

Online Appendix
Marital Matching and Women's Intergenerational Mobility
in the Late-19th and Early-20th-Century U.S.

By Martha J. Bailey^{1,2} and Peter Z. Lin³

June 14, 2024

A. Inverse Propensity Score Weighting for Linked Sample

We generate an inverse propensity score weight for each observation in the LIFE-M sample using the method proposed by Bailey et al. (2020). We first group women in the LIFE-M sample by their birth year and the census year when they and their husbands are observed. The women in each group are then weighted to be a representative sample of the women born in the same year in the relevant census population. For example, we weight the Ohio-born women who were born in the 1910 birth cohort and linked to the 1940 Census to make them a representative sample of all married Ohio-born women born in the 1910 cohort in the 1940 Census. Considering that the analytic samples for different dimensions of homogamy vary significantly, we customize the weights for each sample.

We take the following steps to weight the linked women in each group:

Step 1. We extract Ohio-born women in the birth cohort j from the full-count decennial census t . We apply the same restrictions including (1) age 20-60 in the census; (2) married and co-resided with her husband in the same household.

Step 2. We combine the extracted women from the full-count census with the linked women in the LIFE-M sample.

Step 3. We estimate the following probit model using the pooled LIFE-M census dataset,

$$Pr(\text{sample}_{ijt} = 1 | X_{ijt}) = \Phi(X_{ijt}\beta)$$

where sample_{ijt} is a dummy variable equal to 1 if the observation comes from the linked sample, and equal to 0 if it comes from the census; $\Phi(\cdot)$ is the cumulative distribution function for the standard normal distribution; X_{ijt} is a vector of covariates, including an indicator for specific race (1=White, 0=Non-white), urban residence, farm residence, migration away from birth state, having an occupation associated with positive income score, living with father, living with mother, living with children under age 5, father's nativity (only for the women in the 1900, 1910, and 1920 Census), husband's nativity, husband's occupational income score, and women's name characteristics (length of first name, last name, commonality score of these names). We also include two-way interactions between these covariates.

Step 4. We predict the propensity score of being linked and calculate the inverse propensity score weight as follows:

$$wgt = \left(\frac{1-p}{p}\right) \times \left(\frac{q}{1-q}\right)$$

where p is the predicted propensity score of linkage, and q is the ratio of observations in the LIFE-M sample to the observations in the referenced women population.

Step 5. We repeat this procedure for all other combinations of women's birth cohorts and census years. We also conduct the procedures repeatedly for different marriage homogamy outcomes.

Step 6. As the last step, we augment the weights for sample women in each group to make the weights sum to the referenced women population in the birth cohort and census year. We use this augmented weight to derive descriptive statistics of the overall sample.

We find that the LIFE-M sample is not a representative sample of the analogous population (see table 2, panel A, columns 3, 7 and panel B, columns 3, 8). We find that the LIFE-M samples are over-represented by women who are white, living on farms, and living with children under age 5. Meanwhile, the LIFE-M samples are under-represented by women living in urban areas, born to a foreign-born father, and reporting an occupation associated with a positive occupational income score.

We also verify that the weighted LIFE-M sample is balanced in terms of the observable covariates by woman's birth cohort. To demonstrate this, we estimate the following model with the LIFE-M sample (weighted and unweighted), as well as the referenced population in censuses:

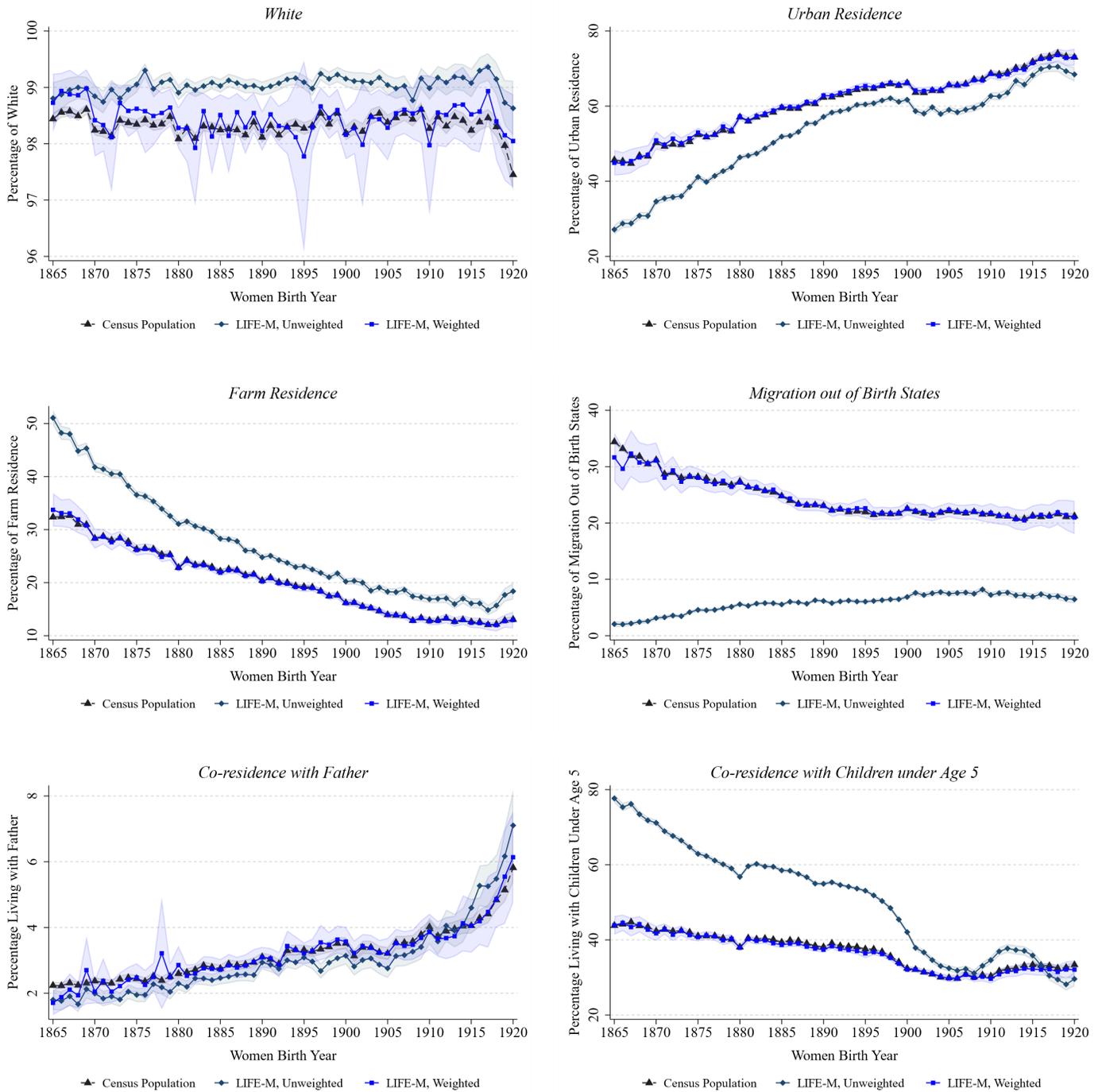
$$Y_{it} = \sum_j \beta_j Cohort_j + q(age_{it}) + \varepsilon_{it}$$

