Northwestern | INSTITUTE FOR POLICY RESEARCH

RESEARCH EXCELLENCE • POLICY IMPACT

IPR Working Paper Series

WP-23-42

Women and the Econometrics of Family Trees

José-Antonio Espín-Sánchez
Yale University

Joseph Ferrie
Northwestern University and IPR

Christopher Vickers
Auburn University

Version: December 18, 2023

DRAFT

Please do not quote or distribute without permission.

Abstract

The researchers present an econometric structure for the analysis of intergenerational mobility that integrates non-linearities, the role of maternal-side effects and the impact of grandparents. They show how previously estimated models are special cases of this general framework and what specific assumptions each embeds. Their analysis of linked U. S. data 1900–40 reveals the extent to which inadequate consideration of assortative mating and the impact of mothers produces misleading conclusions.

The authors thank Jaewon Lee for his help on this paper.

1 Introduction

The economic analysis of intergenerational socioeconomic mobility has advanced dramatically since the first formal model of Becker and Tomes (1979, 1986). Significant effort has been devoted to better estimation of permanent income (Mazumder, 2005), more careful consideration of the stages of the life cycle to compare across generations (Mazumder and Acosta, 2015), the discovery of new ways to exploit historical and contemporary sources to generate links across generations (Abramitzky et al., 2021a) and–more recently–the exploration of potential non-linearities in the relationship between outcomes in one generation and those in the next (Durlauf et al., 2022). The roles of both women and generations prior to the [male] parent in the transmission of advantages to children have received less attention than one would expect in the study of processes of inheritance that are likely inherently both biparental and multigenerational (exceptions are Chadwick and Solon, 2002; Olivetti and Paserman, 2015; Jácome et al., 2021). Whereas they look at different mobility by gender (sons vs daughters), we look at the individual contributions of mothers to their children, and differentially for sons and daughters.

We provide a unified treatment of the econometrics of the transmission of socioeconomic status across generations that integrates the latter three previously neglected dimensions of the empirical analysis of mobility: (1) non-linearities in the links across generations; (2) the role of women; and (3) the role of prior generations. Unlike the Becker and Tomes framework, ours is not a structural model of maximizing behavior subject to constraints. Rather, we present a fully-specified econometric framework–specifically, a three-equation non-linear instrumental variable (NLIV) approach–that both (1) reveals the variety of assumptions embedded in previous mobility studies and (2) demonstrates how different combinations of assumptions permit the identification of a range of structural parameters of interest. Moreover, we show a range of results depending on the data available to the econometrician. With richer data, less restrictive assumptions obtain point-identification. We then apply this framework to data for the U.S. 1900-1940 linked across three generations.[1]

Our results reveal the importance of including women and particularly mothers, in any analysis of intergenerational mobility that occurs in a setting with a degree of assortative mating that is greater than zero and less than one. The absence of complete information on the measured incomes of women who spend more time than men in household production has hampered the inclusion of women in many previous analyses. But we show both how information on the prior generation can overcome this challenge and how much valuable

[1]We do not restrict attention by race; however, Blacks are underrepresented in our sample. It is not possible with our sample size to do a separate analysis for most of the empirical designs, although we include a select number of specifications for Blacks in an Appendix C.

information is lost when women and grandparents are excluded. We also demonstrate how to account for both gendered effects of parental income–allowing the effect of mothers to differ from that of fathers and the effect of grandmothers to differ from that of grandfathers– and differences in the effect of parental income at different generations–allowing the effect of grandparents on parents to differ from those of parents on children.

The conventional measure of father-to-son mobility–the intergenerational elasticity (IGE) of income for fathers in 1900 and sons in 1920–that we find is 0.484, roughly in line with a large number of prior studies for the U.S. Every additional \$1.00 of income for the father is associated with an additional \$0.484 of income for the son. The father's effect falls to 0.167 when we use our three-equation NLIV to account fully for the mother's impact (0.763). Moreover, the mother's contribution to mobility is almost five times larger than the father's. Note that the total parental effect (father plus mother) is closer to 1 than it is to our naive estimate of 0.485, revealing considerably greater immobility once the maternal line is included. We also find a high degree of assortative mating where the correlation between spouses is 0.416. The conventional measure (0.485) is then the sum of a direct effect of the father on the son (0.167) and and indirect effect that measures the effect of the mother (0.763) and the correlation between the parents (0.416), i.e., $0.484 = 0.167 + 0.763 \cdot 0.416$.

We also show substantively and statistically significant results for both assortative mating and differences in transmission across generational pairs (i.e. grandparent-to-parent effects that differ from parent-to-child effects) and additionally that transmission to children can differ based on the gender of the recipient.

Overall, our work demonstrates the high return to taking seriously the econometrics of the relationships underlying the transmission of advantage across generations, the pitfalls of ignoring those complications and the ease with which a proper accounting for them can be adapted to different settings as dictated by data availability, social norms and institutional constraints.

2 Data

In this section we describe the construction of the data, in particular the linking and construction of occupational income scores.

2.1 Linking

The econometric methodology we have described above is sufficiently general to capture the full range of relationships among family members both within a generation (assortative matching) and across generations (direct parent-to-child transmission of financial or genetic or social resources). But that same methodology can also be applied to subsets of the nodes of three-generation family trees. To show how the results the methodology produces can vary by which nodes are used (and implicitly, by which assumptions are made regarding links within and across generations), we need data for the full tree–child, both parents, both sets of grandparents–a total of seven individuals. In some cases we will use maternal uncles as well.

The only datasets sufficiently large to yield a satisfactory number of linkages across three generations are the 100% transcriptions of the U.S. Census of Population. As generations are normally separated by 25 years, but the census is available every tenth year, we have two options in choosing the censuses to link: (1) 1900, 1920 and 1940 with twenty years between adjacent censuses; or (2) 1880, 1910 and 1940 with thirty years between adjacent censuses. An alternative approach would be to use any three-generation family lines that can be found in any of the years 1880, 1900, 1910, 1920, 1930, or 1940.[2] For the present exercise, we have chosen option (1).

We begin with an individual adult in the 1940 Census. We then link that individual to their parents by finding the individual as a child in the 1920 Census. The parents who are adults co-resident with the child in 1920 are then linked back to their own parents in the 1900 Census. Automated linkage across large cross-sectional datasets at different points in time to create intragenerational and multigenerational datasets in this way is now routine. A variety of algorithms have been developed and evaluated since the first work along these lines in the 1930s (Ferrie, 1996; Ruggles et al., 2018; Bailey et al., 2020; Abramitzky et al., 2021b)

All methods result in both some missed links ("efficiency loss," Type I Errors, or false negatives when a link that should have been made was not) and some wrong links ("precision loss," Type II Errors, or false positives when a link that should not have been made was). The approach we have used is essentially that described in Ferrie (1996) with the addition of string comparators to allow for misspellings and mistranscriptions not otherwise captured in name standardization schemes. This modified algorithm corresponds to the ABE-JW method described and evaluated by Abramitzky et al. (2021b) Among the procedures they evaluate that do not use machine learning, this one provides the best trade-off between efficiency loss and precision loss. Both Bailey et al. (2020, p. 1033-4) and Abramitzky et al. (2021b, p. 895-9) show that the ABE-JW procedure matches inferences on intergenerational mobility produced by the most rigorous approaches (manual linkage by trained human researchers

[2]The manuscript schedules of the 1890 U.S. Census have not survived.

and the recently released IPUMS Linked Census Files).[3]

There is, however, one novelty in our linkage process: we have used records from the Social Security Administration's NUMIDENT (Numerical Identification) file to link adult females back to the households were they were children co-resident with their own parents (NARA 2018).[4] This file contains exact place of birth (city or town), exact date of birth (month, day, year) and the full names of both parents, including the mother's pre-marriage surname.[5] In our application it is this last item that is most important: for an adult female in 1940, for example, we can ascertain their father's full name (both given name and surname) from the NUMIDENT file, so even if that 1940 adult female had changed her surname at marriage we would know the family (surnames and given names for both parents, as well as presence of the individual herself) in which to search for her in 1920. Knowledge of the given names of both parents allows us to make this match with high accuracy.[6]

To take the male adult 1920 parent back to his own 1900 household (where his parents–the grandparents of the 1940 adults with which we began the process–can be observed as 1900 adults) is straightforward, as we know the surname for which to search in 1900, his given name, age in 1900 and state of birth.

But for the female adult 1920 parent (whose surname in 1920 would be that of her spouse and not that of her adult 1900 parents) we can use the pre-marriage mother's surname reported in the NUMIDENT for the 1940 adult female to identify the 1900 household in which the 1920 adult female lived as a child, at which point we can observe her parents (the maternal grandparents of the 1940 adult female who was our starting point.[7]

With the 1900 childhood household of the 1920 adult males and females identified, their 1900 male siblings could then be linked forward to 1920. This provides the paternal and maternal uncles of the original 1940 adult female.

---

[3]Given the need for many links across three generations for our purposes, the IPUMS linkages were inappropriate–the overlap between their 1900-20 and 1920-40 linkages was too small to be useful. The time and labor cost of generating tens of thousands of multigenerational links by hand were prohibitive.

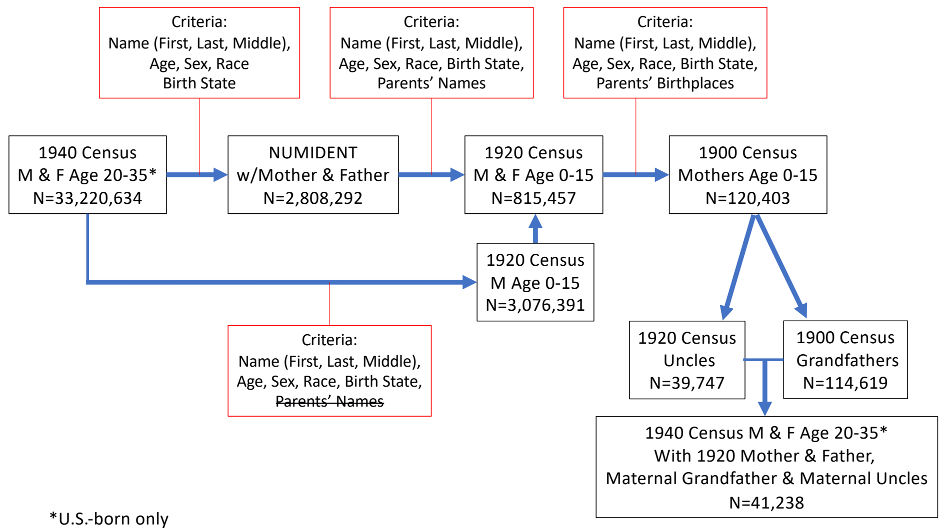
[4]After we began work on this project in 2019, we learned that Althoff et al. (2022) have also used the NUMIDENT to facilitate maternal line linkage. Our linkage procedure was developed independently from theirs.

[5]The version of the file in the possession of the National Archives contains only individuals with a verified death between 1936 and 2007 or who would have been over 110 years old by December 31, 2007. This file records the information provided by individuals as they entered the Social Security system. Though this information is now collected at birth, in the period from the late 1930s through the 1960s it was collected instead when an individual first entered employment in a "covered" industry–until the late 1950s this was primarily any work other than that in agriculture, domestic service, or self-employment. By the end of their work lives, virtually all U S. workers have spent at least some time in covered employment and been thereby at risk to enter the NUMIDENT file.

[6]The given names of both parents also enhances the accuracy of links for males.

[7]Note that, for both males and females observed as 1920 adults, the link back to them as 1900 children will be lower in accuracy than linkage of both males and females observed as 1940 adults back to them as 1920 children because the former links will not rely on parental given names.

Figure 1: Matching



Notes: The figure demonstrates the process of constructing the matches as described in Section 2.1.

Figure 1 presents a flow chart that schematically describes the linkage process and reports the number of observations included at each stage. In Appendix B we discuss the representativeness of the data compared to the general population. Note that because of the stringent data requirements, the sample size is small relative to the population. The link to the NUMIDENT file is successful for under 10% of cases. Conventional links, from one Census to another, have more standard matching rates. For example, the 1920 link backwards to 1900 has a link rate of 15%. Further attrition occurs because all four male members of the tree must have a valid occupation, which is particularly restrictive for the sons.

2.2 Measuring Socioeconomic Status

The IPUMS data that we used in generating our linked samples is based on the full-count files from the U.S. Census of Population from 1900 through 1940. The only measure of socioeconomic standing that is consistently and contemporaneously reported in all of these five census and is therefore available for us to assess intergenerational mobility in socioeconomic status is self-reported occupational title (e.g. "carpenter," "farmer," "surveyor," "laborer").

These titles can be grouped into broad categories that correspond roughly to different levels of socioeconomic status (where "status" is an amalgam of income, education, skill, autonomy and prestige). "Mobility" across generations can then be assessed by measuring the off-diagonal cells in a 2-dimensional matrix of the $M[a]$ collapsed categories in Generation a and the $M[b]$ collapsed categories in Generation b. This approach–often employed in sociology–is used in Long and Ferrie (2013), with four broad categories (white collar, skilled/semiskilled, unskilled, farmer).

Another approach is to assign a value to each of the hundreds of occupational titles that appear in the census manuscripts and analyze differences in these values across generations. IPUMS itself provides such a measure in its OCCSCORE variable. Each occupational title is assigned the median annual income (in hundreds of dollars) in that occupation reported in the 1950 Census of Population.[8] Saavedra and Twinam (2020) document 25 recent studies that adopt this approach and treat the resulting measure of socioeconomic status as cardinal.

The categorical approach suppresses a great deal of potentially useful information (e.g. differences in occupational titles across generations that fall within the same broad category–such as clerk and civil engineer which are both "white collar"–will show up as immobility). The cardinal approach assumes that the distances between scores for any pair of occupations is the same in all years before 1950 as it was in 1950 (e.g. the 1900 ratio of the score for "clerk" to the the score for "carpenter" is the same as that ratio was in 1950).

To remedy the shortcomings of both the categorical and cardinal approaches to the measurement of mobility across generations in socioeconomic status, we have adopted a hybrid approach. We adjust occupational scores in two ways: allowing for heterogeneity by state and computing farmers' income by state and decade.

To allow for heterogeneity across states, we calculate state-specific occscores. These are based on the wage income measure in 1940, but adjusted to be based on total income, a variable not present in 1940. To construct this variable, we first take the ratio of total income to wage income in 1950 among males 18-65 reporting a non-zero total income and wage income and take the mean of this number by occupation. We then inflate the wage income of individuals in 1940 by this ratio with the same sample restriction and then adjust this number so that it is comparable to the original occscores by making the median by occupation the same.[9] We then calculate the median of this number by state.

[8]See Sobek (1995). The choice of 1950 was dictated by data availability: though the 1940 census was the first to report income and therefore the first opportunity to associate median incomes with occupational titles, the income measure was limited to wages and salaries, so the self-employed (proprietors and farmers) were excluded. The 1950 data included both income derived from wages and salaries and income derived from self-employment.

[9]For a small number of occupations, there is no match between the occupations in 1940 and 1950, in which case we use the ratio of total income to wage income for all occupations in 1950 as the ratio.

Farmers present an altogether different challenge. Their annual earnings will fluctuate based on prices, yields and product mix. This occupation was large throughout the period 1900-1950 (though a falling share of total employment): the number of farms was 5.4 million in 1950 and 5.7 million in 1900. U.S. Census Bureau (1960, Series K4). It presents three problems in measuring intergenerational mobility: (1) earnings vary enormously by location (due to differences in climate, soil and crop mix); (2) the socioeconomic status of farmer relative to other occupations changed considerably from 1950 when OCCSCORE is anchored back to 1900; and (3) the distribution of farm earnings is highly unequal in 1950. Based on their 1950 median annual income of $1,430 the IPUMS OCCSCORE for farmers is 14. Based on their 1950 median annual income of $2,076 the IPUMS OCCSCORE for unskilled laborers ("Laborers (n.e.c.)") is 21. This implies a socioeconomic status for unskilled laborers that is 1.5 times greater than that for farmers.

In our main sample, containing both grandparents, 59 percent of grandfathers are farmers in 1900. In Figure 8, we show a violin plot displaying the density of the percentile of farmers in the distribution of adjusted occupational scores for grandfathers. The unadjusted OCCSCORE for farmers, would place the income of all farmers in 1900 at the 4^{th} percentile, i.e., farmers income represents a mass point of over half of the distribution from the 4^{th} to the 73^{th} percentile. Our adjusted OCCSCORES based on state farm income result in a much wider spread of farmers within the overall distribution. Whereas in 1950 farmers are at the bottom of the income distribution, in 1900 the average farmer income in certain states would place farmers at the top quintile of the distribution.

