The her in inheritance: Matching and mobility in two million Quebec marriages, 1800–1970

Matthew Curtis*

December 2021

Abstract

When did strongly assortative marriages begin to increase inequality? Many believe it is a modern development, a consequence of rising female employment and economic empowerment. Using a large novel dataset from Quebec, Canada, I find that marriage was highly assortative as far back as the early nineteenth century. Moreover, matching was not merely between families of similar socioeconomic status, but instead depended on the human capital of the women themselves. Finally, despite their lack of formal employment and legal rights, I show that women mattered as much as their husbands for child outcomes. As a consequence and despite deeply conservative gender norms, the human capital of women mattered both for marriage and for its outcomes long before they entered the labor force.

Keywords Assortative mating, marriage, sorting, human captial, intergenerational mobility

JEL codes J12, J62, N31, N32

*ECARES, Université Libre de Bruxelles. matthew.curtis@ulb.be.

I am grateful for the support and excellent advice of my dissertation committee: Gregory Clark (Chair), Katherine Eriksson, and Santiago Pérez. Special thanks to Michael Dearing, whose generous donation to support economic history at UC Davis helped make this project possible. Thanks to Jean-Sébastien Bournival and Lisa Dillon for help with the dataset. Thanks also to Paula Gobbi, Joanne Haddid, Laura Salisbury, José-Antonio Espín-Sánchez, Zachary Ward, Yuzuru Kumon, and the participants at the Economic History Association Annual Meeting, the All UC Graduate Economic History Workshop, and informal seminar participants at ECARES and UC Davis for helpful feedback and comments. All errors that remain are my own.

1 Introduction

Women had limited employment and educational opportunities until the mid-20th century (Goldin (2006)). As married women began entering the labor force en-masse and higher education gaps closed, in theory there should have been increased inequality due to assortative marriage (Fernández and Rogerson (2001)). Specifically, after women became formally employed, their similarity with their husbands would directly matter for household income inequality. However, the empirical evidence of an increase in inequality due to sorting is mixed (Eika et al. (2019)).

To help explain this contemporary evidence, this paper provides a long-run perspective on assortative marriage. It uses a novel and unusual rich dataset from the Canadian Provence of Quebec 1800–1960, which at the time had relatively conservative gender norms for North America. It finds that, long before contemporary female empowerment, marriage was highly assortative and mattered for inequality through the channel of intergenerational mobility.

To provide empirical evidence to support this claim, this paper sets out to answer three related questions. First, how did the degree of marital assortment evolve over the long run? Estimating sorting with historical data is challenging due to limited information on married women (Olivetti et al. (2020)). Second, did grooms specifically value the human capital of their brides? If not, any estimated correlation between spouses could be incidental, the result of marriages being negotiated between families of similar socioeconomic status (e.g. Puga and Trefler (2014)). Third, how much did mothers influence child outcomes independently from the fathers? As marriages were assortative, it is challenging to untangle the independent effect of a mother from her correlation with the father (Espín-Sánchez et al. (2022)). Together, the answers to these questions demonstrate how assortative marriage increased inequality long before married women entered the labor force.

I am able to answer these questions by using a novel dataset containing millions of marriage records from Quebec 1800–1970.[1] These data have several unique features that

[1] The dataset, the Infrastructure Intégrée Des Microdonnées Historiques de La Population Du Québec

make them particularly suitable to answer these questions. First, Québécoise women retained their family name after marriage and thus can be linked to their parents. Linking married women is much harder in societies such as the United States where women typically take their husbands' surnames (Craig et al. (2020)). Second, the data are close to a complete population registry. Families in the sample are not selected by cohabitation (like in census records) or by living descendants (like in most genealogical datasets). The large sample size and complete family linkages allows me to untangle the underlying mechanisms linking sorting and mobility.

To measure the degree of assortment over time, I develop a new method to estimate the correlation between the human capital of spouses. The data have two complementary proxies for a woman's human capital: her ability to sign her name and the occupational status of her husband or father (Clark (2007) and Craig et al. (2020)). The ability to write is a direct measure of human capital that is consistently available for both men and women. However, the ability to sign one's name is a very low threshold and human capital was increasing dramatically over the sample period. Therefore, to construct more meaningful estimates of sorting over time, I use an indirect measure of female human capital: the occupation of the bride's fathers. Based on a simple model with strong assumptions (but as I show, reasonable ones in this specific context), I propose a novel method to estimate the unobserved correlation between spouses. Using this method, I find that this estimated correlation was surprisingly high — around 0.85 — throughout the entire period of 1800–1970.

Next, I show this sorting was in fact due to matches formed on the human capital of both the groom and the spouse. Any correlation between spouses could, after all, be purely due to matching between families of similar social status or between the groom and his father-in-law. To show that matches were formed based on individual characteristics, I take advantage of the within-family variation in the dataset. Using family-fixed effects, I find that literate

(IMPQ), combines two previous databases, the BALSAC database and the RPQA (IMPQ (2020) and Project Balsac (2020), PRDH (2020)). Individually, the datasets have been used to study questions related to social mobility (Gagnon et al. (2011) and Galor and Klemp (2019)). However, the the IMPQ has newly linked families across the two, allowing a much longer period of time to be studied (Dillon et al. (2018)).

women typically married husbands with greater human capital — 30 percentage points more likely to be literate — than their illiterate sisters. The reverse was true for brothers. Notably, the returns to literacy were symmetric across gender. When it came to marriage, the human capital of women mattered as much as that of men.

Finally, I use remarriages to show that the human capital of women mattered for child outcomes. A child's literacy was as strongly correlated with their mother's literacy as it was with the father's. However, as long as marriages were assortative, this pattern of correlation would still be observed even if the mother had no independent effect on child outcomes. To address this empirical challenge, I use the unusually high frequency of remarriage to control for the father. Both fixed effects and the direct comparison of half-siblings provides evidence that the human capital of mothers had a causal impact on child outcomes independently from that of fathers. Moreover, the estimated effect of mothers was remarkably similar in magnitude to that of fathers.

All together, this is strong evidence that assortative marriage contributed to the overall distributions of human capital and socioeconomic status in the Canadian Province of Quebec. These results also have implications for standard estimates of social mobility. I show that both women and assortative marriage directly influence estimates of rates of mobility even when only data on males are used. A measured increase in father-son correlations, for example, could be entirely driven by an increase in either assortment or in the transmission of maternal human capital.

This paper thus contributes directly to two literatures. First, the paper provides longer-run context to the literature on marriage sorting and inequality. Starting around 1970, developed economies have seen both an increase in female labor force participation rates and in inequality (Goldin (2006) and Piketty and Saez (2014)). This has led to widespread concern that the two are linked, with assortative marriage increasing the dispersion of household incomes. However, economic theory is ambiguous as to the effect of sorting; changing a few assumptions results in very different structural estimates of the effect of sorting on inequality (Kremer (1997), Fernández and Rogerson (2001)). On balance, little empirical evidence has been found that sorting has increased inequality (Greenwood et al. (2014a,b) and Hryshko et al. (2017), Eika et al. (2019).). This paper provides a longer-run context that suggests an explanation: marriage had long been highly assortative. It also adds to the growing literature on historical assortative marriage (Abramitzky et al. (2011a), Craig et al. (2020) and Goñi (2018)).

Second, this paper adds to our understanding of intergenerational mobility over the long run. Until recently, studies of intergenerational mobility have often overlooked women (Black and Devereux (2011)). More recent work has emphasized the need to focus on the mobility of daughters as well as sons (Chadwick and Solon (2002)). However, there are major data challenges to overcome in historical studies of female mobility. Measures of female socioeconomic status are rarely reported, so studies often rely on husbands' incomes or occupations as a proxy (Dribe et al. (2019)). Second, in many societies, linking married daughters to fathers requires either pseudo-linkages on first names or recovering their maiden names from marriage records. (Olivetti and Paserman (2015), Olivetti et al. (2020), Goñi (2018), Craig et al. (2020)). As the records from Quebec have both a proxy for female human capital (signatures) and direct female linkages, they are unusually well suited to considering historical mobility. Moreover, this paper joins Espín-Sánchez et al. (2022) in explicitly considering the role of mothers as well as daughters. In doing so, it adds to the growing literature that considers the role of relatives other than fathers in historical mobility (Long and Ferrie (2018) and Olivetti et al. (2018)).

The structure of the paper is as follows. I start with a discussion of the historical context, arguing that Quebec 1800–1970 is a good setting to consider assortative marriage. Then I describe the data, highlighting its unique strengths and explaining how I construct measures of human capital. Next, I present three related empirical findings. First, I develop a novel method to estimate of the degree of assortment, finding it to be surprisingly high and stable throughout the period. Second, I present evidence that this assortment consisted of matches between individuals, not families. Third, I estimate the independent causal effect of maternal human capital on child outcomes. Then, I discuss the broader implications of these findings for the historical mobility literature. Finally, I conclude: assortment mattered for inequality long before the mid-20th century because women had always played an important role in marriage and mobility.

2 Historical context

While the uniquely high quality data are its main strength, Quebec's historical economy and institutions also make it a useful setting to consider the role of women in assortative marriage and social mobility over the long run. In this section, I argue that while Quebec is in many ways very similar to the rest of North America, it had relatively more conservative gender norms. While its economic development lagged other North American regions, it followed the same trends. For example, it had lower wages until the mid-20th century but the gap was stable over time. (Albouy (2008), Geloso and Lindert (2020)). Before its Quiet Revolution of the 1960's, Quebec was also much less secular than its neighbors.[2] Catholicism asserted significant control over public education and social norms, and deeply conservative beliefs about gender roles were enshrined by law and public policy. Finally, for a population of European descent, the Québécoise had a late demographic transition and unusually large family sizes. Married women would have spent much of their adult life pregnant and raising small children. Altogether, Quebec before the mid-20th century is not a promising time or place to find an important role of female human capital in assortative marriage and child outcomes. As this paper finds just that, it is then likely that it was also the case in other places with higher levels of female empowerment.

[2]The term quiet revolution is also used for the increase in female labor force participation and educational achievement of women in many countries starting in the 1970's (Goldin (2006)).

2.1 The legal rights of women

Women in most historical societies faced systematic legal disadvantages; Quebec was no exception. While Quebec was ceded to the British in 1763, laws pertaining to civil matters remained governed by the Coutume de Paris (Custom of Paris), a codified system of customary French law. Under the Coutume de Paris, and unlike in English-speaking legal traditions, married couples formed a legal entity called the *communauté de biens* (community of property) in which both partners theoretically had equal stakes (Greer (1997)). As a consequence, both the husband and wife were required to sign legal documents,[3] though the husband alone was expected to manage the joint property.

After marriage, women were legally considered incapable, being unable to independently form contracts or initiate lawsuits (Baillargeon (2014)).The reformed Civil Code of Lower Canada, introduced in 1866, only clarified the legal disadvantages of women. While Québécoise women could vote in federal elections after 1918, they could not vote in local elections until 1940 (Tremblay and Roth (2010)).[4] Only after reforms starting in 1964 were women no longer considered legally incapable.

In theory, the law did not discriminate when it came to the inheritance of daughters. The community of property was dissolved by giving the surviving spouse their half and dividing the rest equally amongst the children regardless of gender.[5] Perhaps as a consequence of being unable to write children out of a will, parents had little legal recourse to block a match they disapproved of after the children reached a certain age (Greer (1997)). However, some parents attempted to circumvent the laws by "gifting" property to favored heirs, typically an older son (Greer (1985)).

[3]As mentioned below, this greatly aids the linking of vital records.

[4]Unusually, from the first elections in 1792 until 1849, suffrage was only restricted to individuals meeting age and property requirements; a very small number of women who independently owned property could and did vote based on these criteria. This was considered by 19th century reformers a concerning oversight that needed to be addressed.

[5]One of the few advantages women had was that a widow could renounce the debts of the community of property as it was assumed she was not responsible for their accumulation.

2.2 The demographic regime

Was an unequal partnership in marriage the typical experience for women in Quebec? Before its demographic transition, Quebec had a variant of the European marriage pattern, with earlier marriages and less frequent celibacy than France (Greer (1997)). Most women married, and for most women marriage marked the beginning of many years of pregnancy and childcare. While married, a woman typically gave birth to a child roughly every two years until her forties. Unlike in many historical societies, quick remarriage upon the death of a spouse was common and widows did comparatively well on the marriage market.

One possible factor contributing to this high fertility marriage pattern is that parents and clergy were unlikely to oppose a marriage. It was not costly to start a new household, so parents had little leverage to prevent a match once children were of age. Even if they still required parental consent, the Church could grant exemptions.[6].

While high fertility was common in most settler colonies, Quebec sustained it longer than most. The demographic transition occurred relatively late, only reaching substantial numbers of French-speaking Québécois by the 1920's (Vézina et al. (2014)). Moreover, from first settlement through at least 1835, there appears to have been no attempt of parents to target a specific family size (Clark et al. (2020)). Therefore, at least in the earlier part of the data, married women would spend much of her life pregnant and raising small children.

2.3 Female labor

While the economy of Quebec evolved dramatically from 1800 to 1970, opportunities were persistently limited for married women in the formal labor market. While some women had an important role in their family's business, most women were expected to expected to perform onerous housekeeping labor (Baillargeon (2014)). Before the widespread introduction of labor-saving household devices, simple yet tedious tasks like washing and ironing clothes

[6] Which, at least in the comparable situation of first cousins wishing to marry, they were unusually liberal in granting. This was apparently due to the credible threat of cohabitation or defection to Protestant churches (Greer (1985))

took up vast amounts of time for women (as they did in the US; Greenwood et al. (2003)). Unmarried women in urban areas could work outside the household, but at first most were employed as servants facing the same domestic drudgery in their employer's household (Baillargeon (2014)). A few found employment as educators, first as nuns and and later as secular teachers. As the economy began to industrialize in the 1840's, unmarried women were also employed by factories (typically clothing or tobacco), albeit with substantially lower wages than men. Industrialization also led to the decline of household manufacturing and the rise of the male breadwinner household, further delegating married women to housekeeping labor (similar to other industrializing economies; de Vries (2008)). By the late 19th and early 20th century, occupations dominated by unmarried women emerged such as telephone operators, typists, and secular nurses. However, married women were still expected to be housewives until the 1970's.[7]

2.4 External validity

Overall, how much was Quebec an outlier? In general, it was characteristically a North American economy, albeit one somewhat lagging its neighbors in economic development. Its deeply conservative society delayed the extension of rights to women, but not indefinitely. Its demographic regime was characterized by large family sizes and a delayed demographic transition, but it was still a variant of the European marriage pattern. The role of women in its labor force evolved roughly the same as the rest of North America (Goldin (2006)). If women and assortative marriage mattered for mobility even in conservative Quebec, they probably mattered in neighboring regions. I argue, therefore, that my findings from Quebec are very likely generalizable to the rest of North America (and probably Western Europe as well).