where Y_{it} is the covariate of interest, $Cohort_j$ is a dummy variable for women in cohort j , and $q(age_{it})$ is a quartic function of woman's age in the observed census. After estimating this model, we predict the mean covariate for each cohort at the age of 35. Figures A.1-A.4 plot the predicted mean covariates by woman's birth cohort for women in the LIFE-M samples for age, nativity, and occupation homogamy. For the education homogamy sample, we estimate cohort-specific mean covariates by the following model:

$$Y_{it} = \sum_j \beta_j Cohort_j + \varepsilon_{it}$$

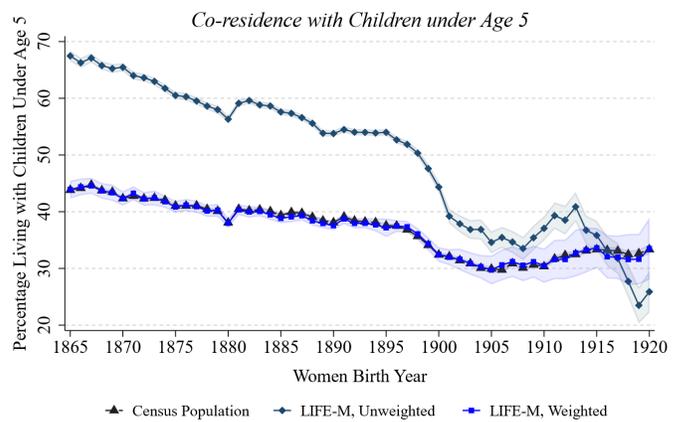
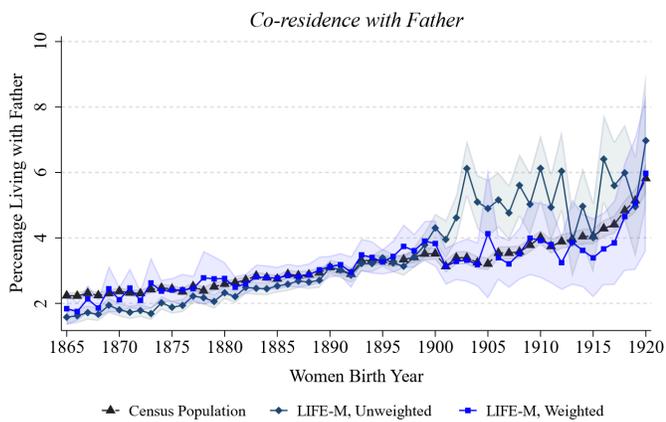
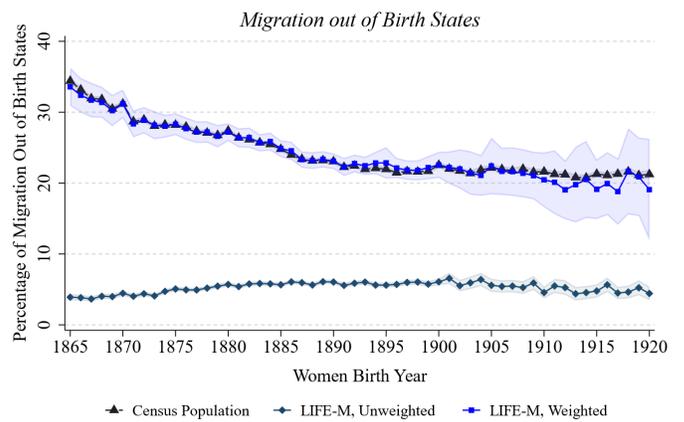
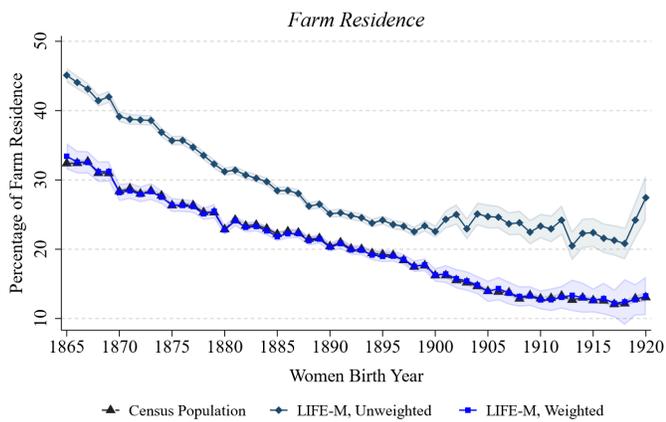
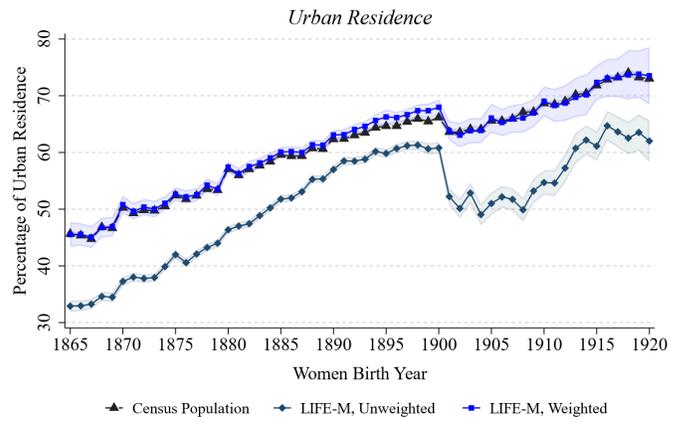
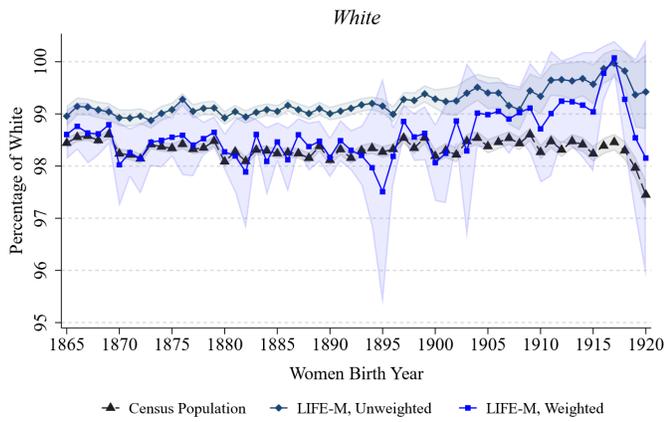
We cannot control for women's age for the education sample because women in this sample were only observed in the 1940 Census, and no additional age variation is available in a single census. For almost all covariates, the LIFE-M samples are not balanced without weighting. Instead, we make the LIFE-M samples indifferent or at least insignificantly different from the referenced population by inverse propensity weights.

