Who married, (to) whom, and where?
Trends in marriage in the United States, 1850-1940*

Claudia Olivetti
Boston College and NBER

M. Daniele Paserman
Boston University and NBER

Laura Salisbury
York University and NBER

E. Anna Weber
Boston University

September 2018

Abstract

This paper presents a novel analysis about marriage in the United States in the late 19th and early 20th centuries, and its relation to socioeconomic status. We document the following facts: 1) Already in the mid-19th Century there was a socioeconomic gradient in marriage rates – men and women born to families in the bottom quartile of the occupational earnings distribution were more likely to marry than those in the top quartile. The gradient had grown steeper by the middle of the 20th Century. 2) The increase in gradient is explained in part by increased income divergence across U.S. regions, together with a regional gradient in marriage rates, and by an increased socioeconomic gradient within regions. 3) Age at marriage follows an inverted U-shape, and exhibits both a socioeconomic and a regional gradient. Both gradients become steeper over time. 4) There is a substantial increase in the degree of assortativeness by socioeconomic status over this period. This is accounted for in about equal measure by regional income divergence, and increased assortativeness within regions. 5) The mean age gap between spouses also declines over time but it explains very little of the change in assortativeness by socioeconomic status. The overall picture is one of a society that was becoming more segmented along the marriage dimension. This increased segmentation is explained only in part by income divergence across geographic regions.

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

*VERY PRELIMINARY AND INCOMPLETE

1 Introduction

The degree to which economic status is passed along generations is key to understanding differences in the extent of inequality across societies and over time. The significant increase in inequality in much of the Western world over the past 40 years has raised the concern that the playing field is not level for everyone, and children born to disadvantaged families may be precluded from climbing the social ladder. The existing empirical evidence supports this view: there is a strong correlation between a country's level of inequality and the intergenerational elasticity, the most commonly used measure of intergenerational mobility (Corak 2013). This has sparked new interest in understanding the process by which economic status is transmitted from parents (and grandparents) to children (Solon 1999; Black and Devereux 2011; Clark, 2014). Within this literature, relatively little attention has been devoted to the role of marriage. Yet marriage, by sorting individuals into families, can play a critical role in the transmission of human capital and wealth across generations. For example, if there is a strong stratification in marriage by social class, this might magnify individual-level intergenerational persistence, because mothers can contribute to their children's human capital accumulation, either directly through their own human capital, or by tapping into their family's financial resources or network connections. Further, as seen in Figure 1 the rate of marriage and the age relationship between spouses have undergone dramatic transformations over the past 150 years.There is evidence of positive educational sorting in the US today, but disagreement remains on whether it has increased in the past 50 years (Mare, 2016; Gihleb and Lang, 2016) and whether it has contributed to increasing income inequality (Greenwood et al., 2014; Eika et al., 2016). Evidence on assortative mating prior to 1940 is more limited, with some evidence that the educational gap between spouses initially rose and then fell over the 20$^{\text{th}}$ Century in the United States (Iyigun and Lafortune, 2016).

In this paper, we contribute to this literature by presenting one of the first studies of the trends and determinants of marriage and its relation to socioeconomic status in the United States in the late 19$^{\text{th}}$ and early 20$^{\text{th}}$ Centuries. In the first part of the paper, we use data from public use Decennial Censuses and an imputation methodology based on first names (Olivetti and Paserman, 2015; Olivetti, Paserman and Salisbury, forthcoming) to document a number of novel facts about marriage in the United States during this period. The core idea behind the imputation methodology is that first names carry information about socioeconomic status. This makes it possible to infer patterns of adult outcomes by socioeconomic status even without the need to match records across census years. A distinct advantage of this methodology is that it works equally well for men and women. Traditional approaches, typically based on linking individuals across Census years based on first and last names, are often silent about women's outcomes, as women can no longer be linked once they marry and change their last name. Another advantage is that it allows to consistently measure outcomes by socioeconomic status over a longer time-horizon.

We document the following five main facts: 1) Already in the mid-19$^{\text{th}}$ Century, marriage was negatively correlated with socioeconomic status: men and 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 in marriage rates becomes steeper over time. For women born in the 1840s, there is at most a 2 percentage point gap in marriage rates between the bottom and top quartiles of the occupational earnings distribution; for women born in the 1900s, the gap in marriage rates had grown to more than 5 percentage points. 2) The increase in the marriage-socioeconomic status gradient is explained in part by increased income divergence across U.S. regions, together with a regional gradient in marriage rates, and by an increased socioeconomic gradient within regions. 3) Age at marriage follows an inverted U-shape, and exhibits both a socioeconomic and a regional gradient. Both gradients becomes steeper over time. 4) There is a substantial increase in the degree of assortativeness over time. The increase in assortativeness is accounted for in about equal measure by regional income divergence, and increased assortativeness within regions. 5) Increased assortativeness by socioeconomic status is accompanied by a decline in the mean age gap between spouses. However, changes in the age gap explain very little of the change in assortativeness by socioeconomic status.

The overall picture is one of a society that was becoming more segmented along the marriage dimension. In part, this increased segmentation was driven by income divergence across geographic regions: by the early 20$^{\text{th}}$ Century, the Northeast had essentially completed its transition to an industrial economy, while the South was still severely lagging behind. However, even within regions, we find a consistent pattern of increased segmentation in marriage rates, and increased assortativeness by economic status.

In the second part of the paper, we investigate which specific factors can explain the divergence in marriage outcomes across socioeconomic quartiles. We link the Decennial Census samples to a variety of social and economic state level measures, such as scholarization rates, urbanization rates, industrial structure and land inequality. We then explore the relationship between these measures, marriage rates and assortative mating. We find that a woman's state of birth explains some of the relationship between SES and rates of marriage, and almost all of the relationship between a woman's SES and that of her spouse. We also provide evidence that differences in urbanization and rates of manufacturing employment explain much of this geographic divergence.

2 Related Literature

[Incomplete] Bailey, Guldi and Hershbein (2014) present rich information on trends in marriage and fertility in the U.S. from 1800 onwards, but have relatively little to say about the relationship between these factors and socio-economic status.

In modern data, most of the literature has focused on educational assortative mating, both in the US and in other countries. There is clear evidence of positive educational sorting in the US but some disagreement on whether sorting has increased in the past 50 years (Mare, 2016; Gihleb and Lang, 2016), and whether it has contributed to increasing income inequality (Greenwood et al., 2014; Eika et al., 2016). Research in sociology suggests that the link between spousal education and equivalized earnings or income might be a crucial determinant of whether educational assortative mating affects inequality (Breen and Salazar, 2011).

We know surprisingly little about the correlation between spouses' *parents'* characteristics in modern data. The only exception is work by Charles et al. (2016), who estimate that the correlation in parental wealth among married spouses is about 0.4.

3 Methodology

The methodology used in this paper is an extension of Olivetti and Paserman (2015, henceforth OP2015). We present here a brief summary.

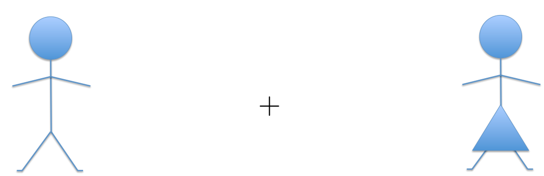
There are two key challenges for our approach. First, we need a way to measure men and women' socioeconomic status in pre-1940 data, before the Census started collecting data on educational attainment. Income and wages are also unavailable before 1940. Finally, only a small fraction of married women worked outside of the household, so that occupational status is also not available. Our solution is to focus on the socioeconomic status of married spouses' *fathers*. In practice, this means studying marriage rates and sorting based on the socioeconomic status of one's family of origin (as in Charles et al., 2016). Fathers' socioeconomic status is measured by the OCCSCORE variable in IPUMS. This variable indicates the median total income (in hundreds of dollars) of persons in each occupation in 1950. More than 200 distinct occupations appear in our data; however, certain occupations appear at high frequency. In particular, farmers comprise approximately 40% of the workforce in 1850, though this declines to 10% by 1940. For this reason, we perform sensitivity analysis with occupational income distributions from different years.