[7]By the 1940's, female occupations start to be reported in the marriage records. The most common female occupation reported is "ménagère" (housekeeper).

3 Data

The Infrastructure Intégrée Des Microdonnées Historiques de La Population Du Québec (IMPQ) is a large new database of family reconstitutions from baptism, burial, and marriage records (IMPQ (2020)). It integrates two previous databases, the BALSAC database and the RPQA (Project Balsac (2020), PRDH (2020)). While it contains data as far back as the founding of the colony, in this paper I use data from a period with frequently reported occupations for men, 1800–1969. While the dataset is still being extended, as of writing it contains 1.4 million unique births, 0.6 million unique deaths, and 2.1 million unique marriages from 1800–1969 (though births and deaths are limited to a particular sub-region after 1849). Moreover, in those records a total of 2.7 million other individuals are mentioned besides the main participants, providing additional observations over time for many people besides their own vital events.

Table 1 presents summary statistics from the dataset. The main unit of observation is a marriage, linked to those of both the groom and the bride's parents. Each observation contains, when available, information about the human capital of the bride, groom, and all four parents. Below, I discuss in detail how the links were constructed, what measures of human capital I use, and how well the data report these measures for women.

3.1 Linked family vital records

Two unusual institutional features of Quebec have resulted in vital records that are particularly easy to link. First, due to the system of community property, both husbands and wives signed their names on all legal documents. Second, women kept their family names when they married. This means both that women can be linked to their fathers and that most vital records have four names on which to link (the first names and last names of both the husband and wife or mother and father).

In both of the databases, links were formed using two similar computer-assisted matching systems (Vézina et al. (2013), Dillon et al. (2018)). These two procedures differ slightly, but they both use all four names to link records and resort to manual linkage in difficult cases.[8] Manual linkages are not necessarily better than automatic linkages; in some applications they produce both more true matches and more false positives (Abramitzky et al. (2019)). However, the fact that the Quebec vital records have four names to match on should increase the accuracy of matching regardless of the method used. Moreover, the parish records of Quebec have survived remarkably intact as local priests were required to send duplicates of all records to their superiors (Dillon et al. (2018)). Therefore, records of almost the entire population survive; this will reduce false positive rates in an analogous way to the linking of full count to full count censuses (Abramitzky et al. (2019)).

3.2 Measures of human capital

The direct measure of human capital, for both men and women, that I use in this paper is the presence of a signature on a marriage record. Signatures have often been used as a proxy for literacy (c.f. A'Hearn et al. (2009)). In Quebec, Catholic churches had long required required both the bride and the groom to sign their marriage records if they were able and the priest to record if they were not (Gagnon et al. (2011)). I code a signature indicator variable as one if the individual signed their marriage record, and zero if they were unable to sign. I omit cases that are either missing or ambiguous. As shown in Figure 1 below, this definition produces a trend that is close to external estimates of literacy.

Was a literacy really a measure of human capital, that is a productive attribute? The qualitative evidence suggests it was. The ability to write is a form of human capital that had always been particularly associated with business activity in Quebec (Greer (1997)). As for reading, it too was likely associated with economic activity in Quebec. As opposed to, say, their majority Protestant neighbors in New England who prioritized literacy education for religious reasons, for Quebec's Catholics reading the Bible was not a religious necessity.

A second proxy for human capital, only available for men, is occupational status. I assign

[8]For example, the BALSAC database standardizes names using the FONEM phonetic algorithm (Bouchard et al. (1981)), whereas the RPQA uses a custom-made name dictionary (Dillon et al. (2018)).

each individual the occupation listed at their first marriage (if any). The occupations are assigned HISCO codes, a classification system designed for comparative studies of historical social mobility (Van Leeuwen et al. (2004)). I then assign various occupational status scores to these HISCO codes aggregated to the three digit level. The primary score I construct is the imputed 1901 earnings, the average yearly earnings reported by men with that occupation in a 5% sample of the 1901 Canadian Census (Canadian Families Project (2002) and Minnesota Population Center (2019)).[9] There are numerous other ways to rank occupations, as discussed in Appendix A3. However, imputed 1901 earnings are easy to interpret (how much the individual would earn, on average, with their occupation in 1901 in Canada), are imputed from data roughly in the middle the time period considered, are at least a proxy for the standard variable of interest in intergenerational mobility studies (lifetime earnings), and produce similar estimates to the other occupational scores. Finally, I argue it is reasonable to assume that the average earnings of an occupation is strongly related to the average level of human capital of those with the occupation. Therefore, the main results in this paper use the imputed 1901 earnings as the primary measure of occupational status.

[9]This requires crosswalking occupations from IPUMS's occupational codes to the original HISCO scheme (Zijdeman (2014)).

Table 1: Summary statistics

Variable	Mean	SD	Min	Max	N obs.
Marriages					
Year	1914	43	1800	1969	2,122,695
Brides					
Sibling order	2.20	1.50	1	17	1,994,133
Marriage number	1.05	0.24	1	7	2,122,695
Signature	0.81	0.39	0	1	1,993,738
Signature, mother	0.67	0.47	0	1	1,670,305
Signature, father	0.61	0.49	0	1	1,662,519
Imputed 1901 earnings, father	357	203	75	2,000	1,010,907
Grooms					
Sibling order	2.26	1.60	1	20	1,983,328
Marriage number	1.10	0.33	1	7	2,122,695
Signature	0.79	0.41	0	1	1,986,399
Signature, mother	0.65	0.48	0	1	1,618,007
Signature, father	0.59	0.49	0	1	1,610,059
Imputed 1901 earnings	417	254	75	2,400	1,184,986
Imputed 1901 earnings, father	358	204	75	2,000	989,747

Note: Sibling order is the order amongst all married siblings by date of first marriage (as birth dates are not reported after 1849). Marriage number is the number of the marriage for the individual. Signature variables are indicators that are one if a signature was recorded, zero if the absence of a signature was recorded, and omitted otherwise. Imputed earnings are the average earnings for men with the individual's occupation in the 1901 Canadian Census sample.

3.3 Are women accurately reported?

Do the vital records accurately record the occupational status and human capital of women? Four extracts of Canadian censuses 1881–1911 and data compiled by Long (1958) for 1920–1960 provide external points of comparison (Dillon et al. (2008), Inwood and Jack (2011), Canadian Families Project (2002), Gaffield et al. (2009), Minnesota Population Center (2019), Long (1958), Killingsworth and Heckman (1986)).[10] Figure 1 compares the fraction of individuals signed their first marriage record to the fraction who self-reported the ability

[10] The census extracts are the 100% 1881 sample, the 5% 1891 sample, the 5% 1901 sample and 1901 oversample, and the 5% 1911 sample.

to write in the censuses. Unlike the censuses, individuals only appear in the marriage records during a specific time in their lives. To account for this, I reweight the census data to match the age distribution of the vital records. As shown in the figure, my estimated literacy rate closely tracks the rate in the census. Two patterns are particularly notable. First, Quebec went from a very low human capital society to a high human capital society from 1800 to 1920. Second, there was actually a gender gap in favor of women from 1850 to 1920.[11]

In contrast, the vital records do a poor job of recording female occupations. Figure 2 shows the employment rate of women by marital status. Here, I reweight the vital records data to match the age distribution in the censuses. Compared to the other sources, the vital records underestimate the employment rate of married women and almost entirely omit unmarried women with occupations. One pattern, however, is clear. While unmarried women often worked outside the home, married women did not begin to work in substantial numbers until the 1940's.

[11] Perhaps this gender gap in favor of women has its roots in the relative effectiveness of teaching nuns in the provence (Magnuson (1992)).

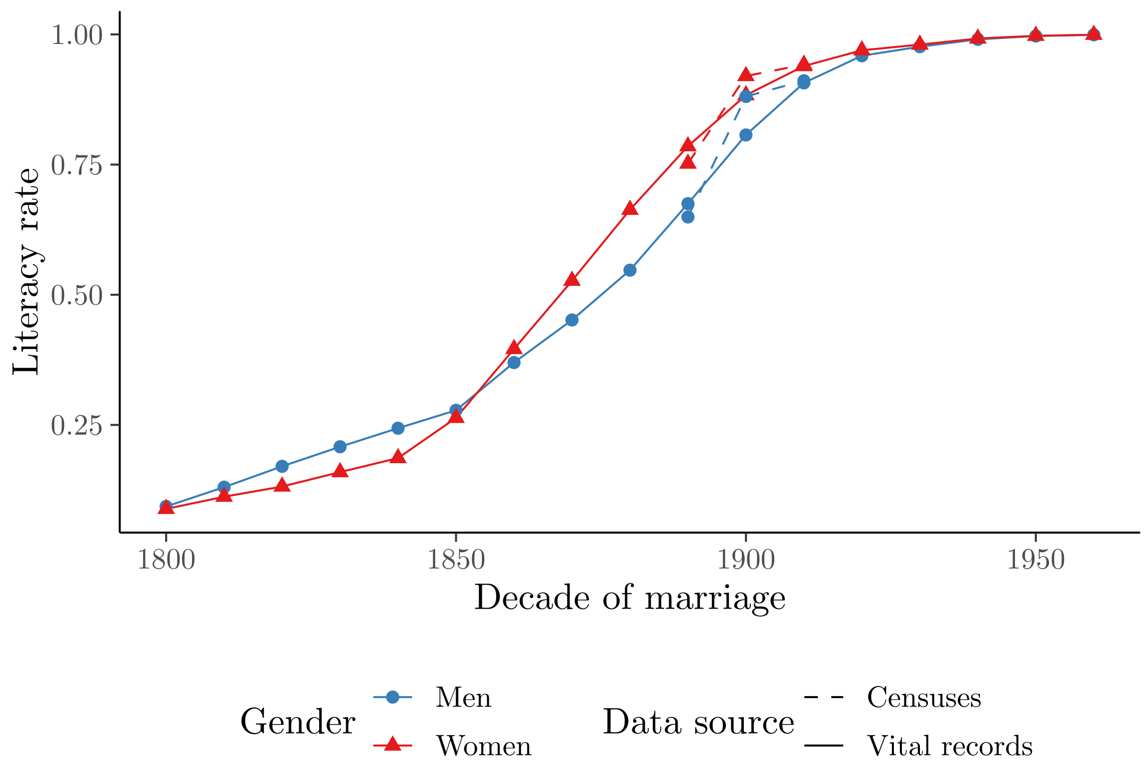

Figure 1: The vital records accurately report the ability to write
Note: The vital record literacy rate is the average of an indicator variable that is one if a signature was recorded, zero if the absence of a signature was recorded, and omitted otherwise. The census record literacy rate is the fraction of individuals who were reported as able to write, reweighted to match the age distribution in the vital records. The two sources broadly agree.

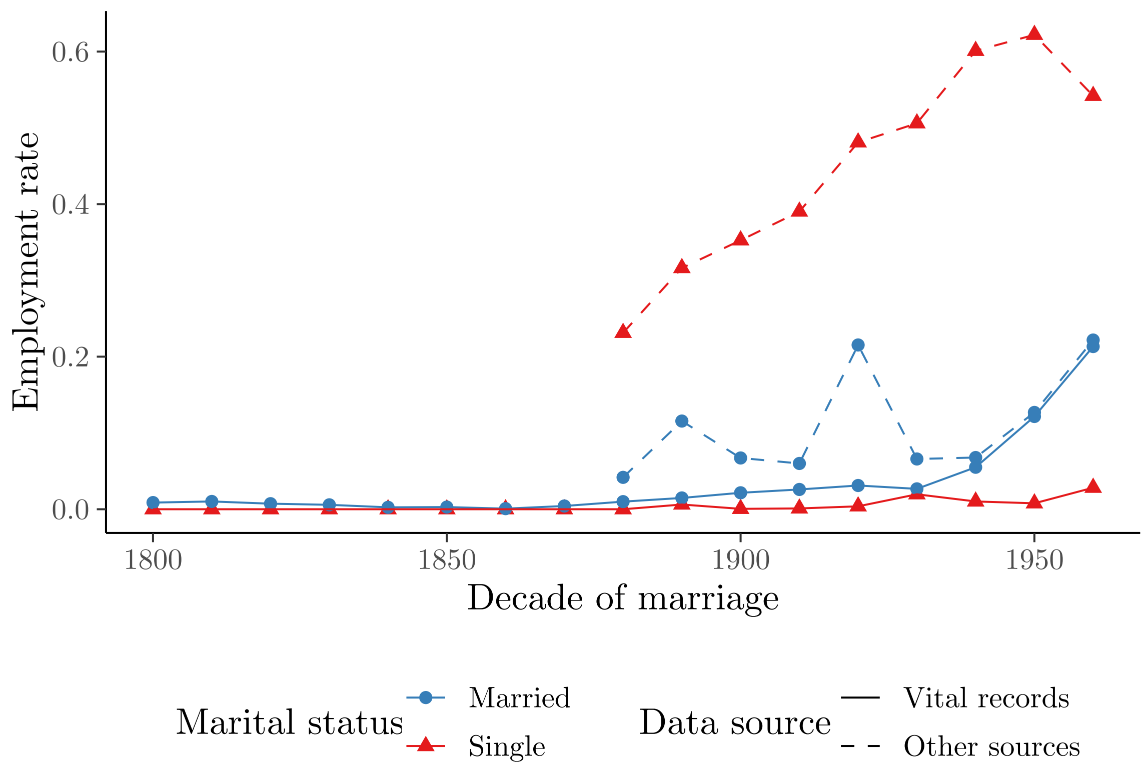

Figure 2: The vital records fail to record female employment
Note: Each woman in the vital records is counted as employed if she ever had a job reported in any record. Otherwise, she is counted as not employed. She is then assigned a year equal to the median of all the years at which she is observed. Then, I compute the average employment rate for each decade, reweighted by age to match the age distribution in the census data. Before 1920, the other sources are census extracts and the rates the fraction of women aged at least 16 with an occupation. After 1920, the other sources are rates compiled by Long (1958), with the rate for married women calculated as an average of the rates for currently married women and widowed or divorced women weighted by the relative frequencies of the two categories in the censuses.