Figure A.1. Balance, Age Sample



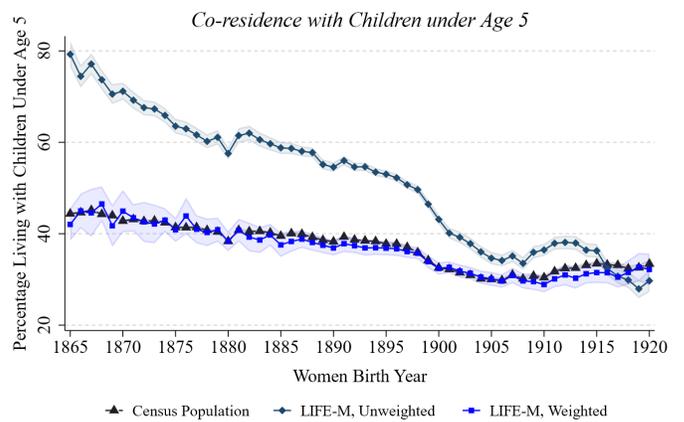
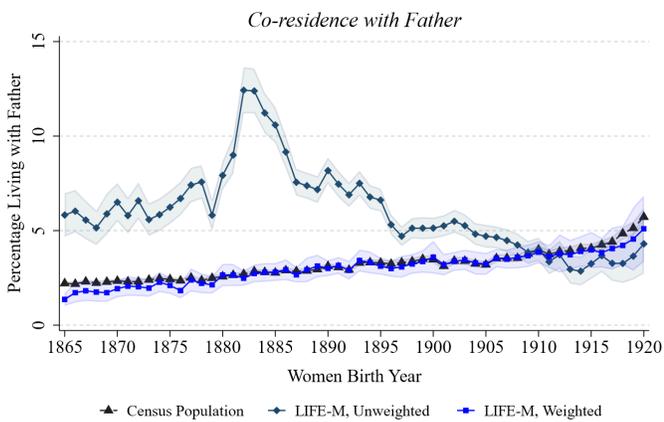
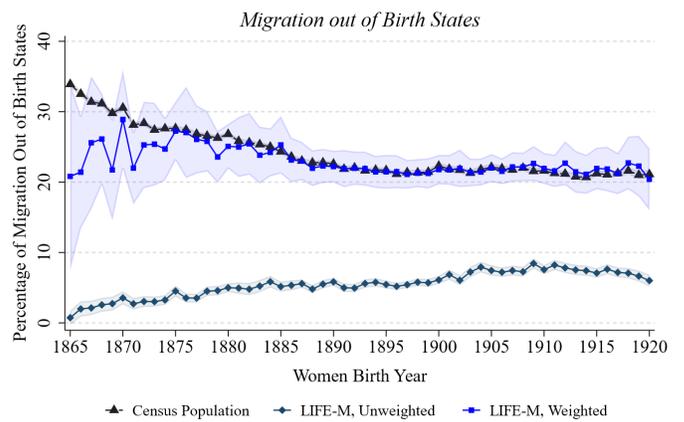
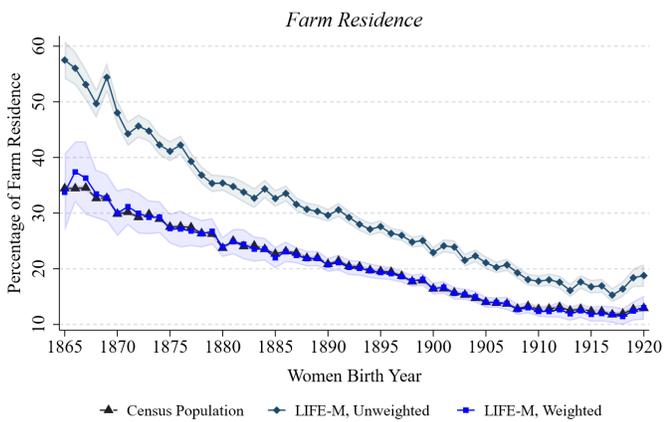
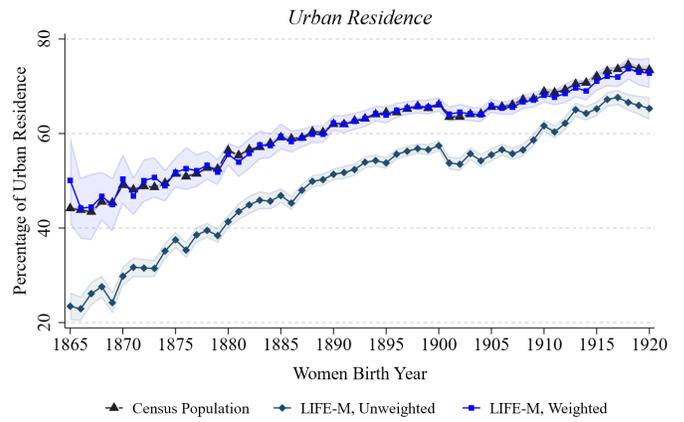
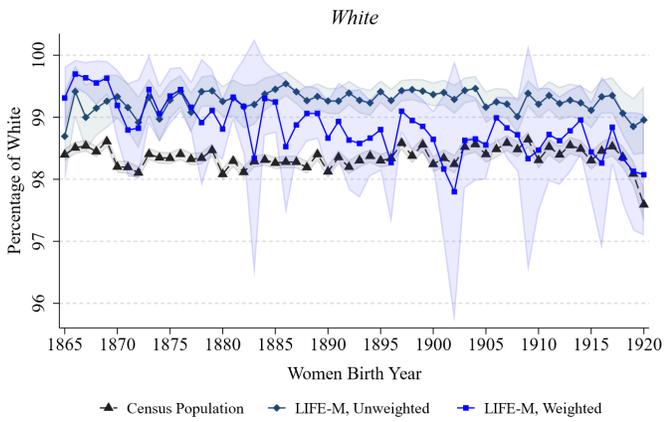
Notes: The figures depict the estimated mean of six covariates indicated in the panel header by wife's birth cohort. These series are age adjusted and projected at age 35 using equation (1). The black curve plots the estimates from corresponding target population in the census, the navy curve plots the unweighted LIFE-M estimates, and the blue curve plots the inverse-propensity score reweighted LIFE-M estimates.

