The Mismeasure of Women: Intergenerational Mobility and the Econometrics of Family Trees*

José-Antonio Espín-Sánchez† Joseph Ferrie‡ Chris Vickers§

March 17, 2025

Abstract

We present an econometric structure for the analysis of intergenerational mobility that integrates the role of the maternal contribution to the transmission of (dis)advantage with the already well-documented paternal contribution. Our structure does this in ways that are consistent with the realities of both mating and reproduction. We show how previously estimated models are special cases of this general framework and what specific assumptions – many of them testable – each embeds. In a new empirical analysis of intergenerational income mobility from 1870 to 1940, we find that mothers contributed considerably more than fathers to the adult incomes of their children. Our analysis reveals the extent to which inadequate consideration of assortative mating and the impact of mothers can produce misleading conclusions regarding mobility levels, trends over time, and mechanisms.

Keywords: Intergenerational Mobility, Gender, Economic History, Non-Linear Instrumental Variables

*We thank Jaewon Lee for his help on this paper. We have benefited from comments from Xiaohong Chen, Marie Connolly, Steven Durlauf, Claudia Goldin, and Elena Manresa. We want to thank seminar participants at Columbia University, Harris School at the University of Chicago, Harvard University, Northwestern University, Stanford University, University of Maryland, and Yale University, and participants at the Economics of Mobility Conference 2023 at NBER.

†Yale University jose-antonio.espin-sanchez@yale.edu

‡Northwestern University and NBER ferrie@northwestern.edu

§Auburn University czvickers@auburn.edu

1 Introduction

The economic analysis of intergenerational socioeconomic mobility has advanced dramatically since the first formal model of Becker and Tomes (1979; 1986). Studies of mobility levels have now been produced for places, populations, and time periods for which none existed previously (Abramitzky et al., 2021; Alesina et al., 2021). Change over time in the rate of mobility has now become an area of intense study (Song et al., 2020; Ward, 2023; Davis and Mazumder, 2024; Nybom and Stuhler, 2024; Jácome et al., 2025). Economists have now employed tools such as Markov chain models and constructs such as the distinction between exchange and circulation mobility long used by sociologists (Blume et al., 2024). Finally, the analysis of the mechanisms that generate correlations in outcomes across generations are being explored (Daruich and Kozlowski, 2020; Mogstad and Torsvik, 2023; Lochner and Park, 2024), along with the the ages at which parental resources best predict child outcomes (Eshaghnia et al., 2025).

Despite this significant progress, however, one glaring deficiency in empirical work on mobility remains the treatment of females, and mothers in particular. Existing research does not take explicit account of the tree-like structure of family relationships that should discipline the modeling of linkages across generations. As a result, a plethora of estimates of intergenerational mobility appear in the literature, with little basis on which to compare them. Of even greater concern, though, is the large number of embedded – and potentially testable – assumptions that undergird many of these exercises.

Many of the responses to this challenge have been ad hoc and unsatisfying. Some studies focus entirely on fathers and omit any influence from the maternal branch of the family tree (Abramitzky et al., 2021). Others focus on parent pairs in which both parents have reported incomes (Raaum et al., 2008; Shepherd-Banigan et al., 2019), which in many cases will be an unrepresentative and evolving subset of the universe of households.[1] Some directly substitute information from the father of the mother herself (the grandfather of the child at the bottom of the family tree whose outcomes are the objects of ultimate interest) for that of the mother (Eriksson et al., 2023; Buckles et al., 2023b), an expedient which, as we show below, is appropriate only under a particular set of circumstances. Still others avoid the problem entirely by pooling the resource such as income observed in the parents' generation and considering the impact of the "household" or "family" income on the child's outcome (Chetty et al., 2020; Carneiro et al., 2021), which abstracts from the possibility that the effect of one parent differs from that of the other. Finally, some studies focus only on outcomes that can generally be observed for both parents such as educational attainment

[1] For example, labor force non-participation among married women age 18-64 residing with their children under age 18 was more than 60% as recently as 1960, and reached its current 30% level only in the 1990s. Calculated using U.S. decennial census data 1900-2020 from Ruggles et al. (2024a,b).

(Card et al., 2022). This focus can be misleading if the outcome for one parent has been shaped by circumstances that weaken the link between the outcome (e.g. years of schooling completed) and the innate characteristic that might be of ultimate interest (e.g. intelligence). Such was undoubtedly the case in the U.S. for cohorts born roughly 1915-1955 when males attained B.A. degrees at twice the rate of females in many years (Goldin et al., 2006, Figure 2).