4 Simple model of marriage and mobility

Below, I develop a model to illustrate how assortative marriage and intergenerational mobility contribute to inequality over the long run. This model, while simple, suggests a new method to measure the degree of marital assortment. Using this method, I show in the following section that assortment was surprisingly high and stable over the period 1830–1969.

4.1 The model

We are interested in a specific measure of socioeconomic status, call it X. In the intergenerational mobility literature, the typical measure of status is lifetime earnings. Here, I consider human capital. This distinction matters because while women often were not employed, they still possessed human capital that could have, under different circumstances, be used to earn a wage.

Following Solon (1992) and Clark and Cummins (2015), assume only an imperfect measure or proxy Y is observed for X. The typical example is proxying for lifetime earnings with the earnings in a specific year. In my context, for some individuals no Y is observed. For the others, Y is observed with classical measurement error:

$$Y_i = X_i + u_i \tag{1}$$

for individual i, where u_i is an error term uncorrelated with X_i.

Then, as in Espín-Sánchez et al. (2022), assume that the human capital of child i, X_i, is inherited depending on the human capital of the child's father X_i^f and of their mother X_i^m:

$$X_i = \beta_f X_i^f + \beta_m X_i^m + e_i \tag{2}$$

where e_i is a random term uncorrelated with the X's. For now, I assume that the effect on children is the same regardless of gender. While this seems a strong assumption, it makes the model much more tractable and I will later provide evidence that it appears reasonable in my context.[12] Moreover, note that both error terms, u_i and e_i, can be allowed to vary by the gender of the child to model the likely situation where non-heritable factors result in gender-biased data quality or outcomes.

Following Chadwick and Solon (2002), assume that the sorting on human capital can be

[12]Though this certainly not the case in every context. For example, while Mazumder (2005) finds very little difference between genders in contemporary US data, Espín-Sánchez et al. (2022) finds that children inherited more strongly from their same-sex parent in pre-modern Spain.

summarized by:

$$corr(X_i^f, X_i^m) = \gamma \tag{3}$$

Now note that if two variables have equal variances, I can re-write the sorting correlation equation as a linear relationship:

$$X_i^m = \gamma X_i^f + v_i^f \tag{4}$$

where v_i^f is an uncorrelated error term. If I substitute this into the intergenerational mobility equation, I get:

$$X_i = (\beta_f + \gamma\beta_m)X_i^f + \gamma\beta_m v_i^f + e_i \tag{5}$$

This can be estimated with a regression:

$$Y_i = \alpha_0 + \alpha_1 Y_i^f + \epsilon_i \tag{6}$$

where $\alpha_1 = (\beta_f + \gamma\beta_m)$. However, as Y_i^f is correlated with u_i^f the estimate is attenuated downwards. Specifically, as:

$$Y_i = (\beta_f + \gamma\beta_m)Y_i^f - (\beta_f + \gamma\beta_m)u_i^f + \beta_m v_i^f + e_i - u_i \tag{7}$$

there is bias of the form:

$$plim\ \hat{\alpha}_1 = (\beta_f + \gamma\beta_m)\frac{\sigma^2_{X_i^f}}{\sigma^2_{X_i^f} + \sigma^2_{u_i^f}} \tag{8}$$

If β_f and γ are greater than zero, then mothers contribute to the observed correlation of X between fathers and sons. Even if mothers are not observed, they therefore could be driving results in the typical intergenerational mobility regression used in the literature.

Section 6 below discusses further the implications of this omitted variable problem.

To conclude, note now that γ, the degree of assortment, is part of the equation determining the intergenerational association between fathers and sons. As long as $\beta_m > 0$ — that is, as long as the human capital of mothers has a direct effect on that of sons — assortment will slow social mobility, increasing overall inequality.[13] Therefore, the empirical agenda of this paper is to estimate γ, provide evidence that the estimate is not spurious, and then show that $\beta_m > 0$.

4.2 Measuring assortment if women are not observed

Typically, assortment is measured by the correlation between the education levels of spouses. However, as shown in Figure 1, the average signature rate changed dramatically during this interval. As the average education level increased, the average education level of individuals who could at least sign their name would have also increased. Therefore, the relationship between the underlying degree of assortment on human capital and the observed degree of assortment on signature rates is likely changing over time.

A more stable proxy for human capital is an individual's occupational status. However, in most of the sample married women have no observed occupational status (and in the mid 20th century when it became more common, they are concentrated in lower status occupations such as housekeeping).

Using the model, I can construct an estimate of the degree of assortment by comparing the correlation between fathers-in-law and sons-in-law to the correlation between sons and fathers. Letting Y_i^{fl} be the observed status of the father-in-law of i:

$$Y_i = \gamma(\beta_f + \gamma\beta_m)Y_i^{fl} - \gamma(\beta_f + \gamma\beta_m)u_i^{fl} + \gamma\beta_m v_i^{fl} + \gamma e_i + v_i - u_i \tag{9}$$

and

[13] Appendix A1 further illustrates this logic, demonstrating the effects β_f and γ on steady-state inequality after many generations.

$$Y_i = (\beta_f + \gamma\beta_m)Y_i^f - (\beta_f + \gamma\beta_m)u_i^f + \beta_m v_i^f + e_i - u_i \tag{10}$$

regressing Y_i on Y_i^{fl} and on Y_i^f, the ratio of the coefficients has the probability limit of:

$$\gamma \frac{\sigma^2_{X_i^{fl}}(\sigma^2_{X_i^f} + \sigma^2_{u_i^f})}{\sigma^2_{X_i^f}(\sigma^2_{X_i^{fl}} + \sigma^2_{u_i^{fl}})} \tag{11}$$

which should be equal to γ if the variance of X_i^f is the same as that of X_i^{fl}. Note that this assumes that the groom is matching only with his bride, not his father-in-law. Appendix A2 provides some evidence that this assumption is valid in this particular context.

Below, I use this ratio method to estimate the degree of assortment in Quebec over time.

5 Empirics

The following section outlines the three main findings of this paper. First, using the simple model in the previous section, I estimate the degree of assortment over time. This estimate is high and fairly stable throughout the period. Then, I provide evidence that this estimated correlation was indeed due to sorting on individual characteristics, not the spurious result of matching between families. Finally, I provide causal evidence that mothers directly affect the outcomes of children. Putting it all together, I argue that assortment increased inequality throughout the period by decreasing social mobility.

5.1 Measuring the degree of marital assortment

Did the degree of assortment for marriages change over time? Figure 3 plots the correlation of spouses' literacy (proxied by signatures) over time. The degree of assortment appears to be relatively stable throughout the 19th century and an "inverted-U" shape in the 20th. However, there are good reasons to be skeptical of this simple measure. The ability to sign one's name is a relatively low bar. An individual who passes that threshold could be barely literate or have decades of schooling. In effect, signature rates are a highly right-censored measure of human capital. As the population approaches near-universal literacy in the early 20th century, this censoring will obscure most of the variation in human capital. A second issue is that even under perfect assortment, if the average signature rate for men and women is different, then the maximum correlation is not one (Liu and Lu (2006)). This maximum correlation changes over time, so how close the observed correlation is to the maximum also changes.

As mentioned above, an alternative measure can be constructed by comparing the correlation of the occupational status of sons-in-law and fathers-in-law to that of sons and fathers. The former are two degrees separate: an intergenerational link from father-in-law to daughter and a marriage link from daughter to son-in-law. The latter has only one degree of separation: an intergenerational link from father to son. In the simple model of marriage and mobility, the ratio of the correlations is equivalent to the degree of assortment. The key assumptions in the model are that children inherit human capital the same regardless of gender, that grooms match directly with their brides and not their fathers-in-law, and that the father-son correlation has the same measurement error as the father-in-law-son-in-law correlation. Relaxing the model's assumptions, the ratio will still hopefully control for trends in intergenerational mobility and leave only the trends in sorting. However, I argue that these strong assumptions have empirical support in this particular context. Appendix A2 shows some evidence that grooms did not match directly with their fathers-in-law. In the third set of empirical results below, I show that inheritance of human capital indeed seems to be symmetric across gender.

Figure 4 shows the correlations used to compute the ratio and Figure 5 plots the ratio measure over time. The estimated correlation in human capital is very high, around 0.85, and it appears to be stable throughout the period (perhaps declining at the very end). The overall trend is similar to that in Figure 3, deviating somewhat in the mid-20th century. However, as mentioned, literacy is both a noisy measure of human capital and becomes much less informative as the average level of education grows. Therefore, together I interpret the two figures as consistent with a story where the degree of assortment is high and stable throughout the entire period.

Appendix A3 presents several robustness checks for the estimation of the degree of assortment over time. The overall conclusion is robust to different measures of occupational status, simulating the within-occupation variation in status, using IV estimates to account for measurement error, and comparing fathers to fathers-in-law directly.

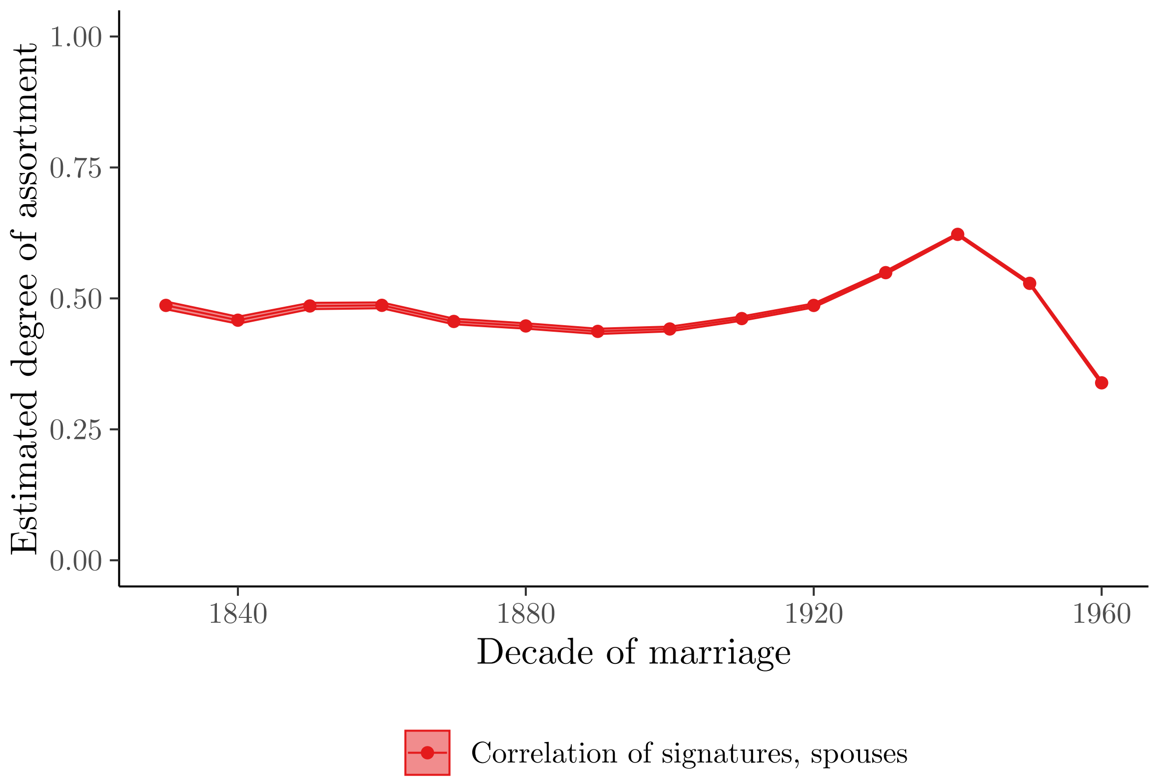

Figure 3: Correlation of spouses' signatures
Note: 95% confidence interval shaded. Signature variables are indicators that are one if a signature was recorded, zero if the absence of a signature was recorded, and omitted otherwise.