The second challenge we face is that of inferring the socioeconomic status of fathers without intergenerationally linked data. The solution, as in OP2015, is to impute father's economic status by *first names*. Specifically we calculate the mean log occupational score of fathers of children aged 0-15 in year t with a given first name, and then impute that score to all adults with that first name in either $t + 20$ or $t + 30$. The distinct advantage of this methodology is that it allows to estimate adult outcomes by the socioeconomic status of the family of origin, even without individually linked data and even if wives change name upon marriage.

The key assumption underlying this methodology is that names carry information about socioeconomic status. There is evidence for this in both modern (Bertrand and Mullainathan, 2004; Fryer and Levitt, 2004) and historical data. In our sample period, between 10 to 17 % of total variation in father's socioeconomic status can be explained by the variation *between* names given to their children (OP 2015).

An illustration of the methodology for calculating assortative mating is given in the following example (see also the illustration below): suppose that high socioeconomic status adults name their children Adam and Abigail, and low socioeconomic status adults name their children Zachary and Zoë. Then, if the Adams marry the Abigails and the Zacharys marry the Zo´es, we would say that there is a high degree of sorting. If, on the other hand, the Adams are equally likely to marry the Abigails or the Zoës, we would conclude that there is a low degree of sorting. This methodology allows one to measure marital sorting by socioeconomic status of the family of origin going all the way back to the middle of the 19$^{\text{th}}$ Century.

Adam & Abigail, age 30-45, 1880



y_m: mean log occ. earnings of fathers of sons 0-15 named Adam, 1850	y_f: mean log occ. earnings of of fathers of daughters 0-15 named Abigail, 1850

It is easy to see how this methodology can be easily applied to measure any adult outcomes: marriage rates, age at first marriage, spousal age gaps, and more.

This methodology suffers from some limitations relative to the measurement of the true level of sorting on fathers' income which could be calculated with individually linked data. First, our estimates are subject to attenuation bias from the measurement error introduced because first names are an imperfect proxy for father's socioeconomic status. It it also possible that our estimates are subject to bias introduced if names themselves have a return on the marriage market correlated with father's socioeconomic status. For example, if high SES status names themselves increase the likelihood of marrying a high SES spouse, this will upward bias our estimated level of assortativeness

An additional possible concern is the interpretation of trends in the case that the informativeness of names changes over time. If names become more informative, our estimated assortativeness will increase even in the absence of a change in the true level of sorting. In Olivetti and Paserman (2015) we find only limited evidence of the increased informativeness of names.

3.1 Data

We use data from the 1850 to 1940 Decennial Censuses of the United States, which contain information on first names. For 1850 to 1930 we use the 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 all the analysis to whites to avoid issues associated with the almost complete absence of blacks in the pre-Civil War sample, and the fact that even in the late cohorts many blacks would have spent a substantial part of their lives as slaves.

Individual level data are available from IPUMS for every decadal Census from 1850 to 1940, with the exception of 1890. This means that we can calculate marriage rates and assortative mating for six cohorts of men and women observed between age 20 and 35 (1870, 1880, 1900, 1920, 1930 and 1940), and five cohorts of men and women observed between age 30 and 45 (1880, 1900, 1910, 1930 and 1940). We focus primarily on the latter, to avoid issues of selection into marriage early in the lifecycle.

We measure socioeconomic status by occupational income, which is the only continuous measure of economic status available throughout the period. We use the OCCSCORE variable from IPUMS, which is based on the median income of workers with a particular occupation in 1950. 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 2010; Olivetti and Paserman 2015); we assign farm income both nationally and at the state level. (2) We assign occupational income using the occupational wealth distribution from 1870 (Ferrie et al 20XX; Olivetti et al 2018). (3) We use LIDO scores (Saavedra and Twinam, 2018).

4 Description of Trends

4.1 Marriage Rates

Figure 2 shows the evolution of the fraction of individual ever married by socioeconomic status, for men and women aged 30-45. A number of points stand out: a) Marriage rates are quite high: between 77 and 86 % for men, and between 85 and 92 % for women; b) the male marriage rate exhibits a marked U-shape; for women, the U-shape is much more muted; c) there is a clear socioeconomic gradient in marriage rates – marriage rates are higher for those born in the bottom quartile of the socioeconomic status distribution; and d) the socioeconomic gradient in marriage rates becomes steeper over time, especially for women.

In Figure 3, we explore differences in marriage rates 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, there has been some divergence in marriage rates by region: for instance, Southern women became much more likely to marry, relative to their Midwestern and Northeastern counterparts.

4.2 Age at First Marriage

We next document patterns in age at first marriage. Here we need to be a bit cautious, because the question on age at first marriage was not asked consistently in each Census. Both the wording of the question and the sample to which the question is administered changes over time. With this caveat in mind, Figure 4 shows the pattern in age at first marriage over time and by socioeconomic status. Female age at first marriage exhibits a slight inverted-U shape, but on the whole the variation over time is dwarfed by the variation across occupational income quartiles. Similarly to marriage rates, age at first marriage also exhibits a clear socioeconomic gradient, which grows steeper over time.

A similar pattern is apparent in Figure 5 , which shows age at first marriage separately by region. The overall shape of the trend is similar across the three regions, but the gap between the different regions becomes larger over time.

4.3 Assortative Mating

We have shown evidence of increasing gaps in marital status and age at marriage by both SES and region, but is there also divergence in who marries whom? In this section, we look at two measures of assortative mating: the fraction women marrying men at different quartiles of parental SES and the correlation in parental SES between spouses.

Figure 6 shows the fraction of women married to men in the bottom and top quartile of the imputed father's income distribution. We observe sorting, far from perfect homogamy, and increasing gaps especially for Q1 women.

Figure 7 shows the further results related to assortative mating. The index of assortative mating is simply calculated as the correlation in log occupation score of the fathers of husbands and wives. For both 20-35 and 30-45 year olds, there is clear evidence of an increase in assortativeness throughout the sample period, with the only exception being a slight dip for cohorts born between 1860 and 1880.

While this analysis shows a clear increase in the correlation between spouses' parens' log occupational incomes, the magnitudes appear quite small. We estimated a correlation coefficient of 0.04-0.09, which is much smaller than recent estimates of the correlation between spouses' educational attainment or parental wealth, which are closer to 0.6 and 0.4, respectively (Ghileb and Lang 2016; Charles et al 2016). However, we argue that this difference in magnitude can be explained by imperfect sorting on own human capital, imperfect transmission of parental human capital to children, and error in the measurement of economic status (see [forthcoming] appendix for details).

In Figure 8, we present correlations between spouses' parental incomes by region of residence. Within regions, the spousal income correlation is lower, and regional trends differ. Within the Northeast, it appears that assortative mating declined, while assortativeness increased within the South. These results suggest that regional income divergence may be important for explaining the overall patterns of assortative mating; we will explore this possibility in more detail below.

4.4 Spousal Age Differences

Figures 9 and 10 look at the age gap between spouses, and how it evolves over time, by socioeconomic status and region. In more traditional societies, women tend to marry older men. Older husbands may be more financially stable, but may also be of lower match quality The spousal age gap declines by almost a full year over the sample period, from an average of about 5 to slightly above 4. The age gap is correlated negatively with socioeconomic status, and this correlation becomes stronger over time. Similarly, there are large regional differences in the age gap – the age gap is largest in the South and smallest in the Northeast. These regional differences also become larger over time.

5 Accounting for Trends

In the previous section, we showed that a number of marriage outcomes – whether to marry, when to marry, and whom to marry – differ by parental socioeconomic status, especially for women. Moreover, we showed that socioeconomic status became more predictive of marriage market outcomes between the mid 19$^{\text{th}}$ and early 20$^{\text{th}}$ centuries. We also showed dramatic regional differences in marriage market outcomes, some of which grew over time.

What explains these patterns? We focus on the role of geography. 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. 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 of the distribution were increasingly concentrated in the South. This alone may have generated an increase in assortative mating at the national level, as well as an increasing socioeconomic gradient in the probability of marriage and age at marriage.

