ECARES

The her in inheritance:
how marriage matching has always mattered, Quebec 1800-1970

Matthew Curtis
ECARES, Université libre de Bruxelles

November 2022

ECARES working paper 2022-38

ECARES
ULB - CP 114/04
50, F.D. Roosevelt Ave., B-1050 Brussels BELGIUM
www.ecares.org

The her in inheritance: how marriage matching has always mattered, Quebec 1800–1970

Matthew Curtis*

September 2022

Abstract

When did marriage become strongly assortative? Is it a recent development, a consequence of increased female employment and a cause of rising inequality? A long run perspective is necessary to answer this question. This paper uses a uniquely suitable database from Quebec 1800–1970 to provide such a perspective. First, it develops a novel method which reveals that marriage was highly assortative as far back as the early nineteenth century. Next, it shows this matching depended on the individual human capital of women, not just on family backgrounds. Finally, it shows that mothers had a causal impact on child outcomes independently from fathers. Thus, despite deeply conservative gender norms, marriage matching — and women — had always mattered for social mobility.

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

JEL codes J12, J62, N31, N32

Access to the data used in this paper is restricted due to ethics regulations. They are available upon request from the BALSAC Project. 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, Lisa Dillon, and Eugénie Capel for help with the dataset. Thanks also to Paula Gobbi, Joanne Haddid, Laura Salisbury, José-Antonio Espın-Sánchez, Zachary Ward, and Yuzuru Kumon for helpful feedback. Thanks finally to seminar and conference participants at UC Davis, the EHA Annual Meeting, the All UC Graduate Economic History Workshop, ECARES, UNamur, and ECORES for their helpful feedback and comments. All errors that remain are my own.

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

1 Introduction

When did spouses start matching closely on their socioeconomic characteristics? This question is important because in recent decades, developed economies have seen an increase in both female labor force participation rates and educational attainment (Goldin, 2006; Piketty and Saez, 2014). This has led to widespread concern that marriage has become more assortative, increasing household inequality. However, the empirical evidence suggests only very modest increases after 1960 (Eika et al., 2019; Greenwood et al., 2014b).[1] Moreover, this increase in sorting has had little direct impact on household inequality (Greenwood et al., 2014a; Hryshko et al., 2017). Why did the economic empowerment of women seem to matter so little for assortment and inequality?

To explain this puzzle, this paper argue we must take a longer-run perspective. To examine marriage sorting before female empowerment, it uses an unusually rich dataset from the Canadian Provence of Quebec 1800–1960. This setting is ideal, not only because of the strenght of the data but also because its similarity to other advanced economies and its deeply conservative gender norms.

I find that, despite limited female labor force participation, marriage had always been highly assortative. There was little room for it to become more so. To explain why, this paper adds to the growing literature on historical female mobility by considering the role of maternal human capital on children (Craig et al., 2020; Olivetti and Paserman, 2015). Marriage had always mattered for inequality because mothers had always mattered for child outcomes.

This conclusion is based on the answers to three related empirical questions. First, how did the degree of marital assortment evolve over the long run? Historical studies are often limited due to missing data on married women (Olivetti et al., 2020). Using a novel method to overcome this, I estimate a correlation between spouses that is surprisingly strong — around 0.85 — and stable over time. This method has since been extended to other contexts

[1] Roughly an increase of 0.05 in correlation for the United States.

by Clark and Cummins (2022), Clark et al. (2022), and Luo (2022).

Second, was matching merely the result of negotiation between families of similar socioeconomic status (e.g. Puga and Trefler, 2014)? Instead, I find evidence that individual human capital mattered. For example, a woman who could sign her name married a man 34 percentage points more likely to sign his name than her sister who could not.

Third, did mothers influence child outcomes directly? As marriages were assortative, it is challenging to untangle the independent effect of a mother (Espín-Sánchez et al., 2022). Using the unusually high frequency of remarriage to control for the father, I find evidence that the human capital of mothers had an independent causal impact on child outcomes. Together, the answers to these questions demonstrate that marriage sorting mattered long before married women held formal employment.

I am able to answer these questions by using millions of linked marriage records from the BALSAC database (Project Balsac, 2020b).[2] These data have several unique features that 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.

This paper thus contributes to the literature on marriage sorting. First, while trends in sorting are well studied after the mid-20th century, there are few studies that extend the analysis further into the past (Clark and Cummins, 2022; Craig et al., 2020; Schwartz and Mare, 2005; Shiue and Keller, 2022). Second, borrowing from the intergenerational mobility literature, it develops a new method to account for attenuation bias in measures

[2]The database has been recently updated to include data from the PRDH (2020) which has been used in the previous economics literature, for example Galor and Klemp (2019).

of assortment (Modalsli and Vosters, 2019; Nybom and Stuhler, 2017; Ward, 2021). By doing so, it relates to studies that explain how changes in the structure of educational categories make simple correlations unreliable (Chiappori et al., 2020; Liu and Lu, 2006). Third, it argues that while assortment might not have substantially increased inequality when female labor force participation rose, it had always increased inequality through social mobility (Eika et al., 2019; Greenwood et al., 2014a; Hryshko et al., 2017). Finally, it adds to the growing literature on the mechanisms of marriage matching, both historical and intergenerational (Abramitzky et al., 2011; Fagereng et al., 2022; Goñi, 2018).

This paper also adds to our understanding of historical intergenerational mobility. Studies of intergenerational mobility have often overlooked women (Black and Devereux, 2011). 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). In many societies, linking married daughters to fathers requires either pseudo-linkages on first names or recovering their maiden names from marriage records (Craig et al., 2020; Goñi, 2018; Olivetti and Paserman, 2015; Olivetti 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; 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 an ideal setting to consider assortative marriage over the long run. 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 matching 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 Quebec 1800–1970 was in many ways very similar to the rest of North America, it had relatively conservative gender norms. This makes it a useful setting to consider the role of women in assortative marriage and social mobility. If they mattered in deeply conservative Quebec, then they surely did so in the rest of North America.

While its economic development lagged other North American regions, Quebec 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). Quebec had lower social mobility than the rest of Canada, but Canada as a whole was more mobile than Europe (Antonie et al., 2022). Before its Quiet Revolution of the 1960's, Quebec was also much less secular than its neighbors.[3] Catholicism asserted significant control over public education and social norms, and deeply conservative beliefs about gender roles were enshrined by law and public policy. For a population of European descent, the Québécoise had a late demographic transition and unusually large family sizes (Vézina et al., 2014). 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.

[3]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, a codified system of customary French law. With some modifications, these customs were incorporated into the 1866 Civil Code of Lower Canada, which was in force until 1994 (McCord, 1867). Under Quebec law, 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,[4] 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).[5] Only after reforms starting in 1964 were married women no longer considered legally incapable.

In theory, the law did not discriminate when it came to the inheritance of daughters. After a married man died, the community of property was dissolved by giving the widow her share and dividing the rest equally amongst the children regardless of gender.[6] 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). Thus, parents could and often did favor a single

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

[5]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.

[6]The widow's claim was gradually weakened over time (Dechêne, 1974).

male heir.

2.2 Marriage and family

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 (Greer, 1985). 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.[7] Another factor was that girls were considered less useful for household production. In poorer families, girls would be encouraged to marry early in order to reduce the burden on her family (Dechêne, 1974).

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, married women would spend much of her life pregnant and raising small children (at least in the earlier part of the period studied).

[7] Which, at least in the similar case of first cousins wishing to marry, they were unusually liberal in granting. This was apparently due to the credible threat of either cohabitation or defection to a Protestant church.

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. Before the widespread introduction of labor-saving household devices, simple yet tedious tasks like washing and ironing clothes 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.[8]

2.4 External validity

Overall, how much was Quebec an outlier? 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 was still a variant of the European marriage pattern. The role of women in

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

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 surely did so in neighboring regions as well. I argue, therefore, that my findings from Quebec are very likely generalizable to the rest of North America (and most likely Western Europe and its other offshoots as well).

3 Data

This paper uses the Project BALSAC[9] database from the Université du Québec à Chicoutimi (Project Balsac, 2020a). The database has been developed since 1971 and contains over 6 million unique individuals from the first European settlement to the present. Notably, it contains all Catholic marriages from 1621 to 1965 (Vézina and Bournival, 2020). Protestant records are less complete, but most of the population were Catholic. These marriages have been linked together to reconstruct families and multigenerational lineages.

The database has recently been expanded with births and deaths through 1849. Records from before 1800, while not used in this paper, were integrated into the BALSAC dataset from the Registre de la population du Québec ancien (RPQA) dataset of the Programme de recherche en démographie historique (PRDH) at the Université de Montréal (PRDH, 2020).[10]

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 the Saguenay-Lac-Saint-Jean 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.

[9]BALSAC is an acronym of the initial regions of Quebec studied.

[10]This dataset may be more familiar to economists; Galor and Klemp (2019).

Table 1 presents summary statistics from the dataset. The main unit of observation is a marriage, linked to both the groom's and the bride's parents. Each observation contains, when available, information about the human capital of the bride, of the groom, and of all four parents. In the rest of this section, I discuss in detail how the links were constructed, what measures of human capital I use, and how reliably 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).

The vital records in the database have been reconstituted into families using computer-assisted linkage. The links are almost entirely based on names, with dates being used to validate links after they are formed (Vézina and Bournival, 2020; Vézina et al., 2018). Names are standardized using the FONEM phonetic program (Bouchard et al., 1981). Manual linkage is used around 20% of cases where there was no unique match. 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 that I use in this paper, for both men and women, 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; the priest was required to record if they were not (Gagnon et al., 2011). I code a signature variable as one if the individual signed their marriage record and zero if they were unable to sign. I omit signatures that are either missing or unrecorded. As shown in Figure 1, this definition produces a trend that is close to external estimates of literacy.

Was literacy human capital in the sense of increasing economic productivity? The qualitative evidence suggests it was. The ability to write had always been associated with business activity in Quebec (Greer, 1997). Table 2 shows the twenty most common occupations for men in the 19th century. Signature rates range widely with a clear occupational hierarchy. Skilled professionals and merchants were highly likely to sign, craftsmen were somewhat less likely to sign, and workers in the primary sector were far less likely to sign. As for reading, it too was likely associated with economic activity. As opposed to their Protestant neighbors who prioritized literacy education for religious ends, Quebec's Catholics would have considered reading the Bible a virtue but not a necessity.

A second proxy for human capital, only reliably available for men, is occupational status. I assign 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. My scores of choice are imputed 1901 earnings. I construct them using a 5% sample of the 1901 Canadian Census for a given occupation (Canadian Families Project, 2002; Minnesota Population Center, 2019). For each occupation, I simply take the average yearly earnings reported by men in Quebec.[11] There are numerous other

[11] This requires crosswalking occupations from IPUMS's occupational codes to the original HISCO

ways to rank occupations, as discussed in Appendix A3. However, imputed 1901 earnings have several advantages. First, they are easy to interpret: they are how much the individual would earn, on average, with their occupation in 1901 in Quebec. Second, they are imputed from data roughly in the middle the time period considered. Third, are at least a proxy for the standard variable of interest in intergenerational mobility studies (lifetime earnings). Fourth, they produce similar estimates to the other occupational scores. Finally, I argue it is reasonable to assume that the average earnings of an occupation was 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.