Figure A.2. Balance, Nativity Sample



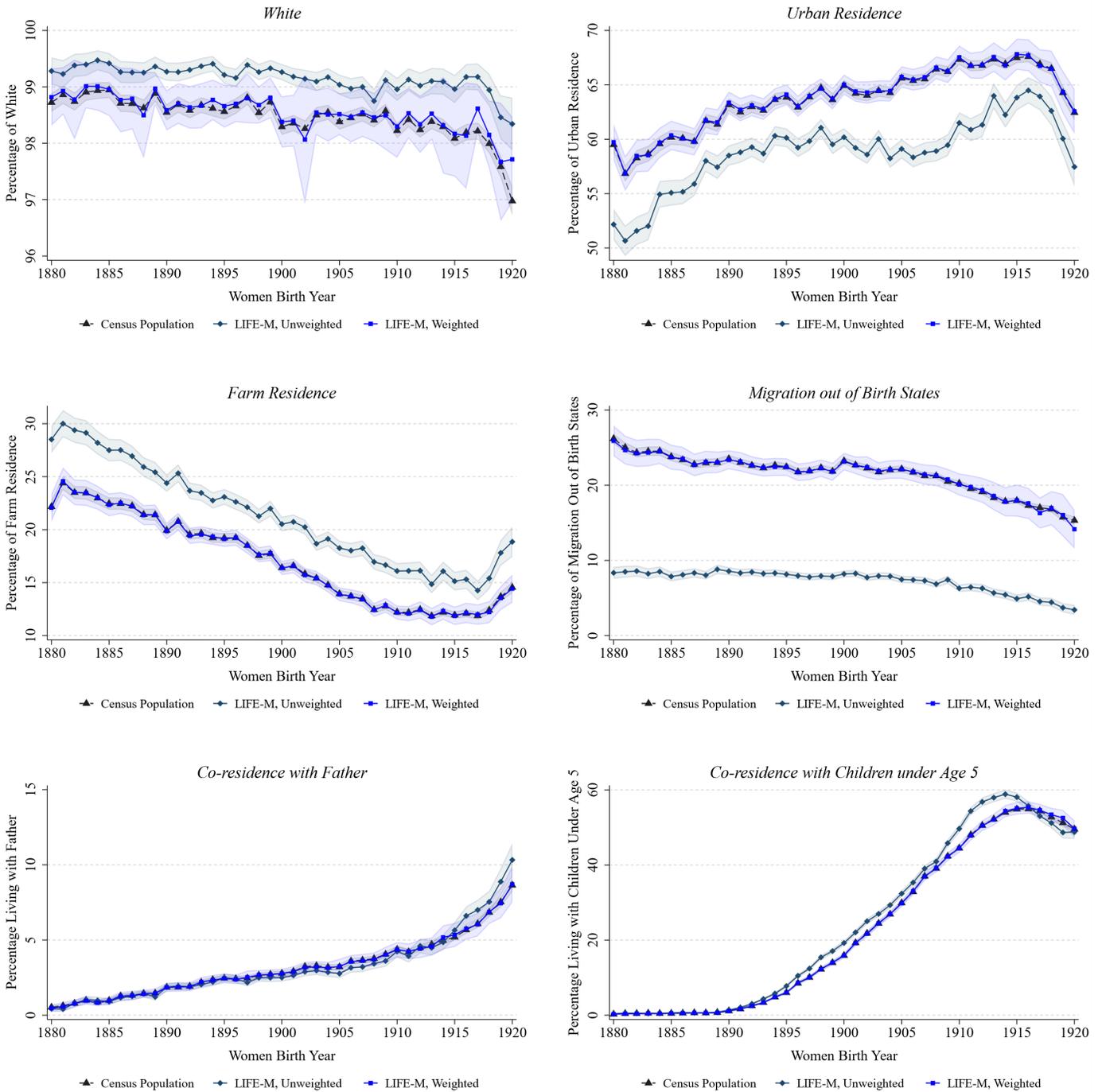
Notes: See Figure A.1 notes

Figure A.3. Balance, Occupational Sample



Notes: See Figure A.1 notes

Figure A.4. Balance, Education Sample

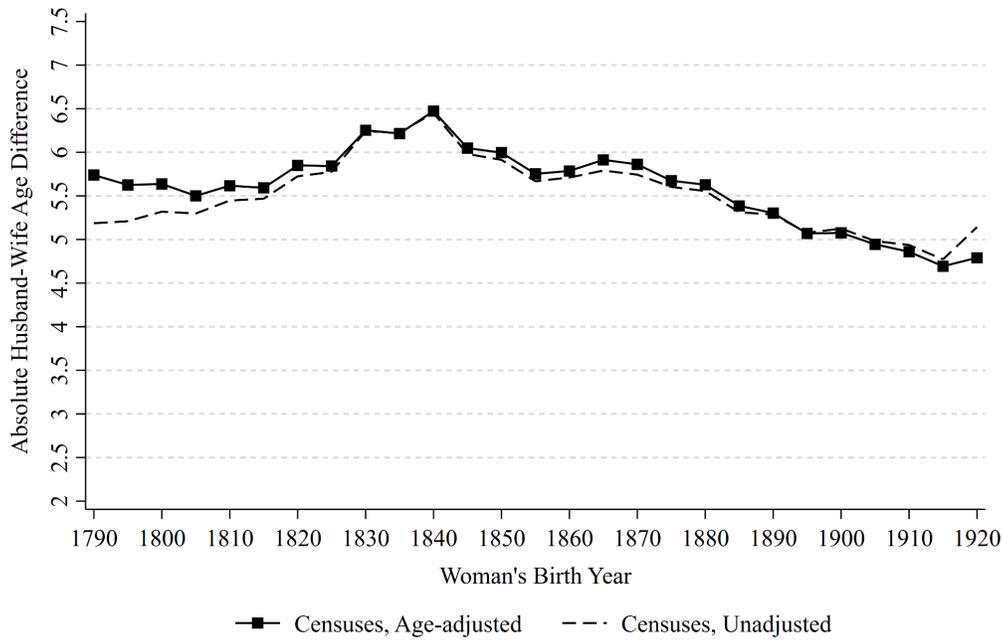


Notes: The figures depict the estimated mean of six covariates indicated in the panel header by wife's birth cohort for the education sample. The black curve plots the estimates from corresponding target population in the census, the navy curve plots the unweighted LIFE-M estimates, and the blue curve plots the inverse-propensity score reweighted LIFE-M estimates.

