This PDF is a selection from a published volume from the National Bureau of Economic Research

Volume Title: The Economic History of American Inequality: New Evidence and Perspectives

Volume Authors/Editors: Martha J. Bailey, Leah Platt Boustan, and William J. Collins, editors

Volume Publisher: University of Chicago Press

Volume ISBNs: 9780226840642 (cloth), 9780226840659 (electronic)

Volume URL: https://www.nber.org/books-and-chapters/economic-history-american-inequality-new-evidence-and-perspectives

Conference Date: April 1, 2023

Publication Date: April 2025

Chapter Title: Marital Matching and Women's Intergenerational Mobility in the Late 19th- and Early 20th-Century US

Chapter Author(s): Martha J. Bailey, Peter Z. Lin

Chapter URL: https://www.nber.org/books-and-chapters/economic-history-american-inequality-new-evidence-and-perspectives/marital-matching-and-womens-intergenerational-mobility-late-19th-and-early-20th-century-us

Chapter pages in book: p. 165–196

5

Marital Matching and Women's Intergenerational Mobility in the Late 19th- and Early 20th-Century US

Martha J. Bailey and Peter Z. Lin

5.1 Introduction

Assortative matching, or marriage homogamy, measures the similarity of traits between a husband and wife. This statistic has long been of great interest to social scientists across disciplines, because it characterizes the functioning of the marriage market, including the scarcity or abundance of potential partners, the complementarity or substitutability of partners' traits in production and consumption (Becker 1973, 1974; Lam 1988; Stevenson and Wolfers 2007), and bargaining power and gender equity within marriage (Kalmijn 1991, 1994; Shorter 1977). Assortative matching also determines long-run economic outcomes, including income inequality between households (Ciscato and Weber 2020; Eika et al. 2019; Fernandez

Martha J. Bailey is a professor of economics and director of the California Center for Population Research at the University of California, Los Angeles, and a research associate of the National Bureau of Economic Research.

Peter Z. Lin is an assistant professor of economics at Western Kentucky University.

This chapter was prepared for a conference volume, *The Economic History of American Inequality*, in honor of Robert Margo. We thank Mengying Zhang, Deniz Gorgulu, Kelsey Figone, and Eric Wang for excellent research assistance. The LIFE-M project was generously supported by the National Science Foundation (SMA 1539228), the National Institute on Aging (R21 AG05691201), the University of Michigan Population Studies Center Small Grants (R24 HD041028), the Michigan Center for the Demography of Aging (MiCDA, P30 AG012846–21), the University of Michigan Associate Professor Fund, the Michigan Institute on Research and Teaching in Economics (MITRE), the Russell Sage Foundation (1911–19560), and the Washington Center for Equitable Growth. We gratefully acknowledge the use of the Population Studies Center's services and facilities at the University of Michigan (P2CHD041028) and the California Center for Population Research at UCLA (P2CHD041022). For acknowledgments, sources of research support, and disclosure of the authors' material financial relationships, if any, please see https://www.nber.org/books-and-chapters/economic-history-american-inequality-new-evidence-and-perspectives/marital-matching-and-womens-intergenerational-mobility-late-19th-and-early-20th-century-us.

and Rogerson 2001; Greenwood et al. 2014; Kremer 1997); the allocation of resources within households (Calvo et al. 2021); and the intergenerational social and economic mobility of children (Aiyagari et al. 2000; Chadwick and Solon 2002; Currie and Almond 2011; Ermisch et al. 2006).

This chapter contributes to the economic history of American inequality by describing the evolution of assortative matching in the US in the late 19th and early 20th centuries, an era characterized by rapid industrialization and economic growth, the dramatic expansion of public education, mass immigration, urbanization, the Great Depression, and war. We examine marital matching in four dimensions: age, nativity, education, and occupational standing. Our analysis uses the full-count 1850–1940 Decennial Censuses (Ruggles et al. 2021) as well as newly released data from the Longitudinal Intergenerational Family Electronic Micro-database (LIFE-M) (Bailey et al. 2022). The LIFE-M data cover the late 19th and early 20th century for individuals born in Ohio and are fundamental to our analysis of the intergenerational mobility of women. Relying on vital records that contain both birth ("maiden") and married surnames, the LIFE-M data contain high-quality, longitudinal links of more than 260,000 women from birth to adulthood. The presence of both birth and married surnames allows LIFE-M to locate both husbands and fathers, facilitating an analysis of women's social and economic mobility in very large samples of US women born between 1865 and 1920. In addition, using both sources of data allows us to compare estimates from linked samples to the census population, reweight the LIFE-M samples to resemble the population of married women, and adjust for selection into marriage by age.

Our results add new insights into the functioning of the marriage market in the 19th and early 20th centuries. Adjusting for the age composition of married women observed in the census, we find that age homogamy changed very little for women born over the 19th century. This stability contrasts with Rolf and Ferrie (2008), who found that age differences between husbands and wives increased sharply, from 2 to 5 years, for women born over the first 30 years of the 19th century. Adjusting for the age composition of married women, however, shows that the age differences between husbands and wives were fairly stable over the 19th century, rising slightly from around 4.5 to 5.2 years in the early 19th century and then returning to a 4.5-year age gap for the duration of the late 19th century. This stability in age homogamy makes the transition to smaller within-couple age gaps in the 20th century exceptional rather than a return to antebellum US marriage patterns (Atkinson and Glass 1985).

A second finding is the significant decrease in marital matching on fathers' nativity in the mid- to late 19th century. As mass immigration to the US transformed the nation, the likelihood that a woman was married to a man whose father was born in the same country as her own father declined from 0.93 to 0.75—a decrease in nativity homogamy suggesting intermarriage helped stir the US melting pot. Then, between 1890 and 1940, rates of nativity homogamy stabilized in the nation overall. But this pattern was not universal. Among Ohio-born women, nativity homogamy increased from around 0.75 to 0.82 between 1885 and 1940, signaling an increase in segregation by father's country of origin in the rapidly industrializing economies of the Midwest.

A third finding is that the association of husbands' and wives' educational attainment changed little for individuals marrying between 1900 and 1940, despite the high school movement increasing the share of Americans graduating from high school, from 10 percent in 1910 to over 50 percent by 1940 (Goldin 1998). But the results for women's intergenerational occupational mobility offer a different picture. After a period of stability in intergenerational mobility for birth cohorts born in the 30 years after the Civil War, we find that the intergenerational persistence between the occupational standing of a woman's husband and her father fell by roughly one-third for cohorts born over the next 30 years. We also find that absolute measures of intergenerational mobility increased for the same cohorts, suggesting considerably more upward mobility for women whose fathers' occupations ranked in the bottom half of the income distribution. The similarity of estimates in the LIFE-M sample to those from unlinked census samples for age and nativity lends credibility to the quality of the links and representativeness of the reweighted samples and these findings for occupational mobility. An important caveat for these findings is that they speak to patterns only in Ohio—a limitation that is being addressed with ongoing linking work to increase the number of states in the LIFE-M data.

These findings for intergenerational occupational mobility agree with but also differ from some estimates in the literature. Similar to findings using a name-based approach (Olivetti and Paserman 2015), linked Massachusetts marriage records (Eriksson et al. 2023), and retrospective surveys (Jácome et al. 2024), our analysis implies that intergenerational persistence fell for women between 1915 and 1940. However, the magnitudes of our estimates of intergenerational persistence differ. Intergenerational persistence for Ohio-born women is twice what Eriksson et al. (2023) measure in Massachusetts for similar cohorts, suggesting considerably more persistence in socioeconomic status in late 19th- and early 20th-century Ohio. This finding is robust to using correlation coefficients or rank-rank coefficients. However, our estimates of intergenerational persistence are lower than in Olivetti and Paserman (2015), who use names rather than occupations, and also lower than in Jácome et al. (2024), who measure mobility in terms of current household income and retrospective reports of father's occupation. Both levels and trends in intergenerational mobility for women in the LIFE-M Ohio data differ from the national estimates using the Census Tree (Buckles et al. 2023), leaving important puzzles for future work.

This chapter contributes a novel historical perspective to a growing body of work examining assortative matching and women's intergenerational mobility in the US in the 19th century (Althoff et al. 2023; Buckles et al. 2023; Eriksson et al. 2023; Olivetti and Paserman 2015) and the late 20th century (Aiyagari et al. 2000; Charles et al. 2013; Eika et al. 2019; Greenwood et al. 2014). It also contributes historical context for understanding changes in marital matching on nativity to modern studies focusing on the rise in interracial/interethnic marriages (Gilbertson 1996; Kalmijn 1998; Qian 1997; Schoen and Thomas 1989). Finally, it adds to recent studies using alternative approaches to studying assortative matching on socioeconomic status by considering assortative matching on fathers' occupational standing and women's own educational attainment separately (Eriksson et al. 2023; Jácome et al. 2024; Olivetti and Paserman 2015).

5.2 An Overview of Assortative Matching and Women's Intergenerational Mobility in US History

Assortative matching and women's intergenerational mobility have been difficult to study in US history, largely due to data limitations. Data linking in historical data has tended to focus on fathers and sons, because men can be linked using their full first and last names (Abramitzky et al. 2021; Collins and Wanamaker 2022; Feigenbaum 2018; Ferrie 1996; Long and Ferrie 2013; Song et al. 2020; Tan 2023; Ward 2021). Census-linking projects have been limited in their ability to follow women over time, because women change their surnames at marriage. Consequently, the availability of large, systematically linked data has expanded the possibilities for studying generational changes for men in the late 19th and early 20th centuries—but less for women.

As an example, the Early Indicators Project, led by Dora Costa, provides an important longitudinal perspective on economic outcomes for men during the middle and late 19th century (Wimmer 2003). The data consist of 39,340 Union Army (UA) soldiers, approximately 6,200 of whom were "Colored Troops." These data measure the date of death and provide rich information on disability, health, use of medical care, and pension receipt for men reaching retirement age in the late 19th century. Through links to the 1850 and 1930 censuses, the UA data also include sociodemographic and economic variables. An important limitation of the UA data is that they consist primarily of men (because women did not serve as soldiers in the Civil War) and most were born in the North.

The Minnesota Population Center's (MPC) Linked Representative Samples (LRS) merge the full-count 1880 census to the 1850–1930 census 1 percent samples (Ruggles et al. 2010). The LRS thus combine economic (e.g., occupation, literacy, labor force participation, homeownership) and demographic (e.g., age, birthplace, race, marital status, number of children) outcomes for around 500,000 people, including both men and women, across the life course. Although large in scale, important limitations of these data are that most women cannot be linked between their birth and married families (due to surname changes) and that intergenerational coverage consists of at most two generations of men (primarily father–son pairs).