What has prevented the full exploitation of the underlying dendriform structure that human biology — through genetic inheritance — and social norms — in the nuclear family and historically low maternal labor force participation — impose on intergenerational mobility is the inability in many circumstances to observe the attributes of the mother that could be passed along to her children. The difficulties posed by failing to take adequate account of the role of mothers in the analysis of mobility can arise in: (1) offering policy prescriptions to enhance mobility, (2) interpreting trends in mobility over time or differences across places, and (3) understanding the mechanisms that generate or impede mobility. For example, we show how differences in assortative mating can vitiate conclusions regarding trends in mobility rates over time and comparisons across contexts when the maternal contribution to the child's outcome is imperfectly observed.

Our work improves upon existing approaches to the empirical analysis of intergenerational mobility in two important ways:

1. It provides an econometric approach to the analysis of mobility that is derived from the constraints that biology imposes on this process. Each individual has 2^n ancestors at each n^{th} generation before their own. This structure forms an inverted triangle with the individual in the current generation at its lower tip. The influences that cascade down from past generations and focus on the current one are linked by both cross-sectional (between fathers and mothers) and time series (from parents to children) correlations, but those correlations are shaped by the biology of human reproduction and are more numerous at generations more distant from the current one. Analysis of such a structure requires a hybrid econometric framework, as it is not a simple process of temporal correlation or cross-sectional dependence or even the combination of these two forms of linkage underlying panel data. Instead, we require a framework that has all of these elements but in addition allows for the exponential growth in the number of cross-sectional linkages as the temporal distance back from the present increases, and imposes a primary causal sequence that moves forward in generational time so earlier generations impact later ones but the opposite is not generally true.

2. This econometric structure lays bare the range of assumptions (regarding assortative mating and the gendered effects of intergenerational transmission, for example) implicit in previous research. These assumptions are not necessarily inappropriate, but a better understanding of

what they actually are will allow us to assess whether they are appropriate across different contexts (over time or across countries, for example), an assessment that may invalidate some comparisons of mobility across contexts. Just as importantly, we demonstrate how these assumptions can actually be tested.

We thus propose a unified treatment of the econometrics of the transmission of socioeconomic status across generations. Unlike the Becker and Tomes (1979, 1986) framework, ours is not a structural model of maximizing behavior subject to constraints, though we do provide a model of this type that is consistent with our econometric structure.[2] Rather, we present a fully-specified econometric framework – a three-equation non-linear instrumental variable (NLIV) approach – that both reveals the variety of assumptions embedded in previous mobility studies and demonstrates how different combinations of assumptions permit the identification of a range of structural parameters of interest. Moreover, we show how a range of results can be revealed depending on the data available to the econometrician.

We then apply this framework to simulated data, for different ranges of the parameters, linked across two or three generations and in a new empirical analysis of U.S. intergenerational income mobility 1870-1940 that exploits several million families followed across two and three generations. We find that mothers contributed considerably more than fathers to the adult incomes of their children. We also show that transmission to children can differ based on the gender of the recipient, and that maternal uncles can be used if data on grandfathers is not available. 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.

The conventional measure of parent-to-child mobility is an intergenerational correlation of some outcome (income, educational attainment, occupation).[3] To see what the father-child correlation yields when the unobserved mother makes a non-zero contribution to the child's outcome, consider an example where the true effect of the father is $\beta_F = 0.3$, the true effect of the mother is $\beta_M = 0.6$, and the degree of assortative mating (the correlation between father and mother) is $\rho = 0.5$. We show below that the estimated father-child intergenerational correlation will be 0.6 (that is, $0.6 = 0.3 + 0.6 \cdot 0.5$), meaning that the estimated father-child intergenerational correlation

[2] Appendix E extends Becker and Tomes (1979) to households with two parents that bargain *à la* Nash. This model informs our econometric approach.