We have used Strauss and Bean (1940, Table 7) for farmer's mean annual earnings 1900-1909 and U.S. Census Bureau (1960, Series K128) for 1910-1950. Both series were converted to a per-farm basis by dividing them by the number of farm operators (interpolated between 1900 and 1909, reported directly from 1910 forward). The resulting 1950 mean value per farm is $3,781–2.7 times the median earnings of farmers, indicting the highly unequal distribution of earnings in this sector, with many farmers with very low earnings and a small number with very high earnings.

As it is not possible to compute the median value back before 1950, we can only compare mean earnings at earlier dates. The ratio of mean 1950 earnings of farmers ($3,781) to mean 1950 earnings of laborers ($2,558) is 1.48. In 1900, this ratio is 1.34 ($674/$503). For a log-normal distribution such as earnings, the ratio of the mean to the median is constant regardless of the mean for a constant variance. As a result, we use the ratio of means between farmers and unskilled workers to adjust the ratio of their medians from 1950 back to 1900.

In Table 1, we show the parameters from a regression of log son income on log father income for pairs with the same occupation, different occupations and for the overall sample, with various adjustments to the OCCSCORE.[10] For unadjusted occscores, fathers and sons with the same occupation have the same OCCSCORE and thus the coefficient is 1. This is particularly problematic for farmers, i.e., 64 percent of farmers in 1900 have son that was a farmer in 1920. Without our new estimates for farmers income, the correlations for those pairs would mechanically be equal to 1. In the various columns, we show the results from unadjusted occscores, adjusting for state differences, adjusting for changes in occupational categories, adjusting for both state and occupational differences and finally allowing also adjustments for farm income. Note that while the final numbers incorporating all of the adjustments give a similar number to the unadjusted OCCSCORE, the numbers for different adjustments can be quite different. For example, we find higher levels of persistence when adjusting for the income at the state level.

Table 1: Elasticities by Income Adjustment

Adjustment	Unadj.	State	Farm	State and Farm
Same Occupations				
Elasticity	1.000	0.978	1.099	1.096
	(.)	(0.004)	(0.009)	(0.009)
N	3,481	3,481	3,481	3,481
Different Occupations				
Elasticity	0.375	0.469	0.221	0.296
	(0.009)	(0.009)	(0.007)	(0.008)
N	20,255	20,255	20,255	20,255
Full Sample				
Elasticity	0.443	0.540	0.366	0.427
	(0.008)	(0.008)	(0.008)	(0.008)
N	23,736	23,736	23,736	23,736

Notes: Coefficients from regressions of log son income on log father income for the sample of matched sons, fathers and grandparents. The columns represent different methods of adjusting income as described in Section 2.2. "State" to the state-specific adjustments and "farm" incorporating farm price adjustments.

[10] Note these variables differ from the standardized ones used in the structural estimation below.

3 Setup

The idea here is simple. Studies of social mobility most often use income data on male parents and their children to assess social mobility as in equation (1). Equation (1) measures the correlation between the (standardized) income of the son (S_i) and the (standardized) income of the father (F_i).[11]

$$S_i = \tilde{\beta}_F F_i + \nu_i \tag{1}$$

The estimation of the parameter of interest $\tilde{\beta}_F$ could be biased if the equation above is misspecified. In particular, one can think of other family members that can have an effect on the characteristics of the child. The first that comes to mind is the mother. Whether we think they directly transmit socioeconomic status, or that they transmit only traits that affect socioeconomic status, mothers will play an essential role. In particular, they are as important as fathers if we think the transmission is due to genetics and usually also as important if we think the transmission is mostly inherited wealth.

Alternatively, we might think that what is transmitted from generation to generation is human capital, or networks, which could be transmitted from father to son, without mothers playing a major role. Even in these instances, grandfathers might play a role in determining the socioeconomic status of the son, even after accounting for the effect of the father (see Long and Ferrie, 2013). In any case, there are reasons to believe that the above equation might be misspecified. Our goal here is to provide a framework where this could be tested and where $\tilde{\beta}_F$ and other parameters of interest could be identified.

This estimation of the parameter of interest $\tilde{\beta}_F$ can be biased if the characteristics of the mother have any effect on the income of the child and the characteristics of the mother and the father are correlated, i.e., if there is assortative mating. In particular, we can think that the income of the son is affected by both the income of the father and the income of the mother (M_i) as reflected in equation (2):

$$S_i = \beta_F F_i + \beta_M M_i + \epsilon_i^S \tag{2}$$

If the data is generated by equation (2) but the econometrician uses an OLS estimator for equation (1), the estimator $\tilde{\beta}_F$ will be biased if $\mathbb{E}\left[F_i\left(\beta_M M_i + \epsilon_i^S\right)\right] \neq 0$ (Espín-Sánchez et

[11] Throughout the paper we consider all variables to have zero mean and unit variance, e.g., $\mathbb{E}[M_i] = \mathbb{E}[F_i] = 0$ and $\mathbb{E}[M_i M_i] = \mathbb{E}[F_i F_i] = 1$.

al., 2022). If equation (2) is well specified we have $\mathbb{E}\left[F_i \epsilon_i^S\right] = 0$ and this condition becomes $\rho\beta_M \neq 0$, where $\rho = \mathbb{E}\left[F_i M_i\right]$. Such an estimator would be

$$\tilde{\beta}_F = \frac{\mathbb{E}\left[F_i S_i\right]}{\mathbb{E}\left[F_i F_i\right]} \tag{3}$$

In general, the OLS estimate $\tilde{\beta}_F$ from equation (1) is a biased estimator of both the effect of the father on the child β_F and the effects of both parents on the child $(\beta_F + \beta_M)$. In one extreme case, where mating is perfectly assortative; i.e., the correlation between maternal and paternal income is 1, i.e., $\rho = 1$, the estimator $\tilde{\beta}_F$ will consistently estimate $(\beta_F + \rho\beta_M)$.[12]

In the another extreme case, where mating is perfectly random, i.e., $\rho = 0$, the estimator $\tilde{\beta}_F$ will consistently estimate β_F, but we would not be able to estimate β_M. These extreme examples are the most optimistic ones for the econometrician. In general, matching will be somewhat assortative ($\rho > 0$) and the estimator $\tilde{\beta}_F$ will produce a number in between β_F and $(\beta_F + \beta_M)$, but we would not know how close the estimator is to either end without knowing the degree of assortment ρ.

We now propose a simple model to estimate consistently all three parameters of interest (β_F, β_M, ρ), using only data on male income. Notice that if we have information on maternal income, we could easily estimate the three parameters of interest just by running an OLS regression on equation (2) and computing the correlation between maternal and paternal income. However, in many historical and contemporary sources maternal income is not available. This gap exists either because only the income of the household head is reported or because women do not report any occupation.[13] The system of equations is then:

$$S_i = \beta_F F_i + \beta_M M_i + \epsilon_i^S \tag{4}$$

$$F_i = \beta_F PGF_i + \beta_M PGM_i + \epsilon_i^F \tag{5}$$

$$M_i = \beta_F MGF_i + \beta_M MGM_i + \epsilon_i^M \tag{6}$$

where PGF_i is the income of the Paternal Grandfather, PGM_i is the income of the Paternal Grandmother, MGF_i is the income of the Maternal Grandfather, MGM_i is the income of the Maternal Grandmother and ϵ_i^S, $\epsilon_i^F$and ϵ_i^M are white noise. In the baseline model, the

[12] For $\tilde{\beta}_F$ to be a consistent estimate of $(\beta_F + \beta_M)$ it is required that $M_i = F_i$ $(\rho = 1)$ for all households. If $M_i = \rho F_i$, the correlation between maternal and paternal income would still be 1, but $\tilde{\beta}_F$ would not be an unbiased estimator of $(\beta_F + \beta_M)$.

[13] The same problem arises when the woman works and the man stays at home, or in any situation when either or both parents do not report income.

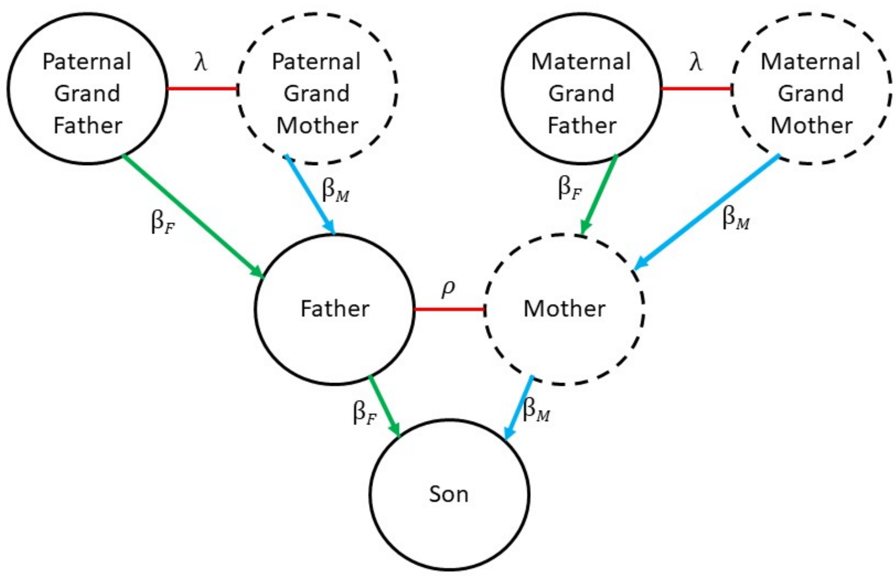
relations that we are trying to estimate are depicted in Figure 2.

- Three (horizontal) relations of assortative mating: Father-Mother, measured by ρ; and paternal grandfather-paternal grandmother and maternal grandfather-maternal grandmother measured by λ, none of which could be estimated directly.

- Three causal male relations (father-son, paternal grandfather-father and maternal grandmother-mother) that are governed by the same parameter β_F and could be estimated if mating was random.

- Three causal female relations (mother-son, paternal grandmother-father and maternal grandmother-mother) that are governed by the same parameter β_M, none of which could be estimated directly.

Figure 2: Structural Parameters.



Notes: The horizontal lines in red represent the degree of assortative matching; the vertical relations in green (arrows) represent the masculine effect on mobility; the vertical relations in blue (arrows) represent the feminine effects on mobility. The solid circles represent individuals (males) with observed income while the dashed circles represent individuals (females) with unobserved income. There are only three terminal nodes in the graph: Father, Mother and Son. Thus, we can only write three equations to represent the model.

We now present the general data generating process of the model, represented by equations (4), (5) and (6). At this point, we do not make any assumptions about the relationship between grandparents. Matrix Σ below represents the variance-covariance matrix among the grandparents. For simplicity, we normalize all variables to have zero mean and unit variance, so the correlation and the cross-products coincide.

$$\Sigma = Var \begin{bmatrix} PGF_i \\ PGM_i \\ MGF_i \\ MGM_i \end{bmatrix} = \begin{bmatrix} 1 & \lambda & a & d \\ \lambda & 1 & b & c \\ a & b & 1 & \lambda \\ d & c & \lambda & 1 \end{bmatrix}$$

where the nuisance parameters measure the four correlation among grandparent pairs, i.e., $a = \mathbb{E}[PGF_i MGF_i]$, $b = \mathbb{E}[MGF_i PGM_i]$, $c = \mathbb{E}[PGM_i MGM_i]$ and $d = \mathbb{E}[MGM_i PGF_i]$.

With this data-generating process, we can see in Figure 2 that there are only three end nodes in the tree: father, mother and son. Therefore, we can only use equations (4), (5) and (6) to estimate the model. In other words, the data is originally generated by the incomes of the four grandparents, with the correlations given by Σ. The transmission from the grandparents to the father and the mother occurs according to the causal relationships in equations (5) and (6), i.e., the income of the father and mother are realized. After those realizations, the transmission from the parents to the son occurs according to the causal relationships in equation (4), i.e., the income of the son is realized.

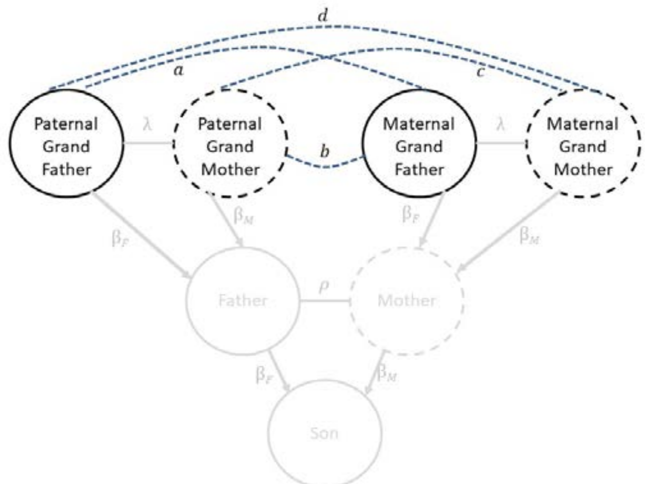
Table 2 summarizes the notation of the model. It is worth discussing the interpretation of the three types of elements here. First, the structural parameters (β_F, β_M, ρ)–shown in Figure 2–are our main parameters of interest. They reflect the parameters of the model and reflect our interest as social scientists in social phenomena. β_F measures the effect of the father's status on the child's status. β_M measures the effect of the mother's status on the child's status. ρ measures assortative mating between the mother and the father, by computing the correlation between the status of the father and the mother.

Second, we have the nuisance parameters (a, b, c, d)–shown in Table (2) and Figure 3.A. We call then nuisance parameters because they are not our main object of interest here. With the exception of $a = \mathbb{E}[PGF_i MGF_i]$, they are not directly observed in the data. Notice, however, that the nuisance parameters measure a more complex relationship between the grandparents than is typically assumed in the literature. Moreover, we can interpret the nuisance parameters, in light of matrix Σ above, as a more general way of thinking of family structure and assortative mating. In other words, the nuisance parameters allow us to think of a more general model of household formation and whether the characteristics of the grandparents affect the mating between the parents in a more nuanced way. We discuss this in detail below in Section 4.

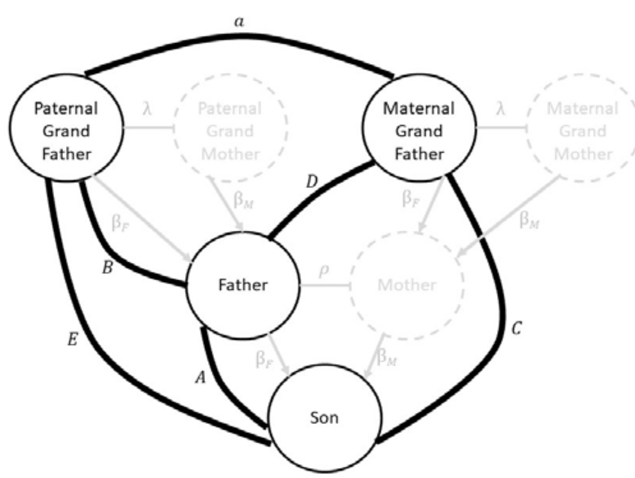
Third, we define the empirical relationships in the data. Because we have four variables

Figure 3: Nuisance Parameters and Empirical Relations.

A. Nuisance parameters.



B. Empirical Relations.



Notes: The solid circles represent individuals (males) with observed income while the dashed circles represent individuals (females) with unobserved income. Panel A. The dashed lines represents the correlations between each of the four grandparents. The rest of the model is represented in light gray. Panel B: The thick solid lines represents all six possible pair-wise empirical relations among the four male members of the family tree.

Table 2: Parameters and Data Relationships.

	Symbol	Definition
Structural Parameters	β_F	Effect of Father on child
	β_M	Effect of Mother on child
	ρ	AM, 2^{nd} gen, $\rho \equiv E\left[F_i, M_i\right]$
	λ	AM, 1^{st} gen, $\lambda \equiv E\left[PGF_i, PGM_i\right]$
Nuisance Parameters	a	$a = \mathbb{E}\left[PGF_i MGF_i\right] = 0.4811$
	b	$b = \mathbb{E}\left[MGF_i PGM_i\right]$
	c	$c = \mathbb{E}\left[PGM_i MGM_i\right]$
	d	$d = \mathbb{E}\left[MGM_i PGF_i\right]$
Empirical Relationships	A	$A = \mathbb{E}\left[S_i F_i\right] = 0.3685$
	B	$B = \mathbb{E}\left[PGF_i F_i\right] = 0.4841$
	C	$C = \mathbb{E}\left[MGF_i S_i\right] = 0.2283$
	D	$D = \mathbb{E}\left[MGF_i F_i\right] = 0.4473$
	E	$E = \mathbb{E}\left[PGF_i S_i\right] = 0.2302$

available (S_i, F_i, PGF_i, MGF_i) we can get six empirical relationships, which correspond to (A, B, C, D, E) and a–shown in Table 2 and Figure 3.A. Below we show how we use these relationships to identify the structural parameters. Notice that the standard models of social mobility only observe two variables (S_i, F_i) and are thus able to obtain one empirical relationship $A = \mathbb{E}\left[S_i F_i\right]$. With that relationship, in our model, the econometrician can only identify $\tilde{\beta_F} \equiv \beta_F + \rho\beta_M$, but not each individual parameter, as shown in Espín-Sánchez et al. (2022).

