880 and 1940 (Panel B). Between market differences in parental log occupational score go from explaining none of the Q1-Q4 gap in spousal log occupational score (for earlier cohorts) to explaining about half of the gap (for cohorts born at the turn of the 20^{th} Century. However, the between-marriage market divergence explains only one third of the increase in the negative socioeconomic gradient in marriage for women (Panel A). We report the coefficients underlying this figure, as well as estimates of $\Delta\gamma_4$ and $\Delta\overline{\gamma}_4$, in Appendix Table C-1.

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.[10] In Figure 4, 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 school enrollment 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

[10]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.

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.[11]

The largest drivers of between-marriage market divergence are the urban population share and the foreign born population share. These characteristics are highly correlated, and thus difficult to separate. We find little evidence that the foreign-born population share drives this relationship. Our results are invariant to including individual controls for foreign-born status, or to restricting the sample to native born individuals. We therefore focus on the urban population share. Urbanization and regional development are closely linked. Virtually no farmers live in cities or large towns, implying that the urban population share is inversely related to the share of the population employed in agriculture. Regional differences in urbanization track regional differences in industrial or occupational structure, which may be why urbanization explains the increasing tendency for men and women from similar socioeconomic backgrounds to marry. This may also explain part of the increase in the socioeconomic gradient in marriage for women. For instance, the South did not urbanize as quickly as the rest of the country, and it had the highest marriage rate for women. Thus, as families with low occupational incomes became more concentrated in the South, the link between marriage and low SES would have strengthened.

Marriage may have decreased more rapidly in urban areas because of housing market conditions – high housing costs raise the costs of setting up a new household. We find little support for this explanation: controlling for housing density (proxied by the number people living in each dwelling unit) or controlling for the average value of farms explain relatively little of the trend in the SES gradient in marriage. [12]

Another possibility is that urbanization increased women's labor market opportunities, leading to a decline in marriage rates. Figure 5 shows three important facts about never-

[11] We acknowledge that we may not be measuring sex ratios perfectly, as sex ratios within socioeconomic brackets may be more relevant.

[12] More direct controls of housing costs, monthly rent and average house value, are only available for the final cohort but explain much less of the level of the SES gap in this year than urbanization or state fixed effects.

married women's economic activity during this period. First, never-married women increasingly participated in the formal labor market. Second, never-married women in urban areas participated in the labor market at almost double the rate of rural women, a gap that is stable to increasing over time. Finally, married women rarely participated in the formal labor market regardless of geographic location. The urban-rural gap in women's labor force participation is likely caused by differences in industrial structure. Goldin (1984) documents the emergence of the clerical sector as an employer of educated single women between 1890 and 1930. In Appendix D, we show that urban single women are increasingly likely to hold higher-status clerical positions during our sample period. In fact, the employment rate of single women explains roughly the same amount of the increase in the socioeconomic gradient in marriage as the urbanization rate and state fixed effects (see Figure 4).

The evidence suggests that the growing socioeconomic gradient in marriage for women was driven in part by an association between urbanization and improved labor market opportunities for women. Women in urban areas were increasingly likely to participate in the formal labor market and faced more appealing employment options. Remaining unmarried thus became more economically attractive over time.[13] As men in more urban states also had higher status occupations, this generated a negative relationship between parental occupational income and the marriage rate. [14]

Overall, there is little increase in assortativeness within marriage markets, which suggests that changes in the marriage market are unlikely to explain changes in inequality or intergenerational persistence during this period. While there is an increasing negative effect of parental SES on the propensity for women to marry, a portion of this is compositional. Daughters of high income fathers were increasingly likely to live in urban areas, and urban residence seems to have made marriage less attractive to women, on account of the more

[13]Salisbury (2017) documents a negative causal effect of outside economic opportunities on women's propensity to marry.

[14]Even within states, women's improved labor market opportunities may strengthen the relationship between parental SES and marriage if high SES daughters are more able to access these occupations; however, we do not test this hypothesis directly.

abundant labor market opportunities for single women. This underscores that the relationship between women's SES and marriage depends on the interaction between women's labor market opportunities and the ability to combine work and family, both of which have changed dramatically over the last 200 years (Goldin 2021).

In Appendix C, we assess the robustness of our results to different ways of calculating occupational income, different samples, and different ways of grouping individuals in order to impute parental SES. All 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 whether these changes are 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.