[3] A closely related measure is the intergenerational elasticity (IGE) of some outcome, which is generally estimated through a parsimonious regression of the log of the child's outcome (e.g. income) on the log of the parent's corresponding outcome. When the outcome in the IGE is multiplied by the ratio of the standard deviation of the outcome in the parent generation divided by the standard deviation of the outcome in the child generation then the IGE and the intergenerational correlation coincide (Mazumder, 2015). In what follows, we will use the intergenerational correlation in occupational income ranks.

is twice as large as the true effect of the father on the child. The intergenerational correlation here is measuring two effects: (1) the direct effect of the father on the child $\beta_F = 0.3$; and (2) the effect of the mother indirectly measured through the father $\rho\beta_M$

The intergenerational correlation we estimate is no longer a single number but instead depends on the parent whose contribution to the child's outcome we observe. To show a single number to summarize how much of the child's outcome is accounted for by the previous generation's outcomes, one can use the total contribution of the parents outcome to the child's outcome.[4] In this case, the total fraction of the variance in child's outcome that can be explained by parental factors is 0.79 (i.e., $0.79 = \sqrt{0.3^2 + 0.6^2 + 0.5 \cdot 2 \cdot 0.3 \cdot 0.6}$), which is larger than either the father-child intergenerational correlation and the mother-child intergenerational correlation but smaller than their sum. The total parental contribution is largest with perfect correlation of the parents' outcome ($\rho = 1$). In that case, the estimated father-child correlation and the true total contribution coincide and are equal to 0.9 (i.e., $0.9 = 0.3 + 0.6$). The father's effect falls to the true value of 0.3 when we use our three-equation NLIV to account fully for the mother's impact (0.6).

Our main contribution is to provide a unified framework so that we can interpret, understand, and contextualize the different correlations that researchers produce with different approaches to data construction.[5] In that sense, our results generalize the seminal contribution of Chadwick and Solon (2002), by estimating all the structural parameters as well as providing more flexible results that are applicable to other datasets. We also use grandfathers' information, in some cases both grandfathers, to estimate the maternal contribution to social mobility in a way that has not been done before. Collado et al. (2023) use a similar approach, but they use direct observation of female outcomes. They use horizontal relations; i.e., sibling and cousin observations. We use vertical information (father-child) to estimate both vertical (mother-child) and horizontal (father-mother) structural relationships.

2 Setup

Studies of social mobility often use outcome data on male parents and their children to assess social mobility as in equation (1), which measures the correlation between the (standardized) outcome of the child (X_i^C) and the (standardized) outcome of the father (X_i^F).

$$X_i^C = \tilde{\beta}_F X_i^F + \nu_i \tag{1}$$

[4] If equation (2) determines the child's outcome, then the father-child outcome correlation is $\mathbb{E}\left[X_i^F X_i^C\right] = \beta_F + \rho\beta_M$. Moreover, the contribution of both parents to the child outcome is $\bar{\beta} \equiv (\beta_F^2 + \beta_M^2 + 2\rho\beta_F\beta_M)^{0.5}$ (Althoff et al., 2024).

[5] Eriksson et al. (2023) provides a comprehensive review of how estimates in the literature relate to our method. Moreover, Eriksson et al. (2023) and Keller and Shiue (2024) are already applying our method to study Massachusetts and China, respectively.

If the goal is merely to present a descriptive assessment, the researcher can use the father-child empirical correlation of outcomes. However, if the goal is to assess causality, or prescribe policy, the estimation of the parameter of interest $\tilde{\beta}_F$ could be biased if equation (1) is misspecified. In particular, one can think of other family members that can have an effect on the outcome of the child. The first that comes to mind is the mother. Whether we think they directly transmit an outcome (e.g., socioeconomic status), or that they transmit only traits that affect that outcome, mothers will play an essential role. In particular, they are strictly as important as fathers if we think the transmission is due to genetics and possibly as important if we think the transmission is mostly inherited wealth or social status. Alternatively, we might think that what is transmitted from generation to generation is human capital or networks related to a particular occupation which could be transmitted from father to child without mothers playing a major role. Even in these instances, grandfathers might play a role in determining the outcome of the child, even after accounting for the effect of the father (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. In particular, we can think that the outcome of the child is affected by both the outcome of the father and the outcome of the mother (X_i^M) as reflected in equation (2):

$$X_i^C = \beta_F X_i^F + \beta_M X_i^M + \varepsilon_i^C \tag{2}$$

X_i^M is a measure of the mother's status measured in the same units as the father's status.[6] It is transmitted to the child according to β_M, but it is not observed by the econometrician.

An alternative interpretation in social mobility studies is that one outcome is observed, but it is a proxy for a latent variable (social status, human capital, networks, genetic endowment) that is being transmitted across generations but not observed. Under this interpretation estimators might suffer from attenuation bias. In Appendix D.2, we discuss whether and how this might affect our estimators.