5.1 Empirical Approach

In this section, we use a regression model to assess the extent to which national trends in marriage are explained by regional income divergence. We also investigate the specific geographic characteristics that explain this divergence. We estimate the following three regression equations at the individual level, using a sample of women ages 30-45.

$$Y_{iqst} = \gamma_{qt} + \delta' Z_{iqst} + u_{iqst} \tag{1}$$

$$Y_{iqst} = \gamma_{qt} + \zeta_{st} + +\delta' Z_{iqst} + u_{iqst} \tag{2}$$

$$Y_{iqst} = \gamma_{qt} + \beta_t' X_{st} + +\delta' Z_{iqst} + u_{iqst} \tag{3}$$

Here, Y_{iqst} is a marriage outcome for individual i, of parental income quartile q, born in state s, observed at time t; γ_{qt} is a parental income quartile-by-year fixed effect; 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 is allowed to affect Y differently in different years; and Z_{iqst} is vector of individual-level controls, including age and foreign born status.

Equation 1 essentially reproduces our results from the previous section. For instance, suppose Y is an indicator for ever having married. If $\gamma_{4,1940} - \gamma_{1,1940}$ exceeds $\gamma_{4,1880} - \gamma_{1,1880}$ in magnitude, this indicates that the socioeconomic gradient in the probability of marriage has grown over time. This method has the advantage of allowing us to test whether the change in the socioeconomic gradient in marriage is statistically significant.

With equations 2 and 3, we are trying to understand what drives changes in γ_{qt}. In equation 2, the fixed effect ζ_{st} will capture level differences in Y across states, as well as changes in the concentration of parents from particular quartiles across states. If the inclusion of these fixed effects substantially affects our estimates of γ_{qt}, this suggests that regional income divergence can partly explain the trends that we document. In equation 3, we replace ζ_{st} with a vector of state-year characteristics, to understand precisely which geographic characteristics are responsible for our findings from equation 2.

5.2 Results

Results from equation 1 are shown in Figure 11. In the left panel, Y is an indicator equal to 1 if the woman has ever been married. In the right panel, Y is the log occupational income of the woman's husband. In both panels, we plot our estimates of γ_{qt} against t. The reference category is the bottom parental income quartile ($q = 1$). These results confirm our previous findings, which did not include individual level controls. There is a negative socioeconomic gradient in the probability of marriage, which appears to grow over time. And, the expected log occupational income of husbands of women from higher parental income quartiles is greater than that of women from lower parental income quartiles; this gap also seems to increase over time.

In Figure 12, we add birth state-by-year fixed effects to these regressions. We plot estimates of q_{4t} against t, with and without these state-year fixed effects. Again, the dependent variable in the left panel is an indicator equal to 1 if the woman has even ben married, and the dependent variable in the right panel is the woman's husband's log occupational income. In both cases, the inclusion of birth state-by-year fixed effects attenuates our estimates of γ_{4t}, indicating that some of the socioeconomic differences in marriage outcomes are driven by the fact that people from particular economic strata are concentrated in particular parts of the country, and marriage market outcomes vary across space. Moreover, the inclusion of state-year fixed effects attenuates the trend in socioeconomic differences in marriage outcomes over time. This suggests that these trends can be partly explained by the increasing concentration of people from high and low income quartiles in different parts of the country. Figure 12 suggests that regional income divergence can explain almost all of the nationwide increase in assortative mating; however, it only explains about half of the increase in the negative socioeconomic gradient in marriage for women. Thus, socioeconomic status does appear to have become more linked to marriage outcomes even within marriage markets.

We now explore which state-level characteristics explain geography's role in these marriage outcomes. Characteristics of interest include the population share in urban areas, access to railroads, male population share, share of employment in manufacturing, scholarization rate, and foreign born share. Table 1 presents the regressions of individual marriage outcomes on these state characteristics along with a set of cohort, age, and birth state fixed effects. These results confirm that these characteristics are correlated with marriage outcomes.

We repeat the regressions presented above, but we include controls for each state-level characteristic interacted with dummies for each cohort. Figure 13 shows the impact of a state's urbanization and rate of manufacturing employment on the probability of ever marrying for women. Figure 14 shows the impact of these same characteristics on women's husbands' log occupational income.

We find that the state urbanization rate explains almost as much of the socioeconomic gradient in marriage and spouse's log occupational income as a full set of birth state-by-year controls. Manufacturing appears to explain some, but not all, of the role of geography.

Results for other state characteristics are shown in Figures 15 and 16. The foreign born share explains a great deal of the growing socioeconomic gradient in marriage and the growth in assortative mating. Access to railroads explains some of these patterns, but not much. The sex ratio and scholarization rate explain very little, if any, of these patterns.

6 Discussion and Conclusion

We document a series of changes in the relationship between parental socioeconomic status and marriage for Americans born between the mid 19$^{\text{th}}$ and early 20$^{\text{th}}$ centuries. Most importantly, high status women became increasingly less likely to marry, relative to their lower status counterparts; and, women became increasingly likely to marry men from similar backgrounds to themselves. Much, but not all, of this pattern can be explained by the increased geographic segregation of men and women from particular parts of the income distribution, or regional income divergence.

Including controls for certain geographic characteristics explains away almost as much of the trend in the relationship between socioeconomic status and marriage as including a full set of birth state-year fixed effects. Specifically, changes in urbanization and employment in manufacturing can explain a lot of the growing socioeconomic gradient in marriage and spousal occupational income. This is intuitive. Urbanization and manufacturing employment are closely linked to industrialization; so, to the extent that differences in the rate of industrialization across states drive regional occupational income divergence, this is exactly what we should see. Moreover, urbanization and manufacturing may have affected marriage and assortative mating *within states* by changing local marriage market conditions.

This paper makes the broader point that marriage markets are typically local. Any study of national trends in marriage behavior or matching should discuss changes in the distribution of people across space as a potential factor contributing to these trends.

References

[1] 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.

[2] 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.

[3] 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.

[4] 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).

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

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

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

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

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

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

[11] 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.

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

[13] 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*.

[14] 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.

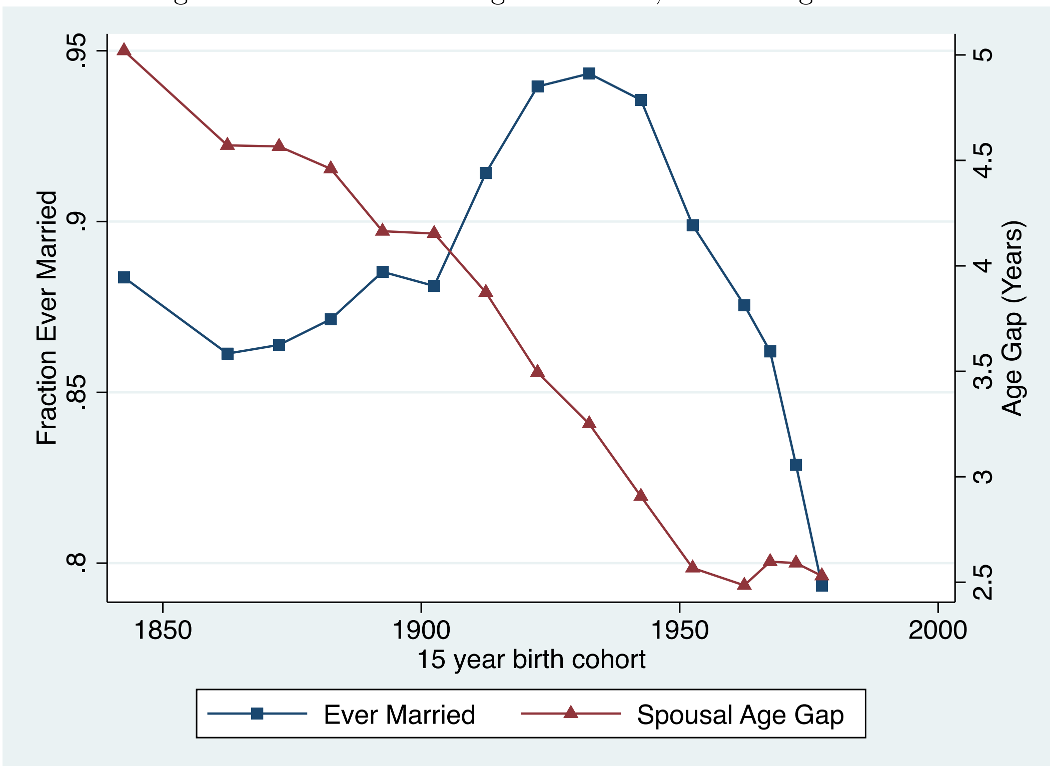
[15] Olivetti, Claudia; M. Daniele Paserman and Laura Salisbury (forthcoming). "Three-generation Mobility in the United States, 1850-1940: The Role of Maternal and Paternal Grandparents." *Explorations in Economic History*.