More recently, MPC released the Multigenerational Longitudinal Panel (MLP), which uses supervised machine learning to link millions of individuals between every pair of adjacent censuses from 1850 to 1940 (Helgertz et al. 2020). The resulting sample sizes range from around 6 million individuals linked between the 1850 and 1860 censuses to 52 million individuals linked between the 1930 and 1940 censuses. MLP's linking strategy is implemented in two steps. First, men are linked between adjacent censuses as individuals. In this step, MLP exploits rich training data and contextual information in the linking process (e.g., place of residence, coresident individuals), in addition to names and basic demographics. This strategy increases match rates while reducing the likelihood of false matches, but the final linked sample overrepresents men who do not move and who retain the same household members. In the second step, the procedure links household members living with the men linked in the first step. This second step helps link women who are coresiding with their spouses and daughters living with their fathers. As in the LRS, census data limitations make it nearly impossible to link women who change households or their names at marriage, which means that MLP contains a nonrepresentative set of women.

Concurrent with the development of MLP, Abramitzky et al. (2020) released census links under the Census Linking Project (CLP), which also links millions of men between every available pair of censuses from 1850 to 1940. Building on the linking approaches in Abramitzky et al. (2012, 2019), CLP uses unsupervised machine learning based on name, race, and year or place of birth information in the census. (Note that the difference between supervised and unsupervised machine-learning methods is that the latter do not use training data to control error rates or optimize performance, resulting in smaller linked samples and higher error rates.) Similar to other census projects, CLP does not link women.

Recent studies have addressed these data limitations for studying women in several ways. Olivetti and Paserman (2015) take the creative approach of using children's first names to impute their childhood socioeconomic status. For example, using the 1850 census, they find that fathers of children named Edward had higher occupational rank on average than fathers of Jesse. Using this information, they compute intergenerational mobility for daughters and sons by correlating imputed occupational status of fathers in childhood and own (or husband's) occupational status in the 1850–1940 decennial censuses. They find that both father–son and father–daughter intergenerational elasticities remained stable around 0.31–0.35 for fathers and sons and 0.34–0.40 for fathers and daughters in the 19th century. The elasticities increased to around 0.49 for both father–son and father–daughter pairs observed between 1900 and 1920, then declined to 0.43 for father–son pairs and to 0.37 for father–daughter pairs between 1920 and 1940.

A more direct approach by Eriksson et al. (2023) uses a Feigenbaum's supervised machine-learning approach to link Massachusetts marriage certificates to the 1850, 1860, 1870, 1880, 1900, 1910, and 1920 censuses to examine the socioeconomic mobility of women in terms of the occupational standing of their husbands and fathers. They linked 38,760 couples (3 percent of all marriage records) to both husbands' and wives' adult and childhood census records at an error rate between 9 percent and 14 percent in the first stage (linking marriage records to postmarriage censuses) and between 6 percent and 10 percent in the second stage (linking postmarriage to premarriage censuses), depending on the marriage cohort.[1] They find intergenerational mobility for women, based on either the occupational income scores or occupational wealth scores of the father and husband, is higher than for men for the 1850–1870 marriage cohorts.[2] By 1880–1900 marriage cohorts, men's mobility increased, but the women's did not, leading rates of intergenerational mobility to converge between the two cohorts.

Buckles et al. (2023) combine census information with a rich and unique set of records on FamilySearch.org, one of the largest, user-created genealogical platforms, to create the "Census Tree." FamilySearch.org information is largely generated by its users, who search the website's trove of information (e.g., vital records, newspapers, cemetery documents, census records) to link their own family's records. The Census Tree combines these user links (which they estimate to be correct around 95 percent of the time) with machine links (which they estimate to be correct around 86 percent–89 percent of the time) to produce a large intergenerational database containing both men and women (Price et al. 2021). Because the final Census Tree differs from the population (in particular, they are less mobile from their birth state; Price et al. 2021, table 7), Buckles et al. (2023) estimate inverse-propensity scores to reweight the data to match the census population following the procedure in Bailey et al. (2020a, 2020b). Their intergenerational mobility estimates for men and women use a variety of occupation scores and an instrumental-variables strategy to account for measurement error in occupational status (Solon 1992; Ward 2021). Their results show that intergenerational persistence for men and women is almost identical, falling from around 0.85 to 0.64 between the cohorts of 1840 and 1890 and remaining roughly constant through the cohort of 1910.

More recently, Jácome et al. (2024) pool multiple retrospective surveys from the second half of the 20th century to characterize the long-run evolution of intergenerational income mobility. Although these surveys are not

1. The linkage process and false positive rates are reported in Table A3.

2. Eriksson et al. (2023) calculate occupational wealth score based on the value of real and all property reported in the 1870 decennial census 1 percent sample.

drawn from the early 20th century, they contain information on cohorts that were born between 1900 and 1920, which overlaps with the youngest cohorts in our analysis sample. An important feature of these survey data is that they contain a fairly representative set of women as adults, who report retrospectively on their fathers' occupations. For men and women born between the 1910s and 1940s, they find that intergenerational income persistence fell, and relative intergenerational mobility rose.

Finally, Bailey et al. (2022) use supervised machine learning and rich features in the Social Security Application Records (SS-5) to link over 1.7 million men and women born in the US between 1910 and 1919 and their parents to the 1940 census at a 3-percent error rate. The SS-5 records contain detailed information on applicants (full birth and married names, sex, race, exact date of birth, state, or country of birth) as well as the full names of both parents. These features allow the analysis to identify parent–child relationships and create nationally representative samples containing all states. The limitation of these records is that they contain very few records before the birth cohorts of 1900, which limits their historical perspective. Importantly for this paper, Bailey et al. (2022) find a high degree of intergenerational persistence in education for both women and men, born from 1910 to 1989.

The estimates of rising occupational mobility and stable rates of educational mobility in Jácome et al. (2024) and Bailey et al. (2022) serve as important points of reference for our estimates of women's occupational and educational mobility.

5.3 New Data to Measure Assortative Matching and Women's Intergenerational Mobility

This chapter uses the 1850–1940 decennial censuses as well as the Longitudinal Intergenerational Family Electronic Micro-database (LIFE-M) to construct new estimates of marital matching in the late 19th and early 20th centuries. This section describes the 1850–1940 decennial censuses and LIFE-M data; our analysis samples for age, nativity, education, and occupational homogamy; and our methods to reweight the linked data.

5.3.1 The Full-Count 1850–1940 Decennial Censuses

For our analysis of age homogamy, we use the 1850–1940 decennial censuses from the Minnesota Population Center's (MPC) Integrated Public Use Microdata Series (IPUMS; Ruggles et al. 2021). We restrict all analytic samples to women who (1) were born in the US, (2) were of marriageable age (20–60) at the time of enumeration, and (3) were coresiding with their husbands. Because we focus on marriage in the US, we exclude from our analysis foreign-born women—many of whom may have married before immigrating to the US. This choice also increases comparability of the census with the LIFE-M data, because foreign-born individuals are not included in the LIFE-M sampling frame of US birth certificates. In addition, many foreign-born women married in their country of origin before immigrating and these marriage outcomes may reflect the dynamics of marriage markets outside the US. Note that although all daughters are US-born, many have fathers who are immigrants.

For our analysis of marital matching on nativity, we use the full-count 1880–1930 censuses, in which individuals reported fathers' birthplaces to the census enumerator. This allows us to consider changes in nativity homogamy for a census population without the need to link individuals.[3]

For our analysis of marital matching on education, we only use the 1940 census, which was the first census in which individuals reported their educational attainment. This analysis is restricted to husbands and wives residing together at the time of enumeration.

5.3.2 The Longitudinal, Intergenerational Family Electronic Micro-Database (LIFE-M)

For each of these analyses, we supplement the census data with the newly available LIFE-M database, which links millions of birth records to other vital records (death and marriage records) and the historical censuses (1880, 1900, 1910, 1920, and 1940 censuses), with the goal of minimizing Type I linking errors while maximizing link rates.[4] The project's supervised learning models target Type I error rates below 3 percent, which are further reduced in thorough cross-checks across multiple record links. The LIFE-M data is particularly useful for examining women's intergenerational outcomes, because vital records contain rich information on women's birth (or "maiden") names, their parents' names, and their spouses' names and allow women to be tracked from their birth to marriage families (Bailey et al. 2022). The large sample of linked women allows us to study the evolution of marriage outcomes for women across birth cohorts. We restrict our analysis to the LIFE-M women who were (1) born in Ohio between 1865 and 1920 (a narrower time frame than our restriction on the census data due to data

3. This question is not available prior to 1880. In the 1940 census, only a sample-line person was asked for their father's birthplace. This means that a woman who is the sample-line respondent will report her own father's birthplace. However, her spouse, who is not a sample-line respondent, will not. Therefore, we cannot observe both father and father-in-law's birthplace for women who do not coreside with their fathers-in-law, which is a selected sample. For instance, in the 1940 census, among 1,255,870 women satisfying our data restrictions and being the sample-line respondent, only 24,927 women (2 percent) coresided with their fathers-in-law.

4. Type I linking errors are cases in which an individual is linked to a different person. LIFE-M does not yet link birth records to the 1890 or the 1930 census. The full-count 1890 census is not available, and the 1930 census is planned for future work. Bailey et al. (2022b) present detailed information on data coverage and linking procedures for interested readers. This analysis excludes LIFE-M's North Carolina links, because of the limited sample sizes for our birth cohorts of interest.

availability), (2) of marriageable age (20–60) when observed in the censuses, and (3) coresiding with their husbands in the linked censuses. The decennial census data provide the target population for our reweighting of the linked samples (Bailey et al. 2020a, 2020b). Our methodology for this reweighting is discussed in later sections.

5.3.3 LIFE-M Sample Sizes for Different Dimensions of Assortative Matching

Table 5.1 reports the number of women in the LIFE-M analytic samples for different measures of assortative matching, as well as the percentage of analogous population covered by the LIFE-M samples. The LIFE-M samples cover between 21 percent (or 79,122 women in the 1900 census) and 30 percent (or 391,643 women in the 1940 census) of the female population that satisfies our sample restrictions. If a woman is linked to more than one census, we include these links as separate observations in our dataset. This data structure allows us to observe women from the same birth cohort at different ages across censuses and, therefore, model marriage patterns across ages.[5]

Panel B shows descriptive statistics for the sample for age homogamy, which is the least restrictive because it requires only the age of a woman and her husband (no information about her father is needed). The sample for nativity homogamy in panel C requires the birthplaces of both a woman's father and her father-in-law. We observe the father's birthplace either from the couple's direct reports (for couples linked to the 1900, 1910, and 1920 censuses) or through a father's own links to any census (where he reported his own birthplace).[6] Panel D shows couples linked to the 1940 census to examine assortative matching by education.

The sample to estimate intergenerational mobility by occupation requires the most information, including both the occupations of a woman's father and her husband.[7] Whereas a husband's occupation is reported directly in the census in which he coresides with his wife, we must additionally link women to their fathers to obtain fathers' occupations. As shown in table 5.1, these additional data restrictions reduce sample sizes: the age sample contains 919,025 observations (panel B, column 5), while the intergenerational mobility sample contains 263,258 observations (panel E, column 5).