Table 1: Summary statistics

Gender		Mean	SD	Min	Max	N
Female	Year	1914	43	1800	1969	2,122,466
	Marriage number	1.05	0.24	1	7	2,122,466
	Signature	0.81	0.39	0	1	1,993,668
	Signature, mother	0.67	0.47	0	1	1,670,151
	Signature, father	0.61	0.49	0	1	1,662,365
	Earnings					0
	Earnings, father	343.89	173.02	70.67	2000.00	960,792
	Sibling order	2.08	1.36	1	14	1,990,380
Male	Year	1914	43	1800	1969	2,122,466
	Marriage number	1.10	0.33	1	7	2,122,466
	Signature	0.79	0.41	0	1	1,986,326
	Signature, mother	0.65	0.48	0	1	1,617,849
	Signature, father	0.59	0.49	0	1	1,609,898
	Earnings	393.46	215.67	70.67	2000.00	1,080,521
	Earnings, father	344.23	172.44	70.67	2000.00	940,873
	Sibling order	2.02	1.32	1	13	1,976,175

Note: Each observation is a marriage, with marriage number the number of the marriage for the relevant participant. Signature variables are indicators that are one if a signature was recorded, zero if the absence of a signature was recorded, and omitted otherwise. Earnings are the imputed annual earnings for the individual's occupation in 1901 Canadian dollars (see text). Sibling order is the order amongst all married siblings by date of first marriage (as birth dates are not reported after 1849).

scheme (Zijdeman, 2014).

Table 2: Twenty most common 19th century occupations ranked by literacy

HISCO	Occupation	Translation	Signature rate	Rank	Percent	Average year
06105	Medecin	Doctor	0.99	18	0.01	1875
45125	Commis marchand	Merchant clerk	0.98	17	0.01	1882
41025	Marchand	Merchant	0.94	5	0.03	1867
79100	Tailleur	Tailor	0.87	13	0.01	1869
93120	Peintre	Painter	0.75	16	0.01	1876
77310	Boucher	Butcher	0.71	15	0.01	1873
77620	Boulanger	Baker	0.62	12	0.01	1868
95410	Menuisier	Carpenter	0.55	3	0.05	1861
83110	Forgeron	Blacksmith	0.52	6	0.02	1863
77120	Meunier	Miller	0.50	20	0.00	1862
80110	Cordonnier	Shoemaker	0.50	4	0.03	1865
76145	Tanneur	Tanner	0.44	19	0.00	1856
98135	Navigateur	Navigator	0.44	7	0.02	1864
95135	Maçon	Mason	0.34	14	0.01	1852
98620	Charretier	Carter	0.33	8	0.02	1870
61110	Cultivateur	Farmer	0.32	1	0.48	1863
43220	Voyageur	Fur trader	0.20	10	0.01	1866
99910	Journalier	Day labourer	0.19	2	0.15	1863
64100	Pecheur	Fisherman	0.14	11	0.01	1874
62105	Laboureur	Laborer	0.05	9	0.01	1819

Note: Observations are grooms at time of their marriage. Occupation titles are taken from the most common within a HISCO code (Van Leeuwen et al., 2004). Signature variables are indicators that are one if a signature was recorded, zero if the absence of a signature was recorded, and omitted otherwise. Percent is the percentage of all non-indeterminate occupations with that HISCO code. The average year is the average year of marriage. In Quebec, *journalier* refers to workers paid by the day whether or not they work in agriculture.

3.3 Reporting of the characteristics of women

Do the vital records accurately report the occupational status and human capital of women? Four extracts of Canadian censuses from 1881–1911 and data compiled by Long (1958) for 1920–1960 provide external points of comparison (Canadian Families Project, 2002; Dillon et al., 2008; Gaffield et al., 2009; Inwood and Jack, 2011; Killingsworth and Heckman, 1986; Long, 1958; Minnesota Population Center, 2019).[12] Figure 1 compares the fraction of individuals who signed their first marriage record to the fraction who self-reported the ability 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 censuses. Two patterns are particularly notable. First, Quebec went from a very low human capital society to a high human capital society between 1800 to 1920. Second, there was actually a gender gap in favor of women between 1850 to 1920.[13]

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 formal employment rate of married women and almost entirely omit unmarried women with occupations. This is especially damning as the censuses very likely underreport female employment as well. One pattern, however, is clear. While unmarried women often worked outside the home, married women did not begin to report formal employment in substantial numbers until the 1940's.

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

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

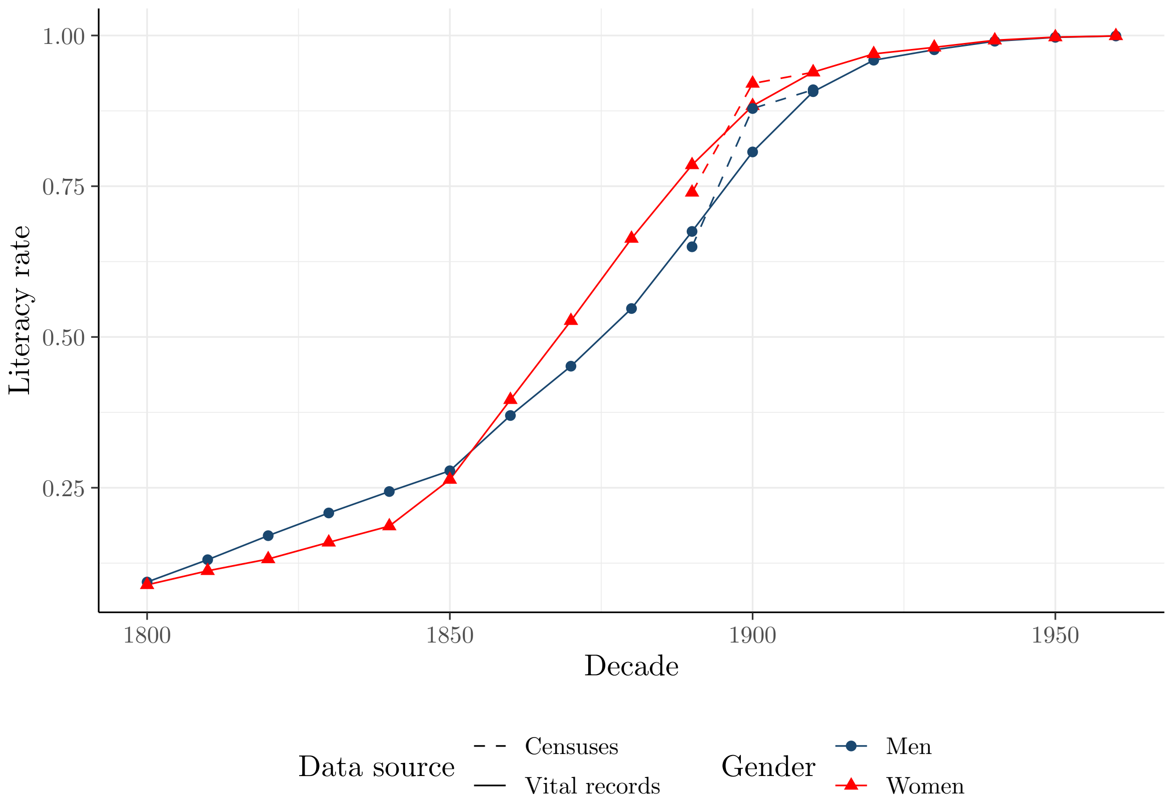

Figure 1: The vital records accurately report the ability to write

Note: For the vital records, literacy is proxied by a signature variable that is one if a signature was recorded, zero if the absence of a signature was recorded, and omitted otherwise. The census 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.

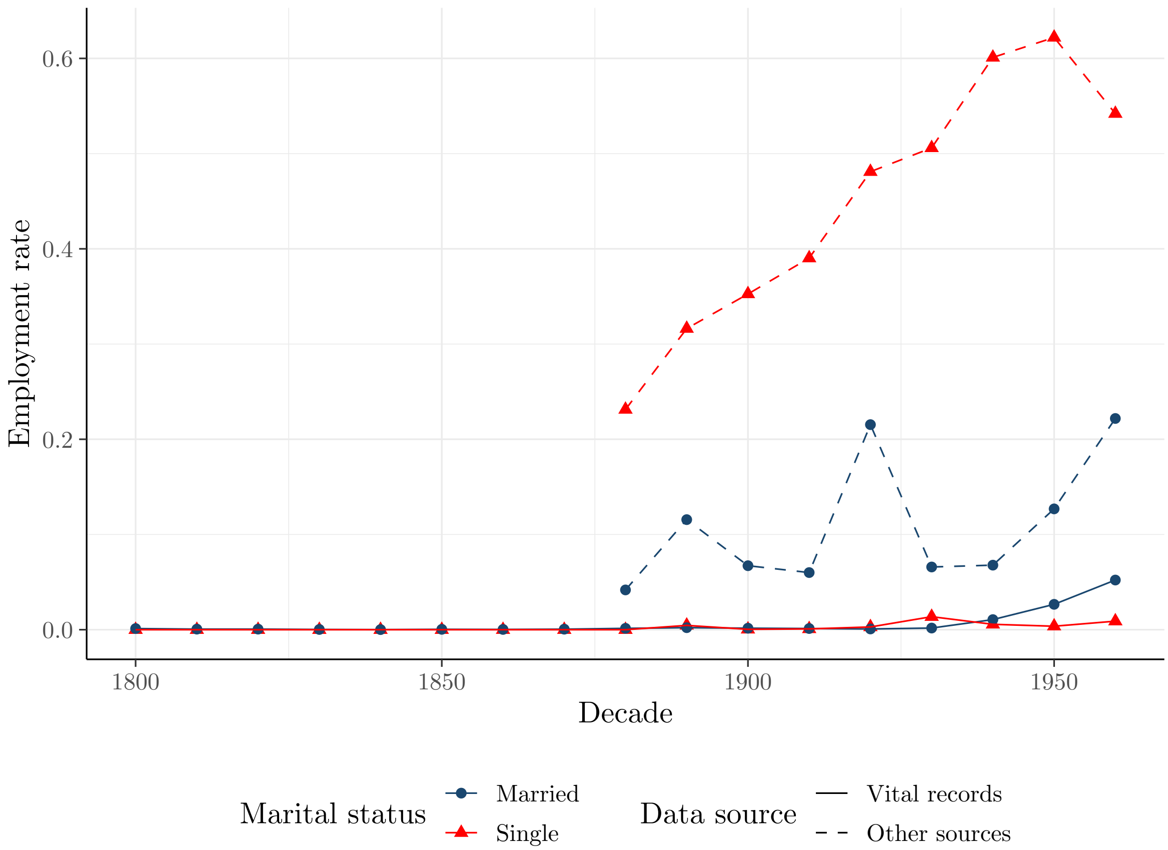

Figure 2: The vital records omit formal female employment

Note: For the vital records, a woman is counted as formally employed if she had an occupation listed in any record. She is 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 employment rate is the fraction of women aged at least 16 with an occupation. After 1920, the other sources are data compiled by Long (1958), with the employment rate for married women calculated as an average of the rate for currently married women and the rate for 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 of 1830–1969.

This ratio method, first developed for this paper, has since been extended to other contexts. Clark and Cummins (2022) similarly finds a degree of assortment of 0.8–0.9 in England 1837–2021. Clark et al. (2022) adapts the method to estimate rates of intergenerational mobility.[14] Finally, Luo (2022) develops a similar model for Imperial China 1614–1854 under the assumption that matching is solely between the father and father-in-law.

4.1 The model

For each individual of type j related to marriage i, let X_i^j be an unobserved measure of socioeconomic status (henceforth referred to as simply "status"). Assume for each type j, X^j has a standard deviation normalized to σ. For example, X_i^j could be the lifetime income percentile rank of all in group j for marriage i.

Also, following Solon (1992) and Clark and Cummins (2015), assume only an imperfect measure or proxy Y is observed for X. In the example example, Y_i^j could be the income in a given year. For some individuals, no Y_i^j is observed.