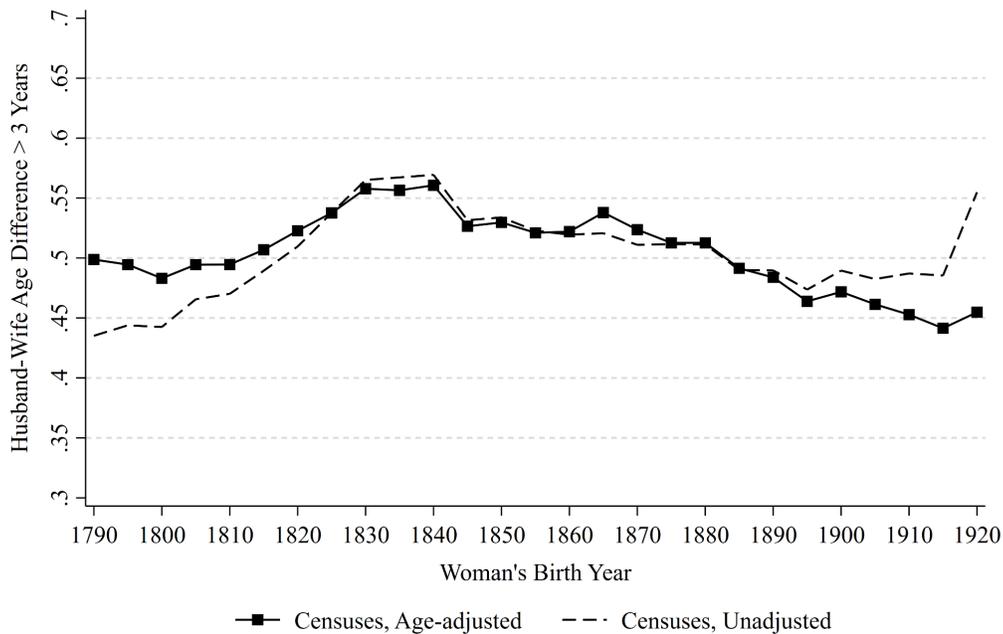
B. Alternative Measure of Age Homogamy and Heterogenous Trends

Figure B.1. Age Homogamy (Alternative Measures) by Wife's Birth Cohort

A. Absolute Husband-Wife Age Difference



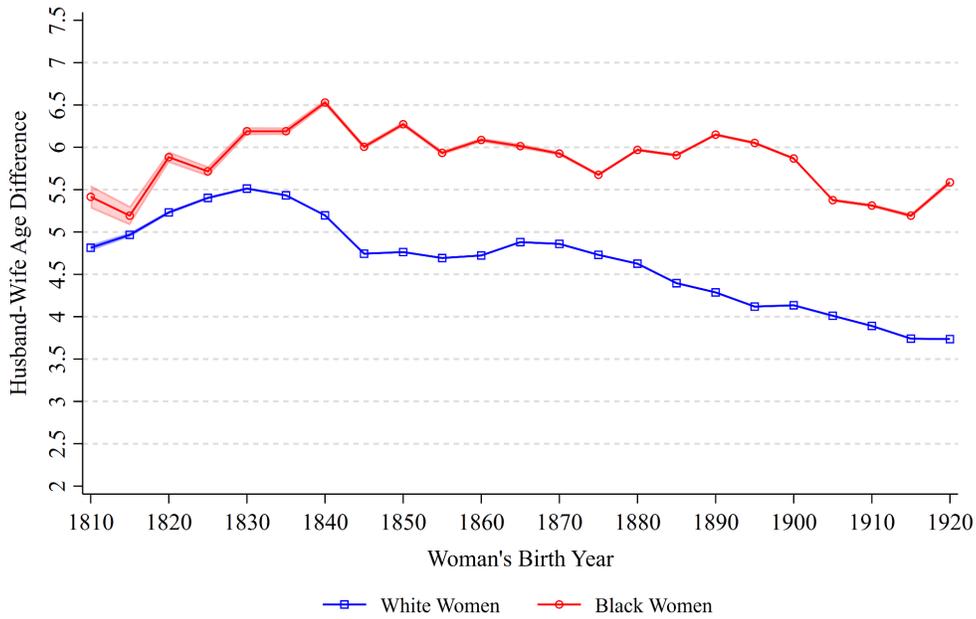
B. Prob(Husband-Wife Age Difference > 3 Years)



Notes: The figures depict alternative measures of age homogamy by a woman's birth cohort: (1) the absolute differences between a husband's and wife's ages; (2) the likelihood that a husband is more than 3 years older than his wife.

Sources: 1850-1940 Censuses (Ruggles et al., 2021).

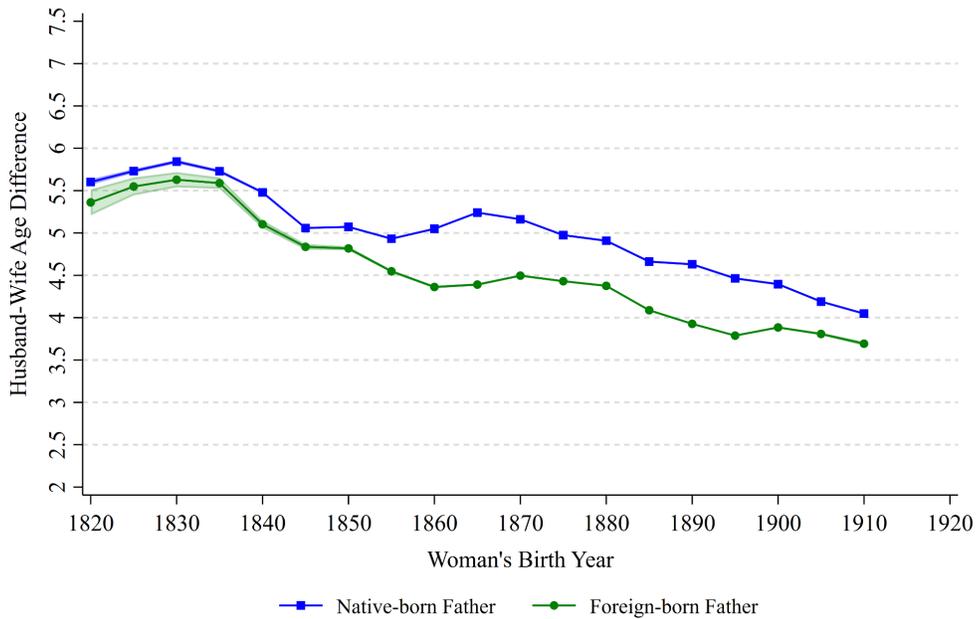
Figure B.2. Age Homogamy (Alternative Measures) by Wife's Birth Cohort and Race



Notes: This figure depicts differential trends of age homogamy by woman's race. We focus on women born after 1810 because Black women were first enumerated in the 1870 Census following the Civil War. We estimate the mean age difference separately for white women (blue line) and Black women (red line).