[16] 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.

[17] Saavedra, Martin Hugo and Twinam, Tate (2018). "A Machine Learning Approach to Improving Occupational Income Scores." Available at SSRN: https://ssrn.com/abstract=2944870 or http://dx.doi.org/10.2139/ssrn.2944870.

[18] 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.

Figure 1: Trends in Marriage Outcomes, Women Ages 30-45



Source: Authors' calculations based on the IPUMS 1% extracts from the Decennial Censuses, 1850-2000 and the 2005 and 2010 American Community Survey samples.

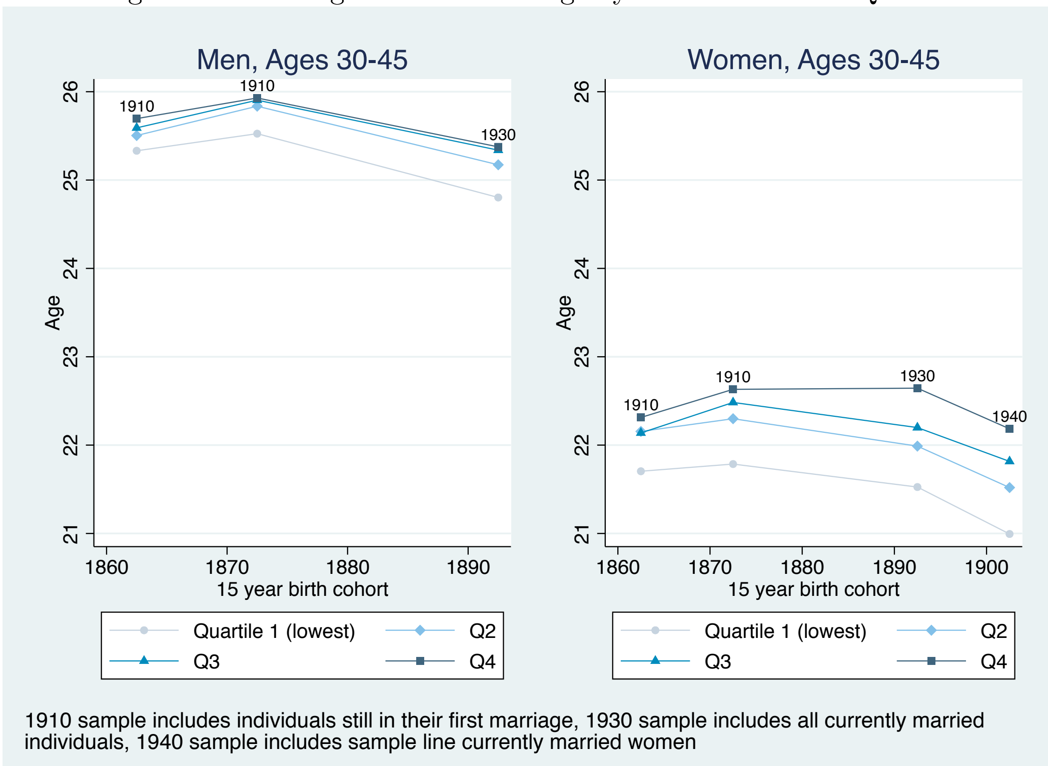
Figure 2: Fraction Ever Married by Parental Occupational Income Quartile

Note: The Figure shows the fraction ever married by birth cohort and parents' quartile of occupational income score. The occupational income score is imputed based on first names. Source: Authors' calculations based on the IPUMS 1% extracts from the Decennial Censuses, 1850-1940.

Figure 3: Fraction Ever Married by Region of Birth

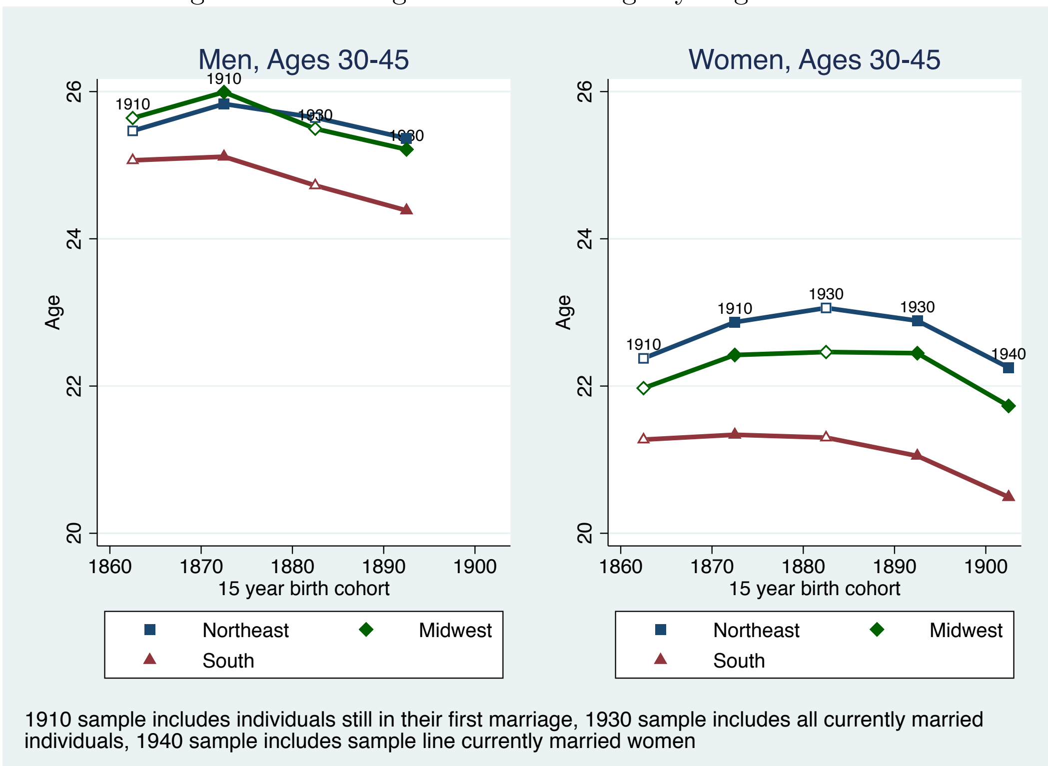
Note: Source: Authors' calculations based on the IPUMS 1% extracts from the Decennial Censuses, 1850-1940.