We show that the large (and well-known) 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. This underscores the close link between women's labor market opportunities and the family. Our paper also shows that, despite a large increase in the correlation between husbands' and wives' SES, the degree of assortativeness *within* marriage markets was quite stable over time. Thus, it is unlikely that changes in marriage market behavior can explain the increase

in intergenerational persistence that occurred over the period.

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. and Peter Z. Lin (2022). "Marital Matching and Women's Intergenerational Mobility in the Late 19th and Early 20th Century U.S.." Mimeo, University of California at Los Angeles.

[6] 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.

[7] 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.

[8] Bertrand, Marianne; Cortes, Patricia; Olivetti, Claudia and Jessica Pan (2021) "Social Norms, Labor Market Opportunities, and the Marriage Gap for Skilled Women." *Review of Economic Studies, 2021*, 88 (4): 1936–78.

[9] Buckles, Kasey; Price, Joseph ; Ward, Zachary and Haley EB Wilbert (2023). "Family Trees and Falling Apples: Historical Intergenerational Mobility Estimates for Women and Men." National Bureau of Economic Research Working Papers, No. 31918.

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

[11] 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.

[12] 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.

[13] 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.

[14] Derenoncourt, Ellora (2022). "Can You Move to Opportunity? Evidence from the Great Migration." *American Economic Review.* 112(2): 369-408.

[15] Eika, Lasse; Mogstad, Magne and Zafar, Basit (2019). "Educational Assortative Mating and Household Income Inequality." *Journal of Political Economy.* 127(6): 2795-2835.

[16] Eriksson, Katherine; Niemesh, Gregory T.; Rashid, Myera and Jacqueline Craig (2023). "Marriage and the Intergenerational Mobility of Women: Evidence from Marriage Certificates 1850-1920." Mimeo, Miami University.

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

[18] Fernandez, Raquel; Rogerson, Richard (2001). "Sorting and Long-Run Inequality." *Quarterly Journal of Economics.* 116(4): 1305–1341.

[19] Fremeaux, Nicolas (2014). "The Role of Inheritance and Labour Income in Marital Choices." *Population.* 69(4): 495-530.

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

[21] Gihleb, Rania and Lang, Kevin (2020). "Educational Homogamy and Assortative Mating Have Not Increased." Polachek, S.W. and Tatsiramos, K. (Ed.) *Change at Home, in the Labor Market, and On the Job (Research in Labor Economics, Vol. 48).* Emerald Publishing Limited, Leeds, pp. 1-26.

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

[23] Goldin, Claudia (1984). "The Historical Evolution of Female Earnings Functions and Occupations." *Explorations in Economic History*, 21(1): 1-27.

[24] 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.

[25] Goldin, Claudia (2021). *Career and Family: Women's Century-Long Journey toward Equity*, Princeton University Press.

[26] 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.

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

[28] 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.

[29] Hwang, Jisoo (2016). "Housewife, "Gold Miss," and Equal: the Evolution of Educated Women's Role in Asia and the U.S." *Journal of Population Economics.* 29: 529–570.

[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] Kremer, Michael (1997). "How Much does Sorting Increase Inequality?" *Quarterly Journal of Economics.* 112(1): 115-139.

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

[33] 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.

[34] 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.*

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

[36] 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.

[37] 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.*

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

[39] 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.

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

[41] Santavirta, Torsten and Stuhler, Jan (2024). "Name-Based Estimators of Intergenerational Mobility." Mimeo., Universidad Carlos III de Madrid.

[42] Salisbury, Laura (2017). "Women's Income and Marriage Markets in the United States: Evidence from the Civil War Pension." *Journal of Economic History*, 77(1): 1-38.

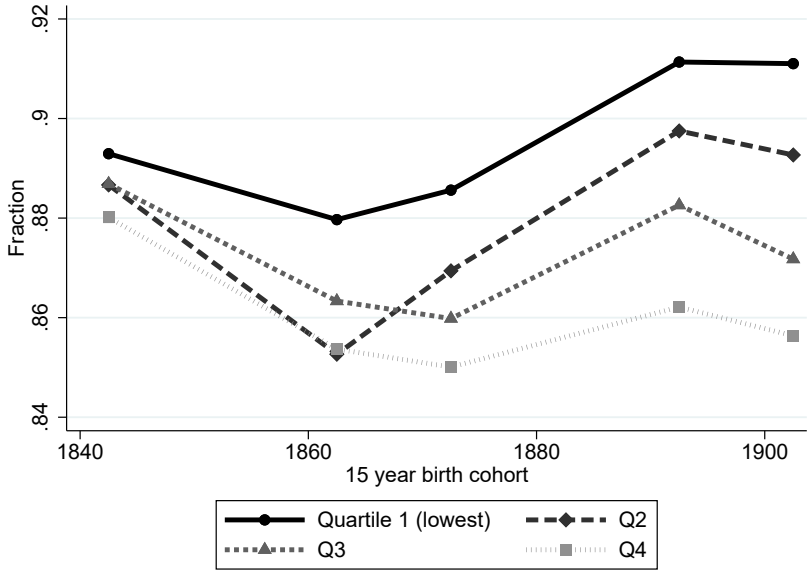
[43] Tan, Hui Ren (2023). "A Different Land of Opportunity: the Geography of Intergenerational Mobility in the Early 20$^{\text{th}}$ US." *Journal of Labor Economics*, 41(1): 77-102.