Table 3 presents a summary of the various propositions, assumptions and identified parameters. These propositions are discussed beginning in Subsection 4.3 below with proofs in Appendix A.

4 Identification

In this section, we discuss the identification of the vector of structural parameters $\beta_F, \beta_M, \lambda, \rho$. In subsection 4.1 we discuss set identification of the structural parameters. We show how the model as specified above, with four variables and six empirical relationships does not provide point identification of the structural parameters. There is one extra degree of freedom. In subsection 4.3, we present identification results when we make assumptions on the nuisance parameters. In subsection 4.2 we discuss identification using two-generation records. In subsection 4.4, we present identification results when we do not make any assumptions on the nuisance parameters. In this case, it is necessary either to use information on other male relatives–such as uncles– or to make assumptions regarding some parameters of the model,

Table 3: Summary of Identification results.

Prop.	Data	Nuisance Assumptions	Structural Assumptions	Point Identified Parameters
Identification using two generations				
Prop. 1	(F, PGF, MGF)	$a = b = c = d$	$\lambda = \rho$	(β_F, β_M, ρ)
Prop. 2	(F, PGF, MGF)		$\beta_M = 0$	(β_F, ρ)
Prop. 3	(F, PGF, MGF)	$a = b = c = d$	$\beta_F = \beta_M$	(β_F, λ, ρ)
Identification using three generations				
Prop. 4	(S, F, PGF, MGF)	$b = d$		$(\beta_F, \beta_M, \lambda, \rho)$
Identification from maternal uncles				
Prop. 5	(S, F, MGF)	$a = b = c = d$	$\lambda = \rho$	$(\beta_F, \beta_M, \rho, a)$
Prop. 6	(S, F, MU, MGF)			$(\beta_F, \beta_M, \lambda, \rho, \gamma)$
Prop. 7	(S, F, MU, PGF)			$(\beta_F, \beta_M, \lambda, \rho, \gamma)$
Prop. 8	(S, F, MU)	$\gamma = 0$	$\lambda = \rho$	(β_F, β_M, ρ)
Identification allowing heterogeneous effects by gender				
Prop. 9	(S, F, PGF, MGF)	$a = b = c = d$		$\left(\beta_F^S, \beta_M^S, \beta_F^D, \beta_M^D, \lambda, \rho\right)$
Prop. 10	(S, F, PGF, MGF)	$b = d = 0$; $a = c$		$\left(\beta_F^S, \beta_M^S, \beta_F^D, \beta_M^D, \lambda, \rho\right)$
Identification allowing heterogeneous effects by generation				
Prop. 11	(S, F, PGF, MGF)	$b = d$		$(\beta_F, \beta_M, \alpha, \rho)$
Cor. 1	(S, F, PGF, MGF)	$a = b = d$	$\alpha_F = \alpha_M$	$(\beta_F, \beta_M, \alpha_F, \lambda, \rho)$
Prop. 12	(S, F, PGF, MGF)	$b = d = 0$; $a = c$		$(\beta_F, \beta_M, \alpha_F, \alpha_M, \lambda, \rho)$
Prop. 13	(S, F, PGF, MGF)	$a = b = d$	$\lambda = \rho$	$(\beta_F, \beta_M, \alpha_F, \alpha_M, \rho)$
Prop. 14	(S, F, PGF, MGF)	$b = d = \sqrt{ac}$		$(\beta_F, \beta_M, \tilde{\alpha}, \lambda, \rho)$
Cor. 2	(S, F, PGF, MGF)	$b = d = \sqrt{ac}$	$\alpha_F = \alpha_M$	$(\beta_F, \beta_M, \tilde{\alpha}, \rho)$

Notes: The estimates for (β_F, β_M, ρ) in Propositions 4, 11 and 14, are identical because they are based on the same moments of the data.
We only use the sixth moment (equation (12)) to estimate structural parameters when we are making assumptions about c, i.e., Propositions 1; 2 and 3 (only for ρ); 5, 9 and 10; and 12 (only for λ).

or both. Finally, in subsections 4.5 and 4.6, we relax respectively the assumptions that the structural parameters are constant across the child gender and constant across generations.

4.1 Set identification

We now discuss what parameters can be identified, depending on the data available. In particular, we focus on the case where data on female characteristics is unknown. We assume that the system formed by equations (4), (5) and (6) is well specified, i.e., that the error term is uncorrelated with the regressor in each equation. In particular, the exclusion restrictions are: 1) $\mathbb{E}\left[F_i\epsilon_i^S\right]=0$; 2) $\mathbb{E}\left[PGF_i\epsilon_i^S\right]=0$; 3) $\mathbb{E}\left[PGF_i\epsilon_i^F\right]=0$; 4) $\mathbb{E}\left[PGF_i\epsilon_i^M\right]=0$; 5) $\mathbb{E}\left[MGF_i\epsilon_i^S\right]=0$; 6) $\mathbb{E}\left[MGF_i\epsilon_i^F\right]=0$; and 7) $\mathbb{E}\left[MGF_i\epsilon_i^M\right]=0$

Notice that we only have 7 exclusion restrictions. Since the son is at the end of the tree, we cannot have any exclusion restriction multiplying S_i by any of the ϵ_i. The first exclusion restriction is the standard one, when we just say that the equation for the son is well specified. Since the father is in the middle of the tree, we can only use one exclusion restriction for F_i, $\mathbb{E}\left[F_i\epsilon_i^S\right]=0$, since F_i would be endogenous to all other epsilons, except the one at the bottom of the tree. The grandfathers are at the top of the tree therefore, PGF_i and MGF_i, are exogenous to the epsilons in the middle of the tree, ϵ_i^F and ϵ_i^M and the one at the end of the tree ϵ_i^S. Using these exclusion restrictions we are using the grandparents as non-linear instruments to construct the relevant moments.

We now combine equations (4), (5) and (6) with the seven exclusion restrictions above to generate six moments.[14] As shown in Appendix A.1, with the exclusion restrictions above we can generate the following moments:
Using equation (4) and $\mathbb{E}\left[F_i\epsilon_i^S\right]=0$ we get

$$\begin{aligned} A &\equiv \mathbb{E}\left[S_iF_i\right] \\ &= \beta_F\mathbb{E}\left[F_iF_i\right]+\beta_M\mathbb{E}\left[M_iF_i\right] \\ &= \beta_F+\rho\beta_M \end{aligned} \tag{7}$$

[14] In each case, we multiply one of the equations for one of the observable variables and take expectations, e.g., we take equation (4) and multiply by F_i, and then take expectations and use $\mathbb{E}\left[F_i\epsilon_i^S\right]=0$ to get equation (7). See Appendix A.1 for details.

Using equation (5) and $\mathbb{E}\left[PGF_i\epsilon_i^F\right]=0$ we get

$$\begin{aligned} B &\equiv \mathbb{E}\left[F_iPGF_i\right] \\ &= \beta_F\mathbb{E}\left[PGF_iPGF_i\right]+\beta_M\mathbb{E}\left[PGM_iPGF_i\right] \\ &= \beta_F+\lambda\beta_M \end{aligned} \tag{8}$$

Using equations (4) and (6) and $\mathbb{E}\left[MGF_i\epsilon_i^M\right]=0$ and $\mathbb{E}\left[MGF_i\epsilon_i^S\right]=0$ we get

$$\begin{aligned} C &\equiv \mathbb{E}\left[MGF_iS_i\right]=\beta_F\mathbb{E}\left[MGF_iF_i\right]+\beta_M\mathbb{E}\left[MGF_iM_i\right] \\ &= \beta_F\mathbb{E}\left[MGF_iF_i\right]+\beta_M\left(\beta_F\mathbb{E}\left[MGF_iMGF_i\right]+\beta_M\mathbb{E}\left[MGF_iMGM_i\right]\right) \\ &= \beta_FD+\beta_M\left(\beta_F+\lambda\beta_M\right) \end{aligned} \tag{9}$$

Using equation (5) and $\mathbb{E}\left[MGF_i\epsilon_i^F\right]=0$ we get

$$\begin{aligned} D &\equiv \mathbb{E}\left[F_iMGF_i\right] \\ &= \beta_F\mathbb{E}\left[PGF_iMGF_i\right]+\beta_M\mathbb{E}\left[PGM_iMGF_i\right] \\ &= \beta_Fa+\beta_Mb \end{aligned} \tag{10}$$

Using equation (4) and $\mathbb{E}\left[PGF_i\epsilon_i^S\right]=0$ and $\mathbb{E}\left[PGF_i\epsilon_i^M\right]=0$ we get

$$\begin{aligned} E &\equiv \mathbb{E}\left[PGF_iS_i\right] \\ &= \beta_F\mathbb{E}\left[PGF_iF_i\right]+\beta_M\mathbb{E}\left[PGF_iM_i\right] \\ &= \beta_F\mathbb{E}\left[PGF_iF_i\right]+\beta_M\left(\beta_F\mathbb{E}\left[PGF_iMGF_i\right]+\beta_M\mathbb{E}\left[PGF_iMGM_i\right]\right) \\ &= \beta_FB+\beta_M\left(\beta_Fa+\beta_Md\right) \end{aligned} \tag{11}$$

Using equations (5) and (6) we get[15]

$$\mathbb{E}\left[F_iM_i\right]=\mathbb{E}\left[\left(\beta_FPGF_i+\beta_MPGM_i\right)\left(\beta_FMGF_i+\beta_MMGM\right)\right]$$

and solving we get

$$\rho=\beta_F^2a+\beta_M^2c+\beta_F\beta_M\left(b+d\right) \tag{12}$$

In equations (10), (11) and (12) we have three terms, the nuisance parameters, that we do not observe in the data, e.g., $b=\mathbb{E}\left[MGF_iPGM_i\right]$, $c=\mathbb{E}\left[PGM_iMGM_i\right]$ and $d=\mathbb{E}\left[MGM_iPGF_i\right]$. Therefore, we have six independent equations and seven unknowns.

We now put out system of equations in light or other econometric estimators used in economics. In particular, we show how our estimator can be seen as a generalization of both

[15] Notice that here we are using $\mathbb{E}\left[MGF_i\epsilon_i^F\right]=\mathbb{E}\left[MGM_i\epsilon_i^F\right]=\mathbb{E}\left[PGF_i\epsilon_i^M\right]=\mathbb{E}\left[PGM_i\epsilon_i^M\right]=0$. In addition to that, for this moment only we are assuming $\mathbb{E}\left[\epsilon_i^F\epsilon_i^M\right]=0$. This implies that mating is done on observables (income) only, i.e., no mating on unobservables. This assumption could be relax and we could estimate $\mathbb{E}\left[\epsilon_i^F\epsilon_i^M\right]$ by assuming, for example, $a=c$.

Instrumental Variables (IV) in cross-sectional data and Time Series data.[16]

Equation (7)–and also equation (8)–can be seen as a standard equation with omitted variables in cross-sectional data where S_i is the dependent variable, F_i is the regressor and M_i is the omitted variable. The standard solution in such case would be to use an instrumental variable that is correlated with the regressor F_i but not with the omitted variable M_i. What we do here instead is to use an instrumental variable that is correlated with the omitted variable M_i in a particular way. For example, Equation (9) is equivalent to using MGF_i as an instrument in equation (4). However, instead of the usual assumption that the instrument is uncorrelated with the omitted variable, i.e., $\mathbb{E}\left[MGF_i M_i\right] = 0$, our model indicates that this correlation is a function of the structural parameters, i.e., $\mathbb{E}\left[MGF_i M_i\right] = \beta_F + \lambda\beta_M$. In that sense, our model is a generalization of the usual IV approach where we put structure on the correlation of our instrument and the omitted variable.

Although mathematically all six moments–that is, equations (7), (8), (9), (10), (11) and (12)–come from the identifying assumptions of the data, there is a qualitative difference between the first three and the last three equations. In the first three equations ((7), (8) and (9)), the exclusion restrictions come from using the error term of a person and the status of her/his father, e.g., ϵ_i^S and F_i, ϵ_i^F and PGF_i, and ϵ_i^M and MGF_i. These moments are analogous to a time series model that follows an $AR(1)$ process. In that case the model is $Y_t = \beta Y_{t-1} + \epsilon_t$ and the exclusion restriction is $\mathbb{E}\left[Y_{t-1}\epsilon_t\right] = 0$. When $\beta_M = 0$ our model is identical to that $AR(1)$ process. In that sense, time series is a particular case of econometric trees where only only line–in this case patrilineal–matters. The availability of the matrilineal line provides extra exclusion restrictions that allow us to identify β_M and ρ. We can think of our estimator as a generalization of the instruments used in time series, where we also use the matrilineal "lags" in the data.

Therefore, in equations equations (7), (8), (9) there are no nuisance parameters. This means that if we have one extra identifying assumption that does not relate to the nuisance parameters, we just need to use the first three equations, together with a new independent assumption and we will get a system of four independent equations and four unknowns. The assumption $\rho = \lambda$ does not provide a new independent equation, because it would make equations (7) and (8) colinear.

In the last three equations (10, 11 and 12), we have the nuisance parameters (b, c, d), as well as the parameters relating to the son's equation (β_F, β_M, ρ). Thus, if we can identify (β_F, β_M, ρ) from other equations, we can use the last three equations to identify (b, c, d).[17]

[16] When we use data maternal uncles we face a problem similar to panel data in that there could be household fixed effects. Our estimator in that case can be seen as a generalization of Arellano and Bond (1991).

[17] The intuition also works conversely: if we can identify (b, c, d), then we can identify (β_F, β_M, ρ). This

Lemma (1) below shows this intuition.

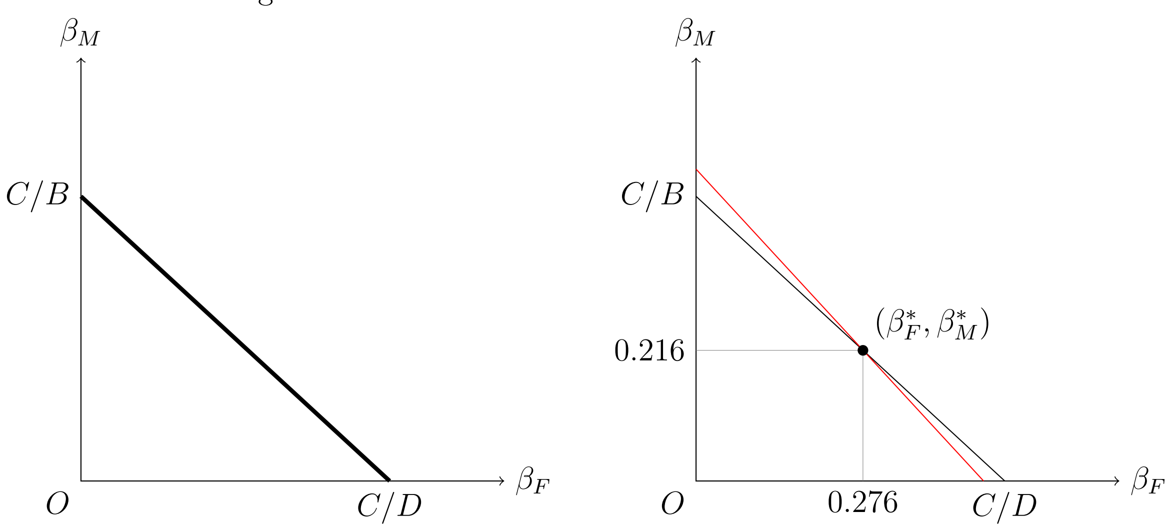
Lemma 1. *If we know* (β_F, β_M, ρ)*, then equations* (10)*,* (11) *and* (12) *identify all nuisance parameters* (b, c, d)*.*

We now discuss set identification using equations (8) and (9) above. Combining both equations we get

$$\beta_M \mathbb{E}\left[F_i PGF_i\right] = \mathbb{E}\left[MGF_i S_i\right] - \beta_F \mathbb{E}\left[MGF_i F_i\right] \tag{13}$$

Thus, we can use two identifying equations to get an equation that is a linear combination of β_F and β_M alone. Figure 4 (left) shows an equation that depends only on β_F and β_M and data, by combining equations (8) and (9). Notice that all the other moments depend on ρ, λ and the nuisance parameters. Thus, without further assumptions, this line is our identified set for β_F and β_M.