Figure 4: Mean Age at First Marriage by Parental Income Quartile



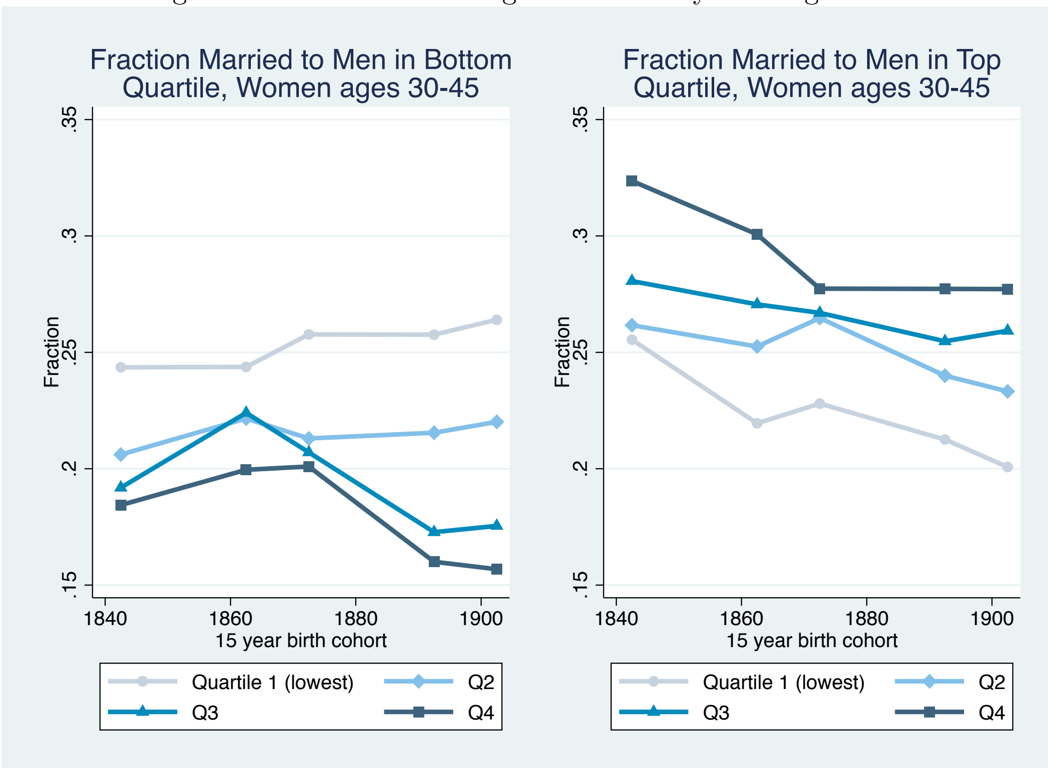
Note: The lines show the mean age at first marriage by socioeconomic status. The numbers above the top line indicate the Decennial Census that is used for the calculation. Note that the sample of indiviuals that are asked about age at first marriage, and the exact wording of the question change over time. Source: Authors' calculations based on the IPUMS 1% extracts from the Decennial Censuses, 1850-1940.

Figure 5: Mean Age at First Marriage by Region of Birth



Note: The lines show the mean age at first marriage by region. The numbers above the top line indicate the Decennial Census that is used for the calculation. Note that the sample of indiviuals that are asked about age at first marriage, and the exact wording of the question change over time. Source: Authors' calculations based on the IPUMS 1% extracts from the Decennial Censuses, 1850-1940.

Figure 6: Assortative Mating - Measured by Marriage Rates

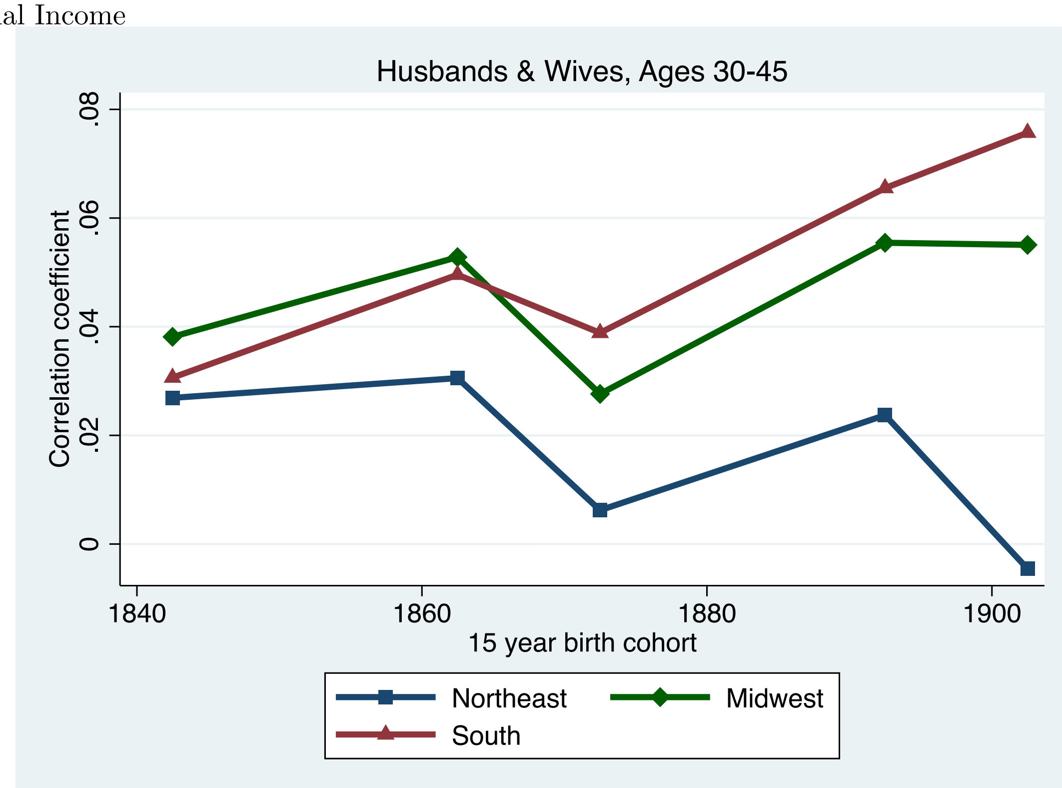


Note: The graph shows the percent of women in each quartile of the socioeconomic status distribution married to men in the bottom (left panel) and top (right panel) of the SES distribution. In each panel, we restrict the sample to couples in which both spouses fall between the specified age range. Source: Authors' calculations based on the IPUMS 1% extracts from the Decennial Censuses, 1850-1940.

Figure 7: Assortative Mating - Measured by Correlation in Parental Log Occupational Income

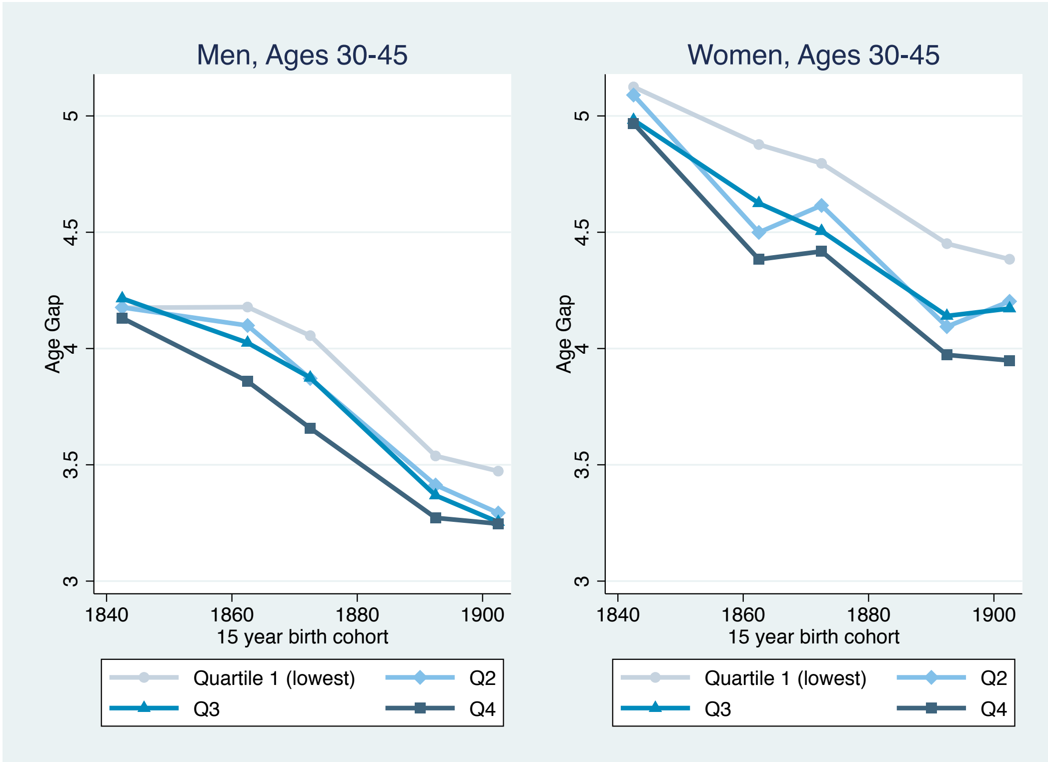
Note: The graph shows the correlation between the log occupational earnings of fathers of husbands and wives. In each panel, we restrict the sample to couples in which both spouses fall between the specified age range. Source: Authors' calculations based on the IPUMS 1% extracts from the Decennial Censuses, 1850-1940.