[44] Wagner, Sander; Boertien, Diederik and Gørtz, Mette (2020). "The Wealth of Parents: Trends over Time in Assortative Mating Based on Parental Wealth." *Demography.* 57: 1809–1831.

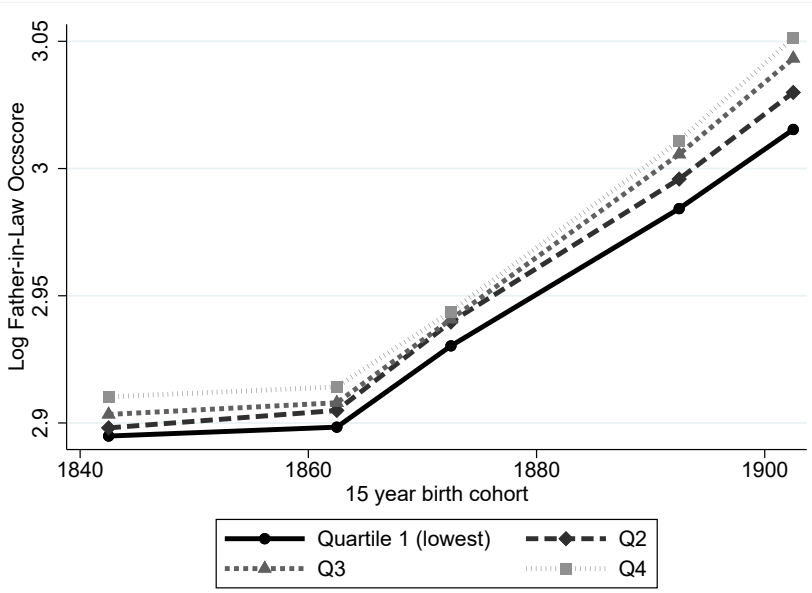
[45] Withrow, Jennifer (2020). ” “The Farm Woman’s Problem”: Farm Crisis in the U.S. South and Migration to the City, 1920-1940.” Mimeo. University of Massachusetts Amherst.