Then, similar to Espín-Sánchez et al. (2022), assume that the status of the groom of i, X_i^g, is inherited depending on the status of his father X_i^f and of his mother X_i^m:

$$X_i^g = \beta_f X_i^f + \beta_m X_i^m + e_i^g \tag{1}$$

[14]Under very similar assumptions, the ratio of the correlation between fathers and fathers-in-law to the correlation of grooms and fathers-in-law will be a measurement error corrected estimate of correlation between fathers and sons.

where e_i^g is a random term uncorrelated with the X_i^j's.

For now, I assume that the effect on children is the same regardless of gender. Thus, the status of the bride of i, X_i^b, is inherited depending on the status of the groom's father-in-law X_i^{fl} and of his mother-in-law X_i^{ml}:

$$X_i^b = \beta_f X_i^{fl} + \beta_m X_i^{ml} + e_i^b \tag{2}$$

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.[15]

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

$$corr(X_i^g, X_i^b) = \gamma \tag{3}$$

As each X_i^j has standard deviation σ, then the coefficient from a simple OLS regression of X_i^g on X_i^g will be the same as the correlation coefficient. Thus:

$$X_i^g = \gamma X_i^b + n_i^b \tag{4}$$

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

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

This could be estimated with a regression:

$$Y_i^g = a + bY_i^f + \epsilon_i \tag{6}$$

[15]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.

where we want to estimate:

$$b = (\beta_f + \gamma\beta_m) \tag{7}$$

However, as we only observe Y_i^g and Y_i^f with measurement error, we instead get:

$$\hat{b} = \theta_f(\beta_f + \gamma\beta_m) \tag{8}$$

where θ_f is a term representing the attenuation bias. If the measurement error is classical, that is if $Y_i^j = X_i^j + u_i^j$ and u_i^j is uncorrelated with Y_i^j, then $plim\ \theta_f = \frac{\sigma^2_{X_i^f}}{\sigma^2_{X_i^f} + \sigma^2_{u_i^f}}$.

Note that if β_f and γ are greater than zero, then mothers contribute to the observed correlation between Y_i^g and Y_i^f. If the parameter of interest is β_f, the direct influence of the father, then the estimate $\hat{b}$ is biased upwards due to the omitted variable of the mother's status. Even if mothers are not observed, they thus could be driving results in the typical intergenerational mobility regression used in the literature. Section 6 below further discusses the implications of omitting the mother.

To conclude, note now that γ, the degree of assortment, is part of the equation determining the intergenerational correlation between fathers and sons. As long as $\beta_m > 0$ — that is, as long as the status of mothers has a direct effect on that of sons — assortment will slow social mobility, increasing overall inequality.[16] 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 The ratio method for measuring assortment

Using the model, I propose a new method to estimate the underlying degree of assortment. The method has two key advantages. First, it accounts for measurement error. Second, it

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

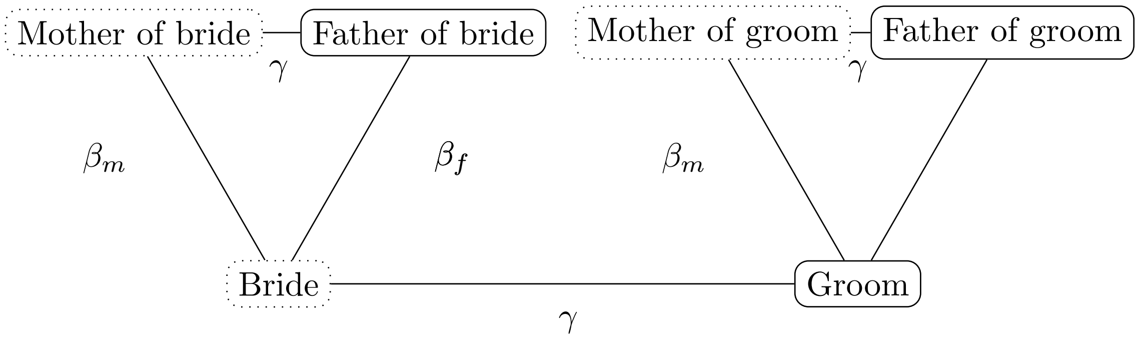

Figure 3: Simple model of assortment

Note: Solid lines represent a direct causal relationship. Solid nodes represent an individual with an imperfect (yet observed) measure of status. Dashed nodes represent an individual with unobserved status.

can be estimated even if women have no observed measures of status. These advantages come at the cost of a few strict assumptions (yet ones that I demonstrate are plausible in my context).

The method compares the correlation between the groom and his father-in-law and the correlation between the groom and his own father. As the standard deviations are normalized, the simple linear regression coefficient is a correlation. Therefore:

$$Y_i^g = \gamma\theta_{fl}(\beta_f + \gamma\beta_m)Y_i^{fl} \tag{9}$$

$$Y_i^g = \theta_f\gamma(\beta_f + \gamma\beta_m)Y_i^f \tag{10}$$

and

where the θ_j's represent the attenuation bias from measurement error.

Regressing Y_i^g on Y_i^{fl} and Y_i^f, the ratio of the coefficients will be:

$$\frac{\gamma\theta_{fl}(\beta_f + \gamma\beta_m)}{\theta_f(\beta_f + \gamma\beta_m)} \tag{11}$$

which is equal to γ in the probability limit if we assume that *plim* θ_{fl} = *plim* θ_f, i.e. we assume that the bias in both regressions is the same.

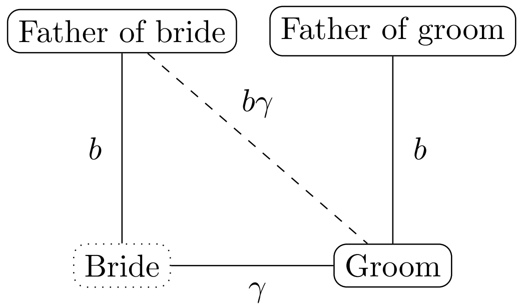

Figure 4: Ratio method model

Note: Signature variables are indicators that are one if a signature was recorded, zero if the absence of a signature was recorded, and omitted otherwise. Dashed lines represent an observed correlation. No correction has been made for measurement error. The pattern observed supports the ratio method assumptions described in the text.

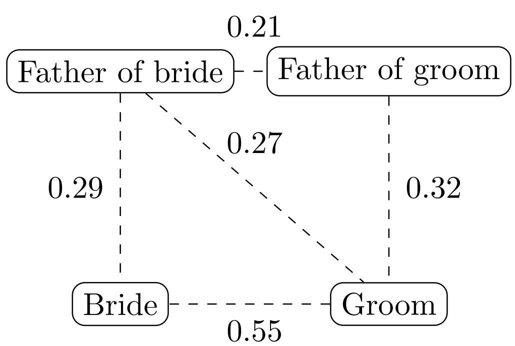

Figure 5: Observed signature correlations

Note: Solid lines represent a direct causal relationship. Dashed lines represent an indirect linkage. Solid nodes represent an individual with an imperfect (yet observed) measure of status. Dashed nodes represent an individual with unobserved status.

To review, the key assumptions for the ratio to be equal to the degree of assortment are:

- Equal inheritance: Latent status is inherited at the same rate regardless of if the child is male or female.

- Equal bias: $plim\ \theta_{fl} = plim\ \theta_f$

Figure 5 shows that the equal inheritance assumption is reasonable my context. In the next section, I discuss the equal measurement error assumption.

4.3 Equal measurement error assumption

One important way in which the errors can be non-classical is if n_i^b is correlated with X_i^{fl}. That is, if the matching equation between the bride and groom has an omitted variable bias from the status of the father-in-law. This would be the case if the groom is matching directly with his father-in-law as well as his bride. Figure 5 shows that the observed correlations of signatures are consistent with this assumption, as the correlation is stronger between the bride and groom than between the groom and father-in-law or the father and father-in-law.

How plausible is assuming that the groom matches only with the bride? Arranged marriages are not a problem as long as those doing the arranging only consider the status of the bride and groom. This assumption is only violated if the status of the groom's in-laws is considered independently from that his spouse inherits. If an elite man's daughter marries well because she benefits from his wealth or social connections, that can be considered inherited status and thus not a problem.

In one potential robustness test, the measurement error in both regressions can be reduced by using a second measure of status as an instrument (Altonji and Dunn, 1991; Modalsli and Vosters, 2019; Ward, 2021). The ratio can then be computed. This method is discussed further in the next section.

5 Empirics

The following section outlines the three main findings of this paper. First, I use the simple model in the previous section to estimate the degree of assortment over time. This estimate is high and stable throughout the period. Then, I provide evidence that this estimated correlation was 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. All considered, 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 6 plots the correlation of spouses' literacy, proxied by signatures. 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 many years 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 described in the previous section, an alternative measure can be constructed by comparing the correlation of the 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 status equally regardless of gender and that the father-son correlation has the same measurement error as the father-in-law-son-in-law correlation. 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 is symmetric across gender.

For my preferred measure of occupational status, I use the imputed annual earnings in 1901 Canadian dollars. To normalize the standard deviations, I compute Spearman's rank correlation coefficients, which is equivalent to estimating the canonical rank-rank mobility regressions (Chetty et al., 2014).

Figure 7 plots the ratio measure over time and Figure 8 shows the correlations used to compute the ratio. The estimated correlation between spouses is very high, around 0.8, and it appears to be stable throughout the period. The overall trend is similar to that in Figure 6 before 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. Together, I interpret the two figures as consistent with a story where the degree of assortment is high and stable throughout the entire period.

A more standard method to account for measurement error is IV regression. IV can be used in the case of classical measurement error if a second measure of the independent variable is available (Solon, 1992; Ward, 2021). This addresses measurement error on the right-hand side which, for simple linear regression models, is sufficient. However, when estimating correlations, both variables are normalized according to the observed distributions. This introduces measurement error on both the right- and left-hand side. As the ratio method uses correlations, a different error correction procedure is necessary. Nybom and Stuhler (2017) proposes one such method.