5. Note that women born after 1900 are only observed in the 1940 census, because they were younger than 20 in 1920.

6. An alternative measure of nativity homogamy compares the birthplaces of a woman's father and her husband. In that case, the analysis requires additional information on the father's birthplace, as the husband's birthplace is always reported in the censuses. See Appendix C, http://www.nber.org/data-appendix/c14841/appendix.pdf.

7. As an alternative measure of occupational homogamy, we also compare the occupations of a woman's father and father-in-law in Appendix D (http://www.nber.org/data-appendix/c14841/appendix.pdf). In this case, the data require additional information on the father-in-law's occupation, which makes the analytic sample even more restrictive.

Table 5.1 **Summary of LIFE-M linked data and analysis samples**

	1850 census (1)	1860 census (2)	1870 census (3)	1880 census (4)	1900 census (5)	1910 census (6)	1920 census (7)	1930 census (8)	1940 census (9)	All censuses (10)
A. Ever-married woman born in the US and ages 20–60 in the census										
All	2,402,578	3,074,331	4,372,836	7,116,727	11,263,010	14,301,494	17,509,586	22,037,814	26,552,781	108,631,120
Born in Ohio, 1865–1920					371,560	671,904	977,037		1,301,988	3,322,489
LIFE-M links					79,122	179,210	293,196		391,643	943,171
% population linked					***21.3%***	***26.7%***	***30.0%***		***30.1%***	***28.4%***
B. Age sample: Panel A and coresident with husband										
All	2,402,578	3,074,331	4,372,798	5,992,634	9,502,522	12,132,365	14,902,751	18,616,633	22,275,760	93,272,372
Born in Ohio, 1865–1920					342,633	609,060	862,236		1,123,879	2,937,808
LIFE-M links					78,078	176,882	290,337		373,728	919,025
% population linked					***22.8%***	***29.0%***	***33.7%***		***33.3%***	***31.3%***
C. Nativity sample: Panel B and nonmissing birthplace of father and father-in-law										
All				5,992,634	9,502,522	12,132,365	14,902,749	18,616,633		61,146,903
Born in Ohio, 1865–1920					342,633	609,060	862,236			1,813,929
LIFE-M links					78,091	176,885	290,331		164,969	710,276
% population linked					***22.8%***	***29.0%***	***33.7%***			
D. Education sample: Panel B and nonmissing education of couple										
All									21,807,116	
Born in Ohio, 1865–1920									1,110,811	
LIFE-M links									368,720	
% population linked									***33.2%***	
E. Occupational intergenerational mobility: Panel B and nonmissing occupations of father and husband										
LIFE-M links					9,669	28,223	68,524		156,842	263,258

Note: The table reports the number of women that satisfy various criteria for different samples. We first report the US-born female population and then the Ohio-born female population satisfying the sample conditions. Then we report the number of women in the LIFE-M who are linked to each census and satisfy the same conditions. Finally, we calculate the percentage of the female population linked through LIFE-M (bold and italic). In panel C, the population in the 1940 census is missing because father's birthplace is only reported for sample-line respondents but not all individuals. Panel E excludes the census population and the link rate because it is unknown how many fathers and husbands have nonmissing information outside the LIFE-M sample. The occupation of the father is only available in the LIFE-M sample, which contains the occupation of the father from his own link to any census between 1880 and 1940.

5.3.4 Representativeness of Linked Samples

One major concern with historical linked samples relates to their representativeness (Bailey et al. 2020a, 2020b), especially because nonrepresentative samples may lead to misleading inferences about population-level intergenerational mobility (Bailey et al. 2020a, 2020b; Jácome et al. 2024). To improve the representativeness of our samples, we create custom weights for each linked sample in table 5.1 using inverse propensity scores for each birth cohort and census year (Bailey et al. 2020a, 2020b; DiNardo et al. 1996; Heckman et al. 1998). Appendix A (http://www.nber.org/data-appendix/c14841/appendix.pdf) describes this procedure in detail.

Table 5.2 presents descriptive statistics for the 1900 to 1940 censuses (panel A, column 1; panel B, columns 1, 6) as well as the unweighted and inverse-propensity-score weighted samples of the LIFE-M data (panel A, columns 2, 4, 6, 8; panel B, columns 2, 4, 7, 9). Differences between the target census population and the weighted LIFE-M samples are in panel A (columns 5 and 9) and panel B (columns 5 and 10), with standard errors listed beneath in parentheses. The unweighted samples are noticeably different from the censuses in almost every characteristic, including individuals' birth years (due to the LIFE-M sampling frame) as well as other characteristics such as husband's occupational income score, coresidence with parents, out-of-birth-state migration, and urban residence, among others. In contrast, the weighted samples are more balanced in terms of these characteristics. Although some of the reweighted means for the LIFE-M data are statistically different from the census, this is due to very large sample sizes: the magnitudes of these differences are very small, especially relative to the unweighted sample differences. Similarity in observed characteristics does not guarantee balance in unobserved characteristics, but the comparability in observed characteristics is reassuring. As an additional point of comparison, we later show that both the magnitudes and trends in our weighted LIFE-M linked samples closely track the results in the census across cohorts when available, whereas the unweighted samples do not (figs. 5A.1–5A.4, http://www.nber.org/data-appendix/c14841/appendix.pdf).

5.4 Statistical Methodology

We characterize historical trends in assortative matching in four dimensions: age, nativity, education, and intergenerational occupational mobility. Ideal data for our analysis would include marital outcomes with educational and occupational histories for all individuals in the US. In practice, we observe only couples who are married and coresiding at a point in time, which means that observed married couples are often different from the population of married couples from the same birth cohort. Figure 5.1 shows this changing selection of the observed couples by age, which presents a key challenge for the analysis. Panel A depicts the average age difference

Table 5.2 **Sample means in the Ohio-born population and LIFE-M data**

	Panel A. Age and Nativity Sample								
	Ohio-born sample table 5.1, panel B	Age sample				Nativity sample			
	1900–40 censuses (1)	LIFE-M unweighted (2)	Diff. (2)–(1) (3)	LIFE-M weighted (4)	Diff. (4)–(1) (5)	LIFE-M unweighted (6)	Diff. (6)–(1) (7)	LIFE-M weighted (8)	Diff. (8)–(1) (9)
Woman's birth year	1888	1889	1.157*** (0.016)	1888	0.00 (0.030)	1885	–3.019*** (0.0154)	1888	0.00 (0.048)
Woman's age	35.52	35.63	0.107*** (0.012)	35.52	0.00 (0.018)	35.25	–0.270*** (0.013)	35.52	0.00 (0.023)
Husband's age	39.53	39.37	–0.160*** (0.013)	39.48	–0.052** (0.020)	39.07	–0.468*** (0.014)	1884	0.089*** (0.025)
Urban residence	0.576	0.513	–0.062*** (0.0006)	0.576	0.0004 (0.001)	0.481	–0.094*** (0.0007)	0.575	–0.0004 (0.0011)
Farm residence	0.218	0.272	0.054*** (0.0005)	0.218	0.0001 (0.001)	0.300	0.082*** (0.0006)	0.219	0.0007 (0.0008)
Migration out of birth state	0.242	0.058	–0.184*** (0.0003)	0.241	–0.001 (0.001)	0.052	–0.190*** (0.0004)	0.238	–0.004** (0.0015)
Coresidence with father	0.031	0.027	–0.004*** (0.0002)	0.031	–0.0004 (0.0003)	0.026	–0.005*** (0.0002)	0.030	–0.0008* (0.0004)
Coresidence with child under 5	0.352	0.459	0.107*** (0.0006)	0.352	0.0004 (0.0008)	0.509	0.157*** (0.0007)	0.353	0.001 (0.001)
Foreign-born husband	0.058	0.047	–0.011*** (0.0003)	0.057	–0.0008 (0.0006)	0.045	–0.013*** (0.0003)	0.055	–0.003*** (0.0008)
Husband's occupational income score	24.81	24.31	–0.500*** (0.014)	24.81	0.007 (0.023)	23.62	-1.184*** (0.016)	24.82	0.013 (0.028)

Note: This table presents means for the population of interest (columns 1, 10, 15), the unweighted LIFE-M samples (columns 2, 6, 11, and 16), the inverse propensity-score reweighted LIFE-M samples (columns 4, 8, 13, and 18). The mean differences between the unweighted linked samples and the target population are reported in columns 3, 7, 12, and 17. The differences between the reweighted linked samples and the target population are reported in columns 5, 9, 14, and 19. See text for more details. ***$p < .01$. **$p < .05$. *$p < .10$. Husbands' occupational income scores are based on the median total income (in hundreds of 1950 dollars) of all persons with that particular occupation in 1950. The occupational scores are provided by IPUMS (Ruggles et al. 2021).

Source: 1880–1940 Full-Count Census Data (Ruggles et al. 2021) and LIFE-M samples (Bailey et al. 2022).

Table 5.2 (cont.)

	Panel B. Education and Occupation Sample									
	Ohio-born Sample table 5.1, panel D	Education Sample				Ohio-born Sample table 5.1, Panel E	Occupation Sample			
	1940 census (10)	LIFE-M unweighted (11)	Diff. (11)–(10) (12)	LIFE-M weighted (13)	Diff. (13)–(10) (14)	1900–40 censuses (15)	LIFE-M unweighted (16)	Diff. (16)–(15) (17)	LIFE-M weighted (18)	Diff. (18)–(15) (19)
Woman's birth year	1897	1897	–0.169***	1897	0.000	1888	1896	7.927***	1888	0.000
			(.0173)		(0.023)			(0.027)		(0.111)
Woman's age	43.17	43.34	0.169***	43.17	0.000	35.37	34.03	–1.342***	35.37	0.000
			(0.0173)		(0.023)			(0.019)		(0.060)
Husband's age	46.53	46.66	0.132***	46.51	0.02	39.28	37.52	–1.765***	39.12	–0.160**
			(0.0193)		(0.025)			(0.020)		(0.064)
Urban residence	0.6348	0.584	–0.051***	0.6350	0.0002	0.572	0.515	–0.057***	0.577	0.005
			(0.0010)		(0.001)			(0.001)		(0.003)
Farm residence	0.1778	0.222	0.044***	0.1780	0.0002	0.225	0.269	0.044***	0.224	–0.001
			(0.0009)		(0.0009)			(0.001)		(0.003)
Migration out of birth state	0.2246	0.080	–0.145***	0.2247	0.0001	0.238	0.055	–0.183***	0.231	–0.007
			(0.0007)		(0.0013)			(0.0005)		(0.005)
Coresidence with father	0.0251	0.023	–0.002***	0.0250	–0.0002	0.031	0.067	0.036***	0.032	0.001
			(0.0003)		(0.0004)			(0.0005)		(0.001)
Coresidence with child under 5	0.1563	0.163	0.007***	0.1562	–0.0001	0.356	0.456	0.100***	0.351	–0.005*
			(0.0008)		(0.0009)			(0.001)		(0.003)
Foreign-born husband	0.0546	0.046	–0.009***	0.0542	–0.0004	0.057	0.034	–0.024***	0.054	–0.003
			(0.0004)		(0.0006)			(0.0004)		(0.002)
Husband's occupational income score	26.65	26.21	–0.442***	26.67	0.015	26.11	25.52	–0.586***	26.08	–0.023
			(.0246)		(0.029)			(0.022)		(0.076)

Note: See notes to table 5.2.