Figure 1: Trends by Woman's Parental SES

(a) Fraction of Women 30-45 Ever Married

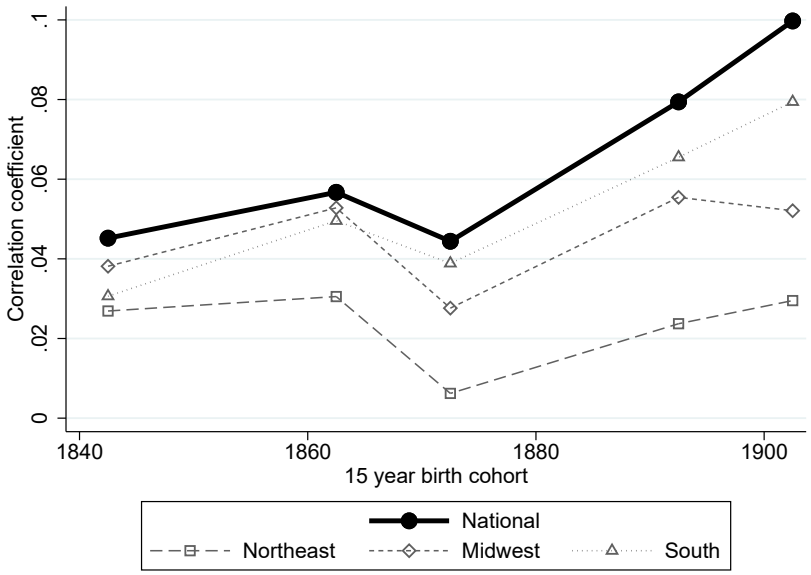


(b) Father-in-Law's SES



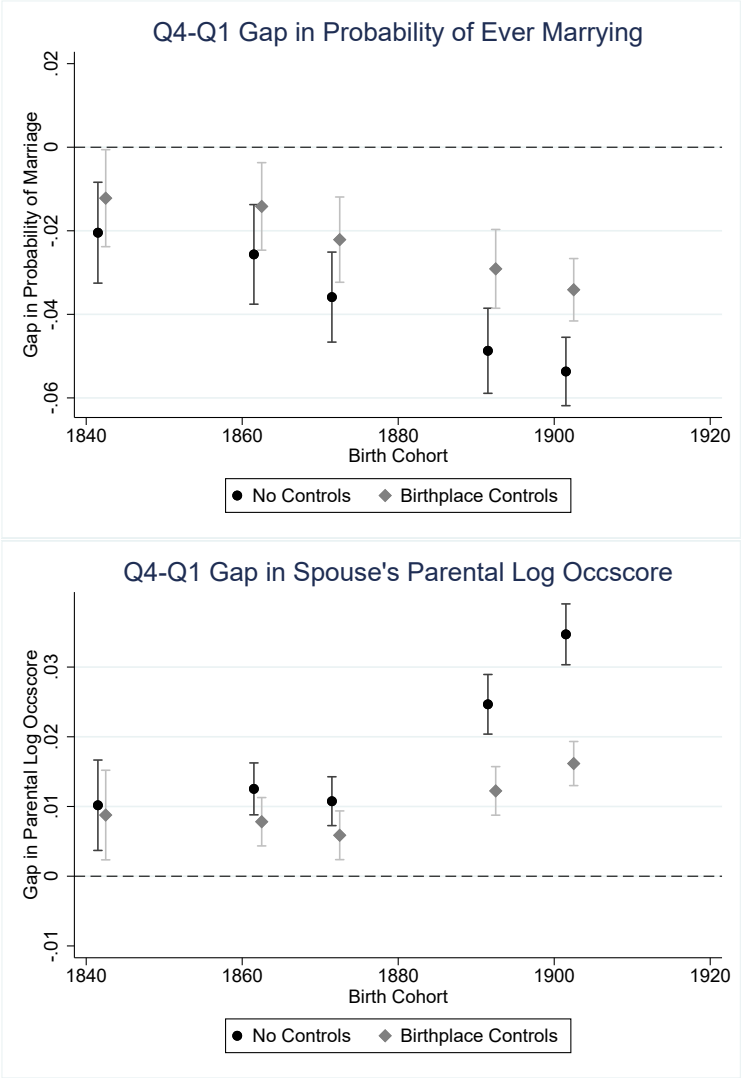
Note: The top panel shows the fraction ever married by birth cohort and parents' quartile of occupational income score. The bottom panel shows the average log occupational score of fathers of husbands aged 30-45 married to wives 30-45 by the woman's parental quartile of occupational income score. The 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



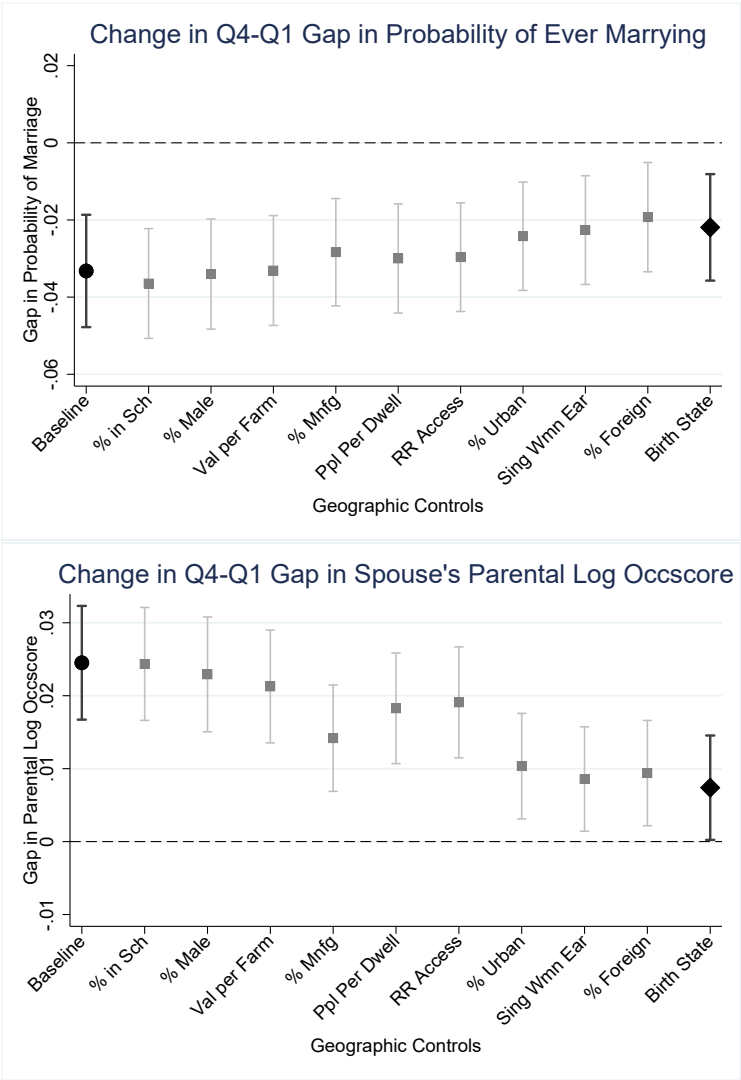
Note: This figure shows the correlation in parental occupational income score between husband and wife. Region assignment is by the wife's region of birth. The occupational income score is imputed based on first names. Source: Authors' calculations based on the 1% IPUMS samples (1850-1930) and a 1% extract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