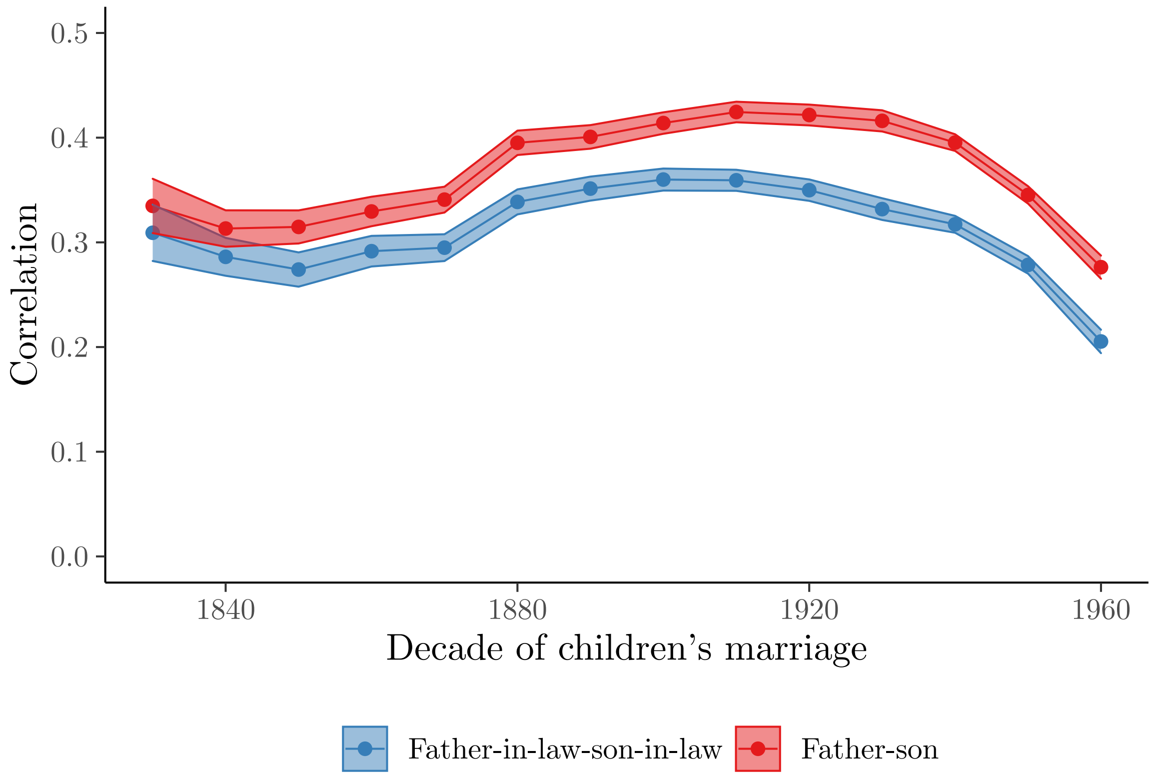

Figure 4: Correlations for ratio method using imputed earnings
Note: 95% bootstrapped confidence intervals shaded (50,000 replications). Imputed earnings are the average earnings for men with the individual's occupation in the 1901 Canadian Census sample. The estimates are ratios of rank-rank regression coefficients, which are equal to the correlation coefficients for the ranked variables assuming no ties in rank (and a reasonable approximation if not) (Chetty et al. (2014)).

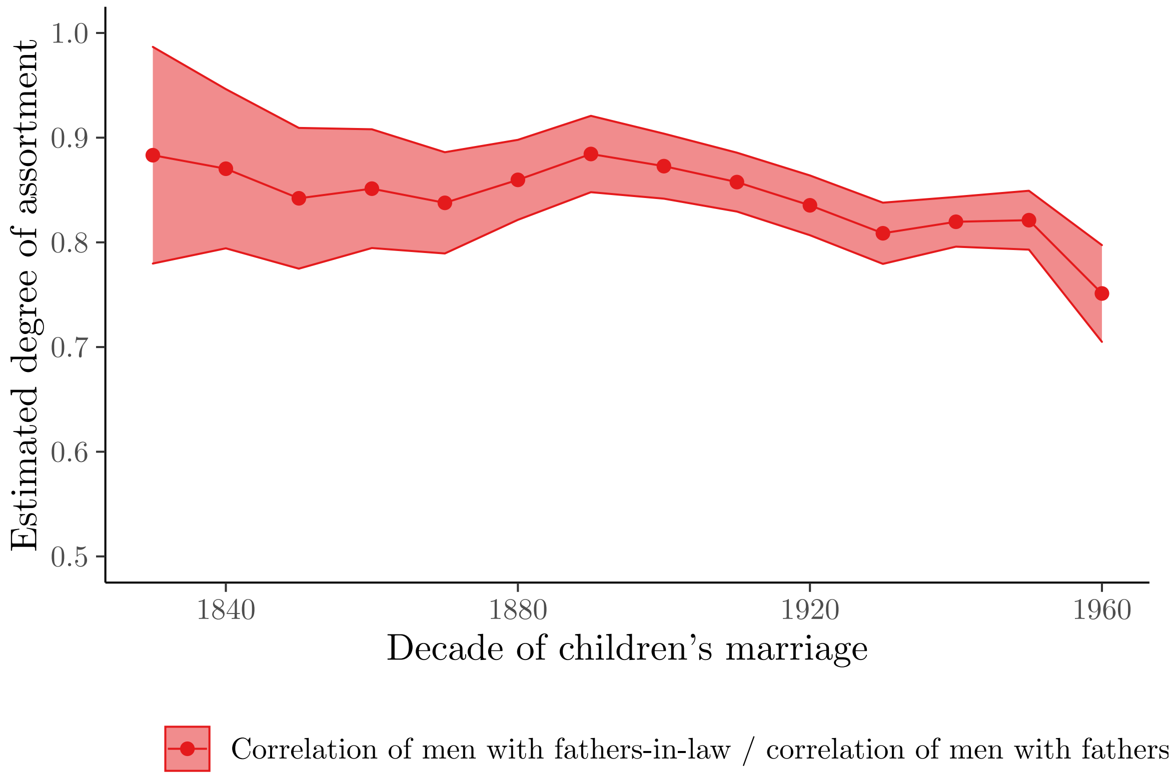

Figure 5: Ratio measure of martial sorting using imputed earnings
Note: 95% bootstrapped confidence intervals shaded (50,000 replications). Imputed earnings are the average earnings for men with the individual's occupation in the 1901 Canadian Census sample. The estimates are ratios of rank-rank regression coefficients, which are equal to the correlation coefficients for the ranked variables assuming no ties in rank (and a reasonable approximation if not) (Chetty et al. (2014)).

5.2 Did spouses match on individual human capital?

One can imagine a society in which marriage matches were not based on the human capital of the brides. For example, perhaps, a society where marriages were negotiated to form an alliance between a husband and his father-in-law with the characteristics of the wife an afterthought at best (Puga and Trefler (2014)).[14] In this hypothetical society, there could still be an observed correlation in the human capital of spouses if a woman's human capital

[14]Marriage as a way of cementing a commercial alliance was not unknown to the early settlers in Quebec. Indeed, marriage *à la façon du pays* ("after the custom of the country") between an indigenous woman and a French fur trader was commonly practiced to cement commercial relationships (Baillargeon (2014)).

is partially determined by her father.

To test if individual characteristics mattered, consider the following fixed effects regression:

$$Y^s_{i,F} = \alpha Y_{i,F} + \phi_F + \beta \mathbf{X}_{i,F} + \epsilon_{i,F} \tag{12}$$

Where $Y_{i,F}$ is a characteristic of individual i of family F, $Y^s_{i,F}$ is the characteristic of spouse s of individual i, ϕ_F are the crucial fixed effects that control for family background, $\mathbf{X}_{i,F}$ is a vector of controls, and $\epsilon_{i,F}$ is an error term. To address any time trends, $\mathbf{X}_{i,F}$ includes fixed effects for both decade and the order of siblings.[15]

In other words, the regression asks if, compared to their siblings, an individual with greater human capital matches with a spouse of greater human capital? If so, α will be positive.

As shown in Table 2 Panel A below, a woman who signed her marriage record married a man with greater human capital than her sisters who did not. Being able to write was associated with an increase in the probability a woman's husband was literate by 30 percentage points, an increase in her husband's imputed earnings by 4%, and an increase in her father-in-law's imputed earnings by 2%. This is evidence that marriage matches took into account individual characteristics. Note that while the family fixed-effect does reduce $\hat{\alpha}$, this does not reveal the degree to which matches are coordinated by families. If matching is only on individual characteristics, the family fixed-effect will still reduce $\hat{\alpha}$ as long as the human capital of sisters is correlated.

What about men and their brothers? As shown in Table 2 Panel B below, men who were able to write also married better. Being able to write associated with an increase in the probability that a man's wife was literate by 28 percentage points and an increase in his father-in-law's imputed earnings by 3%. These estimates are remarkably similar to those for women. The returns to human capital for marriage matching appear to be the same

[15] As I only have date of birth through 1849, I order siblings by the date of their first marriage.

regardless of gender.

What is the economic significance of marriage matching on the individual characteristics of women? If it were not the case, the woman's family could after all still matter for the outcomes of her children. First, the results show that there was a return to education for women in terms of the economic status of the household she formed at marriage. This was the case even if she did not employ her human capital in a formal occupation. Second, it implies a stronger role for assortative marriage in intergenerational mobility, assuming mothers mattered directly for their children's outcomes. Finally, it at least hints that women may have had some agency over the marriage matching process.

Appendix A4 discusses the robustness of these estimates. In particular, I estimate a selection into identification model, which accounts for the fact that families with one literate and one illiterate child are perhaps atypical (Miller et al. (2019)).

Table 2: Marriage matches were determined by individual characteristics

	Dependent variable: Spouse's characteristic					
	Signed		Log imp. earnings		Father's log imp. earnings	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Effect of wife's human capital						
Wife signed	0.49***	0.30***	0.17***	0.04***	0.07***	0.02***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Wife's family FE		X		X		X
Identifying observations		203,284		124,731		108,199
Observations	1,937,871	1,937,871	1,148,769	1,148,769	971,173	971,173
Adjusted R^2	0.60	0.64	0.06	0.38	0.03	0.32
Panel B: Effect of husband's human capital						
Husband signed	0.41***	0.28***			0.11***	0.03***
	(0.00)	(0.00)			(0.00)	(0.00)
Husband's family FE		X				X
Identifying observations		230,364				123,465
Observations	1,928,239	1,928,239			986,398	986,398
Adjusted R^2	0.62	0.64			0.04	0.33

Note: *p<0.10; **p<0.05; ***p<0.01. Family-clustered standard errors in parentheses. The sample excludes individuals with one or more unknown parents. Signature variables are indicators that are one if a signature was recorded, zero if the absence of a signature was recorded, and omitted otherwise. Imputed earnings are the average earnings for men with the individual's occupation in the 1901 Canadian Census sample. Decade and sibling order fixed effects are included in every specification as control variables. Identifying observations are the number of observations from families where at least one child signed and one child did not sign. Note that after adding family fixed effects the estimates are close to symmetrical across gender.

5.3 Do mothers matter directly for child outcomes?

For assortment to matter for social mobility, mothers must matter directly for child outcomes. A mother's literacy is correlated with that of her children even after controlling for that of the father (Table 3). Notably, there appears to be no difference between the associations with children of different genders. However, this pattern could still be observed if the mother did not directly matter for the outcomes of the children. With assortment, if the husband's ability is observed with measurement error, the mother's ability would be correlated with the residual even if its true effect is zero. Therefore, the simple regressions in Table 3 do not identify the effect of the ability of mothers.

In this section, I use two different strategies to determine if mothers have a direct effect on children. I am attempting neither to determine the mechanism of heritability nor to estimate the magnitude of the effect. The underlying process is likely complex; for example, is very plausible that parental human capital is complementary, with the marginal effect of a mother's human capital increasing with the father's. Instead, I aim only to provide causal evidence that mothers had an independent effect on child outcomes. This is sufficient to demonstrate that assortative marriage will have an effect on inequality through the channel of intergenerational mobility.

5.3.1 Controlling for the father with fixed effects

To identify a causal effect, ideally we would control for the father but randomize the mother. A less ideal (yet actually possible) approach is to consider the case where a father has children from more than one marriage. However, this results in two complications. The first is the chance that the children are scarred by whatever event resulted in a second marriage (such as a death or divorce). Assuming this penalty is a constant, it can be controlled for by including fixed effects for the marriage number the children are from. Second, as marriage is assortative on the ability of mothers, the abilities of each wife of the father will be correlated. Therefore, similar to the family fixed effects above, the father fixed effect will absorb part of the effect of the mother's ability.

I regress:

$$Y_{i,f} = \alpha Y_{i,f}^{m} + \phi_f + \beta \mathbf{X}_{i,f} + \epsilon_{i,f} \tag{13}$$

where $Y_{i,F}$ is an outcome of a child i with father f, $Y_{i,f}^{m}$ is a characteristic of the child's mother, ϕ_f are the crucial fixed effects that control for the father, and $\mathbf{X}_{i,f}$ are controls. To address any time trends, $\mathbf{X}_{i,f}$ includes fixed effects for decade, the marriage number of the father, and the order of siblings.

As shown in Table 4 Panel A, even controlling for the father, a mother who could sign her name had children who were 3 percentage points more likely to be able to sign their names and had a 3% higher occupational status score. This direct independent effect, while statistically significant, appears to be very small. However, as shown above, the human capital of spouses are highly correlated. The father fixed will control for most of the characteristics of his wives. Even if one can sign her name and the other cannot, they are likely otherwise very similar, and the father fixed effect controls for that similarity. Moreover, it will also control for any spillovers from one set of children to the other. In other words, after controlling for the father fixed effect, there is only a small residual amount of variation but it is directly attributable to the mother.

For comparison, Table 4 Panel B estimates the effects of the ability of a father controlling for the mother. Notably, the results are very similar to those of the regressions for mothers. Once the correlation between the ability of spouses is accounted for through fixed effects, the direct independent effect of parental human capital appears to be symmetrical across the gender of both the parent and of the child. This is reassuring, as the simple model developed earlier in this paper assumes that children inherit human capital from their fathers at the same rate regardless of gender.

Appendix A5 discusses the robustness of these estimates. In particular, if there is a trend over time in child outcomes, children born after a remarriage would differ from those born before the remarriage even if mothers had no direct independent effect. However, I show that the results are robust to restricting the estimates to a window around the remarriage, suggesting this is not a concern.

Table 3: The association of parental human capital with child outcomes

	Dependent variable:			
	Signed	Signed	Log imp. earnings	
	Daughter	Son	Daughter's-husband	Son
	(1)	(2)	(3)	(4)
Mother signed	0.12***	0.12***	0.06***	0.07***
	(0.001)	(0.001)	(0.002)	(0.002)
Father signed	0.07***	0.09***	0.15***	0.16***
	(0.001)	(0.001)	(0.002)	(0.002)
Observations	1,551,089	1,435,443	938,558	875,264
Adjusted R^2	0.51	0.50	0.06	0.06

Note: *p<0.10; **p<0.05; ***p<0.01. Family-clustered standard errors in parentheses. The sample excludes individuals with one or more unknown parents. Signature variables are indicators that are one if a signature was recorded, zero if the absence of a signature was recorded, and omitted otherwise. Imputed earnings are the average earnings for men with the individual's occupation in the 1901 Canadian Census sample. Fixed effects for decade, marriage number, and sibling order are included in every specification as control variables.