Source: 1880–1940 Full-Count Census Data (Ruggles et al. 2021) and LIFE-M samples (Bailey et al. 2022).

between husbands and wives by the women's birth cohort in the 1850 to 1940 censuses. Within each census year, the average age difference within a couple is largest when the cohort is younger and smaller when the cohort is older. For example, married women born in 1880 who were age 20 in the 1900 census (empty square markers) were more than six years younger than their husbands on average, whereas the same cohort of women age 60 in the 1940 census (cross markers) were only about three years younger than their husbands. This pattern reflects the fact that women who marry at younger ages disproportionately marry older men. It also reflects survival bias in marriages: women who are age 60 are much less likely to be married to a partner who is much older than them, because his mortality risk increases with age. Figure 5A.1 (http://www.nber.org/data-appendix/c14841/appendix.pdf) presents age selection in terms of nativity homogamy. Failing to adjust for this selection could severely bias estimates of age homogamy and, potentially, other measures of marital sorting.

To adjust for selection into marriage by age, we estimate the following linear regression model by ordinary least squares (OLS):

$$Y_{it} = f_k + q(age_{it}) + \varepsilon_{it}, \quad (1)$$

where the dependent variable, Y_{it}, is the marriage outcome of interest, either the husband–wife age difference or a binary variable for same nativity of father and father-in-law for woman i coresiding with her husband in the census year t.[8] We code "same nativity" if a woman's father and father-in-law were born in the same grouping of countries. Considering the border changes in many nations in the late 19th and early 20th centuries, we combine countries that are close to each other geographically and culturally that changed borders into the same country group. While this measure will overstate the share of father-father-in-law nativity homogamy in terms of individual countries, we are more confident that changes in this measure capture real differences in country and culture of origin rather than changes in national borders.[9] We group women into five-year cohort bins, f_k, to reduce

8. We also consider alternative measures, such as the absolute difference in age between husbands and wives and a binary measure of whether a woman is over three years younger than her husband. These results are reported in Appendix B (http://www.nber.org/data-appendix/c14841/appendix.pdf) and change the story of our main analysis little.

9. Appendix C (http://www.nber.org/data-appendix/c14841/appendix.pdf) describes these detailed country groups. We also consider an alternative measure of same-nativity marriage: whether the woman's father and husband had the same birth country group. Although closely related to the baseline measure, the two definitions capture marital sorting in different ways. For instance, if a US-born daughter of a German immigrant married a US-born son of another German immigrant, the marriage is a same-nativity marriage according to our baseline definition. However, it is regarded as a cross-nativity marriage using the second definition because the woman's father was born in Germany and her husband was born in the US. The baseline definition is thereby less affected by immigration shocks and better reflects the intergenerational persistence of nativity preference for partners. These alternative measures yield similar results and are presented in Appendix C (http://www.nber.org/data-appendix/c14841/appendix.pdf).

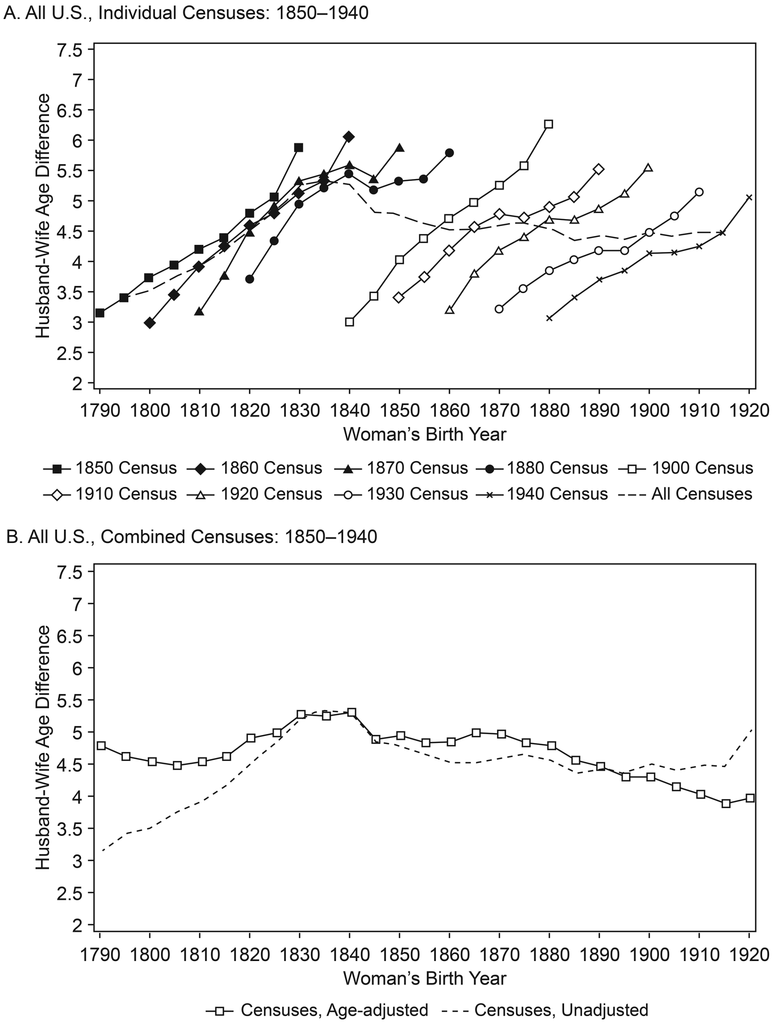


Figure 5.1 Husband–wife age differences, by wife's birth cohort

Note: The figures depict the mean husband-wife age difference (husband's minus wife's age) by wife's birth year. Due to the age heaping in the census and sample sizes in LIFE-M, we group women into five-year birth cohorts and plot the estimates for the midpoint of each five-year birth-year group. Panel A presents the mean age differences by census for the sample in panel A of table 5.1 and also the birth cohort (dashed line). Panel B presents the cohort-specific mean age differences, unadjusted and age-adjusted as described in the text.

Source: 1850–1940 Full-Count Census Data (Ruggles et al. 2021).

noise, and $q(age_{it})$ is a quartic function of the cohort's age in census year t. After estimating this model using the full-count 1850–1940 US decennial censuses (omitting the unavailable 1890 census), we predict outcomes for each birth cohort at a common age—35—thereby adjusting the data for selection into or out of married coresidence at different cohort ages in the census. Standard errors are clustered for dependence at the birth-cohort year level (Moulton 1986).

These census estimates can be compared to those from the LIFE-M sample, which are weighted to reflect observed characteristics in the census using inverse-propensity-score weights (Bailey et al. 2020a, 2020b). Our ability to compare estimates across these samples allows us to assess the quality of the LIFE-M sample as well as assess the external validity of Ohio relative to a nationally representative, unlinked source. Because the censuses are available, the LIFE-M sample is not necessary to examine age or nativity homogamy. However, the comparison of the LIFE-M estimates to census estimates is useful for analyses in which the census cannot be used, for example, for the intergenerational occupational mobility analyses.

Our analysis of marital matching on education builds on a large literature on intergenerational mobility (Black and Devereux 2011; Chetty et al. 2014a, 2014b, 2017; Deutscher and Mazumder 2023). We measure the educational attainment of wives and their husbands in the 1940 census, the first census to report this outcome.[10] Because we observe education in only one census, we cannot adjust by age as we do in the previous analysis. For this reason, the analysis uses women ages 30 to 60 to minimize the effects of selection into marriage by age. Following Greenwood et al. (2014, 2016) and Eika et al. (2019), we examine assortative matching using OLS to estimate the following linear model,

$$(2) \qquad Educ_i^W = \sum_j \gamma_j D_j \times Educ_i + f_j + \varepsilon_i,$$

where $Educ_i^W$ is the educational attainment for wife i; $Educ_i$ is her husband's educational attainment; D_j is a dummy variable for women born in year j; and f_j captures individual birth-year fixed effects. The cohort-specific homogamy coefficient γ_j is a measure of educational homogamy for women of birth cohort j. To account for changes in the marginal distributions of education across time, we also present intergenerational educational correlation estimates, in which $Educ^W$ and $Educ$ in equation 3 are normalized to have a mean of zero and a standard deviation of 1 within cohort groups. Because there are so many ties (identical values) for educational attainment, we do not use the rank-rank approach that is often used for income (Chetty et al. 2014a, 2014b). Standard errors are clustered for dependence at the birth-cohort year level (Moulton 1986).

10. Censuses before 1940 ask about literacy but not detailed educational attainment.

A final analysis uses only the LIFE-M sample to compute intergenerational occupational persistence between a woman's father and her husband using the occupational score of their occupation. For the occupational score, we use those calculated by IPUMS (Ruggles et al. 2021). These represent the median total income (in hundreds of 1950 dollars) of all persons with that particular occupation in 1950.[11] We choose to associate a husband's occupation score rather than women's own, because few women participated in the labor market in this period. This measure has several other advantages as well. First, occupations are readily reported in all historical censuses, which allows us to consider a long period of time. Second, occupation captures a more permanent component of socioeconomic status than income, because it does not experience transitory shocks and is less subject to measurement error. In addition, occupational scores are used in other studies of intergenerational mobility (Eriksson et al. 2023; Olivetti and Paserman 2015), which facilitates straightforward comparisons to important findings in the literature.

We estimate the following specification of intergenerational persistence by OLS,

$$(3)\qquad \log(OccScore)_{it} = \sum_k \gamma_k D_k \times \log(ParentOccScore)_i + \log(ParentOccScore)_i \times q(age_{it}) + f_k + \varepsilon_i,$$

where $\log(OccScore)_{it}$ captures the occupational standing of the husband of woman i, in the observed census year t, and $\log(ParentOccScore)_i$ captures the occupational standing of the woman's father. We group women into cohort bins such that D_k is a dummy variable for women born in cohort $k(k = 1865–1867, 1868–1872, 1873–1877, \ldots, 1912–1917, 1918–1920)$ and f_k is a set of individual cohort fixed effects.[12] A quartic function of the woman's age in the census, $q(age_{it})$, helps capture lifecycle bias and age-based selection into marriage.

The coefficient of interest, γ, is the intergenerational elasticity (IGE)—an estimate of intergenerational persistence that measures how fathers' occupational standing is associated with the occupational standing of their daughters' husbands. A higher value of γ corresponds to higher persistence in

11. The use of occupational income score facilitates comparisons with some studies related to our question of interest, particularly with Olivetti and Paserman (2015) who use the same occupational income score. However, this is a departure from other work such as Collins and Wanamaker (2022), who generate occupational scores based on 1940 wage income information by 3-digit occupation, with some adjustments for self-employed workers and farm workers; these scores are computed separately by race and census division of residence. Song et al. (2020) use a status measure that is based on literacy/education and occupation. Ward (2021) adjusts for racial and regional inequality in his "adjusted Song score" which is based on literacy/education by occupation, race, and region.