Figure 3: Effect of Parental SES and Birthplace on Probability of Ever Marrying and Spouse's Parental Log Occupational Score, Women 30-45



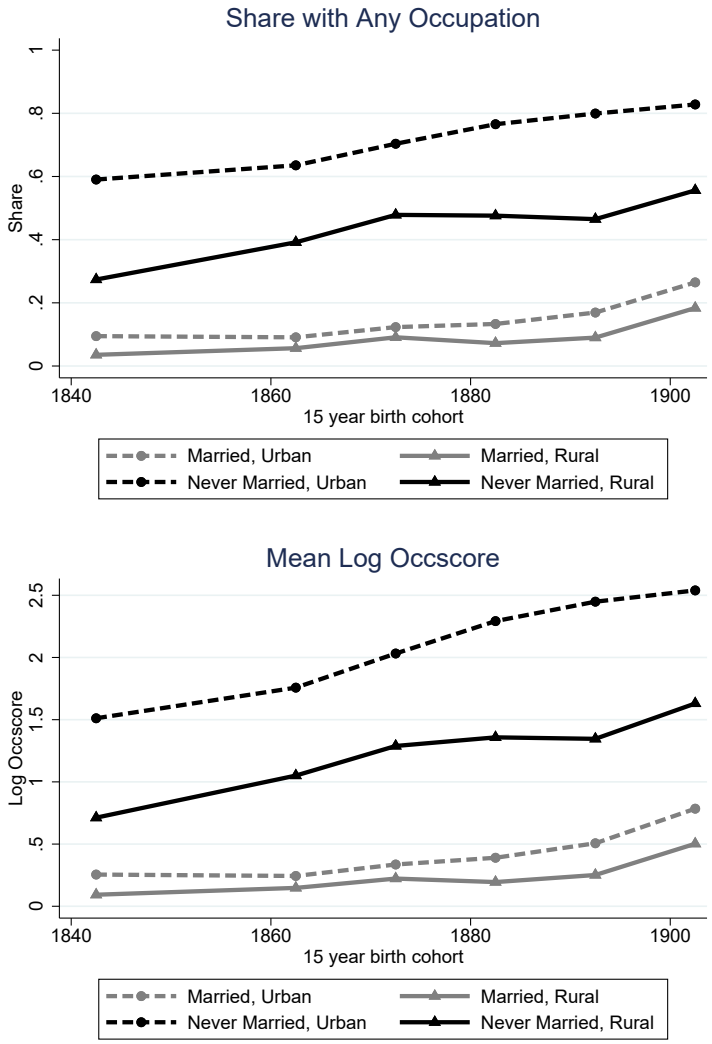
Note: 95% confidence intervals are shown. Plotted coefficients represent the coefficient on the indicator for the top SES quartile, as compared to the bottom SES quartile (the omitted SES category). The Birthplace Controls specification includes a set of dummies for the woman's state of birth interacted 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 4: 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.

Figure 5: Trends in Women's Employment by Marriage and Urban Status, Women 30-45



Note: 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.