Table 4: The effect of parental human capital on child outcomes

	Dependent variable:			
	Signed Daughter	Signed Son	Log imp. earnings Daughter's husband	Log imp. earnings Son
	(1)	(2)	(3)	(4)
Panel A: Controlling for father				
Mother signed	0.02***	0.03***	0.01	0.03***
	(0.01)	(0.01)	(0.01)	(0.01)
Father FEs	X	X	X	X
Identifying observations	18,407	16,058	8,532	7,537
Observations	1,571,362	1,454,557	950,687	886,907
Adjusted R^2	0.68	0.67	0.37	0.41
Panel B: Controlling for mother				
Father signed	0.02***	0.03***	0.03*	0.04**
	(0.01)	(0.01)	(0.02)	(0.02)
Mother FEs	X	X	X	X
Identifying observations	6,488	5,516	2,906	2,385
Observations	1,563,894	1,447,566	946,275	882,625
Adjusted R^2	0.69	0.68	0.37	0.41

Note: *p<0.10; **p<0.05; ***p<0.01. Family-clustered standard errors in parentheses. The sample excludes individuals with one or more unknown parents. Signature variables are indicators that are one if a signature was recorded, zero if the absence of a signature was recorded, and omitted otherwise. Imputed earnings are the average earnings for men with the individual's occupation in the 1901 Canadian Census sample. Fixed effects for decade, marriage number, and sibling order are included in every specification as control variables. Identifying observations are the number of observations with a parent with one spouse who signed and one who did not.

5.3.2 Directly comparing half-siblings

One downside of the father fixed effects approach is that it relies on observing a measure of the ability of the mother. As shown in Table 4, the identifying sample size is quite low: very few parents had two spouses, one of which was literate and one of which was not. Hence, not all the coefficients are significant at the 5% level.

Fortunately, there is another test using parents with more than one marriage that only relies on the characteristics of the children. Consider a pair of children who could be either half-siblings or full siblings. If they share both a mother and a father and the abilities of mothers matter directly, their outcomes should be more correlated than if they share only a father. Again, there is a concern that the event resulted in a second marriage could have harmed the children of the first marriage. Again, assuming the penalty is a constant, fixed effects can control for it.

I estimate the regression:

$$Y_{i,f} = \alpha Y_{j,f} \times I(m_i = m_j) + \beta \mathbf{X_{i,j}} + \epsilon_{i,j,f} \tag{14}$$

Where $Y_{i,f}$ is a characteristic of child i with father f and mother m_i, $Y_{j,f}$ is a characteristic of their half- or full sibling j, $I(m_i = m_j)$ is an indicator that is one if the children share a mother, $\mathbf{X}_{i,j}$ are control variables, and $\epsilon_{i,j,f}$ is an error term. The controls include fixed effects for decade, the order of the siblings, and the marriage number of the father.

The results are shown in Table 5 below. Full siblings are more strongly associated than half-siblings. For example, a daughter signing her name was associated with a 36 percentage point increase in the probability her half-sister could sign her name. However, it was associated with a 42 percentage point increase in the probability her full sister could sign her name. As before, the results are very similar regardless of if I allow mothers or fathers to vary and if I look at daughters or sons.

Appendix A5 discusses the robustness of these estimates. Again, if there is a trend over time in child outcomes, children born after remarriage would differ from those born before a remarriage even if mothers had no direct independent effect. Similar to before, I show that the results are robust to restricting the estimates to a window around the remarriage, suggesting this is not a concern.

Table 5: The effect of parental human capital on half vrs. full siblings

	Dependent variable: Younger sibling's characteristic			
	Signed Daughter	Signed Son	Log imp. earnings Daughter's husband	Log imp. earnings Son
	(1)	(2)	(3)	(4)
Panel A: Controlling for father				
Older sibling's characteristic	0.36***	0.36***	0.22***	0.26***
	(0.00)	(0.00)	(0.01)	(0.01)
Signed × same mother	0.06***	0.05***	0.05***	0.05***
	(0.00)	(0.00)	(0.01)	(0.01)
Observations	2,050,264	1,853,707	839,388	756,645
Adjusted R^2	0.64	0.63	0.11	0.14
Panel B: Controlling for mother				
Older sibling's characteristic	0.36***	0.33***	0.24***	0.22***
	(0.01)	(0.01)	(0.02)	(0.02)
Signed × same father	0.07***	0.08***	0.03*	0.09***
	(0.01)	(0.01)	(0.02)	(0.02)
Observations	1,965,701	1,777,710	806,656	727,123
Adjusted R^2	0.64	0.63	0.11	0.14

Note: *p<0.10; **p<0.05; ***p<0.01. Family-clustered standard errors in parentheses. The sample excludes individuals with one or more unknown parents. Signature variables are indicators that are one if a signature was recorded, zero if the absence of a signature was recorded, and omitted otherwise. Imputed earnings are the average earnings for men with the individual's occupation in the 1901 Canadian Census sample. Fixed effects for decade, marriage number, and sibling order as well as the non-interacted same mother indicator variable are included in every specification as control variables.

6 Discussion

The empirical findings of this paper demonstrate that marriage was strongly assortative in the past and that this had direct consequences for intergenerational mobility. If, as I argue above, findings for Quebec are likely generalizable to other historical populations, this has several implications for the standard approaches to studying intergenerational mobility. In this section, I discuss how overlooking mothers and marriage can lead to misleading conclusions when looking at both father-son intergenerational correlations and when looking at the role of grandfathers.

6.1 Sorting matters for father-son intergenerational correlations

If women directly matter for the outcomes of their children and marriages are assortative, the correlation between characteristics of fathers and sons will be partially determined by the mother.[16] In the simple model in Section 4, the association between fathers and sons is:

$$Y_i = (\beta_m + \gamma\beta_f)Y_i^f + \epsilon_i \tag{15}$$

Note that $\beta_m + \gamma\beta_f$ is often the correlation of interest, as it shows how strongly associated sons are with their fathers. However, it should not be interpreted as the direct effect of the father. If the parents matched on individual characteristics, the mother increases the association through the $\gamma\beta_f$ term. Changes in the observed rates of intergenerational mobility, even if women are not observed, could be driven by changes in marriage matching (γ) or in how strongly mothers influence their children (β_f).

To illustrate this, Table 6 shows the intergenerational elasticity of imputed earnings estimated separately for more and less assorted parents. The less assorted parents are those where only one parents was literate, the more assorted parents those where both parents were either literate or illiterate. The elasticities for the less assorted parents are 0.30 for the sons and 0.28 for daughters (using their husbands' imputed earnings as a proxy). For the more assorted parents, the elasticities are 0.42 for sons and 0.41 for daughters. The more strongly assorted parents have higher estimated rates of intergenerational mobility. It is of course possible that the more and less assorted families are not directly comparable and that the difference is due to some other omitted variable; Appendix A6 addresses this concern.

[16] Espín-Sánchez et al. (2022) makes this point as well.

Table 6: Father-son intergenerational elasticities, more and less assorted marriages

	Dependent variable:			
	Son's log earnings score		Daughter's husband's log earnings score	
	(1)	(2)	(3)	(4)
Father's log earnings score	0.30***	0.42***	0.28***	0.41***
	(0.01)	(0.00)	(0.01)	(0.00)
Parent's differ on signature	X		X	
Parents same on signature		X		X
Observations	27,278	125,094	30,022	129,928
Adjusted R^2	0.06	0.11	0.05	0.12

Note: *p<0.10; **p<0.05; ***p<0.01. Standard errors in parentheses. Signature variables are indicators that are one if a signature was recorded, zero if the absence of a signature was recorded, and omitted otherwise. Imputed earnings are the average earnings for men with the individual's occupation in the 1901 Canadian Census sample. Decade fixed effects are included in every specification as control variables.

6.2 Sorting matters for multigenerational mobility

Many recent studies consider correlations across more than two generations (Clark (2014), Olivetti et al. (2018), Solon (2018), Espín-Sánchez et al. (2022) and Long and Ferrie (2018)). I am also able to estimate multigenerational mobility with the Quebec data, as shown in Table 7 below. Note that when estimated separately, the intergenerational elasticities between grandfathers and grandchildren seem to be the same regardless of if the grandfathers are maternal or paternal. When, however, the partial elasticities are estimated controlling for the log imputed earnings of the other grandfather and the father, there is a larger coefficient for the maternal grandfathers.

Should we interpret this as maternal grandfathers being more important to the outcomes of grandchildren? The answer is no. To illustrate why, refer back to the model in Section 4. If it is directly related to the mother's true status, a grandfather's observed status will have a coefficient biased upwards as the mother is omitted. Likewise, if it is directly related to the father's true status, it will have a coefficient biased upwards if the father is omitted. Controlling for the father's observed status will reduce the bias from omitting the true status of the father much more than it would reduce the bias from omitting that of the mother. As we'd expect the maternal grandfather to be more strongly correlated with the mother, we would therefore expect a larger coefficient than the paternal grandfather after controlling for the father. This is in fact what we observe in Table 7.

This exercise demonstrates how caution must be taken in interpreting intergenerational correlations without accounting for the role of women. It would, at first glance, seem plausible to have found evidence that maternal grandfathers mattered more for the outcomes of children than paternal grandfathers. However, it is merely an artifact of measurement error.

Table 7: Grandfather-grandson intergenerational elasticities

	Dependent variable:		
	Child's status measure		
	(1)	(2)	(3)
Panel A: log imp. earnings (male)			
Maternal grandfather's log imp. earnings	0.30***		0.10***
	(0.00)		(0.00)
Paternal grandfather's log imp. earnings		0.30***	0.13***
		(0.00)	(0.00)
Father's log imp. earnings			0.36***
			(0.00)
Observations	214,856	214,856	214,856
Adjusted R^2	0.04	0.04	0.13
Panel B: Husband's log imp. earnings (female)			
Maternal grandfather's log imp. earnings	0.27***		0.10***
	(0.00)		(0.00)
Paternal grandfather's log imp. earnings		0.28***	0.14***
		(0.00)	(0.00)
Father's log imp. earnings			0.31***
			(0.00)
Observations	219,646	219,646	219,646
Adjusted R^2	0.03	0.04	0.11

Note: *p<0.10; **p<0.05; ***p<0.01. Standard errors in parentheses. Imputed earnings are the average earnings for men with the individual's occupation in the 1901 Canadian Census sample. Decade fixed effects are included in every specification as control variables.

7 Conclusion

In this paper, I construct a simple model of marriage and mobility. It shows that even with no female employment, assortative marriage will increase inequality if the ability of a woman determines whom she marries and the success of her children. To test if this was true in Quebec 1800–1970, I consider a novel dataset containing millions of families reconstructed from vital records. Unusually, married women are linked to their fathers; I use this to develop a new method to estimate the degree of assortment, finding it surprisingly high and stable over time. Next, I find pairs of sisters where only one was able to sign a name. Even though she likely was not in the labor force after she married, I show that the the more educated sister still typically earned an education premium when it came to the socioeconomic status of her husband. Moreover, I show her ability mattered as much as her husband's for the outcomes of their children. As quick remarriage after losing a spouse was the norm, I hold one parent constant and allow the second to vary. Sharing a mother mattered as much as sharing a father for child outcomes. Altogether, I conclude that assortative marriage had always mattered for inequality. It mattered because, despite severe legal and economic disadvantages, women played a major role in mobility and marriage. Overlooking the role of women and marriage would leave our understanding of intergenerational mobility and inequality over the long run incomplete.

References

Abramitzky, Ran, Adeline Delavande, and Luis Vasconcelos (2011a). "Marrying Up: The Role of Sex Ratio in Assortative Matching". *American Economic Journal: Applied Economics* 3.3, pp. 124–157.

— (2011b). "Marrying Up: The Role of Sex Ratio in Assortative Matching". *American Economic Journal: Applied Economics* 3.3, pp. 124–157.

Abramitzky, Ran et al. (2019). *Automated Linking of Historical Data*. Tech. rep. w25825. Cambridge, MA: National Bureau of Economic Research.

A'Hearn, Brian, Jörg Baten, and Dorothee Crayen (2009). "Quantifying Quantitative Literacy: Age Heaping and the History of Human Capital". *The Journal of Economic History* 69.3, pp. 783–808.

Albouy, David (2008). “The Wage Gap between Francophones and Anglophones: A Canadian Perspective, 1970–2000”. *Canadian Journal of Economics/Revue canadienne d’économique* 41.4, pp. 1211–1238.

Baillargeon, Denyse (2014). *A Brief History of Women in Quebec*. Trans. by W. Donald Wilson. Studies in Childhood and Family in Canada. Waterloo, Ontario, Canada: Wilfrid Laurier Univeristy Press.

Black, Sandra E. and Paul J. Devereux (2011). “Recent Developments in Intergenerational Mobility”. In: *Handbook of Labor Economics*. Vol. 4. Elsevier, pp. 1487–1541.

Bouchard, Gérard, Patrick Brard, and Yolande Lavoie (1981). “FONEM : Un Code de Transcription Phonétique Pour La Reconstitution Automatique Des Familles Saguenayennes”. *Population, 36e année* n°6, pp. 1085–1103.

Canadian Families Project (2002). *Canadian Families Project*. Victoria, Canada: University of Victoria.

Caron, Marianne et al. (2017). “Sibling Influence on Access to Marriage: A Comparative Study of Québec and Flanders, 1842–1912”. *Journal of Family History* 42.3, pp. 271–307.

Chadwick, Laura and Gary Solon (2002). “Intergenerational Income Mobility Among Daughters”. *American Economic Review* 92.1, pp. 335–344.

Chetty, Raj et al. (2014). “Where Is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States”. *The Quarterly Journal of Economics* 129.4, pp. 1553–1623.

Clark, Gregory (2007). *A Farewell to Alms: A Brief Economic History of the World*. Princeton, NJ: Princeton University Press.