Figure 4: Set Identification and Point Identification.



Notes: Panel A. The equation comes from using equations (8) and (9). The equation is $\beta_M = \frac{C}{B} - \frac{D}{B}\beta_F = 0.4707 - 0.9221\beta_F$. Panel B. Black line is the same as in panel A. The second equation (red line) comes from using equations (10) and (11) and the assumption $b = d$. The resulting equation is $\beta_M = \frac{E}{D} - \frac{B}{D}\beta_F = 0.5157 - 1.0845\beta_F$.

4.2 Identification Using Two Generations

In this subsection, we present our baseline results. We assume that the econometrician has access to two generations of data–marriage records or similar information–on male socioeconomic status F_i, PGF_i, MGF_i. With three variables, we can only compute three empirical

intuition is the basis for the proofs of the propositions below.

moments: $a = \mathbb{E}\left[PGF_i MGF_i\right]$, $D = \mathbb{E}\left[MGF_i F_i\right]$ and $B = \mathbb{E}\left[PGF_i F_i\right]$. We now show how these three empirical moments, together with the exclusion restrictions, generate three independent equations. These equations, unfortunately, contain nuisance parameters and we cannot identify any structural parameters without further assumptions.

The system of equations using two generations is then:

$$\mathbb{E}\left[F_i PGF_i\right] = \beta_F + \lambda \beta_M \tag{14}$$

$$\mathbb{E}\left[F_i MGF_i\right] = \beta_F a + \beta_M b \tag{15}$$

$$\rho = \beta_F^2 a + \beta_M^2 c + \beta_F \beta_M \left(d + b\right) \tag{16}$$

We now have a system with three equations and seven unknowns: the four structural parameters $(\beta_F, \beta_M, \lambda, \rho)$ and the three unknown nuisance parameters (b, c, d). We now show three possible ways to get identification of some structural parameters when we only have information from two-generation records.

Proposition 1 below shows how one can identify (β_F, β_M, ρ) making assumptions on nuisance parameters, using information on two-generation records alone, i.e., without observing S_i.

Proposition 1. *Suppose F_i, PGF_i, and MGF_i are observed. If we assume $a = b = c = d$ and $\lambda = \rho$, then (β_F, β_M, ρ) is point identified.*

Proposition 1 needs to make the strongest assumption on the nuisance parameters, i.e., that they are all equal. This highlights the importance of having information on three generations, not just two, even if we want to estimate a model where the effects are only across one generation. Having information on the son adds one variable S_i to the system, but in terms of moments, we go from just three empirical relations (a, B, D) to six empirical relations (a, A, B, C, D, E). This extra variable doubles our degrees of freedom and allows us to get a more nuanced picture of intergenerational mobility and assortative mating. Below we show Proposition 2, which uses similar assumptions than Clark et al. (2022), but here we specify the model correctly and obtain a different estimator.

Proposition 2. *Suppose F_i, PGF_i, and MGF_i are observed. If we assume $\beta_M = 0$, then (β_F, ρ) is point identified.*

Notice that Proposition 2 generates a system with three equations and two unknowns. This means that the system is overidentified, or that the assumption $\beta_M = 0$ is testable. On the other hand, one can see that if we do not have information on a we can still estimate (β_F, ρ) by writing $\beta_F = B$ and $\rho = DB$.

There are some recent papers using marriage records to study social mobility and assortative mating (Clark et al., 2022; Curtis, 2022; Clark and Cummins, 2022). The estimator they use are based on the seminal work by Chadwick and Solon (2002, henceforth CS).[18] This method requires two assumptions: i) that the income of the mother causes the income of the father, i.e., $F_i = \rho_0 M_i + \epsilon$; ii) That the parental grandparents' income is uncorrelated with the maternal grandparents' income, i.e., $a = b = c = d = 0$. If this two assumptions hold, then, following CS you can estimate $\rho_0 = D/B$. This is the method used by the papers above. Alternatives to assumption i) give different estimators. In Appendix A.8 we show that if you assume causality in the opposite direction, i.e., $M_i = \rho_1 F_i + \epsilon^{'}$ you get $\rho_1 = B/D$. If you do not assume causality, but rather that the relation between the father and the mother is just correlational, Prop. 2 shows that $\rho = DB$. We think that in most settings the relation between spouses is correlational, not causal. In summary

- CS: $F_i = \rho_0 M_i + \epsilon$, the estimator is $\rho_0 = D/B$.
- Reversed CS: $M_i = \rho_1 F_i + \epsilon^{'}$,the estimator is $\rho_1 = B/D$.
- Non-causal: $\mathbb{E}\left[F_i M_i\right] = \rho$, the estimator is $\rho = BD$.

Whereas i) is an economic assumption, ii) is an empirical assumption. Moreover, if the econometrician is using two-generation records to estimate B and D, she should be able to estimate a. If $a \neq 0$ the CS method is invalid, regardless of the economic setting.

In summary, using two-generation records alone, without further assumptions on structural or nuisance parameters, we cannot identify β_M. If we assume that mothers have no effect on their children, i.e., $\beta_M = 0$, then we can identify β_F and ρ. However, when the relationship between the parents income/status is non-causal, unlike in Chadwick and Solon (2002) and Clark (2023), the estimator is the product of the two empirical relations–$\mathbb{E}\left[PGF_i F_i\right]$ and $\mathbb{E}\left[MGF_i F_i\right]$–not their ratio.

Assuming $\beta_M = 0$ might be too strong in many settings. Proposition 3 below shows how to estimate (β_F, ρ) using data from two generations if you assume $\beta_F = \beta_M$ as in Clark and Cummins (2022). Moreover, using a allow us to also identify λ. This result is somewhat surprising because we only have data in two generations, but we are able to identify the degree of assortative mating on both generations, i.e., we can identify λ and ρ.

Proposition 3. *Suppose F_i, PGF_i, and MGF_i are observed. If we assume $a = b = c = d$ and $\beta_F = \beta_M$, then (β_F, λ, ρ) is point identified.*

[18] For simplicity here, we are restricting attention to the case with $\beta_M = 0$. Chadwick and Solon (2002) compute two parameters: the assortative mating parameter ρ_0, and the total effect parameter $\tilde{\beta}_F = \beta_F + \lambda\beta_M$. Appendix (A.8) shows how the analysis is the same when $\beta_M \neq 0$, and you also get $\rho_0 = D/B$.

4.3 Identification Using Three Generations

We now show that we can get point identification on $(\beta_F, \beta_M, \lambda, \rho)$ when we only make a mild assumption on the nuisance parameters, if we have information on three generations. Unlike in the previous sections, we can now relax the strong assumptions on the nuisance parameters, i.e., $a = b = c = d$, if we have information on three generations, i.e., information on the son S_i. With this fourth variable, we get the full six empirical relationships (a, A, B, C, D, E) and we only need to impose one restriction on the nuisance parameter. In particular, we assume that the correlation among grandparents across genders is equal to the product of the standard deviations, i.e., $\mathbb{E}[PGF_i MGM_i] = \mathbb{E}[MGF_i PGM_i]$, or $b = d$. This is, in our opinion, the weakest assumption on the nuisance parameters.Proposition 4 shows how this assumption generates point identification.

Proposition 4. *Suppose S_i, F_i, PGF_i and MGF_i are observed. If we assume $b = d$, then $(\beta_F, \beta_M, \lambda, \rho)$ is point identified.*

The proof for Proposition 4 is simple. If we combine equations (8) and (9), we get an equation that depends only on β_F and β_M and data. This equation is the black line on the right panel on Figure (4). Using the assumption that $b = d$, we can combine equations (10) and (11) and get an independent equation that depends only on β_F and β_M and data. This equation is the red line on the right panel on Figure 4. Thus, using these four moments and the restriction on the nuisance parameters, we can identify β_F and β_M. Identification on λ and ρ follows using the other moments.

4.4 Identification Using Maternal Uncles

In this section, we discuss identification when we have information on one maternal uncle. In addition to the six moments that we showed above, using data on (S, F, MGF, PGF), we can get additional moments if we have information on one maternal uncle. The equation determining the status of a maternal uncle is

$$MU_i = \beta_F MGF_i + \beta_M MGM_i + \epsilon_i^{MU} \tag{17}$$

With this new equation, we have a new parameter $\gamma \equiv \mathbb{E}\left[\epsilon_i^M \epsilon_i^{MU}\right]$ that measures household fixed effects. Moreover, identification is much simpler now because we have $\mathbb{E}[F_i MU_i] = \rho$. Assortative mating can be computed directly from the data.

Propositions 5 and 6 show identification results when we do not observe PGF_i. In Proposition 5, point identification of the parameters of interest requires strong assumptions on the nuisance parameters and $\lambda = \rho$. However, if we also have information on one maternal uncle, Proposition 6 shows that we can get identification of all parameters without any assumption. Moreover, we can also estimate household fixed effects γ. Notice that adding one maternal uncle to a sample of (S, F, MGF) adds more identification restrictions than adding one paternal grandfather. The reason is that the maternal uncle is closely related to the mother, which is the object here.

Proposition 5. *Suppose S_i, F_i and MGF_i are observed. If we assume $a = b = c = d$ and $\lambda = \rho$, then $(\beta_F, \beta_M, \rho, a)$ is point identified.*

Proposition 6 below shows that we can get identification of all parameters, without any assumption, including household fixed effects γ. Typically one needs to observe the status of two siblings in order to estimate household fixed effects. Here, we can do it without observing the status of a pair of siblings. We observe the status of one sibling, the maternal uncle and some relatives of the other sibling, the mother.

Proposition 6. *Suppose S_i, F_i, MU_i and MGF_i are observed. Then $(\beta_F, \beta_M, \lambda, \rho, \gamma)$ is point identified.*

Proposition 6 uses only four variables and generates six equations and six empirical moments $(A, C, D, \mathbb{E}[F_i MU_i], \mathbb{E}[MGF_i MU_i], \mathbb{E}[MU_i S_i])$. We can use six of these equations, where there are no nuisance parameters and estimate five structural parameters $(\beta_F, \beta_M, \lambda, \rho, \gamma)$. Unlike in the case in Proposition 4, where we have information on the paternal grandfather, but not on the maternal uncle, here we can estimate all structural parameters and γ, with six empirical moments. The difference is that the equations using the paternal grandfather typically involve nuisance parameters, but the equations that we get using the uncles do not. Therefore, for a researcher interested in the structural parameters, the information provided by a maternal uncle is much more valuable.

Proposition 7. *Suppose S_i, F_i, MU_i and PGF_i are observed. Then $(\beta_F, \beta_M, \lambda, \rho, \gamma)$ is point identified.*

Proposition 7 uses all the information available and shows that we can identify all the parameters of interest and still have two extra degrees of freedom. With information on five variables (S, F, MU, PGF, MGF), we can generate empirical moments: the six empirical relations discussed above in Table 2 (A, B, C, D, E, a) and the new four moments with maternal uncles $(\mathbb{E}[F_i MU_i], \mathbb{E}[MGF_i MU_i], \mathbb{E}[MU_i S_i], \mathbb{E}[PGF_i MU_i])$ and ten equations. Out of the ten equations, we can get eight independent equations to estimate eight parameters $(\beta_F, \beta_M, \lambda, \rho, \gamma, b, c, d)$.

Finally, Proposition 8 below shows how we can still get identification on our main parameters of interest (β_F, β_M, ρ), having data only on (S, F, MU). This result requires some assumptions on the structural parameters. However, no assumptions are needed on the nuisance parameters. The reason is that we are not using any data and any equation involving any grandparents.

Proposition 8. *Suppose S_i, F_i, MU_i are observed and $\lambda = \rho$ and $\gamma = 0$, then (β_F, β_M, ρ) is point identified.*

4.5 Gendered Effects

In the baseline model in Section 3, we were implicitly imposing that the effects, even when they were different from the father and from the mother, were the same when the child was a son or a daughter. In this subsection, we present identification results when we allow for gendered effects, but impose restrictions on the nuisance parameters. The system of equations that we are considering is the following

$$\begin{aligned} S_i &= \beta_F^S F_i + \beta_M^S M_i + \epsilon_i^S \\ F_i &= \beta_F^S PGF_i + \beta_M^S PGM_i + \epsilon_i^F \\ M_i &= \beta_F^D MGF_i + \beta_M^D MGM_i + \epsilon_i^M \end{aligned} \tag{18}$$

where β_F^S and β_M^S are the effects of the father and mother on a son, respectively and β_F^D and β_M^D are the effects of the father and mother on a daughter, respectively.

Proposition 9. *Suppose S_i, F_i, PGF_i and MGF_i are observed. If we assume $a = b = c = d$, then $\left(\beta_F^S, \beta_M^S, \beta_F^D, \beta_M^D, \lambda, \rho\right)$ is point identified.*

Proposition 9 above shows that we can identify all six parameters of interest, in the model with gendered effects, if we are willing to impose that all correlations among grandparents' incomes are equal to the observed correlation between the two grandfathers, i.e., $a = b = c = d$. Proposition 10 below shows a similar result, but imposing $b = d = 0$ and $a = c$ instead.

Proposition 10. *Suppose S_i, F_i, PGF_i and MGF_i are observed. If we assume $b = d = 0$ and $a = c$, then $\left(\beta_F^S, \beta_M^S, \beta_F^D, \beta_M^D, \lambda, \rho\right)$ is point identified.*

4.6 Generational Effects

In the baseline model in Section 3, we were implicitly imposing that the mobility effects were the same in both generations. In this subsection, we present identification results when we allow for different generational effects effects, but impose restrictions on the nuisance parameters. The system of equations that we are considering is the following

$$
\begin{aligned}
S_i &= \beta_F F_i + \beta_M M_i + \epsilon_i^S \\
F_i &= \alpha_F PGF_i + \alpha_M PGM_i + \epsilon_i^F \\
M_i &= \alpha_F MGF_i + \alpha_M MGM_i + \epsilon_i^M
\end{aligned}
\tag{19}
$$

where β_F and β_M are the effects of the father and mother in the second generation, respectively and α_F and α_M are the effects of the father and mother in the first generation, respectively.

We now show several results showing sufficient conditions for identification of effects that differ across generations in our model.

Proposition 11. *Suppose S_i, F_i, PGF_i and MGF_i are observed. If we assume $b = d$, then (β_F, β_M, ρ) and α (with $\alpha \equiv \alpha_F + \alpha_M$) is point identified. However, λ, α_F and α_M are not point identified.*

Proposition 11 shows that imposing assumptions on the nuisance parameters, but no assumptions on the structural parameters, is not enough to get point identification here. Nonetheless, identifying $(\beta_F, \beta_M, \alpha)$ could be of interest in many settings. For example, the econometrician might be willing to assume that the coefficients for father and mother for the first generations are equal to each other, i.e., $\alpha_F = \alpha_M$. Corollary 1 below shows that, with this extra assumption, all parameters are identified.

Corollary 1. *Suppose S_i, F_i, PGF_i and MGF_i are observed. If we assume $a = b = d$ and $\alpha_F = \alpha_M$, then $(\beta_F, \beta_M, \alpha_F, \lambda, \rho)$ is point identified.*

Proposition 12 below shows that we can get identification on all the structural parameters without imposing any restrictions on them, if we impose slightly different restrictions on the nuisance parameters. The new restrictions break the dependency across moments that was created by the restrictions imposed in Proposition 11.

Proposition 12. *Suppose S_i, F_i, PGF_i and MGF_i are observed. If we assume $b = d = 0$ and $a = c$, then $(\beta_F, \beta_M, \alpha_F, \alpha_M, \lambda, \rho)$ is point identified.*

One reasonable assumption that one can make is to assume that the degree of assortative mating is constant across generations, i.e., $\lambda = \rho$. Proposition 13 shows that this assumption provides point identification in all the other structural parameters.

Notice that this result contrast with the negative results shown at the beginning of this section. In the simple model, assuming $\lambda = \rho$ did not add any identification to our model, but here, it provides an independent equation, with respect to the results in proposition 11. The reason behind this somewhat surprising result is that in the baseline case, we were imposing that the mobility parameters β_F and β_M were constant across generations. Thus, imposing that also the mating parameters were constant did not add more degrees of freedom. Here, we are allowing the mobility parameters to differ across generations, i.e., $\beta_F \neq \alpha_F$ and $\beta_M \neq \alpha_M$. Thus, assuming that the mating parameters are constant, i.e., $\lambda = \rho$, does add more identification power here.