Sources: 1870-1940 Censuses (Ruggles et al., 2021)

Figure B.3. Age Homogamy (Alternative Measures) by Wife's Birth Cohort and Father's Nativity



Notes: This figure depicts differential trends of age homogamy by a woman's father's nativity (native-born vs. foreign-born). We use the 1880-1930 Censuses (for women born between 1820 and 1910) because they required reporting father's birthplace.

Sources: 1880-1930 Censuses (Ruggles et al., 2021)

C. Measure of Nativity Homogamy and Heterogenous Trends

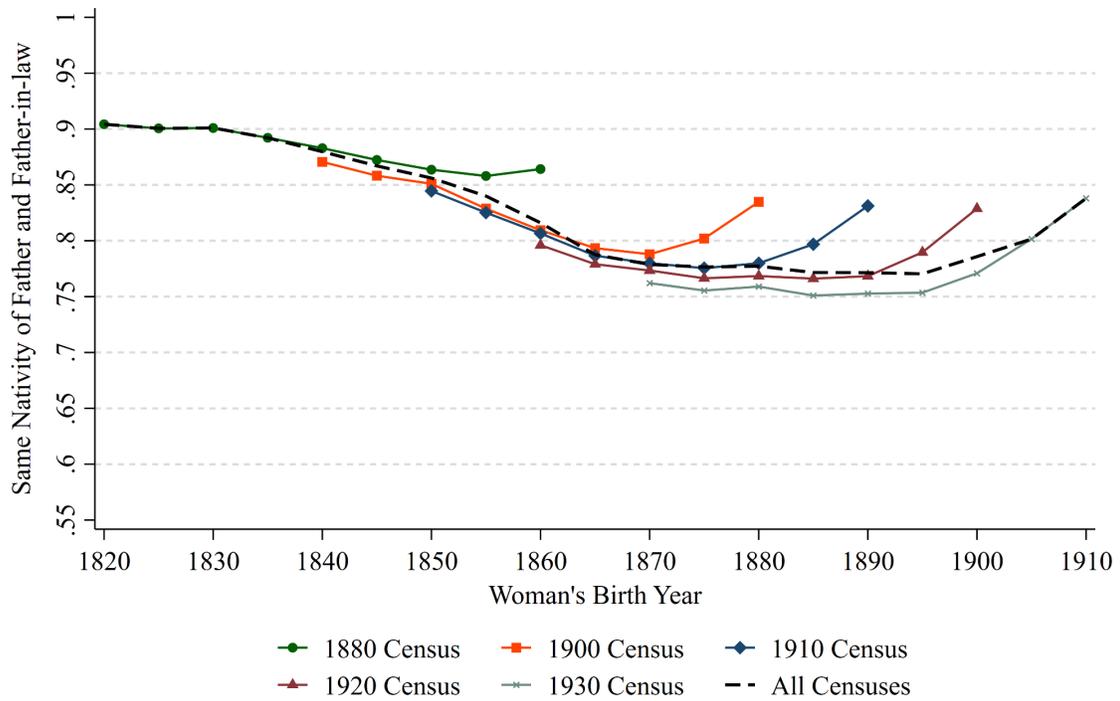
Our primary measure of nativity homogamy for a couple is based on the birth country of the wife's father and her husband's father. Considering the frequent border changes in the early twentieth century (especially for Central and Eastern Europe), we categorize the birth countries (coded by IPUMS) closely connected in terms of geography and culture into multiple groups. Specifically, the groups are:

- (1) All states in the U.S
- (2) U.S. outlying areas: American Samoa, Guam, Puerto Rico, and U.S. Virgin Islands.
- (3) Canada and other North America: St. Pierre and Miquelon and Atlantic Islands.
- (4) Central America: Mexico, Central America, Cuba, and West Indies.
- (5) South America: all South American countries
- (6) North Europe: Denmark, Finland, Iceland, Norway, Sweden.
- (7) Ireland and the United Kingdom: England, Scotland, Wales.
- (8) West Europe: Belgium, France, Liechtenstein, Monaco, Netherlands, Switzerland.
- (9) South Europe: Albania, Andorra, Gibraltar, Greece, Italy, Malta, Portugal, San Marino, Spain, Vatican City.
- (10) Central and Eastern Europe: Austria, Bulgaria, Czechoslovakia, Germany, Hungary, Poland, Romania, Yugoslavia.
- (11) Russian Empire: Estonia, Latvia, Lithuania, and other USSR/Russia territories.
- (12) China
- (13) Japan
- (14) Korea
- (15) Southeast Asia: Cambodia (Kampuchea), Indonesia, Laos, Malaysia, Philippines, Singapore, Thailand, Vietnam.
- (16) Southwest Asia: Afghanistan, India, Iran, Maldives, Nepal.
- (17) Middle East Asia: Bahrain, Cyprus, Iraq, Saudi Arabia, Israel/Palestine, Jordan, Kuwait, Lebanon, Oman, Qatar, Syria, Turkey, and other non-specified Middle East countries.
- (18) Africa
- (19) Oceania: Australia, New Zealand, Pacific Islands.

A woman's father and father-in-law are considered "same nativity" if their birth countries are both included in the same country group as defined above. We treat the father and father-in-law as the same nativity if one (either the father/father-in-law) reports a non-specified European/Asian country and the other reports a specific European/Asian country.

As an alternative measure of nativity homogamy, we also estimate the probability of a couple for which the husband's (rather than father-in-law's) nativity is the same as that of the woman's father. Figure C.2 plots this alternative measure.