12. We chose to center these bins on multiples of five for all but the first and last bins to account for age heaping.

socioeconomic status across generations, or alternatively lower social mobility. Importantly, γ is both affected by the parent–child correlation but also the relative variance in their outcomes (Eika et al. 2019; Gihleb and Lang 2020). To adjust for the fact that the distribution of occupational scores evolves over time, we also present intergenerational correlation estimates, in which log*Occscore* and log*ParentOccScore* in equation 2 are normalized to have a mean of zero and a standard deviation of 1 within cohort groups.

We supplement these log-log estimates using a rank-rank approach (Chetty et al. 2014a, 2014b). For these analyses, we rank father's occupational score relative to all the fathers of the US-born women in cohort j from the most recent census when children were under age 10. For instance, we use the national distribution of fathers' occupational income scores in the 1910 census for women born between 1901 and 1910, when most girls were coresiding with their fathers. We rank husbands by occupational score for a woman in birth cohort j observed in census year t. We then estimate equation 2, replacing log*Occscore* with occupational rankings. A higher rank-rank coefficient suggests stronger marital sorting on occupation and lower intergenerational mobility for women. All regressions are weighted by the inverse propensity-score weights, and standard errors are clustered at the birth-cohort level (Moulton 1986).

Both the rank-rank coefficients and intergenerational elasticities measure relative mobility, without much information on upward or downward mobility. Following Chetty et al. (2017), our third measure is absolute mobility: (1) the mean husband's occupational rank for a woman born to a father ranked below the median in the fathers' occupational score distribution and (2) the mean husband's occupational rank for a woman born to a father ranked above the median. The former measures absolute upward mobility, whereas the latter measures absolute downward mobility. See Deutscher and Mazumder (2023) for an in-depth comparison of these measures.

5.5 Results: Marital Matching in the Late 19th and Early 20th Centuries

We begin with an analysis of age homogamy. Panel B of figure 5.1 compares the age-adjusted and unadjusted (raw) series of husband-wife age differences for cohorts born from 1790 to 1920 and married from around 1810 to 1940. The unadjusted series is based on the simple averages by cohort in the combined censuses. The age adjustment significantly changes the antebellum national trends, correcting the sharp upward rise in husband-wife age differences for the fact that women born earlier in the century are older when they are observed in the 1850 census. The unadjusted series increases from 3.2 years to 5.3 years between the 1790 and 1840 cohorts, whereas the age-adjusted series only increases from 4.7 years to 5.2 years for the same cohorts. After peaking for cohorts born from 1830 to 1840 at 5.2, husband–wife age differences decrease to around 5 years for the 1880 birth cohort—women getting married around the turn of the century. In short, age homogamy in marriage changed more modestly over the 19th century than previously believed—much less than implied by the series unadjusted for age selection (Rolf and Ferrie 2008).[13] The big picture is that relative stability in age homogamy during the 19th century makes the transition to smaller within-couple age gaps beginning in the 20th century appear more exceptional rather than a return to antebellum US marriage patterns.

We also compare changes in age homogamy for the weighted LIFE-M sample to two reference groups: (1) the age-adjusted population for Ohio-born women for the same cohorts from the census and (2) the age-adjusted population for all US-born women. Panel A of figure 5.2 plots these results for women born between 1865 and 1920 and married between roughly 1885 and 1940. Importantly, both the levels and trends in the LIFE-M data track those for the population of Ohio-born women in the census. This finding underscores the ability of high-quality links and inverse-propensity score reweighting to recover population parameters even when linked samples are not representative. It also increases our confidence in the results for occupational sorting that are based only on linked samples when census estimates are unavailable.

Another key finding is that, while trends in age homogamy among Ohio-born women appear similar to changes in the US after 1880, the average husband–wife age difference was around half a year smaller in Ohio—a difference likely due to Ohio's industrialization and economic development relative to the nation. Indeed, panel B of figure 5.2 makes clear the pattern in age differences by level of economic development, with the most developed census region (Northeast) having smaller husband-wife age differences than the least developed census region (South). The Ohio sample from LIFE-M exhibits a smaller age gap on average but follows patterns identical to the Midwest.

We next extend our analysis of marital homogamy to nativity. Figure 5.3 plots the age-adjusted estimates of the likelihood of same-nativity marriages based on birth country groups of a woman's father and father-in-law. The age-adjusted trend shows a continuous decline in same-nativity marriages between the 1820 and 1890 cohorts, roughly married between 1840 and 1930. The probability of a woman marrying a husband from the same nativity group decreased, from 92 percent for the 1820 cohort to around 75 percent for the 1890 cohort. After that, the age-adjusted probability of a same-nativity marriage remained fairly stable between the 1890 and 1910 cohorts (marriages roughly occurring between 1910 and 1940), whereas the unadjusted trend shows significantly increasing same-nativity

13. Appendix Figure B.1, Panels A and B, (http://www.nber.org/data-appendix/c14841/appendix.pdf) shows a similar trend in the absolute age difference and the probability of marrying a husband at least three years older.

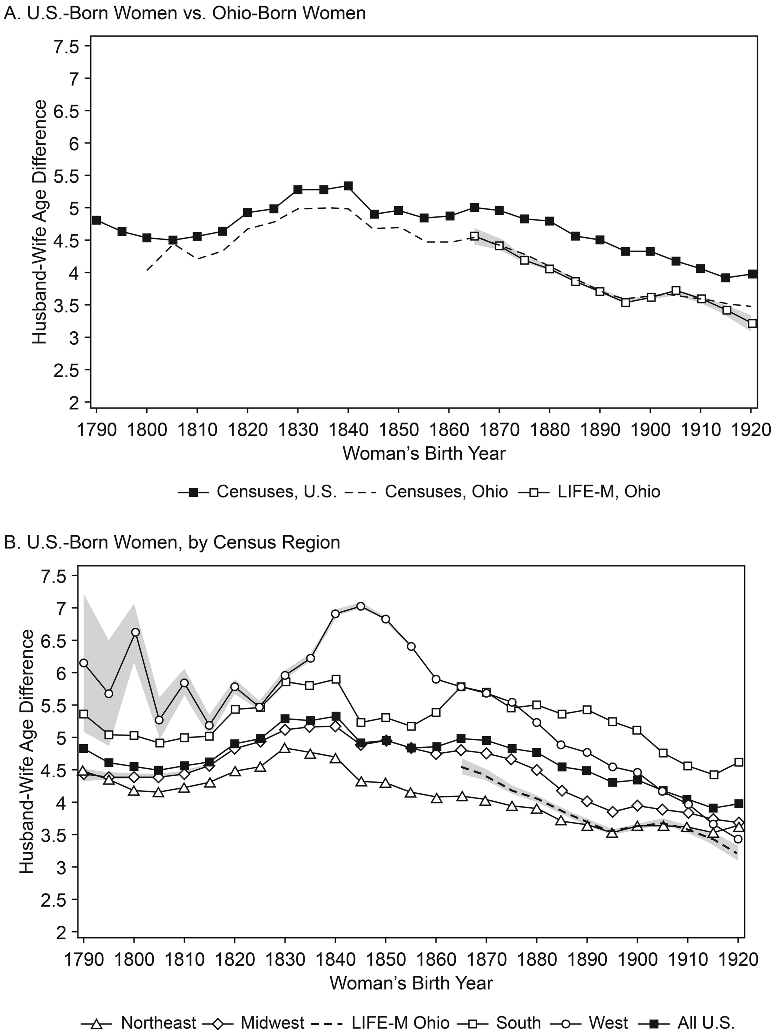
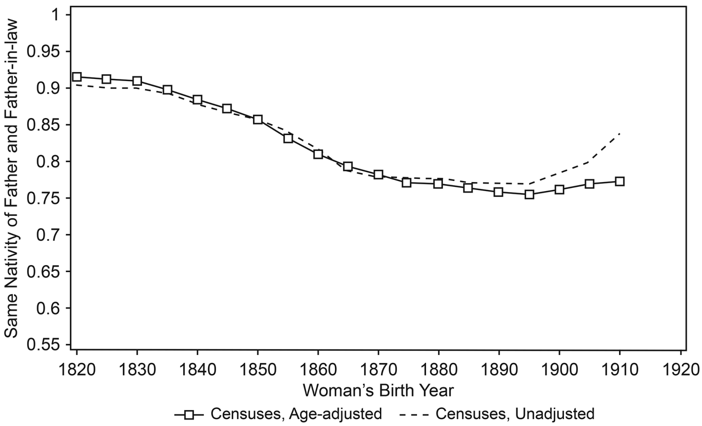


Figure 5.2 Husband–wife age differences, by wife's birth cohort

Note: The figures depict the mean husband-wife age difference by wife's year of birth. Panel A presents the age-adjusted cohort-specific mean age differences for US-born women in the censuses, Ohio-born women in the censuses, and weighted LIFE-M sample of Ohio-born women. The LIFE-M data are weighted using inverse propensity scores as described in the text. Panel B presents the age-adjusted cohort-specific mean age differences for US-born women by their census region of residence along with the Ohio-born women in LIFE-M data. Due to age heaping in the census, we group women into five-year birth cohorts and plot the estimates for the midpoint of each five-year birth-year group. The 95-percent confidence intervals are shown as the shaded area.

Source: 1850–1940 Full-Count Census Data (Ruggles et al. 2021) and LIFE-M samples (Bailey et al. 2022).

A. All U.S.-Born Women, Census Data



B. Ohio-Born Women vs. U.S.-Born Women, Census and LIFE-M Data

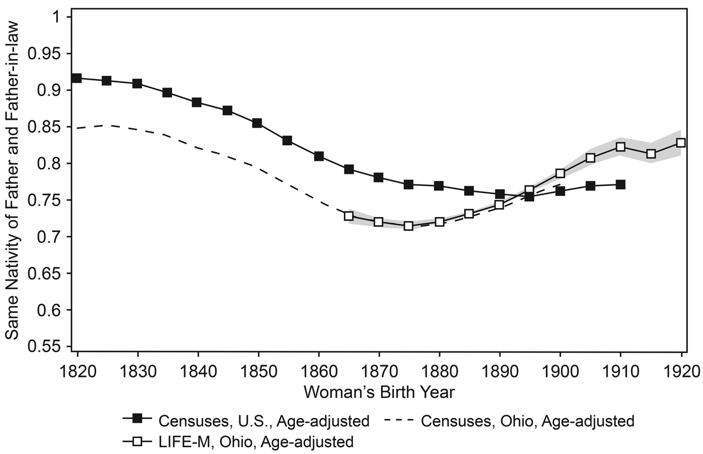


Figure 5.3 Nativity homogamy by father and father-in-law's country group of origin, by wife's birth cohort

Note: The figures show nativity homogamy, defined as a woman's father and her father-in-law being born in the same group of countries. Panel A plots the age-adjusted nativity homogamy by the woman's birth year. (See corresponding series by census and not adjusted by age in Appendix Figure C.1.) Panel B plots the age-adjusted nativity homogamy by the woman's birth year for all US-born women and Ohio-born women in the weighted LIFE-M and census data. The LIFE-M data are weighted using inverse propensity score weights as described in the text. Panel C plots age-adjusted nativity homogamy in census samples by women's census region of residence. Due to age heaping in the census, we group women into five-year birth cohorts and plot the estimates at the midpoint of the group. The 95-percent confidence intervals are shown as the shaded area.