Proposition 13. *Suppose S_i, F_i, PGF_i and MGF_i are observed. If we assume $a = b = d$ and $\rho = \lambda$, then $(\beta_F, \beta_M, \alpha_F, \alpha_M, \rho)$, is point identified.*

Propositions 11 and 12 are two different ways to get point identification in the structural parameters. In both cases, we are adding three restrictions to the nuisance parameters and that allows us to form a system with six equations and six parameters of interest. In Proposition 12 the six equations are independent and thus we can get point identification in all six parameters of interest. In Propositions 11, however, the equations are not independent and we get point identifications in all parameters but α_F and α_M. We get identification in their sum α and we end up with a system of over-identifying restrictions. Proposition 14 below extends this intuition and shows how we can also get point identification on $(\beta_F, \beta_M, \lambda, \rho)$, with a weaker assumption on the nuisance parameters: $b = d = \sqrt{ac}$. The downside is that now we cannot point identify the mobility parameters in the older generation α_F and α_M, or the nuisance parameter c, but we can identify $\tilde{\alpha} \equiv \alpha_F\sqrt{a} + \alpha_M\sqrt{c}$.

Proposition 14. *Suppose S_i, F_i, PGF_i and MGF_i are observed. If we assume $b = d = \sqrt{ac}$, then (β_F, β_M, ρ) and $\tilde{\alpha}$ (with $\tilde{\alpha} \equiv \alpha_F\sqrt{a} + \alpha_M\sqrt{c}$) is point identified. However, α_F, α_M, and λ are not point identified.*

Proposition 14 presents a negative result when we use the assumption $b = d = \sqrt{ac}$. Corollary 1 above shows how the negative result in Proposition 11 can be overcome by a simple restriction on the structural parameters such as $\alpha_F = \alpha_M$. Corollary 2 below shows how the same assumption does not solve the identification issues from Proposition 14.

Corollary 2. *Suppose S_i, F_i, PGF_i and MGF_i are observed. If we assume $b = d = \sqrt{ac}$ and $\alpha_F = \alpha_M$, (β_F, β_M, ρ) and $\tilde{\alpha}$ (with $\tilde{\alpha} \equiv \alpha_F(\sqrt{a} + \sqrt{c})$) is point identified. However, α_F, α_M, and λ are not point identified.*

5 Results

In this section, we first produce reduced-form estimates, and then estimate the structural models developed in Section 4.

5.1 Reduced-form Results

In this subsection, we discuss the usual reduced-form (or proxy) estimates in the literature and how they relate to our structural estimates. The importance of these biases is particularly relevant when one is estimating changes in mobility or assortment over time. The changes in the proxy measure could be due to changes in other structural parameters of interest, such as β_M. In those cases the econometrician would mistake changes over time on the effects of mothers for changes over time on the effects of father or the degree of assortment. First, as we have noted above most of the literature takes the correlation of father and son A (or B) as a measure of β_F (or α_F). In our simple model, we have $A = \beta_F + \rho\beta_M$ (or $A = \beta_F + \rho\beta_M$). This means that A typically overestimates β_F when $\rho > 0$ and $\beta_M > 0$. The larger the assortment and the larger the mother's effect, the large the bias.

Second, some recent papers (Craig et al., 2019; Althoff et al., 2022) compute the correlation between grandparents a as a measure of assortative mating in the parents generation ρ. In our simple model, we have $\rho = \beta_F^2 a + \beta_M^2 c + \beta_F \beta_M (b + d)$. If we assume $a = b = c = d$, we have $\rho = (\beta_F + \beta_M) a$. This means that–even after assuming all nuisance parameters are equal–a is a proxy for ρ but will typically be much smaller. The larger the father or mother's effects, the more severe the bias.

Third, some recent papers (Chadwick and Solon, 2002; Curtis, 2022) write the assortative mating as a causal relationship (the mother causing the status of the father). This creates a measure of assortment equal to D/B. We show in Appendix A.8 that assuming a reverse causal relationship (the father causing the status of the mother) creates a measure of assortment equal to B/D. Our approach, taking the mating relationships as non-causal, is robust to this type of interpretation. Moreover, the closest results (Prop. 2) show that $\rho = BD$.

In Table 4, we show the coefficients from a regression using the two older generations; that is, our fathers and grandparents.[19] We show how even if we regress the father (PGF_i) and father-in-law (MGF_i) on an individual (F_i), we would still not get the proper estimates. In other words one cannot run an OLS regression and identify any of the structural parameters of interest. In our notation that is

$$F_i = \beta_1 PGF_i + \beta_2 MGF_i + \epsilon$$

In simple terms, the reduced-form regression creates two moments

$$B = \beta_1 + \beta_2 a$$
$$D = \beta_1 a + \beta_2$$

[19] For simplicity we restrict attention to the linked sample in which we have linked the son with his father and both grandparents.

Table 4: Intergenerational Mobility: 1900 to 1920

	Father Log Occscore	
	(1)	(2)
PGF	0.443***	0.332***
	(0.007)	(0.008)
MGF		0.253***
		(0.008)
N	23,735	23,735

Notes: Coefficients from regression of standardized income of the fathers on that of his father and his father-in-law. Incomes are adjusted as described in Section 2.2. Sample contains all individuals for whom the father and both grandparents were successfully linked. Standard errors are robust. *, ** and *** represent $p < 0.1$, $p < 0.05$ and $p < 0.01$ respectively.

Assuming $a = b$ the two relevant moments in Proposition 1 are

$$B = \beta_F + \lambda_M$$
$$D = \beta_F a + \beta_M a$$

To make the first two moments coincide we need $\beta_F = \beta_1$, $\beta_2 a = \rho\beta_M$ and $\beta_2 = \beta_M a$. This is true when $\beta_1 = \beta_F$ and $\beta_2 = \beta_M a$ and $\lambda = a^2$. The formulas for β_1 and β_2 are a particular case of our formulas for β_F and β_M in Proposition 1, with $\beta_1 = \beta_F$ and $\beta_2 = \beta_M a$, only when $\lambda = a^2$. When $\lambda = \rho$, Proposition 1 shows that $\rho = D^2/a$. This means that the reduced-form regression consistently estimates β_F in the particular case where $a^3 = D^2$. This is unlikely to hold in reality, except in the case where $\beta_M = 0$. In that case, $a^3 = D^2 = 0$ and $\beta_1 = \beta_F = B$ and $\beta_2 = \beta_M = 0$.[20] In summary, just by looking at the two coefficients from the regression above, we cannot estimate any of the structural parameters. We need to use our proposed GMM method to do so.

5.2 Structural results

In this subsection, we present the structural model estimates. In all the results, we have a system with exactly identified parameters which we estimate using GMM (Hansen, 1982). We use efficient standard errors by using the inverse of the Jacobian of the moments matrix as weighting matrix.

[20] An alternative way of looking at this is looking at the estimates from Proposition 4. The estimates from this reduced-form regression are $\beta_1 = \frac{B-Da}{1-a^2}$ and $\beta_2 = \frac{D-Ba}{1-a^2}$. The formulas for β_F and β_M in Proposition 4 are $\beta_F = \frac{BE-CD}{B^2-D^2}$ and $\beta_M = \frac{BC-ED}{B^2-D^2}$. Thus, for β_1 to be a consistent estimate of β_F we need $E = 1$, $a = C = D$ and $B = 1$. Notice that this will also make β_2 a consistent estimate for β_M. These assumptions however, will imply $\beta_1 = \beta_F = 1$ and $\beta_2 = \beta_M = 0$. In other words, if we use the first two moments from Proposition 1, β_1 and β_2 cannot be consistent estimates of β_F and β_M.

Table 5: Identification using Two Generations

Parameter	Estimate		
	Prop. 1	Prop. 2	Prop. 3
β_F	0.167	0.930	0.465
	(0.025)	(0.015)	(0.007)
β_M	0.763		
	(0.039)		
ρ	0.416	0.416	0.416
	(0.012)	(0.012)	(0.007)

Notes: Parameters identified from two-generation which contain a father and grandparents. Proposition 1 assumes that $\lambda = \rho$ and additionally equality in the nuisance parameters. Proposition 2 assumes that $\beta_M = 0$. Proposition 3 assumes that $\beta_F = \beta_M$ and additionally equality in the nuisance parameters.

In Table 5, we examine identification using data from two generations; that is, social mobility estimates using the father and both grandfathers. Using our baseline specification in Prop. 1, we find a contribution of mothers to social status almost five times larger that that of fathers (0.763 vs 0.167), and a high degree of assortative mating, at 0.416. Propositions requiring either zero contribution from mothers (Prop. 2) or equality of fathers and mothers effects (Prop. 3) show substantially more persistence from fathers. Assuming $\beta_M = 0$ would bias the estimate for β_F up (Prop. 2). Assuming $\beta_M = \beta_F$, when $\beta_M > \beta_F$, would bias the estimate for β_F up (Prop. 3). Prop. 2 uses three moments to estimate two parameters. The estimates displayed in Table 5 use the second and third moment. The missing moment is $B = 0.484 = \beta_F$. This estimate is much lower than the estimate reported in Table 5 (0.484 vs 0.930), this is further evidence that the assumption $\beta_M = 0$ is rejected in our data.

In both cases, the estimate for ρ is the same as in Prop. 1, because we are using the same moments to identify ρ. We conclude from this that allowing for effects to differ by gender is critical for recovering plausible effects of social mobility. Seemingly innocuous assumptions commonly found in the literature, such as $\beta_M = \beta_F$ or $\beta_M = 0$,would produce very different estimates and hide the true size of the mother's contribution.

In Table 6, we show the parameter estimates from point identification of $(\beta_F, \beta_M, \rho, \lambda)$ from Prop. 4, assuming non-gendered effects and equality in two nuisance parameters. We show estimates for the modified OCCSCOREs described in Section 2.2. Using the preferred adjustments, the direct effect of the father is 0.272 vs 0.220 for the mother. However, the point estimate for λ, the degree of assortative mating in the grandparents' generation, is implausibly high, with near perfect assortative mating. We interpret this result as a joint

Table 6: Identification using Three Generations

Parameter	Estimate			
	Prop. 4			
Adjustment	Unadj.	State	Farm	State and Farm
β_F	0.336	0.445	0.328	0.272
	(0.055)	(0.101)	(0.072)	(0.080)
β_M	0.210	0.170	0.108	0.220
	(0.055)	(0.101)	(0.072)	(0.081)
λ	0.320	0.892	1.000	0.963
	(0.180)	(0.096)	(0.142)	(0.064)
ρ	0.000	0.000	0.000	0.437
	(0.259)	(0.594)	(0.662)	(0.207)

Notes: Set identifed parameters. Identification is from the assumtion on the nuisance parameters that $b = d$. Adjustments are to occupational scores as described in Section 2.2.

rejection of the assumptions of the model as specified, in particular the assumption of constant effects across generations ($\beta_F = \alpha_F$ and $\beta_M = \alpha_M$). The raw correlation between fathers and sons ($A = 0.3685$) is significantly lower than that between fathers and paternal grandfathers ($B = 0.4841$). Because equations (7) and (8) only differ on A for B and ρ for λ, the solution implies a low ρ and a high λ. Table 9 below shows α_M to be very high, and much greater than β_M. This is further evidence that the assumption of equal mobility across two generations is rejected by the data. Moreover, this reconciles the high values for β_M obtained in Prop. 1 which uses data from 1900-1920.

In Table 7 we show the results for identification with information on maternal uncles. Prop. 5, which assumes that the nuisance parameters are all identical and the assortative mating parameter is identical across generations to identify using sons, fathers and maternal grandfathers only, delivers reasonable estimates for the main parameters of interest, although the estimate of the nuisance parameter a is large. Prop. 6, using additionally maternal uncles and without these restrictions, delivers implausibly large estimates for the assortative mating parameter in the grandparents generation. This is, similarly as before, due to the relatively high value for $\mathbb{E}\left[MGF_i MU_i\right] = 0.4592$, the solution to the system of equations, given the estimates for β_F and β_M implies a very high λ.[21] This again suggests a joint rejection of the assumptions of the model as specified, in particular the assumption of constant effects across

[21]See equation (28) in Appendix (A).

Table 7: Identification from Maternal Uncles

Parameter	Estimate			
	Prop. 5	Prop. 6	Prop. 7	Prop. 8
a	0.808			
	(0.026)			
β_F	0.307	0.303	0.249	0.166
	(0.012)	(0.017)	(0.028)	(0.025)
β_M	0.246	0.151	0.274	0.479
	(0.025)	(0.032)	(0.061)	(0.075)
γ		0.435	0.191	
		(0.015)	(0.023)	
λ		1.000	0.855	
		(0.163)	(0.115)	
ρ	0.248	0.417	0.414	0.426
	(0.007)	(0.012)	(0.013)	(0.012)

Notes: Estimates from identification based on maternal uncles. Proposition 5 assumes that $\lambda = \rho$ and additionally equality in the nuisance parameters. Proposition 6 and 7 require no assumptions. Proposition 8 assumes that $\lambda = \rho$ and additionally that $\gamma \equiv \mathbb{E}\left[\epsilon_i^M \epsilon_i^{MU}\right] = 0$.

generations.

Prop. 7, which uses instead paternal uncles, yields the same issue of a high λ, using equations (7) and (8), as in Prop. 4. Table 9 shows results relaxing the assumption of constant effects across generations.

Prop. 8 shows a substantially higher effect of mothers than fathers. Note that identification here comes from the assumption of no household fixed effects ($\gamma = 0$), which may be implausible for a variety of reasons and contrasts with our estimates for γ from Prop. 5 and 6.

In Table 8, we show the results from Prop. 9 and Prop. 10 allowing for differences in the status transmission parameters based on the gender of the child. The first, assuming joint equality of the nuisance parameters ($a = b = c = d$), results in reasonable estimates for the assortative mating parameters and high estimates for the maternal contribution to assortative mating. The second, which assumes that the cross moments are zero ($b = d = 0$), delivers very implausible estimates for most parameters, leading us to conclude that in our data, it is unlikely that assortative mating can be described by a process without grandparent

Table 8: Identification Allowing Heterogeneous Effects by Child Gender

Parameter	Estimate	
	Prop. 9	Prop. 10
β_F^D	0.093	0.000
	(0.012)	(5.027)
β_F^S	0.281	0.474
	(0.011)	(0.013)
β_M^D	0.209	1.000
	(0.032)	(5.308)
β_M^S	0.649	0.013
	(0.019)	(0.051)
λ	0.313	1.000
	(0.015)	(4.795)
ρ	0.135	0.005
	(0.012)	(1.089)

Notes: Estimates from propositions allowing for heterogeneous effects by gender. Proposition 9 assumes equality in the nuisance parameters. Proposition 10 assumes that $b = d = 0$ and additionally that $a = c$.

cross-moments.

In Table 9, we show the results from allowing the status transmission parameters to differ across generations. Recall that sons and fathers are less highly correlated in our sample than fathers and grandfathers ($A < B$). Hence, it is reasonable to expect that the structural estimates may be more accurate and stable when allowing for different effects across generations. The estimates for (β_F, β_M, ρ) in Prop. 11, Prop. 13 and Prop. 14, are identical to each other and identical to the estimates in Prop. 4. This is because in all four cases we are using the same moments in the data to identify these three parameters (see Appendix A for details). Prop. 13 then shows the much higher mobility estimates for the earlier generation, attributing the most of the high value of B to a very high effect of mothers α_M. In Propositions 11 and Prop. 14, we cannot individually point-identify α_F and α_M, but we can identify a combined mobility estimate in each case, α and $\tilde{\alpha}$ respectively. As expected, because sons and fathers are less highly correlated in our sample than fathers and grandfathers ($A < B$), the combined estimates–α or $\tilde{\alpha}$–are smaller than $\tilde{\beta} = \beta_F + \rho\beta_M$.

Table 9: Identification Allowing Heterogeneous Effects by Generation

Parameter	Estimate			
	Prop. 11	Prop. 12	Prop. 13	Prop. 14
α	0.465			
	(0.015)			
α_F		0.568	0.138	
		(0.013)	(0.295)	
α_M		0.098	0.792	
		(1.094)	(0.299)	
$\tilde{\alpha}$				0.465
				(0.007)
β_F	0.272	0.346	0.272	0.272
	(0.080)	(0.022)	(0.080)	(0.080)
β_M	0.220	0.169	0.220	0.220
	(0.081)	(0.024)	(0.081)	(0.081)
λ		0.010		0.041
		(0.125)		(0.021)
ρ	0.437	0.157	0.437	0.437
	(0.207)	(0.105)	(0.207)	(0.207)

Notes: Estimates from propositions allowing for heterogeneous effects of parents across different generations. Proposition 11 assumes $b = d$. Proposition 12 assumes that $b = d = 0$ and $a = c$. Proposition 13 assumes that $\lambda = \rho$ and $a = b = d$. Prop. 14 assumes that $b = d = \sqrt{ac}$.