Figure C.1. Probability of Same-Nativity of Father and Father-in-Law, by Wife's Birth Cohort and Census

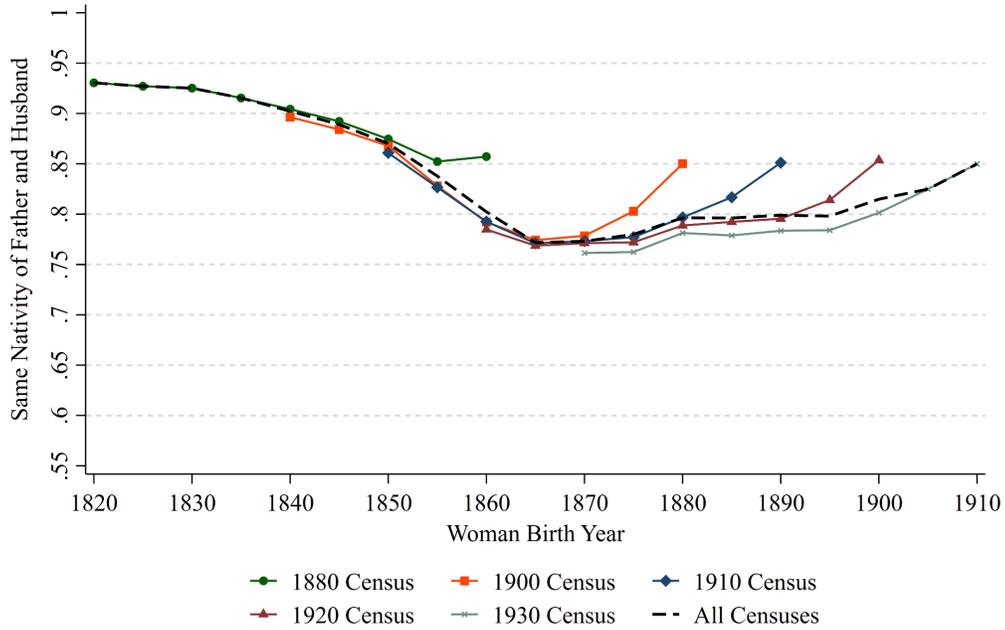


Notes: This figure depicts the likelihood of same-nativity for father and father-in-law by woman's birth cohort in individual censuses between 1880 and 1930.

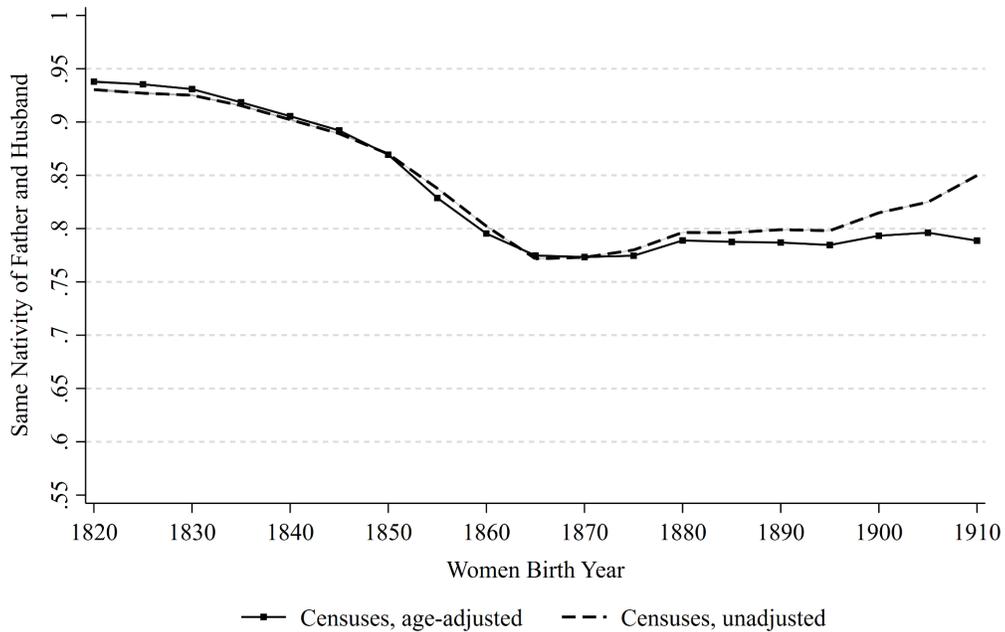
Sources: 1880-1930 Censuses (Ruggles et al., 2021)

Figure C.2. Probability of Same-Nativity of Father and Husband, by Wife's Birth Cohort and Census

A. All U.S. Individual Censuses: 1880-1930



B. All U.S. Combined Censuses: 1880-1930

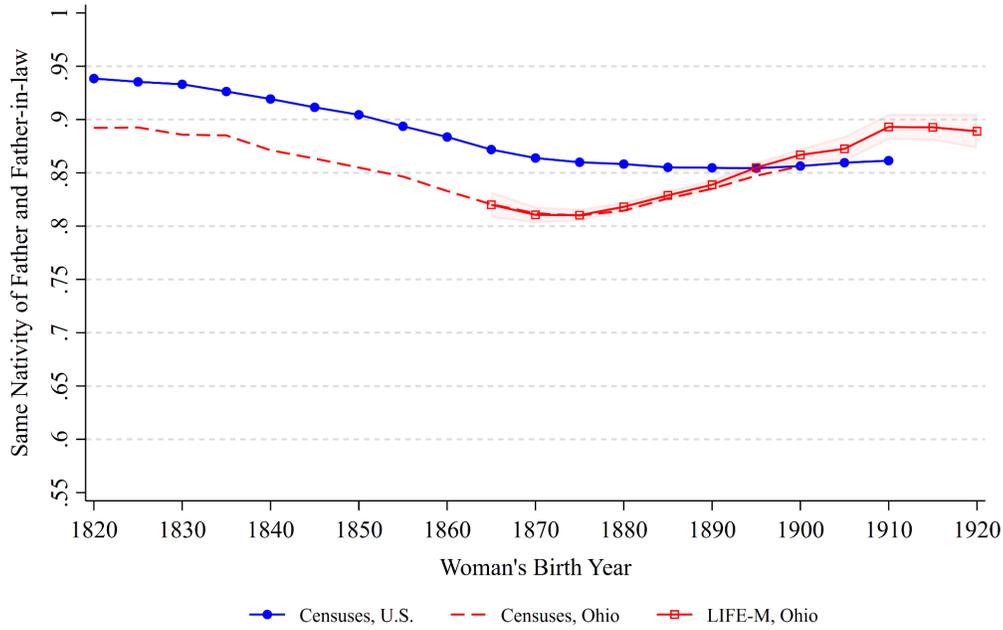


Notes: The figures depict an alternative measure of nativity homogamy: the likelihood of same nativity of a woman's father and her husband. In panel A, we plot the estimated probability by woman's birth cohort in each individual census. In panel B, we plot the estimated probability by the combined census data, with and without age adjustment.