Source: 1880–1930 census data (Ruggles et al. 2021) and LIFE-M samples (Bailey et al. 2022).

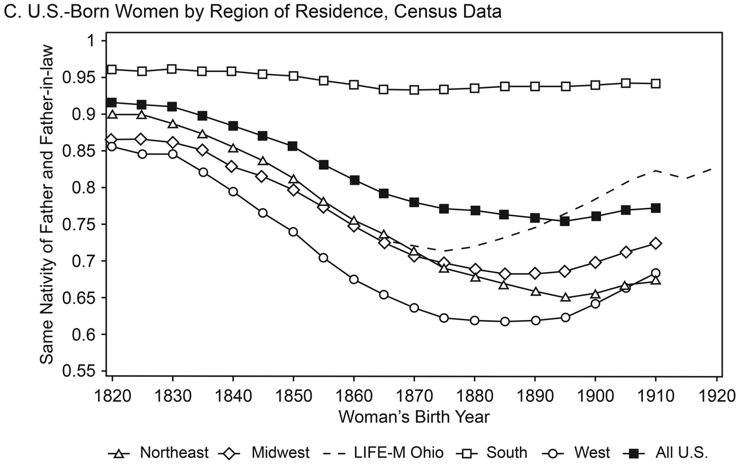


Figure 5.3 (cont.)

marriages for this period. The differences between the unadjusted trend and the age-adjusted trend can also be explained by selection into marriage for more recent cohorts. As Appendix Figure C.1 (http://www.nber.org/data-appendix/c14841/appendix.pdf) shows, women marrying at younger ages (which are the ones we observe for these younger cohorts) tended to marry husbands in the same nativity group, and the age adjustment helps adjust for this tendency.[14]

Panel B of figure 5.3 presents the probability of a same-nativity marriage for Ohio-born women in both the LIFE-M sample and analogous census population data. Similar to the estimates for husband-wife age differences, the age-adjusted LIFE-M estimates and census estimates for Ohio-born women are almost identical—so much so that the dashed line for the census is barely visible in the figure once the LIFE-M data appear for women born after 1865. This similarity again lends credibility to the linked LIFE-M sample's findings and underscores the power of using inverse-propensity reweighting to achieve balance. A second finding, however, is the divergence in the trend for Ohio-born women from US-born women. Among Ohio-born women, nativity homogamy increased from around 0.75 to 0.82 between 1885 and 1940 (cohorts born between 1865 and 1910), signaling an increase in marital sorting by father's country group of origin in the rapidly industrializing economies of the Midwest. Examining regional trends for all

14. Appendix Figure C.2 (http://www.nber.org/data-appendix/c14841/appendix.pdf) shows a similar national trend when defining same-nativity marriages as a woman's father and husband being born in the same group of countries. Similar to the series in Figure 5.3A, we find a significant decline in same-nativity marriage between the 1820 and 1870 cohort, and after that, the probability of a same-nativity marriage remained steady between 1870 and 1910.

US-born women in panel C of figure 5.3 shows only slight increases in nativity homogamy in the broader Midwest census region and West in the early 20th century (cohorts marrying between 1910 and 1930), suggesting that the patterns in Ohio were more the exception than the norm in this period.

Next, we consider changes in marital matching on education, as measured by the association of husbands' and wives' education in the 1940 census using equation 2. Although these trends are not well estimated for cohorts born before 1880, these comparisons are available for all US-born individuals as well as for Ohio-born individuals in the 1940 census and LIFE-M samples.[15] Figure 5.4 shows the cohort-specific slope coefficients (panel A) and correlations (panel B) using equation 2. Notably, the reweighted LIFE-M data again track the census estimates for Ohio-born individuals very closely, which again lends credibility to results using the LIFE-M sample when it cannot be benchmarked in the intergenerational analyses. The results show that the association of husbands' and wives' slope coefficients remained stable for 30 years, for women born between 1880 and 1910 (married between 1900 and 1930), decreasing very slightly for the youngest cohorts (born between 1905 and 1910). Similarly, correlation coefficients increased by only a few points over the period, suggesting very slight increases in assortative matching on education after accounting for the decreasing variance in women's educational attainment relative to men's across cohorts.

In addition, differences between the estimates for the entire US versus Ohio-born residents suggest that Ohio had less assortative matching on education, which is important to keep in mind when considering the external validity of the estimates from the LIFE-M sample. Figure 5.5 shows the correlation coefficients by census region, which indicates that women in the Northeast, Midwest, and West were less assortatively matched on education than women in the South, which raises the estimates for the nation. Like the national trends, however, the regional trends are very stable over time.

In contrast, figure 5.6 shows that women's intergenerational mobility, as measured by fathers' and husbands' occupational standing, was stable in the 19th century and increased meaningfully in the early 20th century. Importantly, these comparisons are not available for all US-born women, so figure 5.6 only uses the LIFE-M sample for women born in Ohio from 1865 to 1920. Panel A plots the cohort-specific rank-rank coefficients, the intergenerational elasticities (IGE, or log-log coefficients), and correlations based on the regressions specified in equation 3. For the 1865 to 1890 cohorts (marriages from 1890 to 1910), we find little change in either the rank-rank or IGE estimates. Assortative matching in terms of husbands' and fathers' occupational standing was fairly stable in the 19th century.

However, intergenerational persistence declined, and mobility increased,

15. The 1940 census is the first to ask this question and the education-mortality gradient makes older cohorts more selected and less representative.

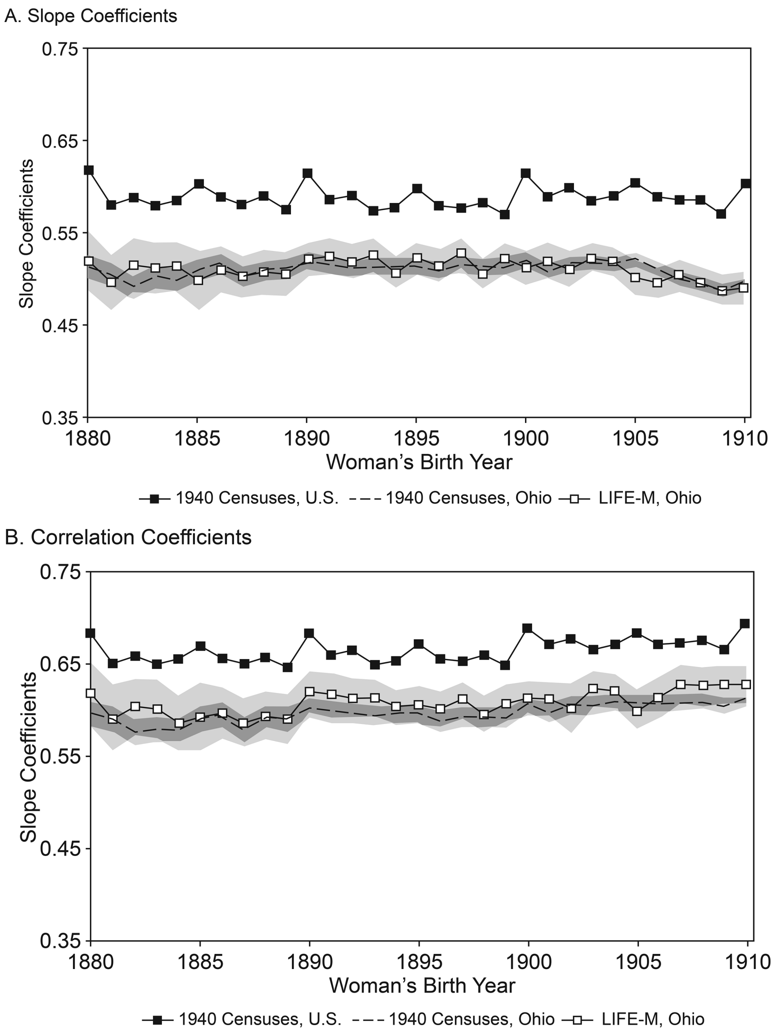


Figure 5.4 Assortative matching by educational attainment, by wife's birth cohort

Note: Panel A depicts education homogamy as captured by regressing a wife's educational attainment on her husband's educational attainment. Educational attainment measures the highest grade completed as reported in the 1940 census. Panel B presents the correlation coefficients. The LIFE-M data are weighted using inverse propensity score weights as described in the text, and 95-percent confidence intervals are shown as the shaded area.

Source: 1880–1940 Full-Count Census Data (Ruggles et al. 2021) and LIFE-M samples (Bailey et al. 2022).

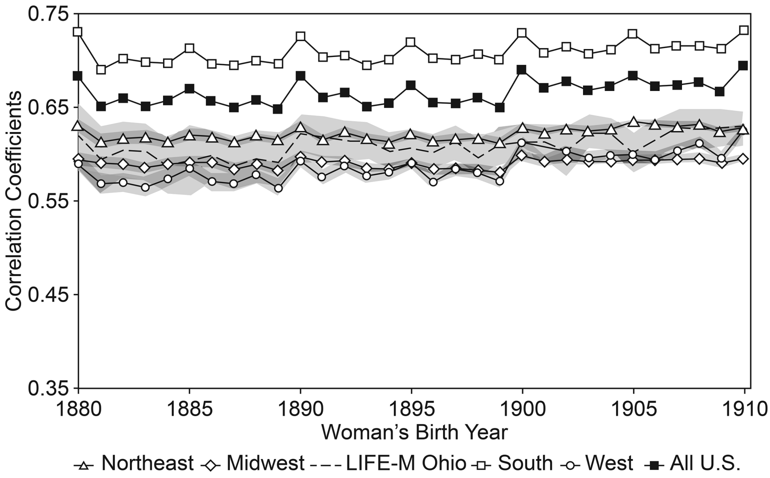


Figure 5.5 Assortative matching by educational attainment, by wife's birth cohort and census region

Note: The series plots the correlation coefficients in educational attainment between husband and wife by wife's birth cohort using the 1940 census. For region-specific coefficients, we group women into five-year cohorts centering on years ending with 5 or 10. For the earliest cohorts, we group the 1880–1882 cohorts and plot them as the 1880 cohort. For the latest cohorts, we group the 1908–1910 cohort and plot their estimates as the 1910 cohort. The shaded area shows the 95-percent confidence intervals. See also notes to figure 5.6.

Source: 1880–1940 Full-Count Census Data (Ruggles et al. 2021).

rapidly for the cohorts born between 1890 and 1920, marriages taking place between 1910 and 1940. Both the IGE and rank-rank coefficients decreased, from 0.32 for the 1890 cohort to 0.15 for the 1920 cohort, which correspond to marriages occurring between 1910 and 1940. Changes in absolute mobility follow similar patterns (panel B). For all cohorts, the average occupational rank of husbands was significantly higher for daughters of above-median occupational rank fathers than for daughters of below-median occupational rank fathers. This is strong evidence for assortative marital matching by socioeconomic status. These patterns remained fairly stable for cohorts born between 1865 and 1890 (marriages from 1885 to 1910), but both upward and downward mobility increased sharply for cohorts born between 1890 and 1920 (marriages from 1910 to 1940).