6 Conclusions

In this article we show that women matters for social mobility, both through their direct effect on their children and indirectly through the high correlation on status between spouses. Our framework could be easily extended in a number of dimensions, three of which we discuss briefly here:

1. The current model does not allow for direct grandparent effects. The effects are indirect through the correlation between grandparents and how grandparents affect the parents, and how parents affect the children. Our techniques could be used to add direct grandparents' effects. This could be done using data on great-grandparents. Similar to our results using data on two generations (marriage data), we would only need data on three generations (Son, Father, MGF, PGF) to estimate direct grandparent effects.

2. In the sixth moment of our main specification (equation (12)) we assume that the error terms in the equation for the father and the mother are uncorrelated, i.e., we are assuming that mating is done only on observables (income). We do not really use this assumption in most of our results, which means that we could relax it and estimate this correlation. This will give us a measure of mating on unobservables. This estimate would be similar in spirit to the household fixed effects γ we estimate using maternal uncles.

3. The literature usually assumes that all nuisance parameters are zero. Our results show how this is rejected by the data. Moreover, we see how even seemingly innocuous assumptions such as $b = d = 0$ (no correlations between PGM and MGF and between PGF and MGM), could lead to extreme and implausible estimates (Prop. 10).

Just in its current form, however, our framework delivers results that shed new light on how different assumptions on the parameters of an econometric model of mobility, especially the often unstated assumptions on nuisance parameters, can affect mobility estimates. We find strong effects on mobility from the maternal side of the family line even without directly observing female labor market outcomes. We observe underlying patterns of assortative mating that generate non-linear links across generations. And we identify a number of previous studies that embody assumptions that we find are rejected by our data and methods.

Our model is econometric, not economic, in nature. We are agnostic here on how households choose investment in their children, or how parents bargain over their resources. This is intentional, as we envision our methodology as one that can be applied to a large class of models of the intergenerational transmission of human capital and intrahousehold bargaining. Finally, our current results are for the whole US, but they could be estimated for different regions and different population subgroups and of course to other countries and periods. We hope that this new technique will be applied to other settings and thereby help us paint a more nuanced picture of social mobility, especially the roles of women and prior generations.

References

Abramitzky, Ran, Leah Boustan, Elisa Jácome, and Santiago Pérez, "Intergenerational mobility of immigrants in the United States over two centuries," *American Economic Review*, 2021, *111* (2), 580–608.

_ , _ , Katherine Eriksson, James Feigenbaum, and Santiago Pérez, "Automated linking of historical data," *Journal of Economic Literature*, 2021, *59* (3), 865–918.

Althoff, Lukas, Harriet Brookes Gray, and Hugo Reichardt, "Intergenerational Mobility and Assortative Mating," 2022.

Arellano, Manuel and Stephen Bond, "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations," *Review of Economic Studies*, April 1991, *58* (2), 277–297.

Bailey, Martha, Connor Cole, and Catherine Massey, "Simple strategies for improving inference with linked data: A case study of the 1850–1930 IPUMS linked representative historical samples," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 2020, *53* (2), 80–93.

Becker, Gary S and Nigel Tomes, "An equilibrium theory of the distribution of income and intergenerational mobility," *Journal of political Economy*, 1979, *87* (6), 1153–1189.

_ and _ , "Human capital and the rise and fall of families," *Journal of labor economics*, 1986, *4* (3, Part 2), S1–S39.

Chadwick, Laura and Gary Solon, "Intergenerational income mobility among daughters," *American Economic Review*, 2002, *92* (1), 335–344.

Clark, Gregory, "The inheritance of social status: England, 1600 to 2022," *PNAS*, 2023, *120* (27).

_ and Neil Cummins, "Assortative Mating and the Industrial Revolution: England, 1754-2021," *CEPR*, 2022, (17074).

_ , _ , and Matthew Curtis, "The Mismeasure of Man: Why IntergenerationalOccupational Mobility is Much Lower thanConventionally Measured, England, 1800-2021," *CEPR workin papers*, 2022, (17346).

Craig, Jacqueline, Katherine Eriksson, and Gregory T. Niemesh, "Marriage and the Intergenerational Mobility of Women: Evidence from Marriage Certificates 1850-1910," *Working Paper*, 2019.

Curtis, Matthew, "The her in inheritance: how marriage matching has always mattered, Quebec 1800-1970," 2022, (2022-38).

Durlauf, Steven N, Andros Kourtellos, and Chih Ming Tan, "The great gatsby curve," *Annual Review of Economics*, 2022, *14*, 571–605.

Espín-Sánchez, José-Antonio, Salvador Gil-Guirado, and Chris Vickers, "La "Doña" è Mobile: The Role of Women in Social Mobility in a Pre-Modern Economy," *The Journal of Economic History*, 2022, *82* (1), 1–41.

Ferrie, Joseph P., "A new sample of males linked from the public use microdata sample of the 1850 US federal census of population to the 1860 US federal census manuscript schedules," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 1996, *29* (4), 141–156.

Hansen, Lars Peter, "Large sample properties of generalized method of moments estimators," *Econometrica*, 1982, *50* (4), 1029–1054.

Jácome, Elisa, Ilyana Kuziemko, and Suresh Naidu, "Mobility for all: Representative intergenerational mobility estimates over the 20th century," Technical Report, National Bureau of Economic Research 2021.

Lam, David and Robert F Schoeni, "Family ties and labor markets in the United States and Brazil," *Journal of Human Resources*, 1994, pp. 1235–1258.

Long, Jason and Joseph Ferrie, "Intergenerational Occupational Mobility in Great Britain and the United States Since 1850," *American Economic Review*, jun 2013, *103* (4), 1109–1137.

Mazumder, Bhashkar, "Fortunate sons: New estimates of intergenerational mobility in the United States using social security earnings data," *Review of Economics and Statistics*, 2005, *87* (2), 235–255.

— **and Miguel Acosta**, "Using occupation to measure intergenerational mobility," *The ANNALS of the American Academy of Political and Social Science*, 2015, *657* (1), 174–193.

Olivetti, Claudia and M. Daniele Paserman, "In the Name of the Son (and the Daughter): Intergenerational Mobility in the United States, 1850 –1940," *American Economic Review*, aug 2015, *105* (8), 2695–2724.

Ruggles, Steven, Catherine A Fitch, and Evan Roberts, "Historical census record linkage," *Annual review of sociology*, 2018, *44*, 19–37.

Saavedra, Martin and Tate Twinam, "A machine learning approach to improving occupational income scores," *Explorations in Economic History*, 2020, *75*, 101304.

Sobek, Matthew, "The comparability of occupations and the generation of income scores," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 1995, *28* (1), 47–51.

Strauss, Frederick and Louis Hyman Bean, *Gross farm income and indices of farm production and prices in the United States, 1869-1937* number 703, US Department of Agriculture, 1940.

U.S. Census Bureau, *Historical Statistics of the United States*, GPO, 1960.

U.S. National Archives and Records Administration (NARA) , *Numerical Identification (NUMIDENT) Files. Record Group 47* U.S. National Archives and Records Administration 2018.

A Proofs of propositions in the paper

A.1 Model Derivation

We now use equations (4), (5) and (6) and exclusion restrictions 1, 3, 5 and 7 above. We get the GMM analog:

$$\begin{aligned} \mathbb{E}\left[F_i\left(S_i-\beta_F F_i-\beta_M M_i\right)\right] &= && 0 \\ \mathbb{E}\left[PGF_i\left(F_i-\beta_F PGF_i-\beta_M PGM_i\right)\right] &= && 0 \\ \mathbb{E}\left[MGF_i\left(M_i-\beta_F MGF_i-\beta_M MGM_i\right)\right] &= && 0 \end{aligned} \tag{20}$$

$$\begin{aligned} \mathbb{E}\left[F_i\left(S_i-\beta_F F_i\right)\right] &= && \mathbb{E}\left[F_i\left(\beta_M M_i\right)\right] \\ \mathbb{E}\left[PGF_i\left(F_i-\beta_F PGF_i\right)\right] &= && \mathbb{E}\left[PGF_i\left(\beta_M PGM_i\right)\right] \\ \mathbb{E}\left[MGF_i\left(\frac{S_i-\beta_F F_i}{\beta_M}\right)\right]-\beta_F &= && \mathbb{E}\left[MGF_i\left(\beta_M MGM_i\right)\right] \end{aligned} \tag{21}$$

we now use that all variables are normalized, i.e., $\mathbb{E}\left[X_i X_i\right]=1$ and get

$$\begin{aligned} \mathbb{E}\left[F_i S_i\right]-\beta_F &= & \beta_M \mathbb{E}\left[F_i M_i\right] &= & \beta_M \rho \\ \mathbb{E}\left[PGF_i F_i\right]-\beta_F &= & \beta_M \mathbb{E}\left[PGF_i PGM_i\right] &= & \beta_M \lambda \\ \frac{\mathbb{E}\left[MGF_i S_i\right]}{\beta_M}-\frac{\beta_F \mathbb{E}\left[MGF_i F_i\right]}{\beta_M}-\beta_F &= & \beta_M \mathbb{E}\left[MGF_i MGM_i\right] &= & \beta_M \lambda \end{aligned} \tag{22}$$

Notice that the first two equations in the system in (21) come directly from equations (4) and (5), while in the last equation we have substituted equation (4) in (6). The dependent variable in equation (6) (M_i) is unobservable but could be expressed as observable variables and estimated parameters using equation (4). Also, instead of writing the equations equal to zero we have written them equaling them to the bias due to unobservable variables to emphasize both that equations (4), (5) and (6) have unobservables and that the bias is the same in equations (4), (5) and (6) when $\rho=\lambda$. The system in (22) reflects a system of three independent equations to estimate four parameters. The expectations in the system are the product of just two variables, which means that all the parameters can be expresses analytically as a function of elements that are easily computed from the data.

We now show the estimator for β_F cannot be identified using patrilineal data only. In particular, if we solve the first two equations, those using only patrilineal data we get

$$\mathbb{E}\left[F_i S_i\right]=\beta_F+\rho\beta_M$$

$$\mathbb{E}\left[PGF_i F_i\right]=\beta_F+\lambda\beta_M$$

This result show that the first two equations are just two ways to estimate the direct effect of a father on his son β_F together with the indirect effect or bias $\rho\beta_M$ (or $\lambda\beta_M$ for the first generation). However, no parameter can be identified using only patrilineal data in this simple setting. If we assume $\lambda = \rho$, we do not get an extra identifying assumption. We would just impose $\mathbb{E}[F_i S_i] = \mathbb{E}[PGF_i F_i]$, which might not be true in the data and we will make the first two moment equations linearly dependent. That is, this assumption reduces the number of parameters by one but also reduces the number of independent equations by one.

We now show how, using all the exclusion restrictions and matrilineal male data, we can get six independent identifying restrictions, to estimate seven parameters. That is, matrilineal male data is the key to get any identification. We would still need an extra assumption to get point identification.

In addition to using the three "standard" moments, we can make a more efficient use of the data that we have. In the previous equations there are only four empirical relations from the data, e.g., A, B, C and D. Notice, however, that there are six empirical relations that we can get from the data, i.e., there are six possible ways to get cross products on the four variables that we have: (S_i, F_i, PGF_i, MGF_i). Thus we are not using information on the other two empirical relations that we have from the data, e.g., a and E. Moreover, if the model specified above in equations (4), (5) and (6) is correct, there are other exclusion restrictions that we can use.

A.2 Set Identification

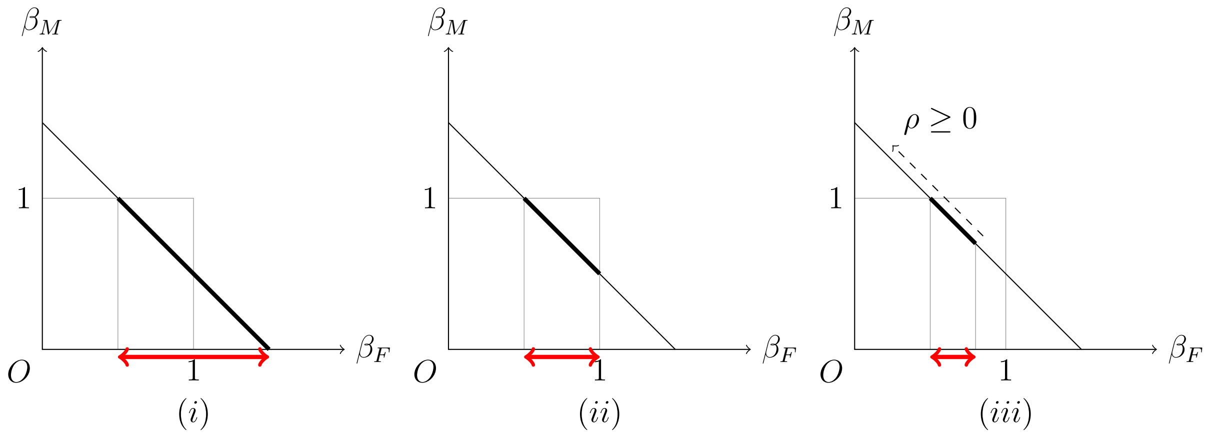
Lemma. *1 If we know (β_F, β_M, ρ), then equations* (11)*,* (10) *and* (12) *identify all nuisance parameters* (b, c, d)*.*

Proof. First, notice that equations (11), (10) and (12) are independent. Second, λ does not appear in any of them. If we know (β_F, β_M, ρ) and the relevant empirical moments from the data, we have a system with three independent equations and three unknowns (b, c, d).

□

Figure 5 below shows set identification restrictions when the equation $\beta_M = \frac{C}{B} - \beta_F \frac{D}{B}$ is partially out of the unit box. In that case, we can get some tighter identification set if we impose that the structural parameters should be positive but smaller than 1. We show the identified set with the following cumulative assumptions on the bottom panels in Figure (5).
i. If we assume $0 \leq \beta_M \leq 1$.
ii. If we also assume $0 \leq \beta_F \leq 1$.
iii. If we also assume $0 \leq \rho \leq 1$.
The width of the identified set gets smaller as $\mathbb{E}[MGF_i F_i]$ gets larger relative to $\mathbb{E}[S_i F_i]$

Figure 5: Set Identification.



A.3 Identification Using Two Generations

Proposition. ***1.*** *Suppose F_i, PGF_i, and MGF_i are observed. If we assume $a = b = c = d$ and $\lambda = \rho$, then (β_F, β_M, ρ) is point identified.*

Proof. With the stated assumptions we can write the system of equations as

$$B = \beta_F + \rho\beta_M$$

$$D = (\beta_F + \beta_M)\, a$$

$$\rho = (\beta_F + \beta_M)^2\, a$$

With the second equation, we can directly get the sum of the mobility effects as a function of the data $(\beta_F + \beta_M) = D/a$. Using this and the third equation, we get

$$\rho = D^2/a.$$

We can now solve for β_F as a function of β_M and data using the second equation. Plugging in this and the formula for ρ in the first equation we get

$$B = \frac{D}{a} - \beta_M + \frac{D^2}{a}\beta_M$$

rearranging we get

$$\beta_M = -\frac{aB - D}{a - D^2}$$

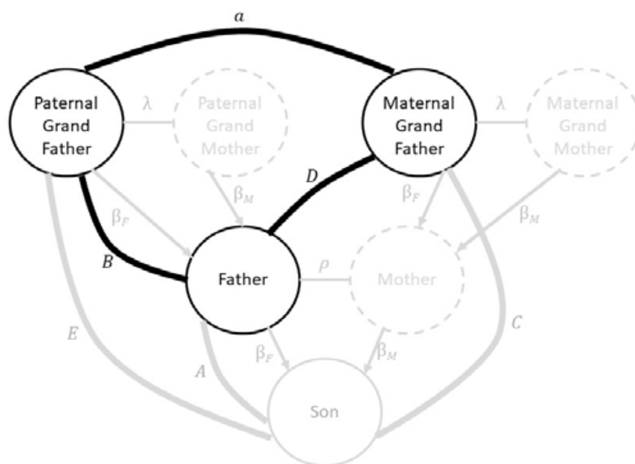
$$\beta_F = \frac{D}{a} + \frac{aB - D}{a - D^2}$$

□

Proposition. ***2.*** *Suppose F_i, PGF_i, and MGF_i are observed. If we assume $\beta_M = 0$, then (β_F, ρ) is point identified.*

Proof. With the stated assumptions we can write the system of equations as

Figure 6: Empirical Relations in Two-generation Records.