— (2014). *The Son Also Rises: Surnames and the History of Social Mobility*. First. The Princeton Economic History of the Western World. Princeton: Princeton University Press.

Clark, Gregory and Neil Cummins (2015). “Intergenerational Wealth Mobility in England, 1858-2012: Surnames and Social Mobility”. *The Economic Journal* 125.582, pp. 61–85.

Clark, Gregory, Neil Cummins, and Matthew Curtis (2020). “Twins Support the Absence of Parity-Dependent Fertility Control in Pretransition Populations”. *Demography* 57.4, pp. 1571–1595.

Craig, Jacqueline, Katherine Eriksson, and Gregory T Niemesh (2020). “Marriage and the Intergenerational Mobility of Women: Evidence from Marriage Certificates 1850-1910”, p. 24.

de la Croix, David and Marc Goñi (2021). “Nepotism vs. Intergenerational Transmission of Human Capital in Academia (1088–1800)”. *SSRN Electronic Journal*.

de Vries, Jan (2008). *The Industrious Revolution: Consumer Behavior and the Household Economy, 1650 to the Present*. New York: Cambridge University Press.

Dillon, Lisa (2010). “Parental and Sibling Influences on the Timing of Marriage, Xviith and Xviiith Century Québec”. *Annales de démographie historique* 119.1, p. 139.

Dillon, Lisa, North Atlantic Population Project, and Minnesota Population Center (2008). *National Sample of the 1881 Census of Canada (Version 2.0).*

Dillon, Lisa et al. (2018). “The Programme de Recherche En Démographie Historique: Past, Present and Future Developments in Family Reconstitution”. *The History of the Family* 23.1, pp. 20–53.

Dribe, Martin, Björn Eriksson, and Francesco Scalone (2019). “Migration, Marriage and Social Mobility: Women in Sweden 1880–1900”. *Explorations in Economic History* 71, pp. 93–111.

Eika, Lasse, Magne Mogstad, and Zafar Basit (2019). “Educational Assortative Mating and Household Income Inequality”. *Journal of Political Economy* 127.6, pp. 2795–2835.

Espín-Sánchez, José-Antonio, Salvador Gil-Guirado, and Chris Vickers (2022). "La "Doña" è Mobile: The Role of Women in Social Mobility in a Pre-modern Economy". *Journal of Economic History (forthcoming)* March, p. 47.

Espín-Sánchez, José-Antonio et al. (2019). "Labor Income Inequality in Pre-Industrial Mediterranean Spain: The City of Murcia in the 18th Century". *Explorations in Economic History* 73, p. 101274.

Fernández, Raquel and Richard Rogerson (2001). "Sorting and Long-Run Inequality". *Quarterly Journal of Economics* November, p. 37.

Gaffield, Chad et al. (2009). *National Sample of the 1911 Census of Canada [Dataset]*. University of Alberta [distributor].

Gagnon, Alain et al. (2011). "Once Were Farmers: Occupation, Social Mobility, and Mortality during Industrialization in Saguenay-Lac-Saint-Jean, Quebec 1840–1971". *Explorations in Economic History* 48.3, pp. 429–440.

Galor, Oded and Marc Klemp (2019). "Human Genealogy Reveals a Selective Advantage to Moderate Fecundity". *Nature Ecology & Evolution* 3.5, pp. 853–857.

Geloso, Vincent and Peter Lindert (2020). "Relative Costs of Living, for Richer and Poorer, 1688–1914". *Cliometrica* 14.3, pp. 417–442.

Goldin, Claudia (2006). "The Quiet Revolution That Transformed Women's Employment, Education, and Family". 96.2, p. 43.

Goñi, Marc (2018). "Assortative Matching at the Top of the Distribution: Evidence from the World's Most Exclusive Marriage Market", p. 90.

Greenwood, Jeremy, Ananth Seshadri, and Mehmet Yorukoglu (2003). *Engines of Liberation*. SSRN Scholarly Paper ID 298479. Rochester, NY: Social Science Research Network.

Greenwood, Jeremy et al. (2014a). "Corrigendum to Marry Your Like: Assortative Mating and Income Inequality". *American Economic Review: Papers and Proceedings* May, p. 2.

Greenwood, Jeremy et al. (2014b). "Marry Your Like: Assortative Mating and Income Inequality". *American Economic Review: Papers and Proceedings* 104.5, pp. 348–353.

Greer, Allan (1985). *Peasant, Lord, and Merchant: Rural Society in Three Quebec Parishes 1740–1840*. Toronto: University of Toronto Press.

— (1997). *The People of New France*. Toronto: University of Toronto Press.

Hryshko, Dmytro, Chinhui Juhn, and Kristin McCue (2017). "Trends in Earnings Inequality and Earnings Instability among U.S. Couples: How Important Is Assortative Matching?" *Labour Economics* 48, pp. 168–182.

IMPQ (2020). *Infrastructure Intégrée Des Microdonnées Historiques de La Population Du Québec*. Trois-Rivières, Québec: Centre Interuniversitaires d'Études Québécoises.

Inwood, Kris and Chelsea Jack (2011). *National Sample of the 1891 Census of Canada*. Guelph, Canada: University of Guelph.

Killingsworth, Mark R and James J. Heckman (1986). "Female Labor Supply: A Survey". In: *Handbook of Labor Economics*. Ed. by Orley C. Ashenfelter and Richard Layard. Vol. 1.

Kremer, Michael (1997). "How Much Does Sorting Increase Inequality?" *Quarterly Journal of Economics* Febuary, p. 25.

Lambert, Paul S. et al. (2013). "The Construction of HISCAM: A Stratification Scale Based on Social Interactions for Historical Comparative Research". *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 46.2, pp. 77–89.

Liu, Haoming and Jingfeng Lu (2006). "Measuring the Degree of Assortative Mating". *Economics Letters* 92.3, pp. 317–322.

Long, Clarence D. (1958). *The Labor Force under Changing Income and Employment*. Princeton, NJ: Princeton University Press.

Long, Jason and Joseph Ferrie (2018). “Grandfathers Matter(Ed): Occupational Mobility Across Three Generations in the US and Britain, 1850–1911”. *The Economic Journal* 128.612, F422–F445.

Magnuson, Roger (1992). *Education in New France*. Montreal [Que.] : McGill-Queen’s University Press.

Mazumder, Bhashkar (2005). “Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data”. *The Review of Economics and Statistics* 87.2, pp. 235–255.

Miller, Douglas, Na’ama Shenhav, and Michel Grosz (2019). *Selection into Identification in Fixed Effects Models, with Application to Head Start*. Tech. rep. w26174. Cambridge, MA: National Bureau of Economic Research, w26174.

Minnesota Population Center (2019). *Integrated Public Use Microdata Series, International: Version 7.2 [Dataset]*. Minneapolis: IPUMS.

Olivetti, Claudia and M. Daniele Paserman (2015). “In the Name of the Son (and the Daughter): Intergenerational Mobility in the United States, 1850–1940”. *The American Economic Review* 105.8, pp. 2695–2724.

Olivetti, Claudia, M. Daniele Paserman, and Laura Salisbury (2018). “Three-Generation Mobility in the United States, 1850–1940: The Role of Maternal and Paternal Grandparents”. *Explorations in Economic History* 70, pp. 73–90.

Olivetti, Claudia et al. (2020). *Who Married, (to) Whom, and Where? Trends in Marriage in the United States, 1850-1940*. Tech. rep. w28033. Cambridge, MA: National Bureau of Economic Research, w28033.

Piketty, Thomas and Emmanuel Saez (2014). “Inequality in the Long Run”. *Science* 344.6186, pp. 838–843.

PRDH (2020). *Programme de Recherche En Démographie Historique. Registre de La Population Du Québec Ancien (RPQA) [Data Set]*. Montréal, Canada: Département de Démographie, Université de Montréal.

Project Balsac (2020). *BALSAC [Data Set]*. Québec, Canada: Université du Québec à Chicoutimi.

Puga, Diego and Daniel Trefler (2014). "International Trade and Institutional Change: Medieval Venice's Response to Globalization". *The Quarterly Journal of Economics* 129.2, pp. 753–821.

Solon, Gary (1992). "Intergenerational Income Mobility in the United States". *The American Economic Review*, pp. 393–408.

— (2018). "What Do We Know So Far about Multigenerational Mobility?" *The Economic Journal* 128.612, F340–F352.

Song, Xi et al. (2020). "Long-Term Decline in Intergenerational Mobility in the United States since the 1850s". *Proceedings of the National Academy of Sciences* 117.1, pp. 251–258.

Tremblay, Manon and Käthe Roth (2010). *Quebec Women and Legislative Representation*. Vancouver: UBC Press.

Van Leeuwen, Marco HD, Ineke Maas, and Andrew Miles (2004). "Creating a Historical International Standard Classification of Occupations an Exercise in Multinational Interdisciplinary Cooperation". *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 37.4, pp. 186–197.

Vézina, Hélène, Danielle Gauvreau, and Alain Gagnon (2014). "Socioeconomic Fertility Differentials in a Late Transition Setting: A Micro-Level Analysis of the Saguenay Region in Quebec". *Demographic Research* 30, pp. 1097–1128.

Vézina, Hélène, Marc St-Hilaire, and Jean-Sébastien Bournival (2013). "The Linkage of Micro Census Data and Vital Records: An Assessment of Results on Quebec Historical Censuses (1852-1911)", p. 25.

Ward, Zachary (2021). *Intergenerational Mobility in American History: Accounting for Race and Measurement Error*. Tech. rep. w29256. Cambridge, MA: National Bureau of Economic Research, w29256.

Zijdeman, Richard (2014). *Oclack - Occupational CLAssifications and CrosswalKs [Github Repository]*.

Appendix

A1 Steady state inequality in the simple model

Two of the main findings of this paper — that sorting was on individual characteristics and that the ability of mothers mattered for child outcomes — individually answer somewhat narrow questions. Together, they imply that a high degree of assortment would have contributed to inequality over the long run. Here, I illustrate the logic behind the claim with the simple model I use to estimate the degree of assortment in Section 4.

To summarize inequality in a given generation, consider the variance of latent status:

$$\sigma^2_{X_i} = (\beta_f)^2\sigma^2_{X_i^f} + (\beta_m)^2\sigma^2_{X_i^m} + 2\beta_f\beta_m(\gamma\sigma_{X_i^f}\sigma_{X_i^m}) + \sigma^2_{e_i} \tag{16}$$

Now define a steady-state equilibrium as when there is no change in inequality from generation to generation:

$$\sigma^2_{X_i} = \sigma^2_{X_i^f} = \sigma^2_{X_i^m} \tag{17}$$

Then:

$$\sigma^2_{X_i} = \frac{\sigma^2_{e_i}}{1 - (\beta_f)^2 - (\beta_m)^2 - 2\gamma\beta_f\beta_m} \tag{18}$$

As the error term u_i is assumed to be independent of X_i, the observed inequality is given by:

$$\sigma^2_{Y_i} = \sigma^2_{X_i} + \sigma^2_{u_i} \tag{19}$$

Unsurprisingly, the more children take after their parents (i.e. the higher the β_f and β_m), the higher the level of steady state inequality. Further, if both β_f and β_m are greater than zero, the degree of assortment γ will increase steady state inequality as well (Figure 6).

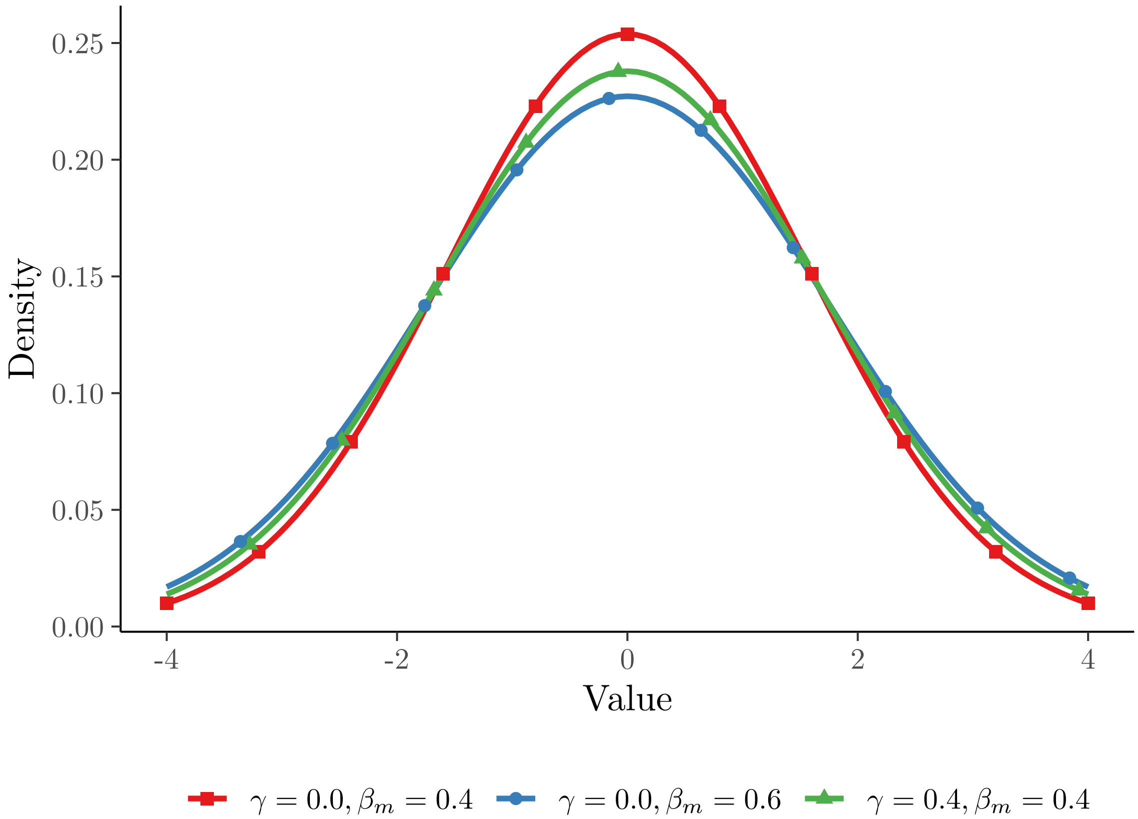

Figure 6: Simulated steady state inequality
Note: Simulated data based on model (see text). γ is the degree of assortment, β_f and β_m are the strength of intergenerational inheritance of latent status from fathers and mothers respectively. I assign e_i (the random component of intergenerational mobility) and u_i (the classical measurement error term) a variance of one in all simulations. As shown by the simulations, increasing γ or β_m increases inequality.