The procedure computes the correlation of $\tilde{x}^*$ and $\tilde{y}^*$, the true ranked variables, when

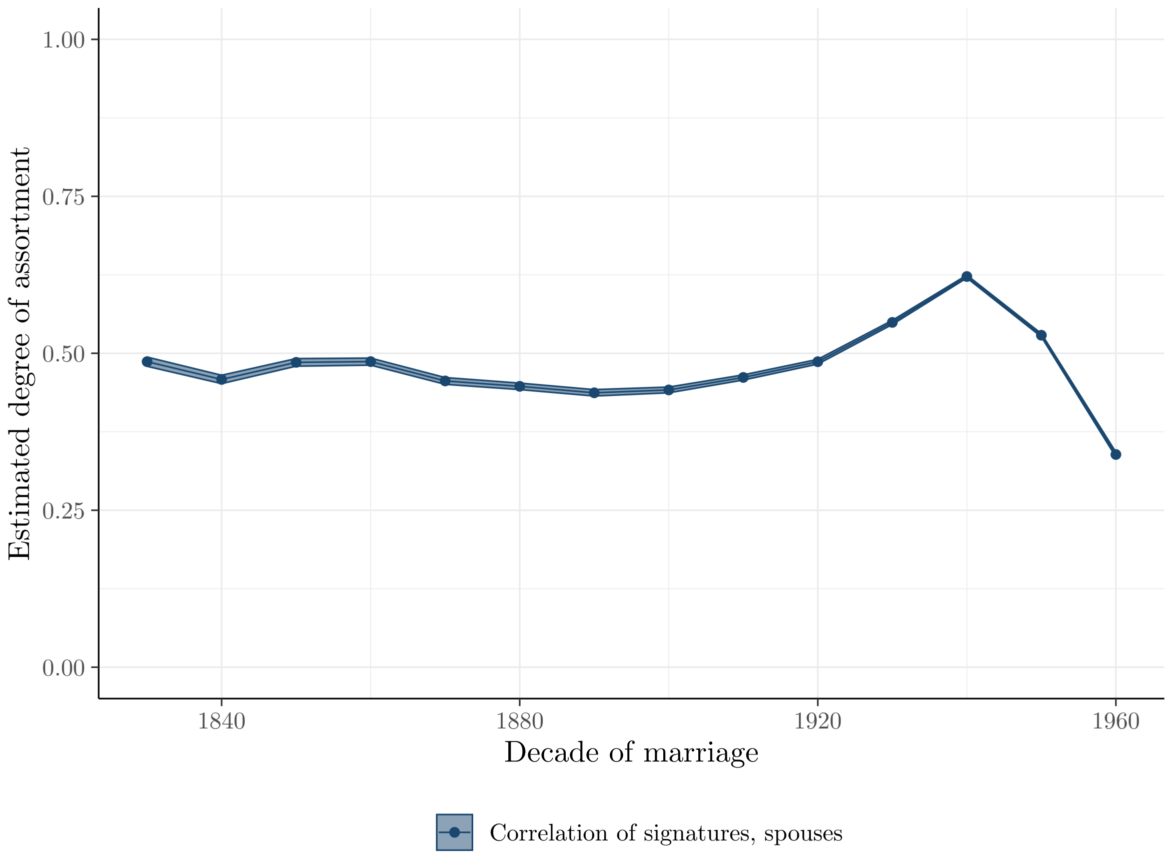

Figure 6: Correlation of spouses' signatures

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 7: Ratio measure of martial sorting using imputed earnings

Note: 95% bootstrapped confidence intervals shaded (10,000 replications). Imputed earnings are the imputed annual earnings for the individual's occupation in 1901 Canadian dollars (see text). Spearman's rank correlations are used (which is equivalent to the correlation of the ranks).

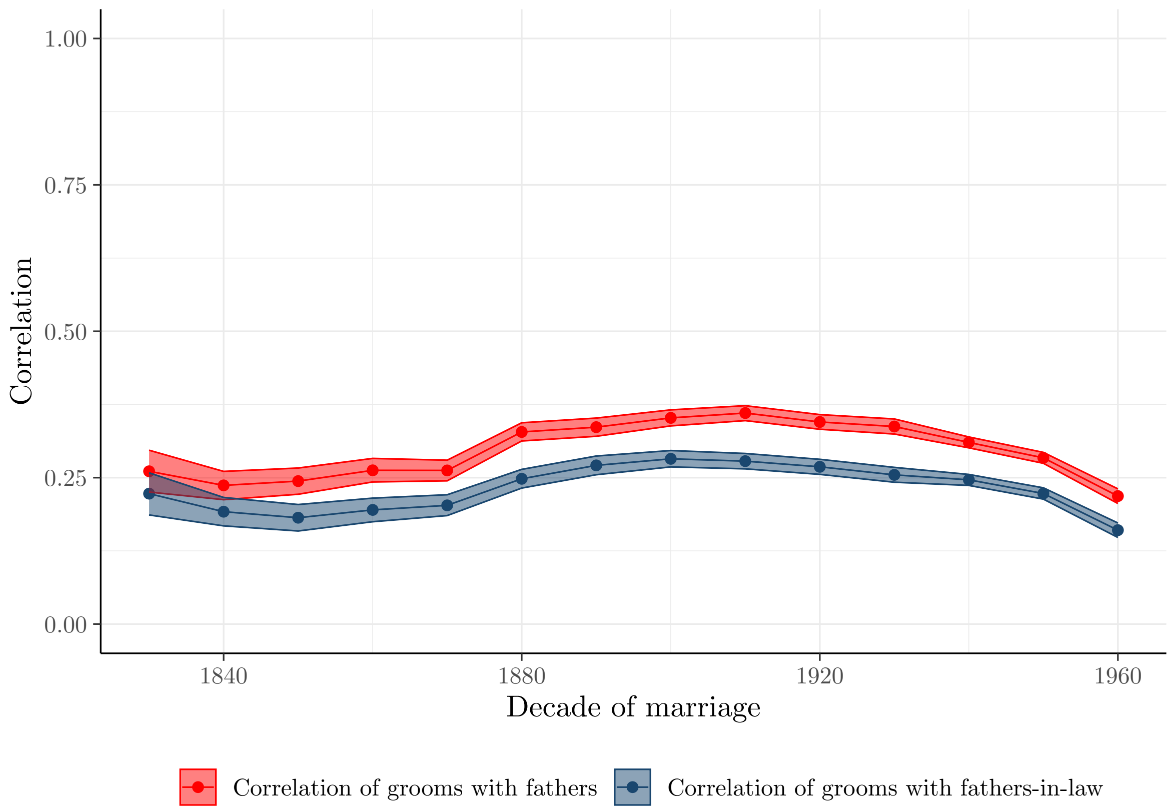

Figure 8: Correlations used to compute ratio

Note: 95% bootstrapped confidence intervals shaded (10,000 replications). Imputed earnings are the imputed annual earnings for the individual's occupation in 1901 Canadian dollars (see text). Spearman's rank correlations are used (which is equivalent to the correlation of the ranks).

only $\tilde{x} = \tilde{x}^* + \tilde{u}$ and $\tilde{y} = \tilde{y}^* + \tilde{v}$ are observed. $\tilde{w}$ and $\tilde{v}$ are the errors in rank. Assume x^* and y^* are measured with classical measurement error. The resulting errors in rank will be non-classical, but can still be addressed using instruments.

First, define λ_y and λ_x as the linear projection of the observed variables on the true variables. Then:

$$\tilde{y} = \alpha_y + \lambda_y \tilde{y}^* + \tilde{w}_y \tag{12}$$

and

$$\tilde{x} = \alpha_x + \lambda_x \tilde{x}^* + \tilde{w}_x \tag{13}$$

where $\tilde{w}_y$ and $\tilde{w}_x$ are now uncorrelated error terms.

If $\rho(x^*, y^*)$ is the true Spearman's correlation coefficient, then (after noting that all the ranked variables have the same variances) the observed Spearman's correlation is:

$$\rho(x, y) = \frac{Cov(\tilde{x}, \tilde{y})}{\sqrt{Var(\tilde{x})(\tilde{y})}} = \frac{\lambda_x \lambda_y Cov(\tilde{x}^*, \tilde{y}^*) + \lambda_x Cov(\tilde{x}^*, \tilde{w}_y) + \lambda_y Cov(\tilde{y}^*, \tilde{w}_x)}{\sqrt{Var(\tilde{x}^*)(\tilde{y}^*)}} = \lambda_x \lambda_y \rho(x^*, y^*) \tag{14}$$

Then, assume we observe additional measures of status $\tilde{x_2} = \tilde{x}^* + \tilde{u_2}$ and $\tilde{y_2} = \tilde{y}^* + \tilde{v_2}$. Assuming $Cor(\tilde{u}, \tilde{u_2})$ and $Cor(\tilde{v}, \tilde{v_2}) = 0$, that is the rank error is uncorrelated across multiple observations of the same individual, then:

$$Cor(\tilde{x}, \tilde{x_2}) = \lambda_x^2 \tag{15}$$

and:

$$Cor(\tilde{y}, \tilde{y_2}) = \lambda_y^2 \tag{16}$$

Thus, with the three sets of correlations above, $\rho(x^*, y^*)$ can be calculated:

$$\rho(x^*, y^*) = \frac{\rho(x, y)}{\sqrt{Cor(\tilde{x}, \tilde{x_2})Cor(\tilde{y}, \tilde{y_2})}} \tag{17}$$

and standard errors can be computed through bootstrapping.

Figure 9 below computes the ratio measure used in Figure 7 but first estimates both correlations using the error correction procedure described above. The instruments used are the occupational status for the occupation reported second closest chronologically to the first marriage.[17] The resulting measure is very similar.

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

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, marriages could be negotiated to form an alliance with the characteristics of the wife an afterthought at best (e.g. Puga and Trefler, 2014).[18] In this hypothetical society, there would still be an observed correlation in the human capital of spouses if a woman's human capital is partially determined by her father.

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

[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.

[18] 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).

Figure 9: Estimated degree of marital assortment, error corrected

Note: 95% bootstrapped confidence intervals shaded (10,000 replications). Earnings scores are the imputed annual earnings for the individual's occupation in 1901 Canadian dollars (see text). Spearman's rank correlations are used (which is equivalent to the correlation of the ranks). To reduce attenuation bias, the correlations are adjusted by procedure proposed by Nybom and Stuhler (2017). This method, similar to instrumental variables regression, employs an additional measure of imputed earnings (using the second closest occupation to the individual's first marriage) for each individual..

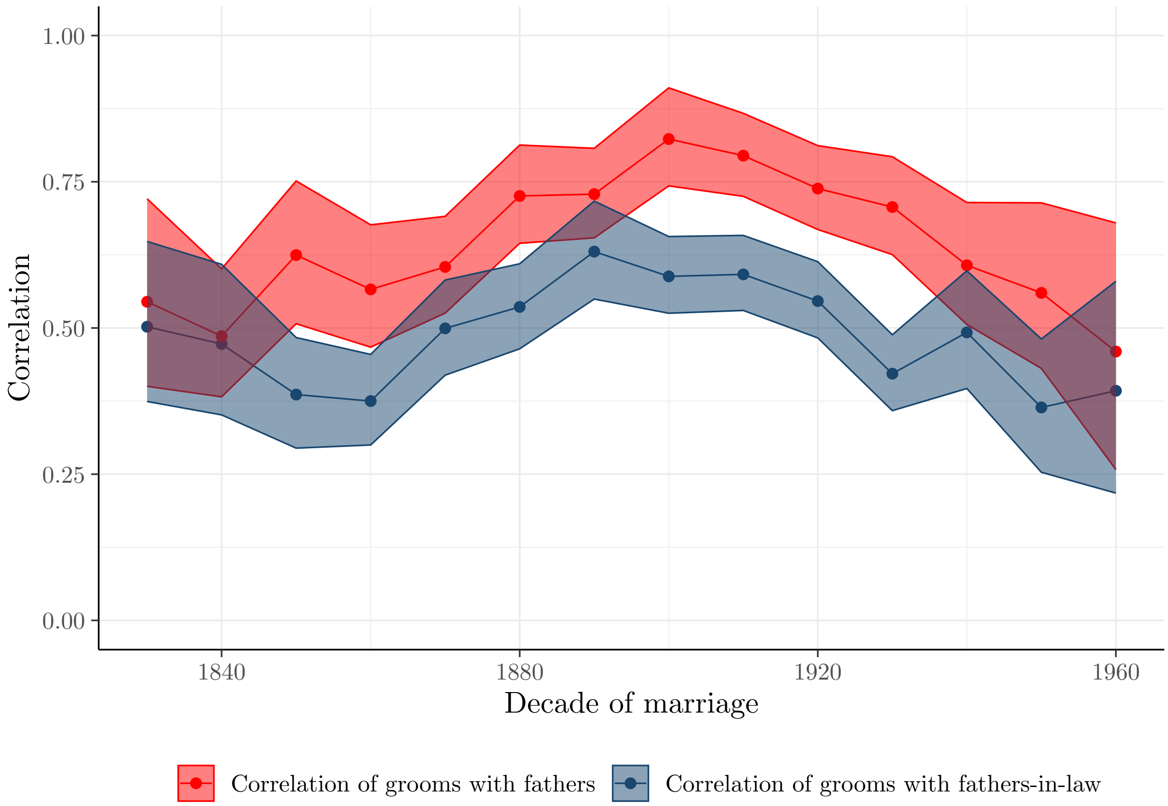

Figure 10: Correlations used to compute ratio, error corrected

Note: 95% bootstrapped confidence intervals shaded (10,000 replications). Earnings scores are the imputed annual earnings for the individual's occupation in 1901 Canadian dollars (see text). To reduce attenuation bias, the correlations are adjusted by procedure proposed by Nybom and Stuhler (2017). This method, similar to instrumental variables regression, employs an additional measure of imputed earnings (using the second closest occupation to the individual's first marriage) for each individual..

$$Y^s_{i,F} = \alpha Y_{i,F} + \phi_F + \beta \mathbf{X}_{i,F} + \epsilon_{i,F} \tag{18}$$

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, $\mathbf{X}_{i,F}$ can include fixed effects for both decade and the order of siblings.[19]

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

As shown in Panels A of Tables 3, 4, and 5, a woman who signed her marriage record married a man with more human capital than her sisters who did not. Being able to write was associated with an increase in the probability that woman's husband was literate by 34 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 were based on 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 Panels B of Tables 3 and 5, men who were able to write also married better. Being able to write was associated with an increase in the probability that a man's wife was literate by 32 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 regardless of gender.