Notes: The solid circles represent individuals (males) with observed income while the dashed circles represent individuals (females) with unobserved income. The thick black solid thick lines represents all three possible pair-wise relations (empirical relations) between the three male members of the family tree.

$$B = \beta_F$$
$$D = \beta_F a$$
$$\rho = \beta_F^2 a$$

This is a system with two unknowns and three equations, so it is overidentified. Similar to Prop. 1, we can use the second and third moment and we have $\rho = D^2/a$. The second moment then gives us $\beta_F = D/a$. The system is overidentified, and the first equation implies $\beta_F = B$. If in the data we have $B \neq D/a$. We should reject the assumption $\beta_M = 0$.

□

Proposition. ***3.*** *Suppose F_i, PGF_i, and MGF_i are observed. If we assume $a = b = c = d$ and $\beta_F = \beta_M$, then (β_F, λ, ρ) is point identified.*

Proof. With the stated assumptions we can write the system of equations as

$$B = \beta_F (1 + \lambda)$$
$$D = 2\beta_F a$$
$$\rho = (2\beta_F)^2 a$$

This is a system with three unknowns (β_F, λ, ρ) and three independent equations, so it is identified. Similar to Prop. 1, we can use the second and third moment and we have $\rho = D^2/a$. From the second equation we get $\beta_F = D/2a$. We can then plug this in the first equation to get $\lambda = B/\beta_F - 1 = 2aB/D - 1$.

□

A.4 Identification Using Three Generations

Proposition. *4. Suppose S_i, F_i, PGF_i and MGF_i are observed. If we assume $b = d$, then $(\beta_F, \beta_M, \lambda, \rho)$ is point identified.*

Proof. We now use three moments to get identification
- $\mathbb{E}\left[PGF_i\epsilon_i^S\right] = 0$ and $\mathbb{E}\left[PGF_i\epsilon_i^M\right] = 0$ - Multiplying equations (4) and (6) by PGF_i and then substituting the definition of M_i,and taking expectations, we get:

$$\mathbb{E}\left[PGF_iS_i\right] - \beta_F\mathbb{E}\left[PGF_iF_i\right] = \beta_M\left(\beta_F a + \beta_M b\right) \tag{23}$$

- $\mathbb{E}\left[MGF_i\epsilon_i^F\right] = 0$ - Multiplying the second equation by MGF_i and taking expectations and using $b = d$, we get:

$$\mathbb{E}\left[MGF_iF_i\right] = \beta_F a + \beta_M b \tag{24}$$

By imposing $b = d$, we can now identify the model. First, we can take equation (24) and insert it in equation (23). This way we get

$$\mathbb{E}\left[PGF_iS_i\right] - \beta_F\mathbb{E}\left[PGF_iF_i\right] = \beta_M\mathbb{E}\left[MGF_iF_i\right] \tag{25}$$

Now we have four equations ((7), (8), (9) and (25)) and four unknowns $(\beta_F, \beta_M, \lambda, \rho)$. The solution to this system is

$$\begin{array}{rcl} \beta_F & = & \frac{BE-CD}{B^2-D^2} \\ \beta_M & = & \frac{BC-ED}{B^2-D^2} \\ \rho & = & \frac{A\left(B^2-D^2\right)-(BE-CD)}{BC-ED} \\ \lambda & = & \frac{B\left(B^2-D^2\right)-(BE-CD)}{BC-ED} \end{array}$$

Notice that equations (8) and (9) combine create an equation on β_F and β_M only. This together with equation (25) above, creates a system with two equations and two unknowns β_F and β_M. We can then use equation (7) to solve for ρ and equation (8), to solve for λ.

□

A.5 Identification Using Maternal Uncles

Using similar assumptions as above can get the following moments using MU_i

$$\mathbb{E}\left[MU_iS_i\right] = \beta_F\mathbb{E}\left[F_iMU_i\right] + \beta_M\left(\beta_F^2 + \beta_M^2 + 2\lambda\beta_F\beta_M + \gamma\right) \tag{26}$$

$$\mathbb{E}\left[F_iMU_i\right] = \rho \tag{27}$$

$$\mathbb{E}\left[MGF_iMU_i\right] = \beta_F + \lambda\beta_M \tag{28}$$

$$\mathbb{E}\left[PGF_iMU_i\right] = \beta_F a + \beta_M d \tag{29}$$

where $\gamma \equiv \mathbb{E}\left[\epsilon_i^M\epsilon_i^{MU}\right]$ measures household fixed effects.

With the full set of five variables we can observe ten empirical moments: the six empirical relations discussed above in Table 2 (A, B, C, D, E, a) and the new four moments with maternal uncles $(\mathbb{E}[F_i MU_i], \mathbb{E}[MGF_i MU_i], \mathbb{E}[MU_i S_i], \mathbb{E}[PGF_i MU_i])$; and compute ten equations: the original six moments plus the new four above.

Proposition. *5.* *Suppose S_i, F_i and MGF_i are observed. If we assume $a = b = c = d$ and $\lambda = \rho$, then $(\beta_F, \beta_M, \rho, a)$ is point identified.*

Proof. First, we impose $a = b = c = d$ and $\rho = \lambda$ in the system of equations above and we get

$$A = \beta_F + \rho\beta_M$$

$$C = \beta_F D + \beta_M (\beta_F + \rho\beta_M)$$

$$D = (\beta_F + \beta_M) a$$

$$\rho = (\beta_F + \beta_M)^2 a$$

Notice that now, because we do not observe PGF, we do not observe B, E and a. We can take the last two equations and write $\rho = D(\beta_F + \beta_M)$. This equation and the first two form a system of three independent equations and three unknowns (β_F, β_M, ρ). Thus, we can identify all three structural parameters. If the reader is interested, we can then use the third or second equations above to estimate a. The reason we can get four independent equations here with only three empirical moments (A, C, D) is that the fourth equation is quadratic, non linear, in the parameters.

□

Proposition. *6.* *Suppose S_i, F_i, MU_i and MGF_i are observed. Then $(\beta_F, \beta_M, \lambda, \rho, \gamma)$ is point identified.*

Proof. The moments that do not use PGF_i are equations (7), (9) and (10). Equation (10) includes two nuisance parameters that are unobserved here (a, b). There are two unobservable nuisance parameters in the same equation so we cannot identify them. Since we are not interested in the nuisance parameters, we do not need to use this equation.
In addition to equations (7), (9), we can use equations (26), (27) and (28). This is a system of five independent equations and five unknowns. We now show that the equations are indeed independent and how to solve the system.
We can substitute ρ, which is directly observable in equation (27) into equation (7). We can also substitute $\beta_F + \lambda\beta_M$ from equation (28) into the equation (9). Equations (7) and (8) become

$$A = \beta_F + \beta_M \mathbb{E}[F_i MU_i]$$

$$C = \beta_F D + \beta_M \mathbb{E}[MGF_i MU_i]$$

This is a system with two equations and two unknowns β_F and β_M, so they are both identified. With these two parameters we can go to equation (28) and solve for λ. We can use equation (26) and identify the household fixed effects γ.

□

Proposition. *7.* *Suppose S_i, F_i, MU_i and PGF_i are observed. Then $(\beta_F, \beta_M, \lambda, \rho, \gamma)$ is point identified.*

Proof. The moments that do not use MGF_i are equations (7), (8) and (11). Equation (9) includes two nuisance parameters that are unobserved here (a, d). However, this would not be a problem as we see below.
In addition to equations (7), (8) and (11), we can use equations (26), (27) and (29). These six equations form a system of five independent equations, when we substitute equation (29) into the equation (11). Thus, this is a system of five independent equations and five unknowns. We now show that the equations are indeed independent and how to solve the system.
We can substitute ρ, which is directly observable in equation (27) into equation (7). We can also substitute $\beta_F a + \beta_M d$ from equation (29) into the equation (11). Equations (7) and (11) become

$$A = \beta_F + \beta_M \mathbb{E}\left[F_i MU_i\right]$$

$$E = \beta_F B + \beta_M \mathbb{E}\left[PGF_i MU_i\right]$$

This is a system with two equations and two unknowns β_F and β_M, so they are both identified. With these two parameters we can go to equation (8) and solve for λ. We can use equation (26) and identify the household fixed effects γ.

□

Proposition. *8.* *Suppose S_i, F_i, MU_i are observed and $\lambda = \rho$ and $\gamma = 0$, then (β_F, β_M, ρ) is point identified.*

Proof. We can generate three moments (26), (27) and (28). Assuming $\lambda = \rho$ and $\gamma = 0$ we get

$$A = \beta_F + \rho\beta_M$$

$$\mathbb{E}\left[F_i MU_i\right] = \rho$$

$$\mathbb{E}\left[MU_i S_i\right] = \beta_F \mathbb{E}\left[F_i MU_i\right] + \beta_M \left(\beta_F^2 + \beta_M^2 + 2\rho\beta_F\beta_M\right)$$

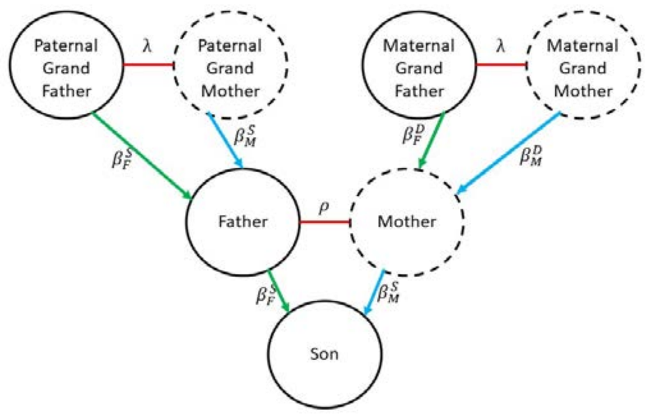
This is a system with three equations and three unknowns (β_F, β_M, ρ). The second equation identifies ρ directly. We can then use the first equation to write β_F as a function of β_M and substitute that into the last equation. Then we only need to solve for a cubic equation on β_M.

□

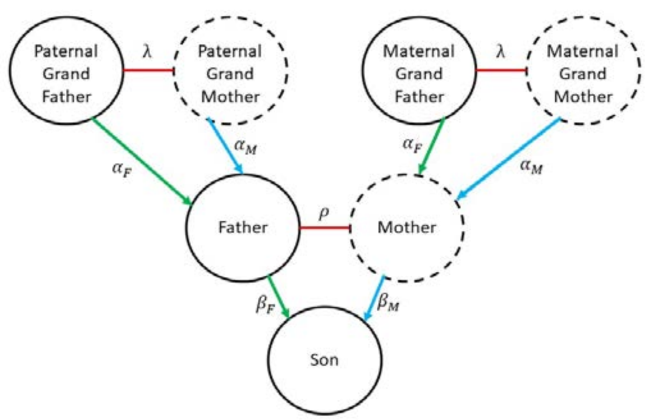
A.6 Gendered Effects

Figure 7: Family Trees for Gendered and Generational Effects.

A. Gendered Effects.



B. Generational Effects.



Notes: The horizontal lines in red represent the degree of assortative matching; the vertical relations in green (arrows) represent the masculine effect on mobility; the vertical relations in blue (arrows) represent the feminine relations on mobility. The solid circles represent individuals (males) with observed income while the dashed circles represent individuals (females) with unobserved income.

Following the same steps as before, we get the following set of moments.

$$A = \beta_F^S + \rho\beta_M^S \tag{30}$$

$$B = \beta_F^S + \lambda \beta_M^S \tag{31}$$

$$C = \beta_F^S D + \beta_M^S \left(\beta_F^D + \lambda \beta_M^D\right) \tag{32}$$

$$D = \beta_F^S a + \beta_M^S b \tag{33}$$

$$E = \beta_F^S B + \beta_M^S \left(\beta_F^D a + \beta_M^D d\right) \tag{34}$$

$$\rho = \beta_F^S \beta_F^D a + \beta_F^D \beta_M^S b + \beta_F^S \beta_M^D d + \beta_M^S \beta_M^D c \tag{35}$$

The system above shows six equations and nine parameters: six structural parameters $\left(\beta_F^S, \beta_M^S, \beta_F^D, \beta_M^D, \lambda, \rho\right)$ and three nuisance parameters (b, c, d). Thus the system is not point identified. To get point identification we need at least three independent restrictions in the parameters. Proposition 9 below shows a set of sufficient conditions for point identification of gendered effects in our model.

Proposition. *9.* *Suppose S_i, F_i, PGF_i and MGF_i are observed. If we assume $a = b = c = d$, then $\left(\beta_F^S, \beta_M^S, \beta_F^D, \beta_M^D, \lambda, \rho\right)$ is point identified.*

Proof. First, we impose $a = b = c = d$ in the system of equations above. Notice that a is observable. Second, take the system of six equations above and notice that: ρ only appears in equations (30) and (35); and λ only appears in equations (31) and (32). We can take equation (30), solve for ρ and substitute it in equation (35) and take equation (31), solve for λ and substitute in equation (32). We get

$$\frac{A - \beta_F^S}{\beta_M^S} = \left(\beta_F^S + \beta_M^S\right)\left(\beta_F^D + \beta_M^D\right) a$$

$$C = \beta_F^S D + \beta_M^S \beta_F^D + \left(B - \beta_F^S\right) \beta_M^D$$

The two equations above, together with equations (33) and (34) form a system of four independent equations with four unknowns $\left(\beta_F^S, \beta_M^S, \beta_F^D, \beta_M^D\right)$. Once we solve this system, we can just use equation (30) to solve for ρ and equation (31) to solve for λ.

□

Proposition. *10.* *Suppose S_i, F_i, PGF_i and MGF_i are observed. If we assume $b = d = 0$ and $a = c$, then $\left(\beta_F^S, \beta_M^S, \beta_F^D, \beta_M^D, \lambda, \rho\right)$ is point identified.*

Proof. First, we impose $b = d = 0$ and $a = c$ in the system of equations above. Notice that a is observable. Second, take the system of six equations above and notice that: ρ only appears in equations (30) and (35); and λ only appears in equations (31) and (32). We can take equation (30), solve for ρ and substitute it in equation (35) and take equation (31), solve for λ and substitute in equation (32). We get

$$\frac{A - \beta_F^S}{\beta_M^S} = \left(\beta_F^S \beta_F^D + \beta_M^S \beta_M^D\right) a$$

$$C = \beta_F^S D + \beta_M^S \beta_F^D + \left(B - \beta_F^S\right) \beta_M^D$$

The two equations above, together with equations (33) and (34) form a system of four independent equations with four unknowns $\left(\beta_F^S, \beta_M^S, \beta_F^D, \beta_M^D\right)$. Once we solve this system, we can just use equation (30) to solve for ρ and equation (31) to solve for λ.

□

A.7 Generational Effects

Following the same steps as before, we get the following set of moments.

$$A = \beta_F + \rho\beta_M \tag{36}$$

$$B = \alpha_F + \lambda\alpha_M \tag{37}$$

$$C = \beta_F D + \beta_M \left(\alpha_F + \lambda\alpha_M\right) \tag{38}$$

$$D = \alpha_F a + \alpha_M b \tag{39}$$

$$E = \beta_F B + \beta_M \left(\alpha_F a + \alpha_M d\right) \tag{40}$$

$$\rho = \alpha_F^2 a + \alpha_F\alpha_M \left(b + d\right) + \alpha_M^2 c \tag{41}$$

Proposition. ***11.*** *Suppose S_i, F_i, PGF_i and MGF_i are observed. If we assume $b = d$, then (β_F, β_M, ρ) and α (with $\alpha \equiv \alpha_F + \alpha_M$) is point identified. However, λ, α_F and α_M are not point identified.*

Proof. First, we impose $b = d$ in the system of equations above. Notice that a is observable. We can substitute equation (37) into equation (38) and substitute equation (39) into equation (40). We have now a system with two equations and two unknowns (β_F, β_M).

$$C = \beta_F D + \beta_M B$$

$$E = \beta_F B + \beta_M D$$

We can go to equation (36) and identify ρ. This is as far as we can get. In equations (39), (40) and (41) we can only identify α but not each component. In equations (37) and (38) we can only identify $(\alpha_F + \lambda\alpha_M)$. Thus we have two independent equations, say (37) and (39), for three unknowns λ, α_F and α_M.