A2 Evidence that grooms did not match with fathers-in-law

I can directly test if the matching is between husbands and fathers-and-laws. For fathers-in-law who die before 1849, are their sons-in-law who married before their death different than those married after? As shown in Table 8, there appears to be no difference. In other words, if husbands are matching with their fathers-in-law, they don't seem to mind if their father-in-law is deceased before their marriage.

Table 8: Marriage matching appears not to have been between husbands and fathers-in-law

	Dependent variable:	
	Signed Son-in-law	Log imp. earnings Son-in-law
	(1)	(2)
Married after father-in-law's death	−0.01	0.00
	(0.01)	(0.00)
Family FEs	X	X
Observations	83,988	147,641
Adjusted R^2	0.37	0.70

Note: *p<0.10; **p<0.05; ***p<0.01. Family-clustered standard errors in parentheses. Signature variables are indicators that are one if a signature was recorded, zero if the absence of a signature was recorded, and omitted otherwise. Imputed earnings are the average earnings for men with the individual's occupation in the 1901 Canadian Census sample. Fixed effects for decade, marriage number, and sibling order are included in every specification as control variables. Deaths are only observed before 1849, so only sons-in-law of men who died before 1849 are included.

A3: Robustness of estimates of sorting

Figure 7 below estimates the ratio measure estimate of the degree of assortment using a number of different occupational status scores. All the occupational scores give roughly the same picture of the overall level and trend of assortment. Moreover, the occupational status score used in the main results based on imputed 1901 earnings is about in the middle of the distribution of estimates.

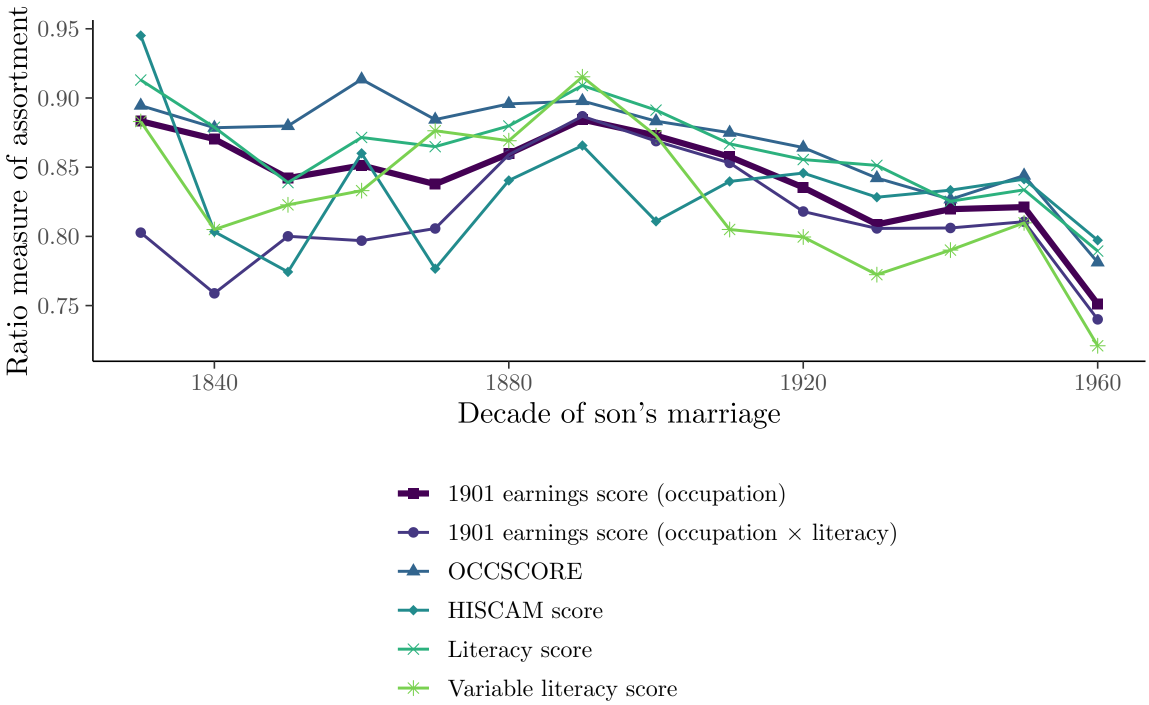

Figure 7: Alternative occupational status scores

Note: The estimates are ratios of rank-rank regression coefficients, which are equal to the correlation coefficients for the ranked variables assuming no ties in rank (and a reasonable approximation if not) (Chetty et al. (2014)). 1901 imputed earings are the average earnings for a man with that occupation in the 1901 Canadian Census sample. 1901 imputed earnings with signature also uses literacy (proxied by signatures) to impute the earnings. OCCSCORE is the IPUMS imputed earnings which is based on 1950 US Census earnings (Minnesota Population Center (2019)). HISCAM is the universal HISCAM score, a social distance based ranking of 19th century occupations (Lambert et al. (2013)). Occupational literacy scores are the share of men with that occupation in the 1890's in the vital records who could sign their name. Variable occupational literacy scores are computed for each decade using the method in Song et al. (2020): for each occupational category and decade, the score is the sum of the percentile rank of each educational group (signed and not signed) weighted by the share of the occupation in that category. This is essentially a reweighted version of the average signature rate in that group that accounts for the varying rate of signatures over time.

Using occupational status scores introduces three potential sources of bias. The first source of bias is classical measurement error, as the true socioeconomic status of the individuals is necessarily measured with error when using occupation as a proxy. The second source of bias arises when comparing two individuals with the same occupation. In this case, there is non-classical measurement error as both are assigned the status scores with the same measurement error (Espín-Sánchez et al. (2019)). A third source of error is sample selection bias, which arises if the underlying human capital is correlated with the probability of reporting an occupation that has a score (de la Croix and Goñi (2021)).

The main results of this paper attempt to control for the first source of bias by using the ratio method described above. A more standard method used in the literature is IV regression using a second measure of occupational status to reduce attenuation bias (Solon (1992) and Ward (2021)). This approach requires no assumption about the relative magnitudes of attenuation bias on the numerator and denominator, but only reduces the attenuation bias. While not using IV for causal identification, it also still faces the standard IV tradeoff where the estimator is inherently biased, though hopefully less so than regular OLS regression. Figure 8 below computes the ratio measure used in Figure 5 but first estimates both correlations using an instrumental variable regression. The instrument used is the occupational status for the occupation second closest chronologically to the first marriage.[17] The resulting measure is very similar.

[17] Recall, the dataset often contains individuals who are not the primary subject of the vital event. For example, a man might have an occupation reported at his child's wedding.

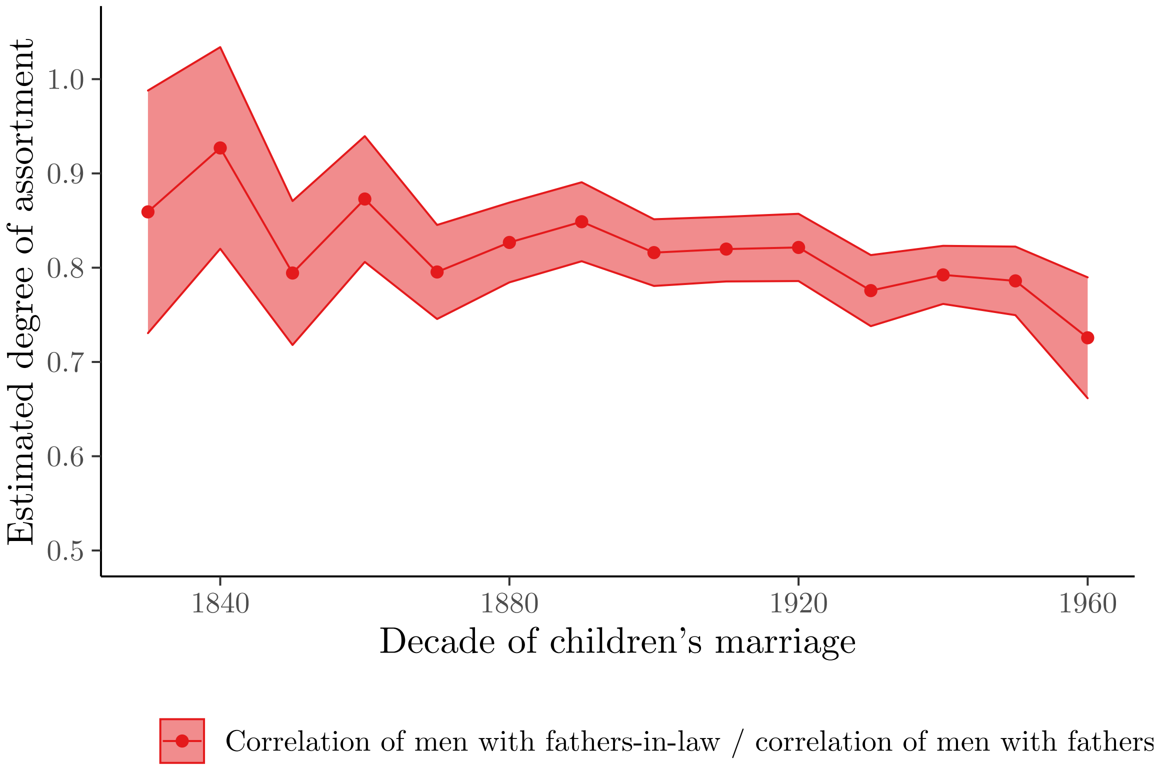

Figure 8: Estimated degree of marital assortment, IV
Note: 95% bootstrapped confidence intervals shaded (50,000 replications). Imputed earnings are the average earnings for men with the individual's occupation in the 1901 Canadian Census sample. The estimates are ratios of rank-rank regression coefficients, which are equal to the correlation coefficients for the ranked variables assuming no ties in rank (and a reasonable approximation if not) (Chetty et al. (2014)). To reduce attenuation bias, the dependent variable in the regressions are instrumented using a second measure of imputed earnings (using the second closest occupation the the individual's first marriage).

One way to estimate a lower bound robust to the second source of non-classical measurement error that arises when using occupational status is to simulate the underlying distribution of within-occupation status. For example, for the correlation between fathers' and sons' occupational earning scores, each individual with a given occupation can be assigned a draw from a log-normal distribution fit to the earnings data. This is a lower bound as fathers and sons likely have correlated earnings even after controlling for occupation.

However, it avoids the upwards bias from the non-classical measurement error. Figure 9 below shows the ratio measure of occupational status using this randomization method.

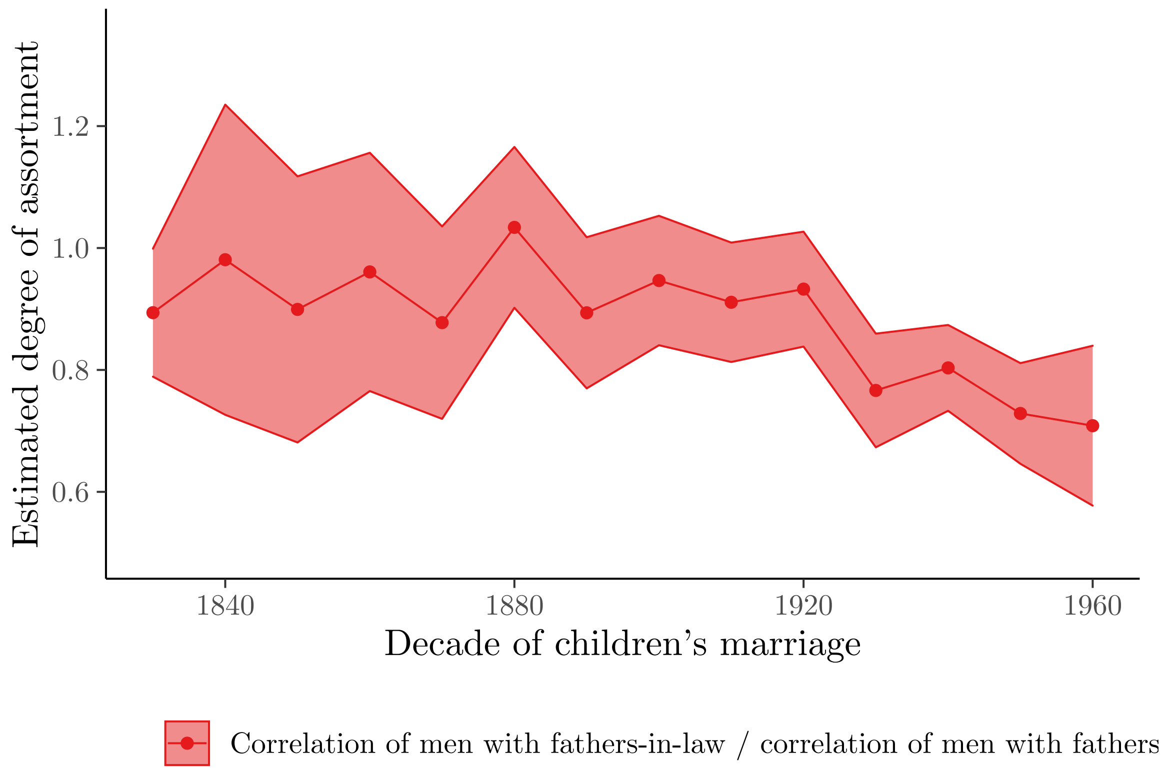

Figure 9: Estimated degree of marital assortment, randomized occupational earnings
Note: 95% bootstrapped confidence intervals shaded (10,000 replications). Imputed earnings are draws from a log normal distribution fit on the earnings for men with the individual's occupation in the 1901 Canadian Census sample (Espín-Sánchez et al. (2019)). The estimates are ratios of rank-rank regression coefficients, which are equal to the correlation coefficients for the ranked variables assuming no ties in rank (and a reasonable approximation if not) (Chetty et al. (2014)). Each individual received a new draw each bootstrap replication, so the confidence intervals are large. Regardless, the overall magnitude and trend remains very similar.