What is the economic significance of 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

[19] Since I only have date of birth through 1849, I order siblings by the date of their first marriage. In the subsample with birth dates, the marriage order rank and the birth order rank have correlations of 0.85 for women and 0.86 for men.

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 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).

5.3 Do mothers matter directly for child outcomes?

For assortment to matter for social mobility, mothers must have a direct causal influence on child outcomes. A mother's literacy is associated with that of her children even after controlling for that of the father (Table 6). Notably, there appears to be no difference between the associations with children of different genders.[20] 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 6 do not identify the causal effect of the ability of mothers.

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. 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.

[20] Recall that this is an assumption of the ratio method used to estimate the degree of assortment.

Table 3: Family fixed effects, literate spouse

	Panel A: Groom's signature			
	(1)	(2)	(3)	(4)
Bride's signature	0.77***	0.34***	0.34***	0.31***
	(0.0008)	(0.002)	(0.002)	(0.002)
N	1,850,379	1,850,379	192,626	1,850,379
Family FE		X	X	X
Sample restriction			X	
Controls				X

	Panel B: Bride's signature			
	(1)	(2)	(3)	(4)
Groom's signature	0.70***	0.32***	0.32***	0.30***
	(0.001)	(0.002)	(0.002)	(0.002)
N	1,764,632	1,764,632	200,105	1,764,632
Family FE		X	X	X
Sample restriction			X	
Controls				X

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. Column 2 is my preferred specification. In Column 3, to illustrate the size of the identifying variation, the sample is restricted to just families where at least one sibling signed and one did not. Note that after adding family fixed effects the estimates are close to symmetrical across gender.

Table 4: Family fixed effects, spouse earnings score

	Panel A: Groom's earnings score			
	(1)	(2)	(3)	(4)
Bride's signature	0.21***	0.04***	0.04***	0.03***
	(0.0009)	(0.002)	(0.002)	(0.002)
N	997,453	997,453	85,184	997,453
Family FE		X	X	X
Sample restriction			X	
Controls				X

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. Earnings scores are the natural logarithm of the imputed annual earnings for the individual's occupation in 1901 Canadian dollars (see text). Column 2 is my preferred specification. In Column 3, to illustrate the size of the identifying variation, the sample is restricted to just families where at least one sibling signed and one did not. Note that after adding family fixed effects the estimates are close to symmetrical across gender.

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 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 (likely a death). Assuming this penalty is a constant, it can be controlled for by including fixed effects for the number of the marriage 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:

$$Y_{i,f} = \alpha Y_{i,f}^{m} + \phi_f + \delta_{mar_f} + \epsilon_{i,f} \quad (19)$$

Table 5: Family fixed effects, father-in-law's earnings score

	Panel A: Groom's father's earnings score			
	(1)	(2)	(3)	(4)
Bride's signature	0.12***	0.02***	0.02***	0.01***
	(0.0009)	(0.002)	(0.002)	(0.002)
N	879,845	879,845	67,009	879,845
Family FE		X	X	X
Sample restriction			X	
Controls				X

	Panel B: Bride's father's earnings score			
	(1)	(2)	(3)	(4)
Groom's signature	0.14***	0.03***	0.03***	0.02***
	(0.0008)	(0.002)	(0.002)	(0.002)
N	851,979	851,979	74,295	851,979
Family FE		X	X	X
Sample restriction			X	
Controls				X

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. Earnings scores are the natural logarithm of the imputed annual earnings for the individual's occupation in 1901 Canadian dollars (see text). Column 2 is my preferred specification. In Column 3, to illustrate the size of the identifying variation, the sample is restricted to just families where at least one sibling signed and one did not. Note that after adding family fixed effects the estimates are close to symmetrical across gender.

where $Y_{i,F}$ is an outcome of a child i with father f, $Y^m_{i,f}$ is a characteristic of the child's mother, ϕ_f are the crucial fixed effects that control for the father, δ_{mar_f} are fixed effects to control for the marriage number of the father, and $\epsilon_{i,j,f}$ is an error term.

As shown in Table 7 and 8, even controlling for the father, a mother who could sign her name had children who were 1–3 percentage points more likely to be able to sign their names. 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 effect 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. 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.

I also estimate 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.

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 the third columns of Tables 7, 8, and 9, the

Table 6: Association between parental human capital and child outcomes

	Daughter Signature (1)	Son Signature (2)	Son-in-law Earnings score (3)	Son Earnings score (4)
Signature of mother	0.30***	0.32***	0.06***	0.06***
	(0.001)	(0.001)	(0.001)	(0.001)
Signature of father	0.15***	0.18***	0.14***	0.15***
	(0.001)	(0.001)	(0.001)	(0.001)
N	1,551,082	1,435,435	853,451	795,217

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. Earnings scores are the log imputed annual earnings for the individual's occupation in 1901 Canadian dollars (see text).

identifying variation is quite small. 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:

$$Y_{i,f} = \alpha Y_{j,f} \times I(m_i = m_j) + \delta_{mar_f} + \epsilon_{i,j,f} \tag{20}$$

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

Table 7: Parental human capital and daughters' literacy

	Panel A: Daughter's signature			
	(1)	(2)	(3)	(4)
Mother's signature	0.41***	0.01**	0.01**	0.01***
	(0.001)	(0.005)	(0.005)	(0.005)
N	1,571,351	1,571,351	18,280	1,571,351
Father FE		X	X	X
Marriage number FE		X	X	X
Sample restriction			X	
Controls				X

	Panel B: Daughter's signature			
	(1)	(2)	(3)	(4)
Father's signature	0.34***	0.02**	0.02**	0.02***
	(0.001)	(0.007)	(0.007)	(0.007)
N	1,563,885	1,563,885	6,485	1,563,885
Mother FE		X	X	X
Marriage number FE		X	X	X
Sample restriction			X	
Controls				X

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. In Column 3, to illustrate the size of the identifying variation, the sample is restricted to just parents who had at least one spouse who signed and one who did not. Controls include marriage year and sibling marriage order fixed effects (as birth dates are not reported after 1849).

Table 8: Parental human capital and sons' literacy

	Panel A: Son's signature			
	(1)	(2)	(3)	(4)
Mother's signature	0.45***	0.03***	0.03***	0.03***
	(0.001)	(0.006)	(0.006)	(0.006)
N	1,454,550	1,454,550	16,127	1,454,550
Father FE		X	X	X
Marriage number FE		X	X	X
Sample restriction			X	
Controls				X

	Panel B: Son's signature			
	(1)	(2)	(3)	(4)
Father's signature	0.39***	0.02***	0.02***	0.02***
	(0.001)	(0.008)	(0.008)	(0.008)
N	1,447,558	1,447,558	5,513	1,447,558
Mother FE		X	X	X
Marriage number FE		X	X	X
Sample restriction			X	
Controls				X

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. In Column 3, to illustrate the size of the identifying variation, the sample is restricted to just parents who had at least one spouse who signed and one who did not. Controls include marriage year and sibling marriage order fixed effects (as birth dates are not reported after 1849).

Table 9: Parental human capital and sons' earnings

	Panel A: Son's earnings score			
	(1)	(2)	(3)	(4)
Mother's signature	0.17***	0.03***	0.03***	0.02**
	(0.001)	(0.009)	(0.009)	(0.01)
N	806,117	6,720	6,720	806,117
Father FE		X	X	X
Marriage number FE		X	X	X
Sample restriction			X	
Controls				X

	Panel B: Son's earnings score			
	(1)	(2)	(3)	(4)
Father's signature	0.19***	0.03	0.03	0.02
	(0.001)	(0.02)	(0.02)	(0.02)
N	802,127	802,127	2,131	802,127
Mother FE		X	X	X
Marriage number FE		X	X	X
Sample restriction			X	
Controls				X

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. Earnings scores are the natural logarithm of the imputed annual earnings for the individual's occupation in 1901 Canadian dollars (see text). In Column 3, to illustrate the size of the identifying variation, the sample is restricted to just parents who had at least one spouse who signed and one who did not. Controls include marriage year and sibling marriage order fixed effects (as birth dates are not reported after 1849).

mother, δ_{mar_f} are fixed effects to control for the marriage number of the father, and $\epsilon_{i,j,f}$ is an error term.

The results are shown in Table 10. Full siblings are more strongly associated than half-siblings. For example, a daughter signing her name was associated with a 60 percentage point increase in the probability her half-sister could sign her name. However, it was associated with a 72 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. Similar to before, 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. I show that the results are robust to restricting the estimates to a window around the remarriage, suggesting this is not a concern.

6 Discussion

The empirical findings of this paper demonstrate that marriage was strongly assortative in the past, with direct consequences for intergenerational mobility. As I argue above, these findings for Quebec are generalizable to other 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 the role of grandfathers.

6.1 Matching matters for father-son intergenerational correlations

If women directly matter for the outcomes of their children and marriages are assortative, the correlation between fathers and sons will be partially determined by the mother.[21] In

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

Table 10: The effect of parental human capital on half vrs. full siblings

	Panel A: Controlling for father			
	Daughter Signature (1)	Son Signature (2)	Son-in-law Earnings score (3)	Son Earnings score (4)
That of (half) sib.	0.60***	0.62***	0.26***	0.33***
	(0.005)	(0.005)	(0.009)	(0.01)
” × same mother	0.12***	0.08***	0.05***	0.05***
	(0.005)	(0.005)	(0.009)	(0.01)
N	2,050,712	1,838,040	721,719	649,203
Marriage number FE	X	X	X	X

	Panel B: Controlling for mother			
	Daughters Signature (1)	Sons Signature (2)	Sons-in-law Earnings score (3)	Sons Earnings score (4)
That of (half) sib.	0.65***	0.64***	0.29***	0.28***
	(0.008)	(0.009)	(0.02)	(0.02)
” × same father	0.06***	0.06***	0.01	0.10***
	(0.009)	(0.009)	(0.02)	(0.02)
N	1,974,765	1,770,143	695,604	625,937
Marriage number FE	X	X	X	X

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. Earnings scores are the natural logarithm of the imputed annual earnings for the individual's occupation in 1901 Canadian dollars (see text).

the simple model in Section 4, the association between fathers and sons is:

$$X_i^g = (\beta_m + \gamma\beta_f)X_i^f + e_i^g \tag{21}$$

$\beta_m + \gamma\beta_f$ 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 demonstrate this, Table 11 estimates the intergenerational elasticity of imputed earnings separately for more and less assorted parents. The less assorted parents are those where only one parents was literate and the more assorted parents those where both parents were either literate or illiterate. The elasticities for the less assorted parents are 0.31 for the sons and 0.26 for daughters (using their husbands' imputed earnings as a proxy). For the more assorted parents, the elasticities are 0.43 for sons and 0.37 for daughters. The more strongly assorted parents have higher estimated rates of intergenerational mobility. It is 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.