Figure 8: Assortative Mating by Region - Measured by Correlation in Parental Log Occupational Income



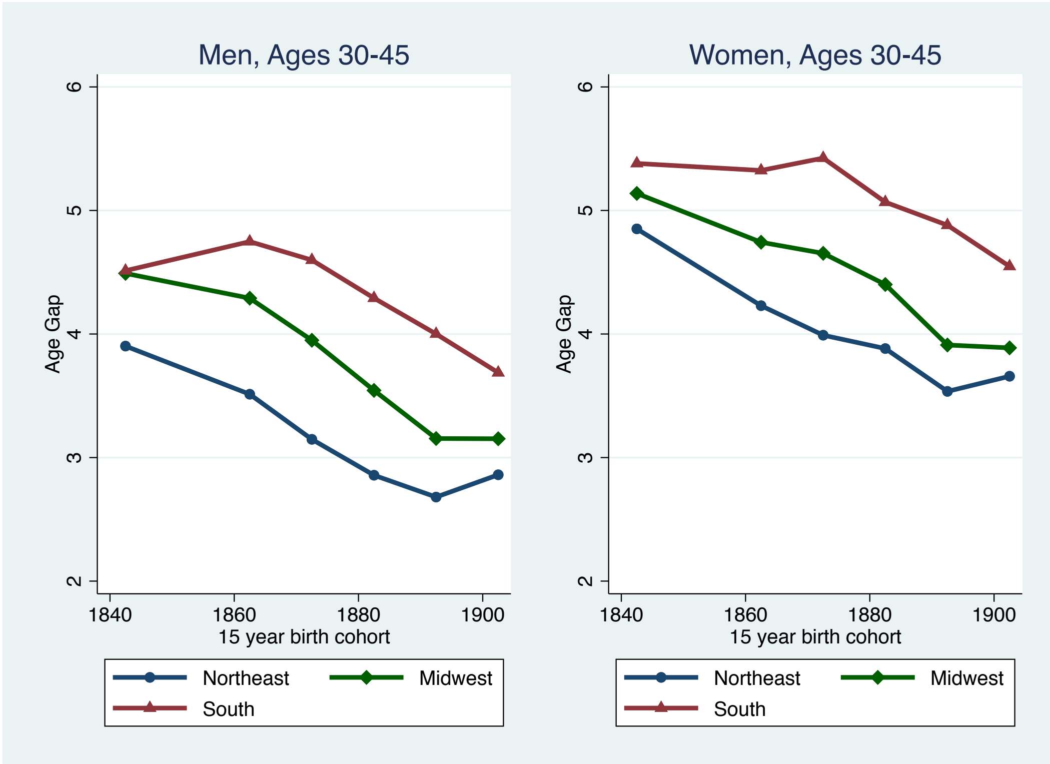
Note: The graph shows the correlation between the log occupational earnings of fathers of husbands and wives, within regions. We restrict the sample to couples in which both spouses fall between the specified age range. The region is defined by the wife's birth region. Source: Authors' calculations based on the IPUMS 1% extracts from the Decennial Censuses, 1850-1940.

Figure 9: Spousal Age Difference (Husband - Wife) by Parental Income Quartile



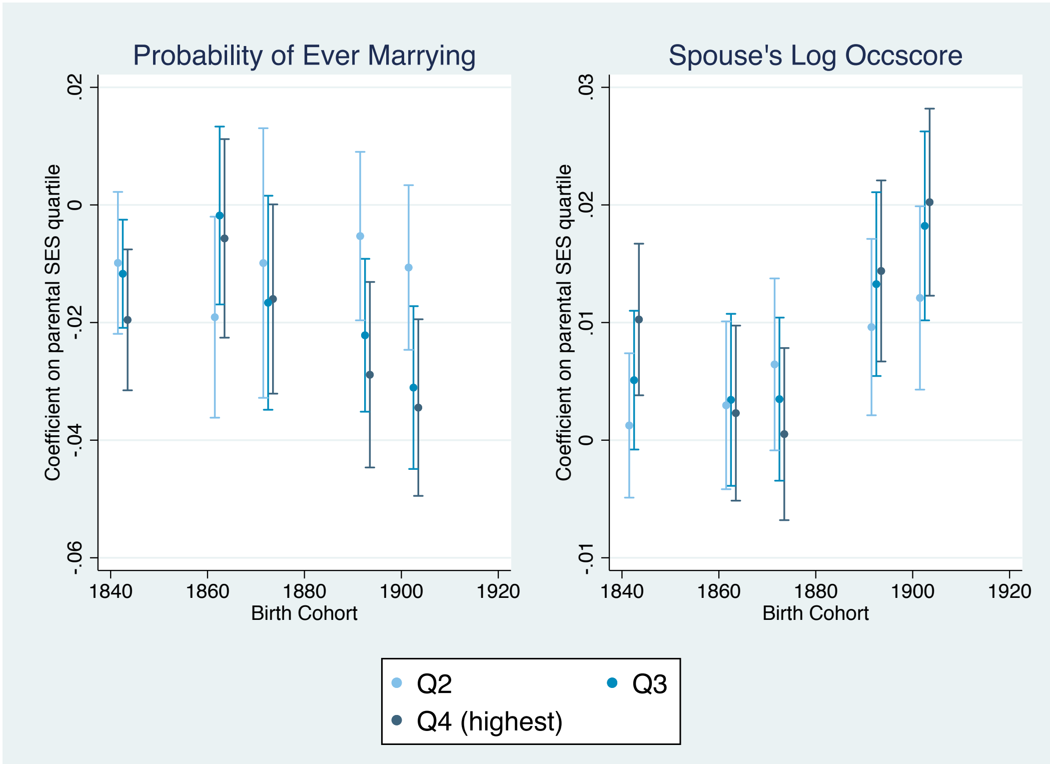
Note: The lines represent the mean age gap (husband's age minus wife's age) by socioeconomic status. Source: Authors' calculations based on the IPUMS 1% extracts from the Decennial Censuses, 1850-1940.

Figure 10: Spousal Age Difference (Husband - Wife) by Region



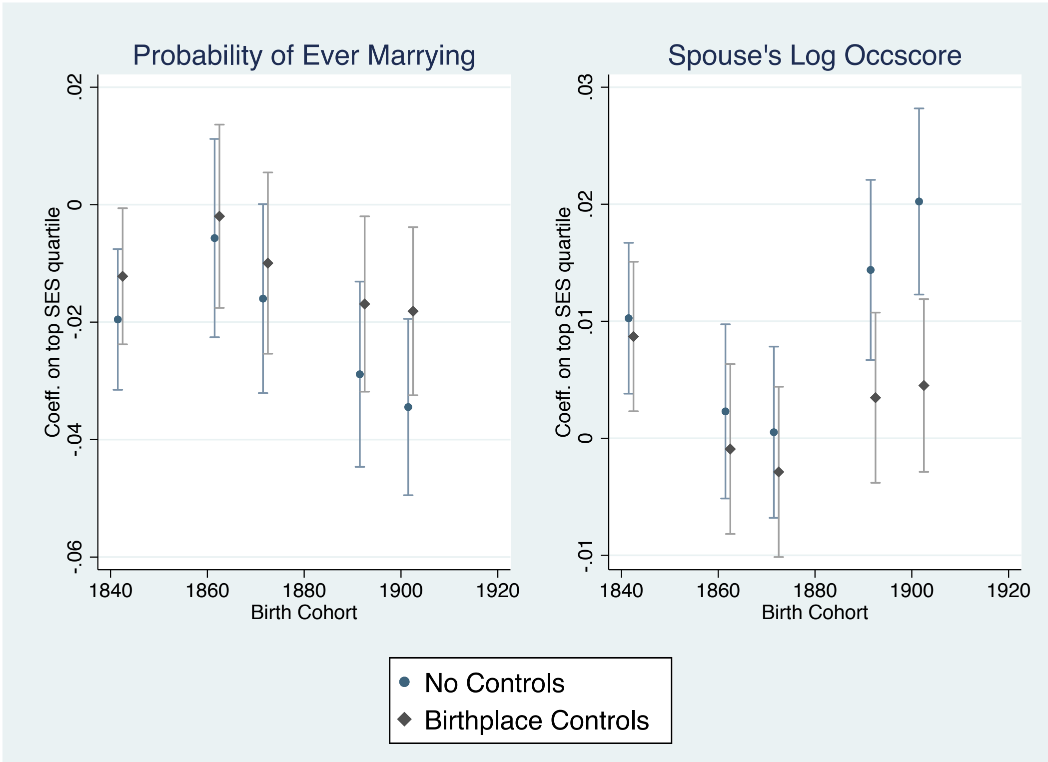
Note: The lines represent the mean age gap (husband's age minus wife's age) by region of birth. Source: Authors' calculations based on the IPUMS 1% extracts from the Decennial Censuses, 1850-1940.