Although we cannot compute estimates for other census regions, Figure 5.7 compares our estimates to those from other studies. Our estimates of Ohio-born women are slightly lower in level but compare favorably to Olivetti and Paserman (2015), especially for their Midwest sample. The fact that names are stickier than occupations, which can be upgraded over one's lifetime, may explain why occupational homogamy appears lower and economic mobility appears higher using occupational measures. Differences from levels in

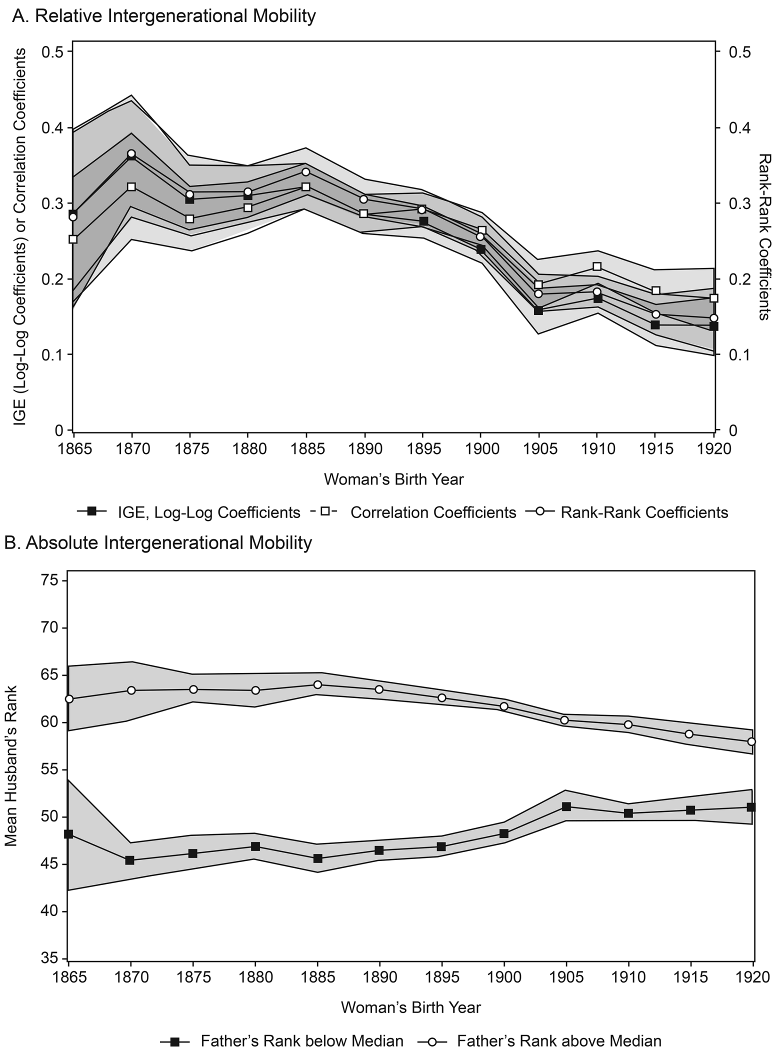


Figure 5.6 Intergenerational mobility by husband's and father's occupation score, by wife's birth cohort

Note: The figures depict changes in occupational homogamy by wife's year of birth according to the relationship between her father's and husband's occupational income scores, which are based on the median income by occupation in the 1950 Census. Panel A characterizes relative mobility in terms of log-log and rank-rank coefficients derived from regressing the log/rank of husband's occupational score on the log/rank of father's occupational score. Panel B plots absolute upward and downward mobility by plotting the husband's occupational rank for women whose fathers fall below or above the national median. We group women into five-year birth cohorts and plot the estimates for the midpoint of each five-year birth-year group. The LIFE-M data are weighted using inverse propensity score weights as described in the text. The 95-percent confidence intervals are shown as the shaded area.

Source: LIFE-M samples (Bailey et al. 2022).

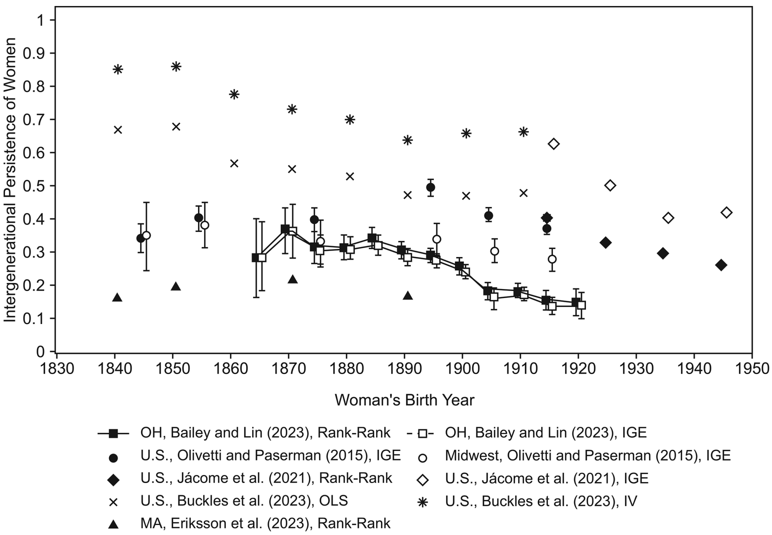


Figure 5.7 Estimates from different studies of intergenerational mobility, by wife's birth cohort

Note: The figure depicts estimates of intergenerational persistence of women in this paper (the squares), as well as the other three related works. The asterisks and cross marks refer to the estimates by Buckles et. al. (2023). The circles refer to the estimates by Olivetti and Paserman (2015). Their estimates are based on a child's first name and pseudo-linking between a father's occupational income score and a husband's occupational income scores, as defined by median total income (in hundreds of 1950 dollars) of all persons with that particular occupation in 1950 (Ruggles et al. 2021). Their estimates are calculated by census years, and we realign them on the woman's birth cohort by subtracting 25 from the census years. The diamonds refer to the estimates by Jácome et al. (2024), which uses surveys reporting fathers' occupations to create occupational-income scores and daughters' family incomes. Their estimates are calculated by the woman's birth cohort and are plotted accordingly. The triangles are estimates from Eriksson et. al. (2023), which are based on links between Massachusetts marriage certificates from 1850 to 1915 to the 1850, 1860, 1870, 1880, 1900, 1910, and 1920 censuses. We plot their estimates based on 1950 occupational income scores for best comparison with our estimates. They do not report confidence intervals so these are not reported here. We translate their estimates for the marriage cohorts 1850–1870, 1860–1880, 1880–1900, and 1900–1920 to the birth cohorts of the 1840s, 1850s, 1870s, and 1890s.

Source: LIFE-M samples (Bailey et al. 2022).

Jácome et al. (2024) likely reflect the fact that retrospective reports of fathers' occupations may reflect his most persistent occupation rather than his work at only one point in time, as measured in the census.[16] Said another way, our measures of occupation from the census may mismeasure socioeconomic standing for much of childhood relative to retrospective reports due to transitory factors or life cycle biases (Mazumder 2005; Solon 1992).

16. See Ward (2021) for an in-depth discussion of this source of measurement error in occupations in historical census data and Haider and Solon (2006) for a discussion of lifecycle bias more generally.

It is harder to interpret the differences in levels from Eriksson et al. (2023). Like Eriksson et al. (2023), this chapter also uses occupational measures but has lower match rates (7 percent to 9 percent) and higher linking error rates (9 percent to 13 percent) than in the LIFE-M data. The former could make their data less representative and the latter could attenuate intergenerational elasticities. In addition, intergenerational occupational mobility may differ between Massachusetts and Ohio, as suggested by stratifications by region for our age, nativity, and education results.

The largest differences in levels and trends emerge between our estimates and those of Buckles et al. (2023), who use the Census Tree data, a different occupational-income score definition, and an instrumental variables approach to account for measurement error in occupational status. These differences in estimates remain open questions for future research to resolve.

To complement these findings, we also examine marital sorting by occupational standing of a woman's father and father-in-law, instead of the husband. The level of marital sorting by parents' socioeconomic outcomes can reflect the relative strength of ascribed and acquired traits in the marriage market (Charles et al. 2013). We measure the sorting by fathers' occupations by estimating equation 3, but we replace the husband's occupational standing with that of his own father. Appendix Figure D.1 (http://www.nber.org/data-appendix/c14841/appendix.pdf) plots the results. We find a similar trend between the 1870 and 1890 cohorts, and the rank-rank coefficients decreased significantly from 0.35 for the 1890 cohort to around 0.25 for the 1910 cohort. The decline is smaller than that of rank-rank coefficients between father and husband (fig. 5.6, panel A), suggesting that the increase in intergenerational mobility of women was caused by both decreasing marital sorting on parents' socioeconomic status and increasing intergenerational mobility of husbands.

Overall, the trends in figure 5.7 reinforce the conclusion that occupational homogamy changed little in the late 19th century for marriages starting between 1880 and 1900 (cohorts born between 1860 and 1880). However, for women born in the early 20th century intergenerational persistence decreased and economic mobility increased.

5.6 Conclusion

This chapter contributes to the economic history of American inequality by characterizing the evolution of marital matching during the late 19th and early 20th century, the eras of mass immigration, rapid industrialization and economic growth, urbanization, the Great Depression, and war. We find that age homogamy changed very little during the 19th century, which makes the rapid transition to smaller within-couple age gaps in the 20th century appear exceptional rather than a return to antebellum US marriage patterns. As mass immigration to the US transformed the nation, the likelihood that a woman was married to a man whose father was born in the same group of countries as her father declined rapidly—a decrease in nativity homogamy, suggesting intermarriage helped stir the US melting pot. From 1900 to 1940, women's intergenerational mobility in terms of her husband's occupational standing relative to her father's increased, whereas the association of husbands' and wives' educational attainment changed little. As the high school movement transformed America's public school landscape, we conclude that women's own educational attainment remained a powerful force in shaping their socioeconomic status in adulthood. Understanding how these trends shaped—and were themselves shaped—by the demographic transition, rapid industrialization, and the transformation of women's paid work remains for future research.

References

Abramitzky, R., L. P. Boustan, and K. 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.

Abramitzky, R., L. Boustan, K. Eriksson, S. Pérez, and M. Rashid. 2020. Census Linking Project: Version 2.0 [dataset]. https://censuslinkingproject.org.

Abramitzky, R., L. Boustan, E. Jácome, and S. Pérez. 2021. "Intergenerational Mobility of Immigrants in the United States over Two Centuries." *American Economic Review* 111 (2): 580–608.