6.2 Matching matters for multigenerational mobility

Many recent studies consider correlations across more than two generations (Clark, 2014; Espín-Sánchez et al., 2022; Long and Ferrie, 2018; Olivetti et al., 2018; Solon, 2018). I am also able to estimate multigenerational mobility with the Quebec data, as shown in Table 12 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. However, when the partial elasticities are estimated controlling for the log imputed earnings of the other grandfather and of the father, there is a larger

Table 11: Father-son intergenerational elasticities, more and less assorted marriages

	Son's earnings score		Daughter's husband's earnings score	
	(1)	(2)	(3)	(4)
Father's earnings score	0.31***	0.43***	0.26***	0.37***
	(0.01)	(0.00)	(0.01)	(0.00)
N	77,944	395,627	83,019	421,290
Parents differ on signature	X		X	
Parents the same on siganture		X		X

Note: *p<0.10; **p<0.05; ***p<0.01. Signature variables are indicators that are one if a signature was recorded, zero if the absence of a signature was recorded, and omitted otherwise. Earnings scores are the natural logarithm of the imputed annual earnings for the individual's occupation in 1901 Canadian dollars (see text).

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 one would 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 what we observe in Table 12.

This exercise demonstrates how caution must be taken in interpreting intergenerational correlations without accounting for the role of women. It would at first 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 omitted variable bias and measurement error.

Table 12: Grandfather-grandson intergenerational elasticities

	Panel A: Son's earnings score		
	(1)	(2)	(3)
Maternal grandfather's earnings score	0.28***		0.12***
	(0.00)		(0.00)
Paternal grandfather's earnings score		0.28***	0.10***
		(0.00)	(0.01)
Father's earnings score			0.36***
			(0.00)
N	439,068	429,426	164,696
	Panel B: Daughter's husband's earnings score		
	(1)	(2)	(3)
Maternal grandfather's earnings score	0.25***		0.13***
	(0.00)		(0.00)
Paternal grandfather's earnings score		0.25***	0.11***
		(0.00)	(0.00)
Father's earnings score			0.29***
			(0.00)
N	489,868	479,496	181,443

Note: *p<0.10; **p<0.05; ***p<0.01. Family-clustered standard errors in parentheses. Earnings scores are the natural logarithm of the imputed annual earnings for the individual's occupation in 1901 Canadian dollars (see text).

7 Conclusion

In this paper, I construct a simple model of marriage and mobility. It shows that even absent female participation in the labor force, 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. I show that 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

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.

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

Abramitzky, Ran et al. (2019). "Automated Linking of Historical Data". NBER Working Paper w25825. Cambridge, MA: National Bureau of Economic Research.

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.

Altonji, Joseph G. and Thomas A. Dunn (1991). *Relationships Among the Family Incomes and Labor Market Outcomes of Relatives*. Working Paper.

Antonie, Luiza et al. (2022). "Intergenerational Mobility in a Mid-Atlantic Economy: Canada,1871–1901." *The Journal of Economic History*.

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.

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.

Butts, Kyle and John Gardner (2022). *{did2s}: Two-Stage Difference-in-Differences*.

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.

Chadwick, Laura and Gary Solon (2002). "Intergenerational Income Mobility Among Daughters." *American Economic Review* 92.1.

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.

Chiappori, Pierre-André, Monica Costa Dias, and Costas Meghir (2020). “Changes in Assortative Matching: Theory and Evidence for the US”. NBER Working Paper 26932.

Clark, Gregory (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.

— (2022). “Assortative Mating and the Industrial Revolution: England, 1754-2021”. CEPR Discussion Paper DP17074. London: Centre for Economic Policy Research.

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

— (2022). “The Mismeasure of Man: Why Intergenerational Occupational Mobility Is Much Lower than Conventionally Measured, England, 1800-2021”. CEPR Discussion Paper DP17346. London: Centre for Economic Policy Research.

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

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.

Dechêne, Louise (1974). *Habitants et Marchands de Montréal Au XVIIe Siècle.* Paris: les Editions Plons.

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

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.

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

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

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.

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.

Fagereng, Andreas, Luigi Guiso, and Luigi Pistaferri (2022). “Assortative Mating and Wealth Inequality”. CEPR Discussion Paper DP17148. London: Centre for Economic Policy Research.

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.

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

Gardner, John (2022). *Two-Stage Differences in Differences.*

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

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

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

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.

— (2014b). "Marry Your Like: Assortative Mating and Income Inequality." *American Economic Review: Papers and Proceedings* 104.5.

Greer, Allan (1985). *Peasant, Lord, and Merchant : Rural Society in Three Quebec Parishes, 1740-1840.* Toronto ; Buffalo : 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.

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.

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.

Liu, Haoming and Jingfeng Lu (2006). “Measuring the Degree of Assortative Mating.” *Economics Letters* 92.3.

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.

Luo, Xizi (2022). “Parental Dictates: Marriage Sorting and Social Mobility in Imperial China, 1614-1854.” Working Paper.

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.

McCord, Thomas (1867). *The Civil Code of Lower Canada*. Montreal: Dawson Brothers.

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

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

Modalsli, Jørgen and Kelly Vosters (2019). "Spillover Bias in Multigenerational Income Regressions." *Journal of Human Resources.*

Nybom, Martin and Jan Stuhler (2017). "Biases in Standard Measures of Intergenerational Income Dependence." *Journal of Human Resources* 52.3.

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.

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.

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

Piketty, Thomas and Emmanuel Saez (2014). "Inequality in the Long Run." *Science* 344.6186.

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

— (2020b). "BALSAC [Datatset]". 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.

Schwartz, Christine R. and Robert D. Mare (2005). "Trends in Educational Assortative Marriage from 1940 to 2003." *Demography* 42.4.

Shiue, Carol H and Wolfgang Keller (2022). "Marriage Matching over Five Centuries in China."

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

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

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.

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.

Vézina, Hélène and Jean-Sébastien Bournival (2020). "An Overview of the BALSAC Population Database. Past Developments, Current State and Future Prospects." *Historical Life Course Studies* 9.

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.

Vézina, Hélène et al. (2018). "The Linkage of Microcensus Data and Vital Records: An Assessment of Results on Quebec Historical Population Data (1852–1911)." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 51.4.

Ward, Zachary (2021). *Intergenerational Mobility in American History: Accounting for Race and Measurement Error*. Working Paper.

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 characterize inequality in a given generation, consider the variance of latent status (relaxing the normalization assumption):

$$\sigma^2_{X_i^c} = (\beta_f)^2\sigma^2_{X_i^f} + (\beta_m)^2\sigma^2_{X_i^m} + 2\beta_f\beta_m(\gamma\sigma_{X_i^f}\sigma_{X_i^m}) + \sigma^2_{e_i^c} \tag{22}$$

where c is the child of f and m.

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

$$\sigma^2_{X_i^c} = \sigma^2_{X_i^f} = \sigma^2_{X_i^m} \tag{23}$$

Then:

$$\sigma^2_{X_i^c} = \frac{\sigma^2_{e_i^c}}{1 - (\beta_f)^2 - (\beta_m)^2 - 2\gamma\beta_f\beta_m} \tag{24}$$

Assuming any measurement error is classical:

$$Y_i^c = X_i^c + u_i^c \tag{25}$$

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

$$\sigma^2_{Y_i^c} = \sigma^2_{X_i^c} + \sigma^2_{u_i^c} \tag{26}$$

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. If both β_f and β_m are greater than zero, the degree of assortment γ will increase steady state inequality as well (Figure A.1).

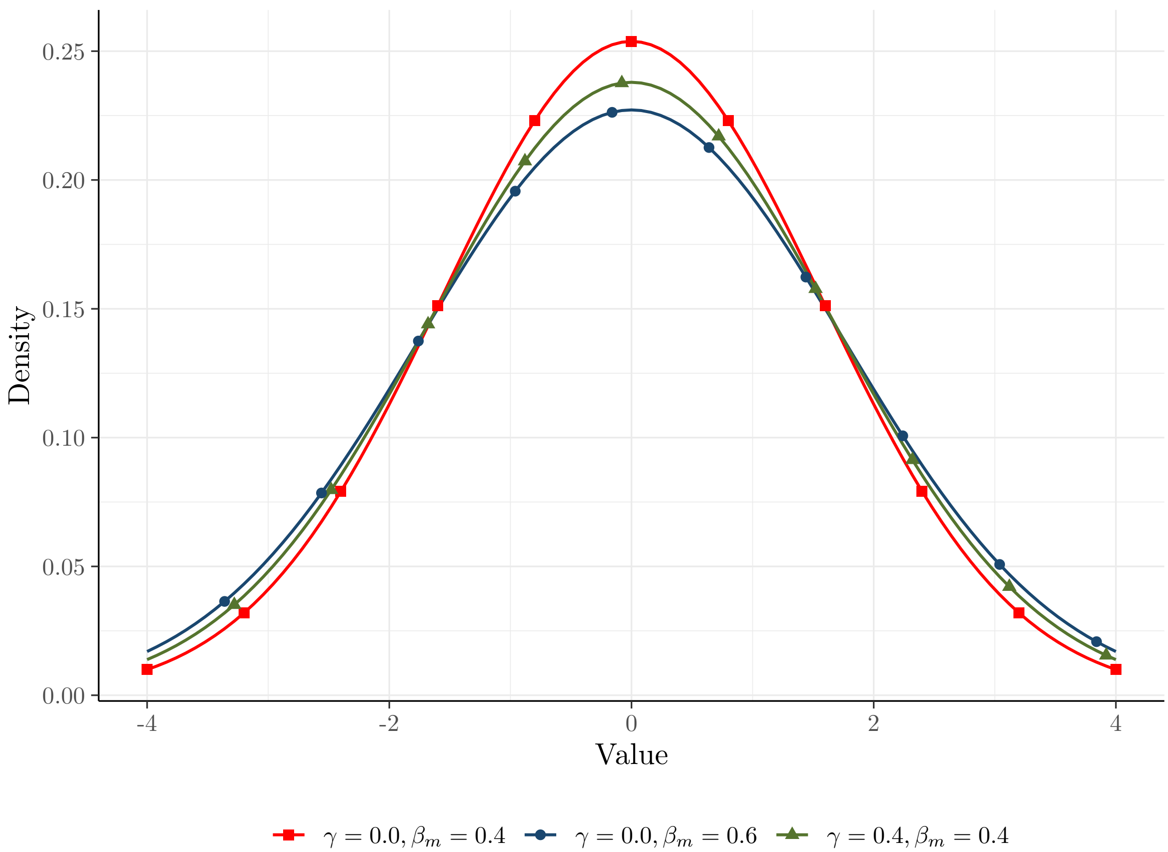

Figure A.1: Simulated steady state inequality

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 have different human capital than those married after? As shown in Table A.1, 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 A.1: Marriage matching and father-in-law's death

	Husband signed	
	OLS	2sDiD
	(1)	(2)
Father of bride dead	0.003	0.004
	(0.005)	(0.004)
N	73,218	53,455
Bride family FE	X	X
Bride sib order FE	X	X

Note: *p<0.10; **p<0.05; ***p<0.01. Standard errors in parentheses. Column 1 has family-clustered standard errors. Column 2 is estimated using two-stage difference-in-differences and has bootstrapped standard errors with 50,000 replications (Butts and Gardner, 2022; Gardner, 2022). Signature variables are indicators that are one if a signature was recorded, zero if the absence of a signature was recorded, and omitted otherwise. As deaths are only observed before 1849, only the sons-in-law of men who died before 1849 are included.