Figure 11: Effect of Parental SES on Probability of Ever Marrying and Spouse's Log Occupational Income, Women 30-45



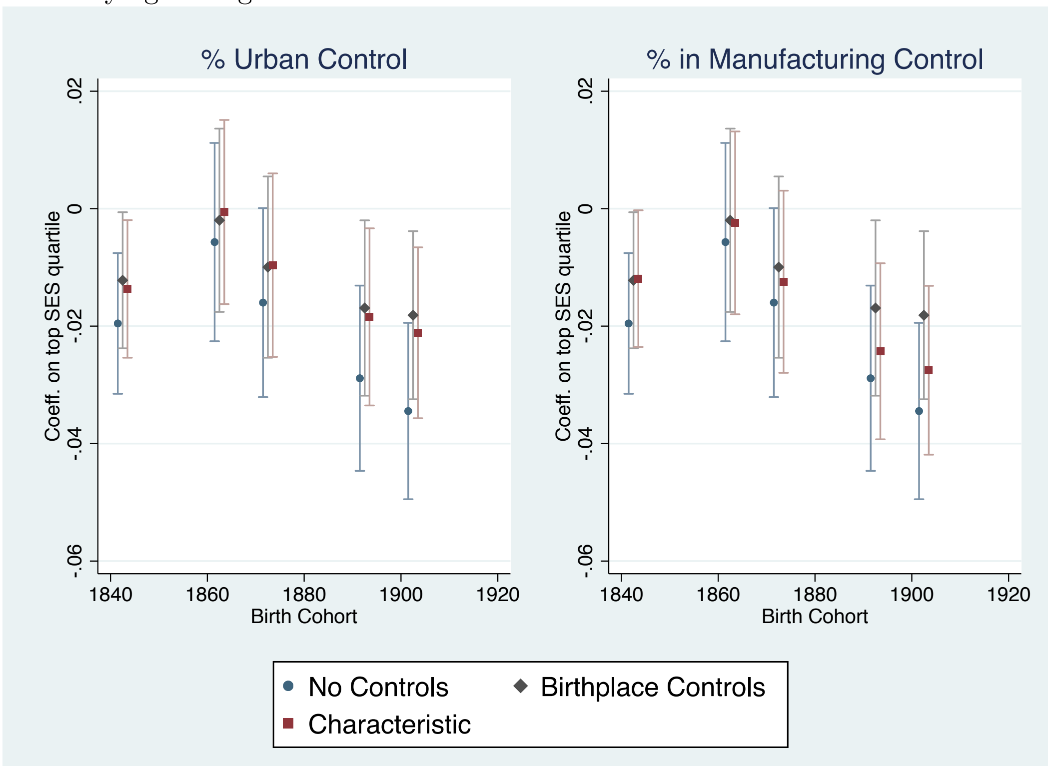
Note: 95% confidence intervals are shown. Plotted coefficients represent the difference from Q1, the omitted category. Source: Authors' calculations based on the IPUMS 1% extracts from the Decennial Censuses, 1850-1940.

Figure 12: Effect of Parental SES and Birthplace on Probability of Ever Marrying and Spouse's Log Occupational Score, Women 30-45



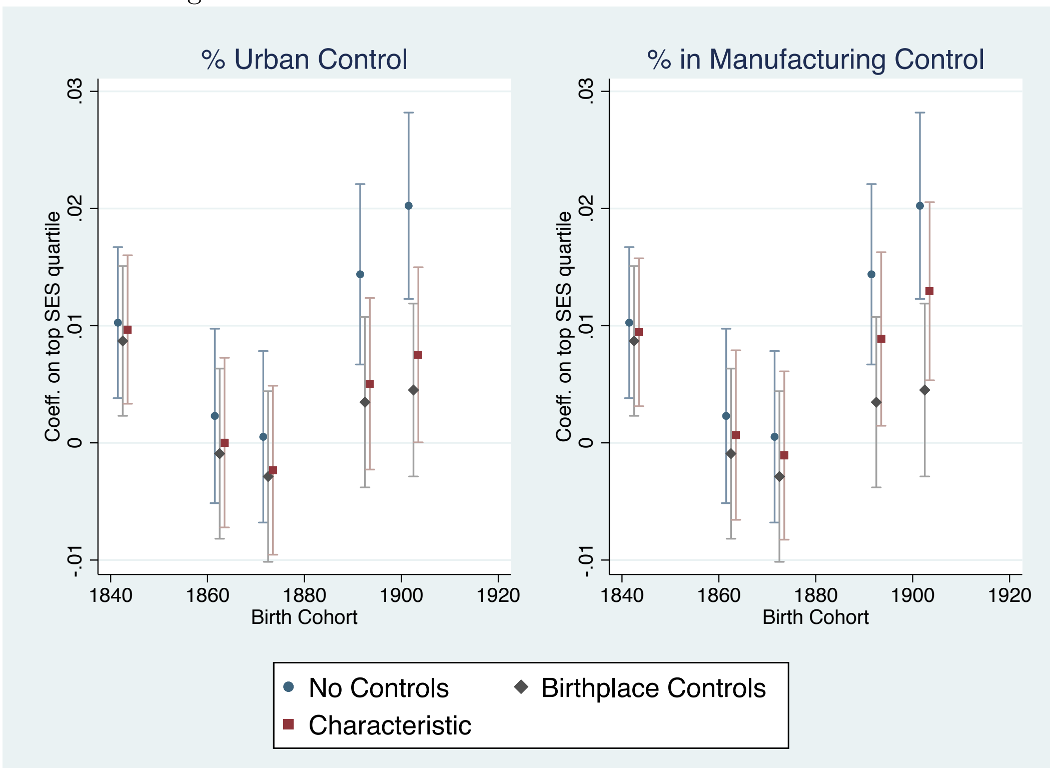
Note: 95% confidence intervals are shown. Plotted coefficients represent the coefficient on the indicator for the top SES quartile, as compared to the bottom SES quartile (the omitted SES category). The Birthplace Controls specification includes a set of dummies for the woman's state of birth interacted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

Figure 13: Effect of Parental SES, Birthplace, Urbanization, and Manufacturing on Probability of Ever Marrying among Women 30-45