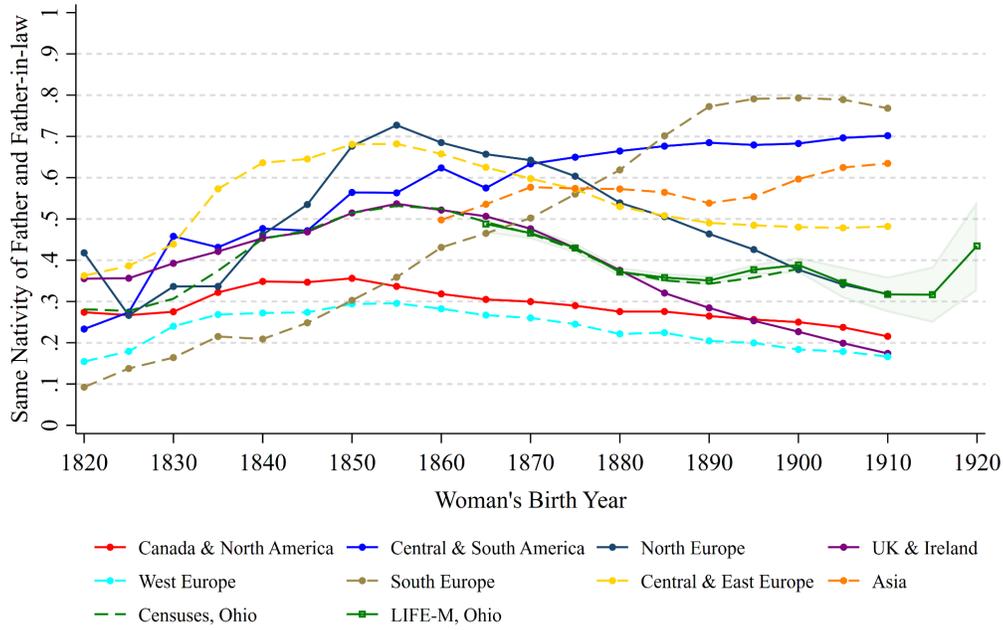
Sources: 1880-1930 Censuses (Ruggles et al., 2021)

Figure C.3. Probability of Same-Nativity of Father and Husband, by Wife's Birth Cohort and Father's Country of Origin

A. Native-born Father



B. Foreign-born Father

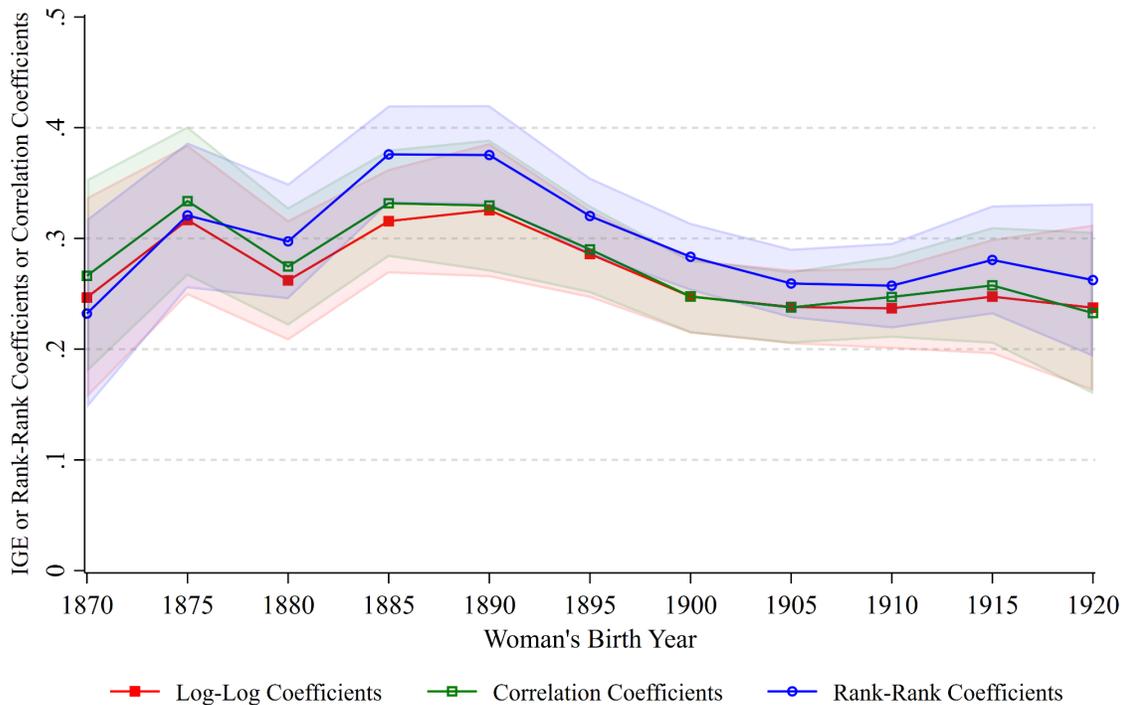


Notes: The figures depict differential trends of nativity homogamy by a woman's father's nativity. Panel A plots the trends for women whose fathers are U.S.-born; Panel B plots the trends for women whose fathers are foreign-born. In Panel B we separate daughters of foreign-born fathers into eight groups depending on fathers' country of origin. We also plot the Ohio-born women whose fathers were foreign-born in the LIFE-M sample and Censuses.

Sources: 1880-1930 Census data (Ruggles et al., 2021) and LIFE-M samples (Bailey et al., 2022a)

D. Alternative Measures of Occupational Homogamy and Heterogenous Trends

Figure D.1. Assortative Matching by Father's and Father-in-Law's Occupational Score, by Wife's Birth Cohort



Notes: The figures depict changes in occupational homogamy by women's year of birth according to the relationship between her father's and father-in-law's occupational income scores, which are based on the 1950 Census occupational scores. We characterize the level of assortative matching by the log-log and rank-rank coefficients derived from regressing the log/rank of father's occupational score on the log/rank of husband's occupational score. We also plot the correlation coefficients derived from regressing the standardized log of father's occupational score on the standardized log of husband's occupational score. We group women into five-year birth cohorts for a more accurate and smoother trend. Estimates are weighted by inverse propensity scores weights and 95-percent confidence intervals are shown as the shaded area.

Sources: LIFE-M samples (Bailey et al., 2022a)