A3: Robustness of estimates of sorting

Figure A.2 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.

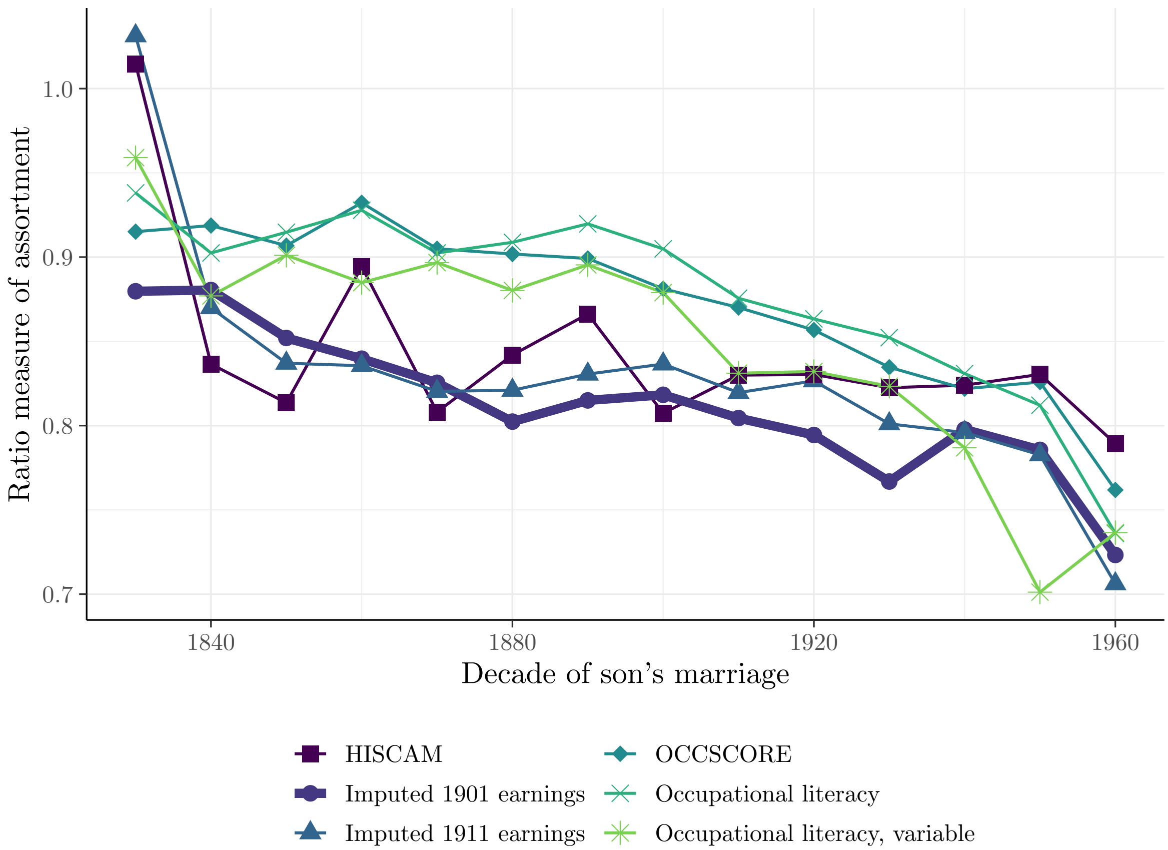

Figure A.2: Alternative occupational status scores

Note: 1901 imputed earings are imputed annual earnings for the individual's occupation in 1901 Canadian dollars (see text). 1901 imputed earnings with signature also uses literacy (proxied by signatures) to impute the earnings. OCCSCORE is the IPUMS imputed earnings score, 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 average signature rate by occupational category that accounts for the varying rate of signatures over time. All the measures roughly agree.

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

One way to estimate a lower bound robust to the second source of measurement 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. Figure A.3 shows the ratio measure of occupational status using this randomization method.

Figure A.5 estimates just the correlation between fathers and fathers-in-law using the same error correction model mentioned as before. 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 correlation. Regardless, the trend follows a very similar pattern over time to my preferred ratio method.

A4 Robustness of sorting on individual human capital

The estimates in Table 3 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

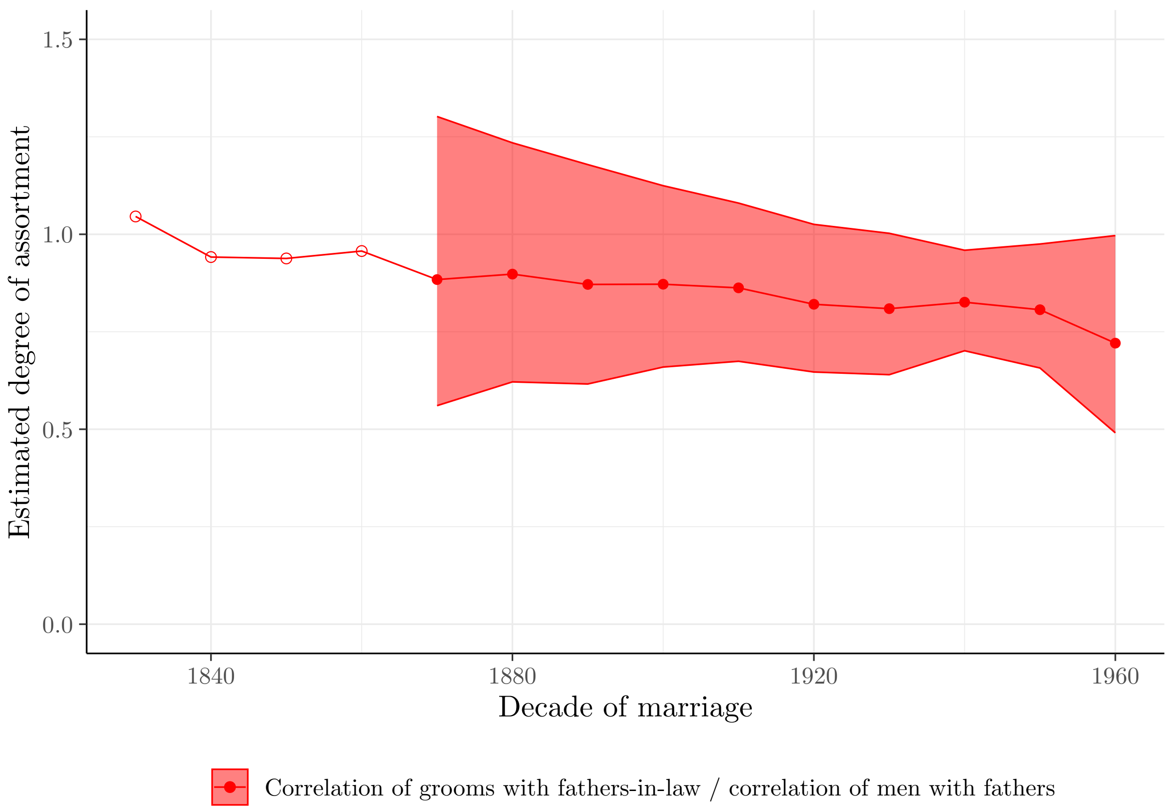

Figure A.3: Estimated degree of marital assortment, resampled earnings

Note: 95% bootstrapped confidence intervals shaded (10,000 replications). Confidence intervals exceeding 1.5 are not displayed in order to aid comparisons with Figure 7. Resampled earnings are draws from a log normal distribution fit on the earnings scores (Espín-Sánchez et al., 2019).The earnings scores are imputed annual earnings for the individual's occupation in 1901 Canadian dollars (see text). Spearman's rank correlations are used (which is equivalent to the correlation of the ranks). The overall magnitude and trend remains very similar.

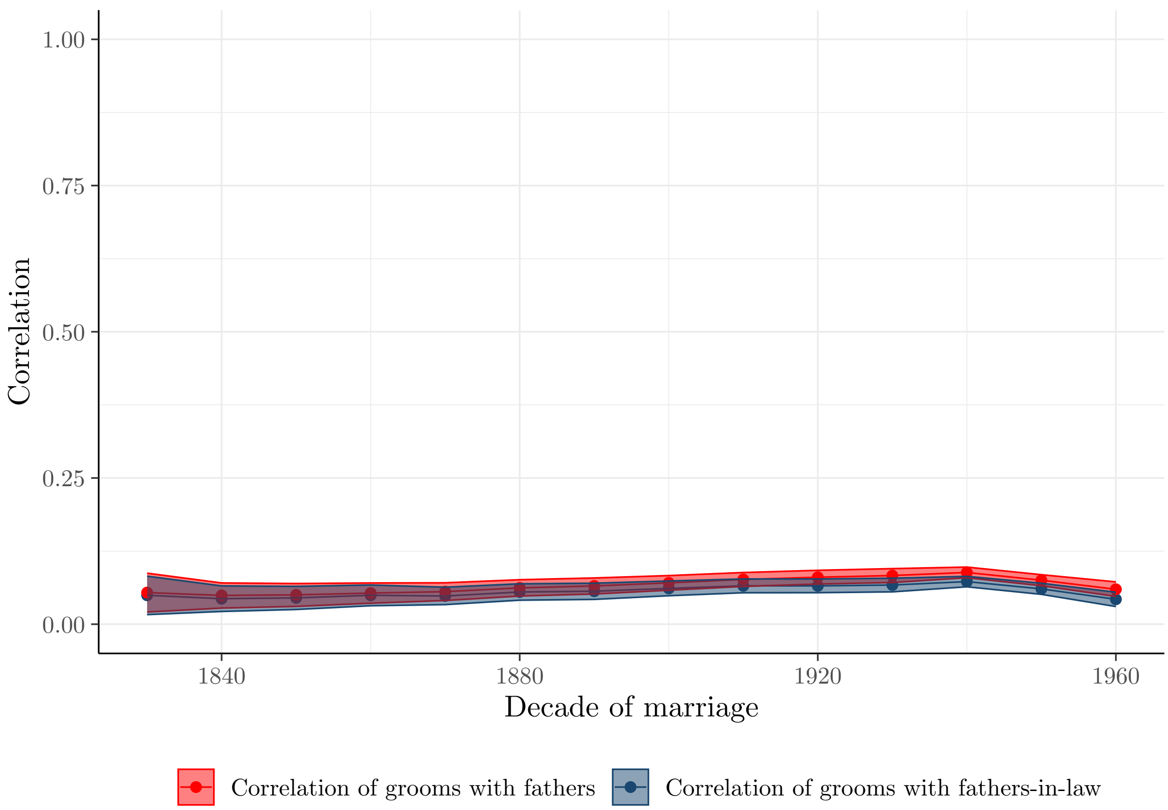

Figure A.4: Estimated degree of marital assortment, randomized occupational earnings

Note: 95% bootstrapped confidence intervals shaded (10,000 replications). Resampled earnings are draws from a log normal distribution fit on the earnings scores (Espín-Sánchez et al., 2019). The earnings scores are imputed annual earnings for the individual's occupation in 1901 Canadian dollars (see text). Spearman's rank correlations are used (which is equivalent to the correlation of the ranks). The correlations are much smaller, which is expected as the resampling method is a lower bound.