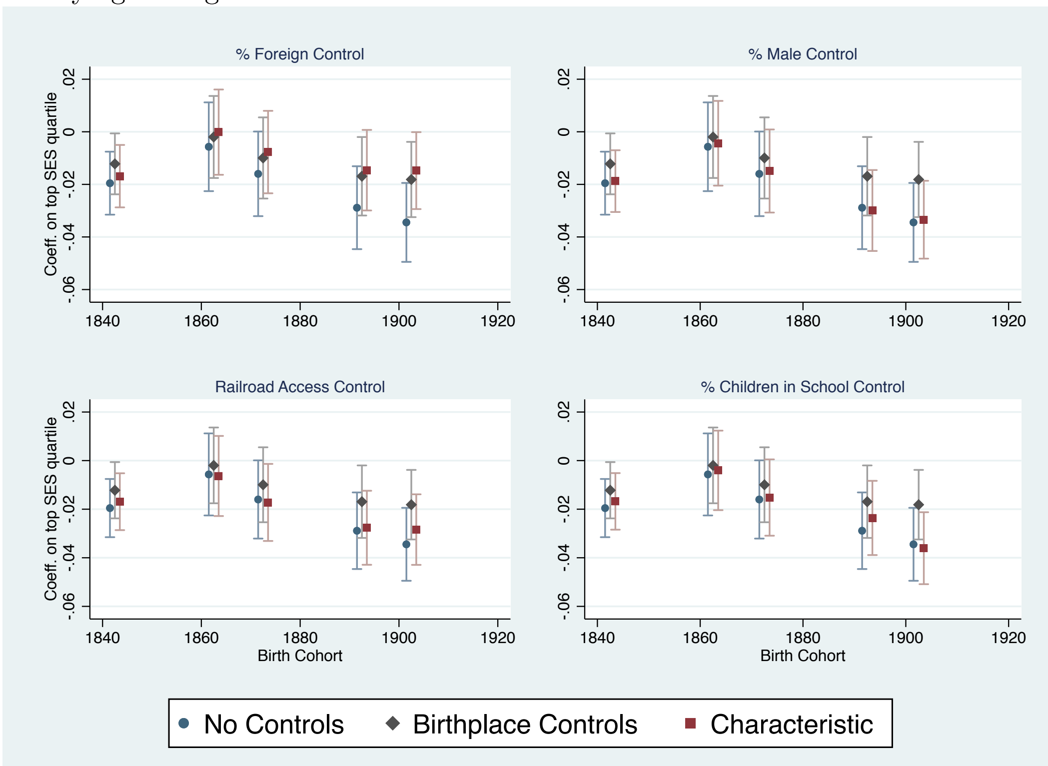
Note: 95% confidence intervals are shown. Plotted coefficients represent the coefficient on the indicator for the top SES quartile, as compared to the bottom SES quartile (the omitted SES category). The Birthplace Controls specification includes a set of dummies for the woman's state of birth interacted with cohort dummies. The % Urban Control specification includes a control for the percent of the woman's birth state population in urban areas 10 years prior to the year of observation, interacted with cohort dummies. The % in Manufacturing Control specification includes a control for the percent of the labor force employed in manufacturing in the woman's birth state 10 years prior to the year of observation, interacted with cohort dummies. Source: Authors' calculations based on the IPUMS 1% extracts from the Decennial Censuses, 1850-1940; IPUMS 100% population databases from the Decennial Censuses, 1850, 1880, and 1900-1940; and ICPSR summary census data, 1850-1940.

Figure 14: Effect of Parental SES, Birthplace, Urbanization, and Manufacturing on Spouse's Log Occscore among Women 30-45



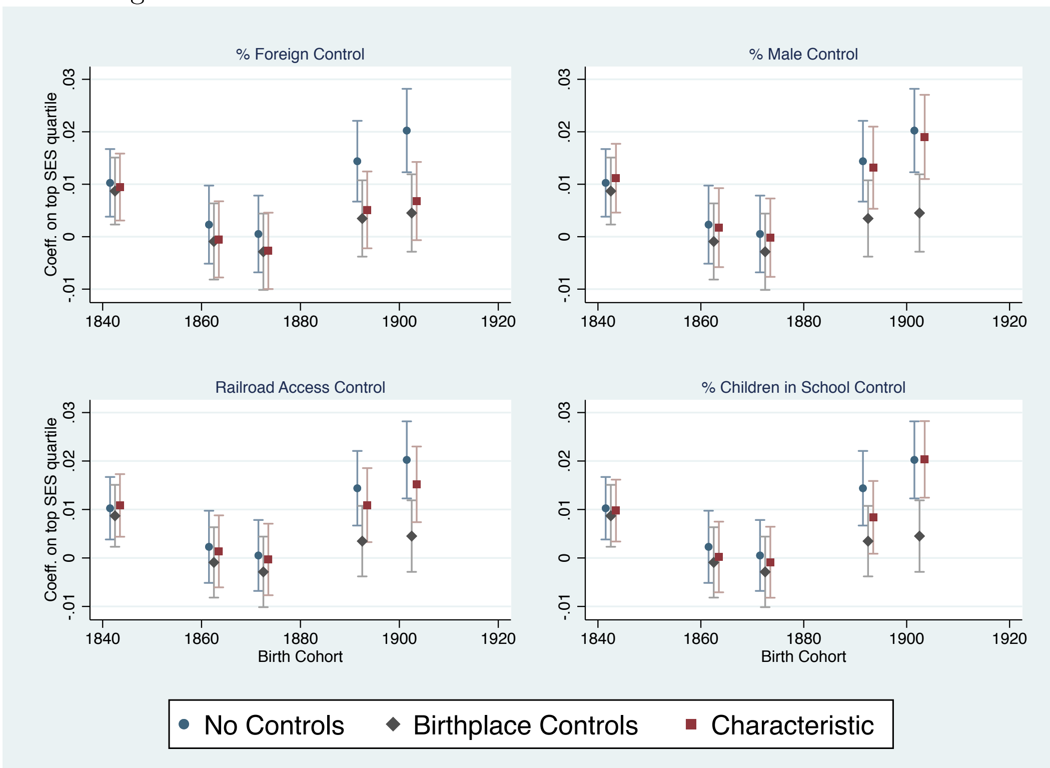
Note: 95% confidence intervals are shown. Plotted coefficients represent the coefficient on the indicator for the top SES quartile, as compared to the bottom SES quartile (the omitted SES category). The Birthplace Controls specification includes a set of dummies for the woman's state of birth interacted with cohort dummies. The % Urban Control specification includes a control for the percent of the woman's birth state population in urban areas 10 years prior to the year of observation, interacted with cohort dummies. The % in Manufacturing Control specification includes a control for the percent of the labor force employed in manufacturing in the woman's birth state 10 years prior to the year of observation, interacted with cohort dummies. Source: Authors' calculations based on the IPUMS 1% extracts from the Decennial Censuses, 1850-1940; IPUMS 100% population databases from the Decennial Censuses, 1850, 1880, and 1900-1940; and ICPSR summary census data, 1850-1940.

Figure 15: Effect of Parental SES, Birthplace, and State Characteristics on the Probability of Ever Marrying among 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 IPUMS 1% extracts from the Decennial Censuses, 1850-1940; IPUMS 100% population databases from the Decennial Censuses, 1850, 1880, and 1900-1940; and ICPSR summary census data, 1850-1940.

Figure 16: Effect of Parental SES, Birthplace, and State Characteristics on Spouse's Log Occscore among 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 IPUMS 1% extracts from the Decennial Censuses, 1850-1940; IPUMS 100% population databases from the Decennial Censuses, 1850, 1880, and 1900-1940; and ICPSR summary census data, 1850-1940.

Table 1: State Characteristics and Marriage Outcomes

	Ever Married	Husband's Father's Income
Father's Income	-0.0791***	0.0306***
	(0.00562)	(0.00405)
Percent Urban	-0.0679***	0.0237***
	(0.0219)	(0.00840)
Land w/in 15 mi of RR	-0.0177**	0.00834**
	(0.00725)	(0.00386)
Pct Male Among Whites 15-30	1.055***	0.0688
	(0.139)	(0.0572)
Pct Lab Force in Manufacturing	0.0752	-0.0931*
	(0.135)	(0.0476)
Pct of Whites 5-20 in School	0.0240	-0.0286***
	(0.0200)	(0.00985)
Pct Foreign Among Whites	-0.0117	0.0703***
	(0.0526)	(0.0130)
Gini of Farm Size	0.0430**	0.00857
	(0.0191)	(0.00843)
Observations	330731	164188

All regressions control for a full set of state of birth dummies and time dummies.
Standard errors clustered by birth state in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$