Figure 10 below estimates just the correlation between fathers and fathers-in-law using the same IV strategy mentioned above. This measure is more typically used in the literature (Craig et al. (2020)). However it is not a direct measure of the correlation between spouses. If the matching is at least partially on the characteristics of the groom and bride, the true correlation between spouses will likely be higher than this ratio. Regardless, the trend follows a very similar pattern over time.

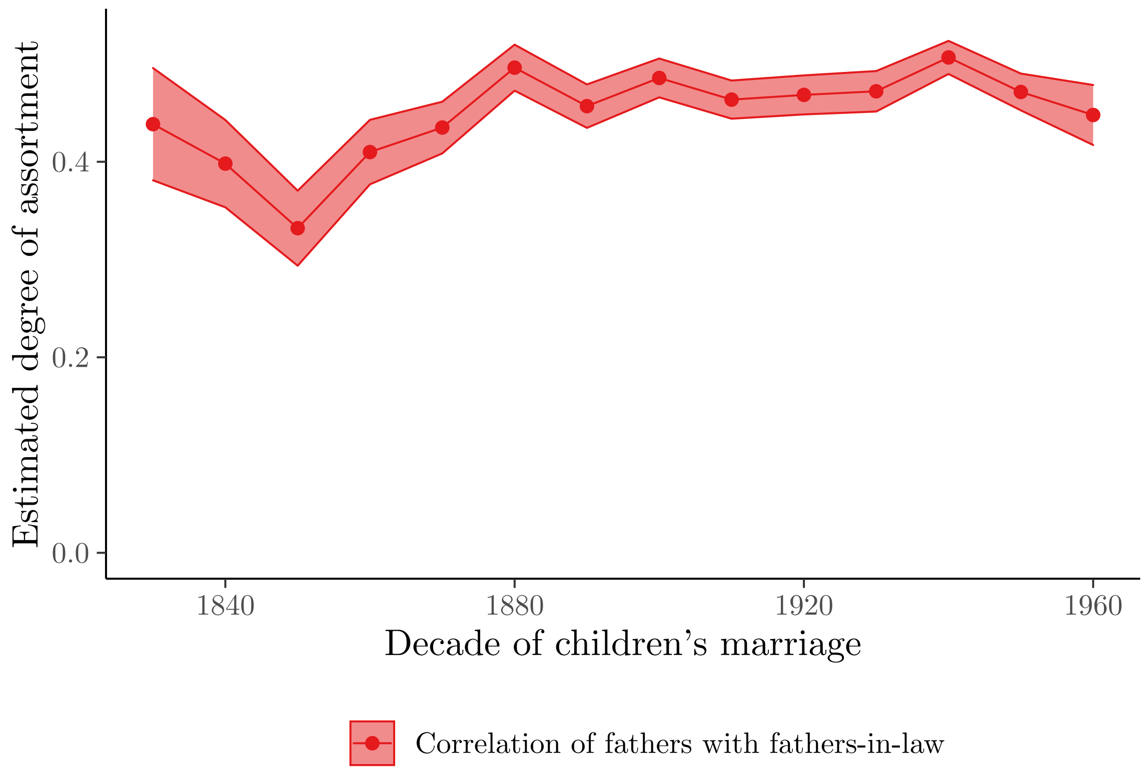

Figure 10: Father-father-in-law correlation, IV
Note: 95% confidence intervals shaded. Imputed earnings are the average earnings for men with the individual's occupation in the 1901 Canadian Census sample. The estimates are rank-rank regression coefficients, which are equal to the correlation coefficients for the ranked variables assuming no ties in rank (and a reasonable approximation if not) (Chetty et al. (2014)). To reduce attenuation bias, the dependent variable in the regressions are instrumented using a second measure of imputed earnings (using the second closest occupation the the individual's first marriage).

A4 Robustness of sorting on individual human capital

The estimates in Table 2 Columns 2, 4, and 6 are identified using family fixed effects. This means that families where one child signed and one child did not sign are the ones driving the results. The reported "identifying observations" are thus the number of individuals in families with at least one literate and one illiterate child.[18] The estimated coefficients are an average treatment effect of individuals in these "treated" families being able to sign. However, it is possible that these families have unusual characteristics; a more interesting average treatment effect is, perhaps, that for the entire population.

One method of estimating this population-wide effect is to estimate the effect separately for each treated family and use a weighted average of the effects (Miller et al. (2019)). The weights are inverse propensity scores, estimated from a logistic regression of an indicator for being treated regressed on observed family characteristics using the entire sample and normalized to sum to one. For this to be a true average treatment effect, the method does come at the cost of several fairly strict assumptions.[19] It is also only reweighting based on selection into identification on observables. Any unobservable characteristic that makes these families unique will not be accounted for. However, regardless of assumptions, it is still a useful exercise to see if the estimates are robust to reweighting.

Here, I estimate the propensity scores using indicator variables for the parent's signatures, the mother's decade of first marriage, the mother's borough of first marriage, the denomination of the parish where the mother first got married, and the number of married children in the family. Missing values are included as an additional category for each indicator variable. As shown in Table 9 below, there is still a positive and significant marriage premium for literacy.

[18] Strictly speaking, as there are decade and sibling order fixed effects, a family where no child signs or all children sign still contributes some to the estimated effect. After demeaning by decade and sibling order, the signature indicator variable will have some within family variation in most cases.

[19] The assumptions: 1. There is no selection into treatment within groups. 2. Conditional on observables, there is no selection into treatment between groups based on heterogenous effects. 3. The logistic regression is the correct functional form. 4. There is a non-zero probability of treatment for every value of observable.

Table 9: Marriage selection, reweighting for selection into identification

	Dependent variable: Husband's characteristic					
	Signed		Log imputed earnings		Father's log imputed earnings	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Effect of wife's human capital						
Wife signed	0.31***	0.20***	0.02***	0.03***	0.04***	0.05***
	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Re-weighted		X		X		X
Wife's family FE	X	X	X	X	X	X
Decade FE	X	X	X	X		X
Observations	1,937,871	1,937,871	971,173	971,173	1,148,769	1,148,769
Panel B: Effect of husband's human capital						
Wife signed	0.28***	0.35***			0.03***	0.11***
	(0.00)	(0.01)			(0.00)	(0.00)
Re-weighted		X				X
Husband's family FE	X	X			X	X
Decade FE	X	X			X	X
Observations	1,928,239	1,928,239			982,166	982,166

Note: *p<0.10; **p<0.05; ***p<0.01. Family-clustered standard errors in parentheses. The sample excludes individuals with one or more unknown parents. Signature variables are indicators that are one if a signature was recorded, zero if the absence of a signature was recorded, and omitted otherwise. Imputed earnings are the average earnings for men with the individual's occupation in the 1901 Canadian Census sample. Fixed effects for decade, marriage number, and sibling order are included in every specification as control variables. Re-weighted estimates are constructed by estimating the effect separately for each family and then taking the weighted average of the effects. The weights are inverse propensity score weights constructed by running a logistic regression of an indicator for if a family had at least one child who signed and one who did not on indicator variables for the parent's signatures, the mother's decade of first marriage, the mother's borough of first marriage, the decade, and the number of married children in each family. Missing values are included as an additional category for each indicator variable in the logistic regression.

A5 Robustness of estimates of effects of parental human capital

The identifying assumption for the analysis in Table 4 is that the human capital of children of parents who remarry did not change over time faster than those of parents who do not remarry. Table 10 replicates the analysis except it drops children of parents who remarry if they were more than one sibling away from a half-sibling in the order of siblings.[20]

[20] As elsewhere, I order siblings by date of first marriage as I do not have birth dates for the entire sample.

This is analogous to restricting the sample to children born on either side of the remarriage, which should limit the importance of differential time trends. The results are very similar.

Table 10: Effect of parents, fixed effects with window

	Dependent variable:			
	Signed Daughter	Signed Son	Log imp. earnings Daughter's husband	Log imp. earnings Son
	(1)	(2)	(3)	(4)
Panel A: Controlling for father				
Mother signed	0.02***	0.03***	0.01	0.03**
	(0.01)	(0.01)	(0.01)	(0.01)
Father FEs	X	X	X	X
Identifying observations	18,303	16,130	8,440	7,438
Observations	1,559,975	1,445,605	944,017	881,588
Adjusted R^2	0.68	0.67	0.37	0.41
Panel B: Controlling for mother				
Father signed	0.02***	0.03***	0.03*	0.04**
	(0.01)	(0.01)	(0.02)	(0.02)
Mother FEs	X	X	X	X
Identifying observations	6,486	5,500	2,799	2,289
Observations	1,522,830	1,411,662	921,530	860,897
Adjusted R^2	0.69	0.68	0.37	0.41

Note: *p<0.10; **p<0.05; ***p<0.01. Family-clustered standard errors in parentheses. The sample excludes individuals with one or more unknown parents. Signature variables are indicators that are one if a signature was recorded, zero if the absence of a signature was recorded, and omitted otherwise. Imputed earnings are the average earnings for men with the individual's occupation in the 1901 Canadian Census sample. Fixed effects for decade, marriage number, and sibling order are included in every specification as control variables. Children of parents who remarry are only included if they were in a window of one away from a half-sibling in the order of siblings.

Table 11 replicates Table 5 except it drops children of parents who remarry if they were more than two siblings away from that a half-sibling in the order of siblings. The logic is the same as above, except I widen the band as the regression compares siblings to full siblings. Again, the results are very similar.

Table 11: Effects of parents, half-siblings with window

	Dependent variable: Younger sibling's characteristic			
	Signed Daughter	Signed Son	Log imp. earnings Daughter's husband	Log imp. earnings Son
	(1)	(2)	(3)	(4)
Panel A: Controlling for father				
Older sibling's characteristic	0.38***	0.36***	0.23***	0.26***
	(0.00)	(0.00)	(0.01)	(0.01)
Signed × same mother	0.05***	0.05***	0.04***	0.05***
	(0.00)	(0.00)	(0.01)	(0.01)
Observations	1,953,016	1,853,707	800,284	756,645
Adjusted R^2	0.63	0.63	0.11	0.14
Panel B: Controlling for mother				
Older sibling's characteristic	0.36***	0.33***	0.24***	0.22***
	(0.01)	(0.01)	(0.02)	(0.02)
Signed × same father	0.07***	0.08***	0.02	0.09***
	(0.01)	(0.01)	(0.02)	(0.02)
Observations	1,946,719	1,777,710	799,342	727,123
Adjusted R^2	0.64	0.63	0.11	0.14

Note: *p<0.10; **p<0.05; ***p<0.01. Family-clustered standard errors in parentheses. The sample excludes individuals with one or more unknown parents. Signature variables are indicators that are one if a signature was recorded, zero if the absence of a signature was recorded, and omitted otherwise. Imputed earnings are the average earnings for men with the individual's occupation in the 1901 Canadian Census sample. Fixed effects for decade, marriage number, and sibling order as well as the non-interacted same mother indicator variable are included in every specification as control variables. Children of parents who remarry are only included if they were in a window of two or less away from the nearest half-sibling in the order of siblings.

A6 Robustness of effect of sorting on intergenerational elasticity

One concern with Table 6 is that families where only one parent was literate were selected on some omitted factor that decreases intergenerational mobility. One way to overcome this endogeneity is to find a variable that changes the degree of assortment of the parents' marriage and only matters for the outcome of the children through the degree of assortment. One plausible variable that meets these criteria is the fraction of the mother's older siblings who are female (Abramitzky et al. (2011b), Caron et al. (2017), and Dillon (2010)). Unfortunately, I do not observe ages in most of the sample, I instead use the percentage of the mother's siblings who got married before her who are female.

The gender of children should be, at least at birth, as good as random, especially as there is no evidence of parity-dependent fertility control (Clark et al. (2020)). Why should this matter for sorting? One could imagine a scenario where a set of sisters has multiple potential suitors of similar characteristics in their neighborhood or social network. As more of the sisters marry, the remaining sisters will have to be less picky. It is possible that older sisters have a different effect on younger sisters compared to older brothers.[21] However, if it merely changes the human capital of the younger sister, who then matches accordingly, it shouldn't introduce bias.

As shown in Table 12, the fraction female decreases the association between the signature rates of spouses and decreases the intergenerational elasticity between fathers and sons. This is exactly what we'd expect if the mother directly mattered for the outcomes of children.

[21] In fact, this is fairly likely. In preliminary research I have conducted for another project, I find that before 1849, the fraction of older siblings that are male increases the rate of infant mortality for younger sisters.

Table 12: Father-son intergenerational elasticities, more and less assorted marriages

	Dependent variable:	
	Mother signed	Log earnings score
	(1)	(2)
Fraction female (married before mother)	0.01***	0.13***
	(0.002)	(0.03)
Father signed	0.37***	
	(0.002)	
Father signed X fraction	−0.01***	
	(0.003)	
Father's log earning score		0.41***
		(0.004)
Father's log earning score X fraction		−0.02***
		(0.005)
Sibling marriage order FEs	X	X
Decade FEs	X	X
Observations	379,251	390,145
Adjusted R^2	0.55	0.13

Note: *p<0.10; **p<0.05; ***p<0.01. Family-clustered standard errors in parentheses. Signature variables are indicators that are one if a signature was recorded, zero if the absence of a signature was recorded, and omitted otherwise. Earning scores are the average earnings for the individual's occupation in Quebec based on the 1901 Canadian Census sample. The fraction is the share of the mother's siblings who married before her that were female. Decade fixed effects are included in every specification as control variables.