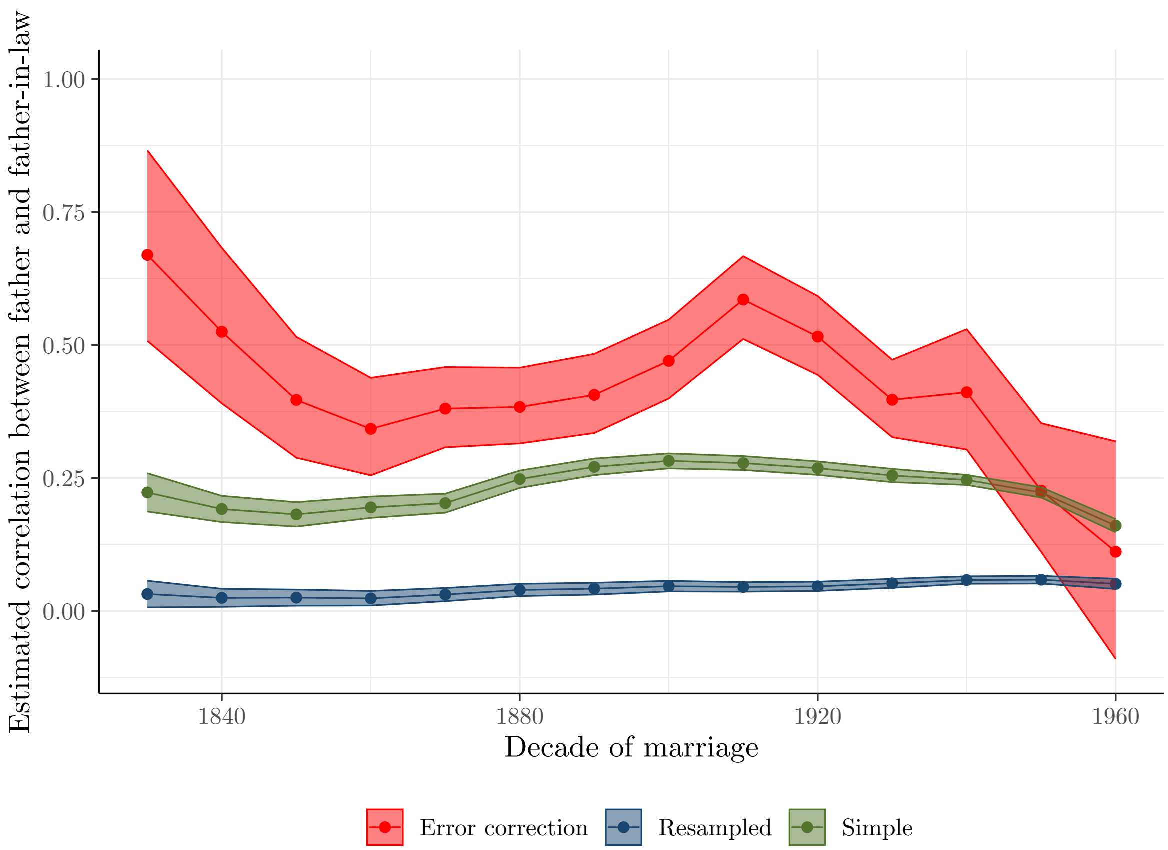

Figure A.5: Father-father-in-law correlations

Note: 95% bootstrapped confidence intervals shaded (10,000 replications). The earnings scores are the imputed annual earnings for the individual's occupation in 1901 Canadian dollars (see text). Resampled earnings are draws from a log normal distribution fit on the earnings scores (Espín-Sánchez et al., 2019). To reduce attenuation bias, the correlations are adjusted by procedure proposed by Nybom and Stuhler (2017). This method, similar to instrumental variables regression, employs an additional measure of imputed earnings (using the second closest occupation to the individual's first marriage) for each individual.. Spearman's rank correlations are used (which is equivalent to the correlation of the ranks).

estimated coefficients are an average treatment effect of individuals in these "treated" families being able to sign. It is possible that these families have unusual characteristics.

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.[22] It is also reweighting based only on observables; any unobservable characteristic that makes these families unique will not be accounted for. 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 mother's signature, decade of first marriage, borough of first marriage, denomination, and the number of married children. Missing values are included as an additional category for each indicator variable. As shown in Table A.2, there is still a positive and significant marriage premium for literacy.

A5 Robustness of estimates of effects of parental human capital

The identifying assumption for the analysis in Tables 7, 8, and 9 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. Tables A.3, A.4, and A.5 replicate the analysis except they drop children of parents who remarry if they were more than two positions away in the order of sibling marriages from a half-sibling.[23] 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.

[22]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.

[23]As before, since I only have date of birth through 1849, I order siblings by the date of their first marriage. In the subsample with birth dates, the marriage order rank and the birth order rank have correlations of

Table A.2: Marriage selection, reweighting for selection into identification

	Panel A		
	Groom Signature (1)	Groom Earnings score (2)	Groom's father Earnings score (3)
Bride's signature	0.41***	0.04**	0.02*
	(0.009)	(0.02)	(0.009)
N	1,850,379	997,453	879,845
Family FE	X	X	X

	Panel B		
	Bride Signature (1)		Bride's father Earnings score (3)
Groom's signature	0.37***		0.05***
	(0.009)		(0.01)
N	1,843,748		893,398
Family FE	X		X

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. Earnings scores are the imputed annual earnings for the individual's occupation in 1901 Canadian dollars (see text). Reweighted 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, and the number of married children of the same gender in each family. Missing values are included as an additional category for each indicator variable in the logistic regression.

Table A.3: Parental human capital and daughters' literacy, window

	Panel A: Daughter's signature			
	(1)	(2)	(3)	(4)
Mother's signature	0.38***	0.02*	0.02*	0.02**
	(0.001)	(0.009)	(0.009)	(0.01)
N	1,061,386	1,061,386	4,153	1,061,386
Father FE		X	X	X
Marriage number FE		X	X	X
Sample restriction			X	
Controls				X

	Panel B: Daughter's signature			
	(1)	(2)	(3)	(4)
Father's signature	0.33***	0.02*	0.02*	0.02**
	(0.001)	(0.01)	(0.01)	(0.01)
N	1,070,541	1,070,541	1,702	1,070,541
Mother FE		X	X	X
Marriage number FE		X	X	X
Sample restriction			X	
Controls				X

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. Column 2 is my preferred specification. In Column 3, to illustrate the size of the identifying variation, the sample is restricted to just parents who had at least one spouse who signed and one who did not. Children of parents who remarry are excluded if they are more than two away in the order of sibling marriages from a half-sibling. Controls include marriage year and sibling marriage order fixed effects (as birth dates are not reported after 1849).

Table A.4: Parental human capital and sons' literacy, window

	Panel A: Son's signature			
	(1)	(2)	(3)	(4)
Mother's signature	0.45***	0.03***	0.03***	0.03***
	(0.001)	(0.006)	(0.006)	(0.006)
N	1,454,550	1,454,550	16,127	1,454,550
Father FE		X	X	X
Marriage number FE		X	X	X
Sample restriction			X	
Controls				X

	Panel B: Son's signature			
	(1)	(2)	(3)	(4)
Father's signature	0.39***	0.02***	0.02***	0.02***
	(0.001)	(0.008)	(0.008)	(0.008)
N	1,447,558	1,447,558	5,513	1,447,558
Mother FE		X	X	X
Marriage number FE		X	X	X
Sample restriction			X	
Controls				X

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. Column 2 is my preferred specification. In Column 3, to illustrate the size of the identifying variation, the sample is restricted to just parents who had at least one spouse who signed and one who did not. Children of parents who remarry are excluded if they are more than two away in the order of sibling marriages from a half-sibling. Controls include marriage year and sibling marriage order fixed effects (as birth dates are not reported after 1849).

Table A.5: Parental human capital and sons' earnings, window

	Panel A: Son's earnings score			
	(1)	(2)	(3)	(4)
Mother's signature	0.18***	0.03*	0.03*	0.03
	(0.001)	(0.02)	(0.02)	(0.03)
N	552,941	552,941	1,325	552,941
Father FE		X	X	X
Marriage number FE		X	X	X
Sample restriction			X	
Controls				X

	Panel B: Son's earnings score			
	(1)	(2)	(3)	(4)
Son's signature	0.20***	0.02	0.02	0.02
	(0.001)	(0.03)	(0.03)	(0.03)
N	557,423	557,423	539	557,423
Mother FE		X	X	X
Marriage number FE		X	X	X
Sample restriction			X	
Controls				X

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. Earnings scores are the natural logarithm of the imputed annual earnings for the individual's occupation in 1901 Canadian dollars (see text). Column 2 is my preferred specification. In Column 3, to illustrate the size of the identifying variation, the sample is restricted to just parents who had at least one spouse who signed and one who did not. Children of parents who remarry are excluded if they are more than two away in the order of sibling marriages from a half-sibling. Controls include marriage year and sibling marriage order fixed effects (as birth dates are not reported after 1849).

Table A.6 replicates Table 10. Again, the results are very similar.

Table A.6: Effects of parents, half-siblings within window

	Panel A: Controlling for father			
	Daughter Signature (1)	Son Signature (2)	Son-in-law Earnings score (3)	Son Earnings score (4)
That of (half) sib.	0.65***	0.63***	0.28***	0.34***
	(0.01)	(0.01)	(0.03)	(0.03)
" × same mother	0.08***	0.08***	0.05*	0.06**
	(0.01)	(0.01)	(0.03)	(0.03)
N	424,506	397,151	147,932	139,889
Marriage number FE	X	X	X	X
	Panel B: Controlling for mother			
	Daughters Signature (1)	Sons Signature (2)	Sons-in-law Earnings score (3)	Sons Earnings score (4)
That of (half) sib.	0.63***	0.62***	0.32***	0.32***
	(0.02)	(0.02)	(0.03)	(0.04)
" × same father	0.10***	0.09***	0.004	0.08**
	(0.02)	(0.02)	(0.03)	(0.04)
N	439,980	411,454	153,525	145,188
Marriage number FE	X	X	X	X

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. Earnings scores are the imputed annual earnings for the individual's occupation in 1901 Canadian dollars (see text). Children of parents who remarry are excluded if they are more than two away in the order of sibling marriages from a half-sibling.

A6 Robustness of effect of sorting on intergenerational elasticity

One concern with Table 11 is that families where only one parent was literate were selected on some omitted factor that decreases intergenerational mobility. One way to overcome

0.85 for women and 0.86 for men.

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., 2011; Caron et al., 2017; Dillon, 2010). As I do not observe ages in most of the sample, I instead consider the sex composition of the mother's siblings who got married before her.

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.[24] However, if it merely changes the status of the younger sister, who then match accordingly, it shouldn't introduce bias.

As shown in Table A.7, the sex composition decreases the association between the signature rates of spouses and decreases the intergenerational elasticity between fathers and sons. This is exactly what we would expect if the mother directly mattered for the outcomes of children.

[24]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 A.7: Father-son intergenerational elasticities, more and less assorted marriages

	Mother signed	Son's earnings score
	(1)	(2)
Father signed	0.64***	
	(0.00)	
Share female	0.01	0.12***
	(0.01)	(0.04)
Father signed × share female	-0.01**	
	(0.01)	
Father's earnings score		0.43***
		(0.00)
Father's earnings score × share female		-0.02***
		(0.01)
N	519,887	519,887

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 natural logarithm of the imputed annual earnings for the individual's occupation in 1901 Canadian dollars (see text). The share female is the fraction of the mother's siblings who married before her that were female.