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 $\mathbb{E}\left[X_i^F\left(\beta_M X_i^M + \varepsilon_i^C\right)\right] \neq 0$ (Espín-Sánchez et al., 2022). If equation (2) is well specified we have $\mathbb{E}\left[X_i^F \varepsilon_i^C\right] = 0$ and this condition becomes $\rho\beta_M \neq 0$, where $\rho = \mathbb{E}\left[X_i^F X_i^M\right]$ is the extent of assortative mating in the couple formed by the father and the mother. Such an estimator would be

$$\tilde{\beta}_F = \frac{\mathbb{E}\left[X_i^F X_i^C\right]}{\mathbb{E}\left[X_i^F X_i^F\right]} \tag{3}$$

In general, the OLS estimate $\tilde{\beta}_F \equiv \beta_F + \rho\beta_M$ from equation (1) is a biased estimator of both

[6]In our formal notation, individuals are identified by their position in the family: F=father, M=mother, C=child, S=son, D=daughter, PU=paternal uncle (father's brother), MU=maternal uncle (mother's brother), PF=paternal grandfather, PM=paternal grandmother, MF=maternal grandfather, MM=maternal grandmother.

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 ($\rho = 1$), the estimator $\tilde{\beta}_F$ will consistently estimate $(\beta_F + \beta_M)$. In the other extreme case, where mating is perfectly random ($\rho = 0$), 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 ($1 > \rho > 0$) and the estimator $\tilde{\beta}_F$ will lie between β_F and $(\beta_F + \beta_M)$, but we would not know how close the estimator is to either end of this range without knowing the degree of assortment ρ.

Figure 1 shows the total parental contribution as a function of the father-child correlation. Each curve represents a different value for β_M, and each panel is for a different value of ρ. When mothers do not have an effect on children ($\beta_M = 0$), the father-child correlation is a consistent estimate for the effect of the father on the child (β_F). This is shown as the 45-degree line. When mothers have an effect, the total contribution will be larger than the father-child correlation. The vertical distance between a given curve and the 45-degree line measures the bias in the total contribution if it is measured using the father-child correlation. When the effect of the mother (β_M) is larger, the bias would be larger, as we can see the lines measuring the total effects going up for larger values of β_M. When assortative mating (ρ) is lower, the bias is also larger. We can see how the lines in panel B are substantially higher than those in panel A, even for modest changes in ρ. Moreover, when mothers have an effect, the implied total contribution could exceed 1, e.g., $\rho = 0.3$, $\beta_M = 0.6$, and $\mathbb{E}[X_i^F X_i^C] > 0.9$, which is infeasible. In other words, if $\rho = 0.3$ and $\beta_M = 0.6$, a father-child correlation of 0.9 would imply an infeasibly high effect of the father on the child $\beta_F > 1$.

In this paper, we restrict attention to the model depicted by equation (2), as the simplest natural extension of equation (1). The system of equations is then:[7]

$$X_i^C = \beta_F X_i^F + \beta_M X_i^M + \varepsilon_i^C \tag{4}$$

$$X_i^F = \beta_F' X_i^{PF} + \beta_M' X_i^{PM} + \varepsilon_i^F \tag{5}$$

$$X_i^M = \beta_F' X_i^{MF} + \beta_M' X_i^{MM} + \varepsilon_i^M \tag{6}$$

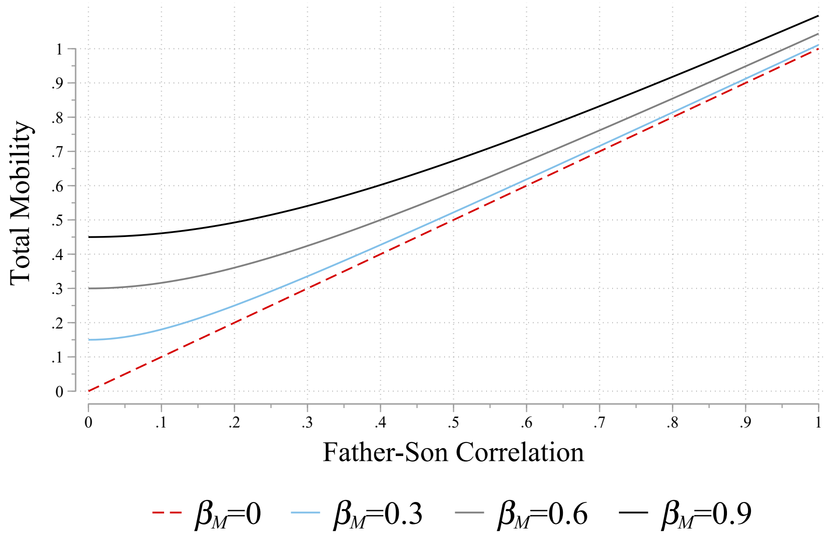
where X_i^{PF} is the outcome of the paternal grandfather, X_i^{PM} is the outcome of the paternal grandmother, X_i^{MF} is the outcome of the maternal grandfather, X_i^{MM} is the outcome of the maternal grandmother, and ε_i^C, ε_i^F and ε_i^M are the corresponding error terms. In the baseline model, the relations that we are trying to estimate are depicted in Figure 2.