□

Corollary. ***1.*** *Suppose S_i, F_i, PGF_i and MGF_i are observed. If we assume $a = b = d$ and $\alpha_F = \alpha_M$, then $(\beta_F, \beta_M, \alpha_F, \lambda, \rho)$ is point identified.*

Proof. We can follow the same steps as in Proposition 11 to get identification on (β_F, β_M, ρ). Unlike before, we have now an extra assumption $\alpha_F = \alpha_M$. Moreover, we are imposing $a = b$ now. We can use equation (39) and get $D = 2\alpha_F a$ to identify α_F. We can then use equation (37) to identify λ.

□

Proposition. ***12.*** *Suppose S_i, F_i, PGF_i and MGF_i are observed. If we assume $b = d = 0$ and $a = c$, then $\Theta \equiv (\beta_F, \beta_M, \alpha_F, \alpha_M, \lambda, \rho)$ is point identified.*

Proof. First, we impose $b = d = 0$ and $a = c$ in the system of equations above. Notice that a is observable. Second, take the system of six equations above and notice that: ρ only appears in equations (36) and (41); and λ only appears in equations (37) and (38). We can take equation (36), solve for ρ and substitute it in equation (41) and take equation (37), solve for λ and substitute in equation (38). We get

$$\frac{A - \beta_F}{\beta_M} = \left(\alpha_F^2 + \alpha_M^2\right) a$$

$$C = \beta_F D + \beta_M B$$

With the assumption here, equation (39) identifies α_F directly, i.e, $\alpha_F = D/a$. If we substitute this in equation (40), we get

$$E = \beta_F B + \beta_M D$$

This, together with the second equation above forms a system with two equations and two unknowns and identifies β_F and β_M. Using the values for α_F, β_F and β_M, together with the first equation above, we can identify α_M. Finally, we can use the mobility parameters α_F, α_M, β_F and β_M and using equations (36) and (37), we get the mating parameters ρ and λ.

□

Proposition. *13.* *Suppose S_i, F_i, PGF_i and MGF_i are observed. If we assume $a = b = d$ and $\rho = \lambda$, then $(\beta_F, \beta_M, \alpha_F, \alpha_M, \rho)$, is point identified.*

Proof. First, we impose $a = b = d$ in the system of equations above. Notice that a is observable. Second, take the system of six equations above and notice that λ only appears in equations (37) and (38). We can take equation (37), solve for λ and substitute in equation (38). We get

$$C = \beta_F D + \beta_M B$$

We can take equation (39) and substitute in equation (40) to get

$$E = \beta_F B + \beta_M D$$

The previous two equations form a system of two equations and two unknowns (β_F, β_M). In fact, this is the same system depicted in Figure 4 (right). Now we can just use equation (36) to solve for ρ.

$$A = \beta_F + \rho\beta_M$$

Now imposing $\rho = \lambda$ means that we can use equations (37) and (39) for form the system below.

$$B = \alpha_F + \rho\alpha_M$$

$$D = (\alpha_F + \alpha_M)\, a$$

The previous two equations form a system of two equations and two unknowns (α_F, α_M).

□

Proposition. *14.* *Suppose S_i, F_i, PGF_i and MGF_i are observed. If we assume $b = d = \sqrt{ac}$, then (β_F, β_M, ρ) and $\tilde{\alpha}$ (with $\tilde{\alpha} \equiv \alpha_F\sqrt{a} + \alpha_M\sqrt{c}$) is point identified. However, α_F, α_M, and λ are not point identified.*

Proof. We can substitute equation (37) into equation (38) and get

$$C = \beta_F D + \beta_M B$$

Using $b = d$ we can substitute equation (39) into equation (40) and get

$$E = \beta_F B + \beta_M D$$

These two equations above form a system with two equations and two unknowns β_F and β_M. Solving for β_F and β_M and using equation (36) we can solve for ρ.
If we would write $\rho = \lambda$ we can write equation (37) as

$$B = \alpha_F + \rho\alpha_M$$

This gives us one equation to identify α_F and α_M. Notice, however, that α_F and α_M appear in all other equations as $\tilde{\alpha} \equiv \alpha_F\sqrt{a} + \alpha_M\sqrt{c}$.[22] This means we have only one independent equation to estimate $\tilde{\alpha}$ and equation (37) that relates α_F and α_M. Without further assumptions on c we do not get point identification.

□

Corollary. *2. Suppose S_i, F_i, PGF_i and MGF_i are observed. If we assume $b = d = \sqrt{ac}$ and $\alpha_F = \alpha_M$, then (β_F, β_M, ρ) and $\tilde{\alpha}$ (with $\tilde{\alpha} \equiv \alpha_F\left(\sqrt{a} + \sqrt{c}\right)$) are point identified. However, α_F, α_M, and λ are not point identified.*

Proof. First, we impose $b = d = \sqrt{ac}$ and $\alpha_F = \alpha_M$ in the system of equations above. Notice that a is observable. Second, take the system of six equations above and notice that: ρ only appears in equations (36) and (41); and λ only appears in equations (37) and (38). We can take equation (36), solve for ρ and substitute it in equation (41) and take equation (37), solve for λ and substitute in equation (38). We get

$$\frac{A - \beta_F}{\beta_M} = \alpha_F^2\left(\sqrt{a} + \sqrt{c}\right)^2 = \frac{\alpha_F^2}{a}\left(a + \sqrt{ac}\right)^2$$

$$C = \beta_F D + \beta_M B$$

The previous two equations, together with equations (39) and (40), below, form a system with four equations and four unknowns.

$$D = \alpha_F\left(a + \sqrt{ac}\right)$$

$$E = \beta_F B + \beta_M\alpha_F\left(a + \sqrt{ac}\right)$$

We can take the last two equations and get $E = \beta_F B + \beta_M D$. This together with the second equation above ($C = \beta_F D + \beta_M B$) creates a system of two equations and two unknowns and identifies β_F and β_M. With β_F and β_M and using equation (36), we can identify ρ.
Notice that α_F appears three times in the equations above, but each time appears in the same form $\alpha_F\left(a + \sqrt{ac}\right)$ thus, we only have one independent equation for α_F and c. The other time that α_F appears is in equations (12) and 13. In both cases it appears together with λ as $\alpha\left(1 + \lambda\right)$. Thus, we have two independent equations to estimate three parameters (α_F, λ, c). Thus, without further assumptions, α_F, λ and c, are not point identified.

□

[22] Equation (39), dividing both sides by $\sqrt{a}$, can be written as $\frac{D}{\sqrt{a}} = \tilde{\alpha}$ and similarly for equation (40).

A.8 Chadwick and Solon (2002)

We now relate Chadwick and Solon (2002) analysis and put it into the context of our model. For simplicity and because they are implicitly assuming this, we look at the case $\beta_M = 0$.[23] There are three equations in Chadwick and Solon (2002). In our notation and removing the intercepts, they are first,

$$M_i = \beta_F MGF_i + \epsilon_{wi} \tag{42}$$

where $M_i = logE_{wi}$ denotes the log earnings of the wife, $MGF_i = y_{0i}$ denotes the maternal grandfather's log income and $\beta_F = \beta_w$ denotes the effect of a father's earnings in her daughters earnings, or the intergenerational elasticity. Second,

$$\rho_0 = Corr\left(M_i, F_i\right) \tag{43}$$

where $\rho_0 = \gamma$ measures the degree of assortative mating and $F_i = logE_{hi}$. Their final equation is

$$F_i = \beta_h MGF_i + \epsilon_{hi} \tag{44}$$

This last equation comes from combining equation (42) above, with this equation below from Lam and Schoeni (1994)

$$F_i = \rho_0 M_i + \epsilon \tag{45}$$

We then substitute equation (42) into equation (45) and get

$$F_i = \rho_0 \beta_F MGF_i + \left(\rho_0 \epsilon_{wi} + \epsilon\right)$$

In this case, as Chadwick and Solon (2002) argue, $\beta_h = \rho_0 \beta_F$ and $\epsilon_{hi} = \rho_0 \epsilon_{wi} + \epsilon$.[24] In this equation, we get $\beta_h = \mathbb{E}\left[F_i MGF_i\right] \equiv D$ if $\mathbb{E}\left[MGF_i \epsilon_{hi}\right] = 0$. This implies

$$\rho_0 = D/B$$

We have $\mathbb{E}\left[MGF_i \epsilon_{wi}\right] = 0$ by assumption that the equation (42) is well specified. However, it is not clear that $\mathbb{E}\left[MGF_i \epsilon\right] = 0$. First, it is an assumption inconsistent with the exercise that the mother's earnings are causing the father's earnings. $\mathbb{E}\left[MGF_i \epsilon\right] = 0$ implies either: i) that the father's earnings are not correlated with her own parents earnings ($\beta_F = \beta_M = 0$); or ii) that his own parents earnings are not correlated with anything correlated with the mother's earnings ($a = b = c = d = 0$). Since the goal of the exercise is to show social mobility, we focus on the second case. This hypothesis is testable by computing whether $a = \mathbb{E}\left[PGF_i MGF_i\right] \neq 0$. It is rejected in our sample,

In summary, our methodology differs from Chadwick and Solon (2002) in two respects. The first is an economic distinction. The model assumes that the maternal grandparents transmit their status to their daughter (M_i) but then the father status is not affected by her parents, or if it is, that the father family tree is completely orthogonal to the mother and to her side of the tree. The math is correct, but it is implicitly assuming $a = b = c = d = 0$. While this is a common implicit assumption, it may not hold and is rejected in our data.

[23] When $\beta_M \neq 0$ the analysis below is still correct, but using $\tilde{\beta_F} = \beta_F + \lambda \beta_M$ instead of β_F.

[24] If $\beta_M \neq 0$, we can write $F_i = \rho_0 \beta_F MGF_i + \rho_0 \beta_M MGM_i + \left(\rho_0 \epsilon_{wi} + \epsilon\right)$, multiplying both sides by MGF_i, taking expectations and using the exclusion restrictions we get $\mathbb{E}\left[F_i MGF_i\right] = \rho_0 \beta_F \mathbb{E}\left[MGF_i MGF_i\right] + \rho_0 \beta_M \mathbb{E}\left[MGM_i MGF_i\right]$, substituting the values in the expectations we get $D = \rho_0\left(\beta_F + \lambda\beta_M\right) = \rho_0 \tilde{\beta_F}$. As shown in the main body of the paper, we have $B = \tilde{\beta_F}$. Therefore, when $\beta_M \neq 0$, we have $\rho_0 = D/B$.

The second distinction is econometric in nature. They model assortative mating as a causal relation as in equation (45) above. While they model the father status as determined by the mother status, one could also write the mother status as determined by the father:

$$M_i = \rho_1 F_i + \epsilon^{'} \tag{46}$$

If we repeat the exercise using equation (46) instead of equation (45) we get.

$$F_i = \frac{1}{\rho_1} M_i + \frac{\epsilon}{\rho_1} = \frac{\beta_F}{\rho_1} MGF_i + \frac{\epsilon_{wi}}{\rho_1} + \frac{\epsilon}{\rho_1} = \frac{\beta_F}{\rho_1} MGF_i + \left(\frac{\epsilon_{wi}}{\rho_1} + \frac{\epsilon}{\rho_1} \right)$$

Multiplying by MGF_i and taking expectations we get $\mathbb{E}\left[F_i MGF_i\right] = \frac{\beta_F}{\rho_1}$. Solving for ρ_1 we get

$$\rho_1 = B/D$$

In other words, if we asume that is the mother causing the father status, we also get a ratio moment, but inverted.

One could think that equations (46) and (45) are equally valid to represent the correlation between M_i and F_i expressed in equation (43). However, using equation (46) instead of equation (45) gives us a different equation for the relationship between F_i and MGF_i. We can use equation (43) to represent the assortative mating relationship, which is a horizontal non-causal relationship. Using either equation (45) or equation (46) is problematic since they give different results and are implicitly assuming a causal relationship.

B Data Appendix

In this section, we analyze the representativeness of our sample. We restrict attention to the sample with both grandparents and the father, but do not impose the requirement to additionally link any uncles. We then do a logistic regression of the probability of being successfully linked in the complete count Census for the corresponding year. We restrict the complete count Census to be males between the ages of 18 and 79, but do not impose any additional sample restrictions. As controls, we use whether the individual lives in an urban area, whether they are white, age, dummies for nativity and broad occupational category dummies for white collar occupations, middle skilled occupations and low skilled occupations, with farming the excluded category.[25] The results are in Table 10, where we report the odds ratios for the regressions. The linked sample size is 23,737.
In column (1), we report the results for the paternal grandfather's generation. The matched sample is more likely to be rural, substantially more likely to be white and more likely to be a farmer than the overall population. There is no significant difference between the matched sample and the population in nativity. Column (2), showing the maternal grandparents, is very similar. In column (3) analyzing the fathers, the proportion of white-collar occupations is more similar to the rest of the population. Although the requirement to have linked grandparents of course biases the sample towards being American born, interestingly there still are individuals who report being born overseas. The overrepresentation of rural and white individuals remains. In column (4), we look at the representativeness of sons. In contrast to older generations, here farmers are underrepresented relative to the overall population. The linked sons are also disproportionately younger.

The most important adjustment to the occupational scores is for farmers, particularly for the grandfathers' generation. For the grandfathers of the main sample containing both grandparents, 59 percent of grandfathers are farmers. In Figure 8, we show a violin plot displaying the density of the percentile of farmers in the distribution of adjusted occupational scores for paternal grandfathers. The unadjusted OCCSCORE would be at the 4^{th} percentile, with farmers a mass point of over half of the distribution, between the 4^{th} and the 63^{th} percentiles. The adjusted OCCSCORE based on state farm income result in a much wider spread of farmers within the overall distribution.

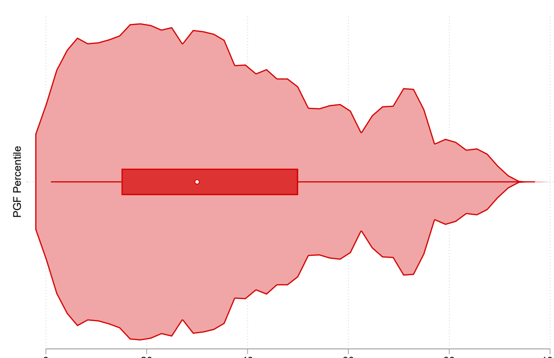
[25] More specifically, for the nativity dummies we use one for whether either or both parents were foreign born (but they themselves were born in America) and one for whether they themselves were born outside the United States. Individuals with no foreign born parents and born in the United States are the excluded category.

Table 10: Logistic Regression of Probability of Matching

	(1) PGF	(2) MGF	(3) Father	(4) Son
Urban	0.407***	0.417***	0.458***	0.535***
	(0.008)	(0.008)	(0.008)	(0.007)
White	3.599***	3.679***	3.922***	3.403***
	(0.155)	(0.158)	(0.169)	(0.131)
Age	1.011***	1.008***	0.956***	0.853***
	(0.001)	(0.001)	(0.001)	(0.001)
Parents foreign	1.168***	1.191***	0.983	0.274***
	(0.024)	(0.024)	(0.018)	(0.032)
Born foreign	1.143***	1.065***	0.182***	1.000
	(0.022)	(0.020)	(0.008)	(.)
White collar	0.564***	0.560***	0.807***	4.188***
	(0.015)	(0.015)	(0.018)	(0.103)
Middle skill	0.753***	0.803***	0.817***	3.409***
	(0.016)	(0.016)	(0.016)	(0.080)
Laborer	0.402***	0.442***	0.387***	2.629***
	(0.009)	(0.010)	(0.009)	(0.061)

Notes: Odds ratios from logistic population of matching an individual. Sample restricted to males between 18 and 79. *, ** and *** represent $p < 0.1$, $p < 0.05$ and $p < 0.01$ respectively.

Figure 8: Farmers Percentile in Income Distribution, Paternal Grandfathers.



C Estimates Using Information on Blacks

In this subsection, we show the results for the subsample of linked Blacks. Because of the small sample size, we are only able to estimate a small number of the propositions with any degree of precision. Notice that for the overall sample, which is 97.9% white, we have $A < B$, which creates estimates for λ close to 1. For Blacks, we have $B > A$. This creates estimates for λ close to zero.

Table 11: Black Subsample

Parameter	Estimate			
	Prop. 1	Prop. 2	Prop. 3	Prop. 4
β_F	0.335	0.200	0.335	0.397
	(0.887)	(0.097)	(0.040)	(0.072)
β_M	0.214	0.420		
	(0.884)	(0.216)		
λ	0.000			0.000
	(4.187)			(0.219)
ρ	0.284	0.305	0.037	0.104
	(2.976)	(0.047)	(0.012)	(0.068)

Notes: Estimates for Propositions 1-4 for Blacks.