Abramitzky, R., R. Mill, and S. Pérez. 2019. "Linking Individuals Across Historical Sources: A Fully Automated Approach." NBER Working Paper No. 24324. National Bureau of Economic Research.

Aiyagari, S. R., J. Greenwood, and N. Guner. 2000. "On the State of the Union." *Journal of Political Economy* 108 (2): 213–44.

Althoff, Lukas, Harriet Brookes Gray, and Hugo Reichardt. 2023. "The Missing Link: Women and Intergenerational Mobility." https://hugoreichardt.github.io/pdf/igm_women.pdf.

Atkinson, M. P., and B. L. Glass. 1985. "Marital Age Heterogamy and Homogamy, 1900 to 1980." *Journal of Marriage and the Family* 47: 685–91.

Bailey, M. J., Connor Cole, Morgan Henderson, and Catherine Massey. 2020a. "How Well Do Automated Methods Linking Perform? Evidence from the LIFE-M Project." *Journal of Economic Literature* 58 (4): 997–1044.

Bailey, M., C. Cole, and C. Massey. 2020b. "Simple Strategies for Improving Inference with Linked Data: A Case Study of the 1850–1930 IPUMS Linked Representative Historical Samples." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 53 (2): 80–93.

Bailey, M., P. Z. Lin, A. R. S. Mohammed, P. Mohnen, J. Murray, M. Zhang, and A. Prettyman. 2022. *LIFE-M: The Longitudinal Intergenerational Family Electronic Micro-Database*. Inter-University Consortium for Political and Social Research. https://doi.org/10.3886/E155186.

Bailey, Martha J., A. R. Shariq Mohammed, and Paul Mohnen. 2023. "US Educational Mobility in the Early Twentieth Century." UCLA Working Paper, November.

Bailey, M., P. Z. Lin, A. R. S. Mohammed, P. Mohnen, J. Murray, M. Zhang, and A. Prettyman. 2023. "The Creation of LIFE-M: The Longitudinal, Intergenerational Family Electronic Micro-Database Project." *Historical Methods*: *A Journal of Quantitative and Interdisciplinary History* 56 (3): 138–59.

Becker, G. S. 1973. "A Theory of Marriage: Part I." *Journal of Political Economy* 81 (4): 813–46.

Becker, G. S. 1974. "A Theory of Marriage: Part II." *Journal of Political Economy* 82 (2): S11–S26.

Black, S. E., and P. J. Devereux. 2011. "Recent Developments in Intergenerational Mobility." *Handbook of Labor Economics* 4b: 1487–1541.

Buckles, Kasey, Joseph Price, Zachary Ward, and Haley E. B. Wilbert. 2023. "Family Trees and Falling Apples: Historical Intergenerational Mobility Estimates for Women and Men." NBER Working Paper No. 31918. National Bureau of Economic Research.

Calvo, P., I. Lindenlaub, A. and Reynoso. 2021. "Marriage Market and Labor Market Sorting." NBER Working Paper No. 28883. National Bureau of Economic Research.

Chadwick, L., and G. Solon. 2002. "Intergenerational Income Mobility Among Daughters." *American Economic Review* 92 (1): 335–44.

Charles, K. K., E. Hurst, and A. Killewald. 2013. "Marital Sorting and Parental Wealth." *Demography* 50 (1): 51–70.

Chetty, R., D. Grusky, M. Hell, N. Hendren, R. Manduca, and J. Narang. 2017. "The Fading American Dream: Trends in Absolute Income Mobility since 1940." *Science* 356 (6336): 398–406.

Chetty, R., N. Hendren, P. Kline, and E. Saez. 2014a. "Where Is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States." *Quarterly Journal of Economics* 129 (4): 1553–1623.

Chetty, R., N. Hendren, P. Kline, E. Saez, and N. Turner. 2014b. "Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility." *American Economic Review: Papers & Proceedings* 104 (5): 141–47.

Ciscato, E., and S. Weber. 2020. "The Role of Evolving Marital Preferences in Growing Income Inequality." *Journal of Population Economics* 33 (1): 307–47.

Collins, W. J., and M. H. Wanamaker. 2022. "African American Intergenerational Economic Mobility Since 1880." *American Economic Journal: Applied Economics* 14 (3): 84–117.

Currie, J., and D. Almond. 2011. "Human Capital Development Before Age Five." In *Handbook of Labor Economics*, 4:1315–1486. Elsevier.

Deutscher, Nathan, and Bhashkar Mazumder. 2023. "Measuring Intergenerational Income Mobility: A Synthesis of Approaches." *Journal of Economic Literature* 61 (3): 988–1036.

DiNardo, J., N. Fortin, and T. Lemieux. 1996. "Labor Market Institutions and the Distribution of Wages, 1973–1992: A Semi-Parametric Approach." *Econometrica* 64 (5): 1001–44.

Eika, L., M. Mogstad, and B. Zafar. 2019. "Educational Assortative Mating and Household Income Inequality." *Journal of Political Economy* 127 (6): 41.

Eriksson, K., G. T. Niemesh, M. Rashid, and J. Craig. 2023. "Marriage and the Intergenerational Mobility of Women: Evidence from Marriage Certificates 1850–1910." Working Paper. https://niemesgt.github.io/files/WomenMobilityMarriage.pdf.

Ermisch, J., M. Francesconi, and T. Siedler. 2006. "Intergenerational Mobility and Marital Sorting." *Economic Journal* 116 (513): 659–79.

Feigenbaum, J. J. 2018. "Multiple Measures of Historical Intergenerational Mobility: Iowa 1915 to 1940." *Economic Journal* 128 (612): F446–F481.
Fernández, R., and R. Rogerson. 2001. "Sorting and Long-Run Inequality." *Quarterly Journal of Economics* 116 (4): 1305–41. http://www.jstor.org/stable/2696460.
Ferrie, J. P. 1996. "A New Sample of Males Linked from the Public Use Microdata Sample of the 1850 U.S. Federal Census of Population to the 1860 U.S. Federal Census Manuscript Schedules." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 29 (4): 141–56.
Gihleb, Rania, and Kevin Lang. 2020. "Educational Homogamy and Assortative Mating Have Not Increased." *Research in Labor Economics* 48: 1 - 1 (26).
Gilbertson, G. A. 1996. "Hispanic Intermarriage in New York City: New Evidence from 1991." *International Migration Review* 16:445–59.
Goldin, C. 1998. "America's Graduation from High School: The Evolution and Spread of Secondary Schooling in the Twentieth Century." *Journal of Economic History* 58 (2): 345–74.
Greenwood, J., N. Guner, G. Kocharkov, and C. Santos. 2014. "Marry Your Like: Assortative Mating and Income Inequality." *American Economic Review* 104 (5): 348–53.
Greenwood, J., N. Guner, G. Kocharkov, and C. Santos. 2016. Technology and the changing family: A unified model of marriage, divorce, educational attainment, and married female labor-force participation. *American Economic Journal: Macroeconomics* 8 (1): 1–41.
Haider, S., and G. Solon. 2006. "Life-Cycle Variation in the Association Between Current and Lifetime Earnings." *American Economic Review* 96 (4): 1308–20.
Heckman, James J., Hidehiko Ichimura, Jeff Smith, and Petra Todd. 1998. "Characterizing Selection Bias Using Experimental Data." *Econometrica* 66 (5): 1017–98.
Helgertz, J., S. Reggles, J. Warren, C. Fitch, R. Goeken, J. D. Hacker, M. A. Nelson, J. Price, E. Roberts, and M. Sobek. 2020. *IPUMS Multigenerational Longitudinal Panel: Version 1.0* [dataset]. IPUMS.
Jácome, E., I. Kuziemko, and S. Naidu. 2024. "Mobility for All: Representative Intergenerational Mobility Estimates over the 20th Century." *Journal of Political Economy* 132 (12).
Kalmijn, M. 1991. "Shifting Boundaries: Trends in Religious and Educational Homogamy." *American Sociological Review* 56 (6): 786.
Kalmijn, M. 1994. "Assortative Mating by Cultural and Economic Occupational Status." *American Journal of Sociology* 100 (2): 422–52.
Kalmijn, M. 1998. "Intermarriage and Homogamy: Causes, Patterns, Trends." *Annual Review of Sociology* 24 (1): 395–421.
Kremer, M. 1997. "How Much Does Sorting Increase Inequality?" *Quarterly Journal of Economics* 112 (1): 26.
Lam, D. 1988. "Marriage Markets and Assortative Mating with Household Public Goods: Theoretical Results and Empirical Implications." *Journal of Human Resources* 23 (4): 462.
Long, J., and J. Ferrie. 2013. "Intergenerational Occupational Mobility in Great Britain and the United States Since 1850." *American Economic Review* 103 (4): 1109–37.
Moulton, B. R. 1986. "Random Group Effects and the Precision of Regression Estimates." *Journal of Econometrics* 32 (3): 385–97.
Olivetti, C., and M. D. 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.

Price, Joseph, Kasey Buckles, Jacob Van Leeuwen, and Isaac Riley. 2021. "Combining Family History and Machine Learning to Link Historical Records." *Explorations in Economic History* 80: 1–28.
Qian, Z. 1997. "Breaking the Racial Barriers: Variations in Interracial Marriage Between 1980 and 1990." *Demography* 34 (2): 263–76.
Rolf, K., and J. Ferrie. 2008. "The May–December Relationship Since 1850: Age Homogamy in the U.S." Working paper.
Ruggles, S., C. A. Fitch, R. Goeken, J. D. Hacker, M. A. Nelson, E. Roberts, S. Megan, and M. Sobek. 2021. *IPUMS Ancestry Full Count Data: Version 3.0* [dataset]. IPUMS.
Schoen, R., and B. Thomas. 1989. "Intergroup Marriage in Hawaii, 1969–1971 and 1979–1981." *Sociological Perspectives* 32 (3): 365–82.
Shorter, E. 1977. *The Making of the Modern Family*. Basic Books.
Solon, Gary. 1992. "Intergenerational Income Mobility in the United States." *American Economic Review* 82 (3): 393–408.
Song, X., C. G. Massey, K. A. Rolf, J. P. Ferrie, J. L. Rothbaum, and Y. Xie. 2020. "Long-Term Decline in Intergenerational Mobility in the United States since the 1850s." *Proceedings of the National Academy of Sciences* 117 (1): 251–58.
Stevenson, Betsey, and Justin Wolfers. 2007. "Marriage and Divorce: Changes and Their Driving Forces." *Journal of Economic Perspectives* 21 (2): 27–52.
Tan, H. R. 2023. "A Different Land of Opportunity: The Geography of Intergenerational Mobility in the Early Twentieth-Century United States." *Journal of Labor Economics* 41 (1): 77–102.
Ward, Z. 2021. "Intergenerational Mobility in American History: Accounting for Race and Measurement Error." NBER Working Paper No. 29256. National Bureau of Economic Research.
Wimmer, L. T. 2003. "Reflections on the Early Indicators Project: A Partial History." In *Health and Labor Force Participation over the Life Cycle: Evidence from the Past*, edited by Dora L. Costa, 1–11. University of Chicago Press.