- Three (horizontal) relations of assortative mating: $X_i^F \sim X_i^M$, measured by ρ; and $X_i^{PF} \sim X_i^{PM}$ and $X_i^{MF} \sim X_i^{MM}$ measured by ρ', none of which could be estimated directly.

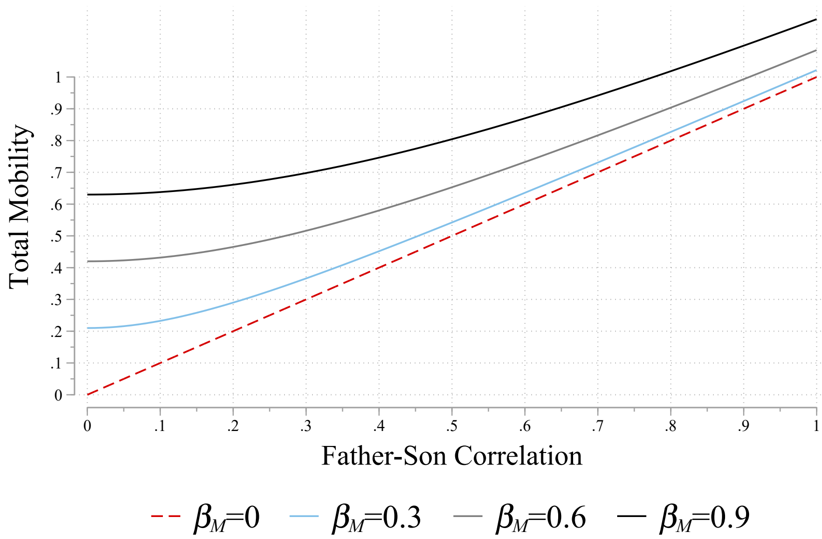
[7] The method developed here can be easily extended to richer models that allow for direct transmission from grandparents, interaction effects between the parents, or direct diagonal effects of uncles. The equations presented above could be easily extended by adding observable characteristics such as age, state of birth, or number of siblings in the household. The results would remain true if we add any observable and exogenous variables to any equation.

Figure 1: Total Contribution as a function of Father-Child Correlation.

A. With medium assortative mating ($\rho = 0.5$).



B. With low assortative mating ($\rho = 0.3$).



*Not*es: The father-child outcome correlation is $\mathbb{E}\left[X_i^F X_i^C\right] = \beta_F + \rho\beta_M$, or $\beta_F = \mathbb{E}[X_i^F X_i^C] - \rho\beta_M$. The contribution of both parents to the child outcome is $\bar{\beta} \equiv (\beta_F^2 + \beta_M^2 + 2\rho\beta_F\beta_M)^{0.5}$. We can write $\bar{\beta}$ as a function of β_M and ρ and get $\bar{\beta} = [(\mathbb{E}[X_i^F X_i^C] - \rho\beta_M)^2 + \beta_M^2 + 2\rho\beta_M(\mathbb{E}[X_i^F X_i^C] - \rho\beta_M)]^{0.5}$. We plot the total contribution as a function of β_M for $\rho = 0.5$ (Panel A) and $\rho = 0.3$ (Panel B).

- Three causal male relations ($X_i^F \sim X_i^C$, $X_i^{PF} \sim X_i^F$ and $X_i^{MF} \sim X_i^M$) that are governed by β_F (or β_F') and could be estimated if mating was random.
- Three causal female relations ($X_i^M \sim X_i^C$, $X_i^{PM} \sim X_i^F$ and $X_i^{MM} \sim X_i^M$) that are governed by β_M (or β_M'), none of which could be estimated directly.

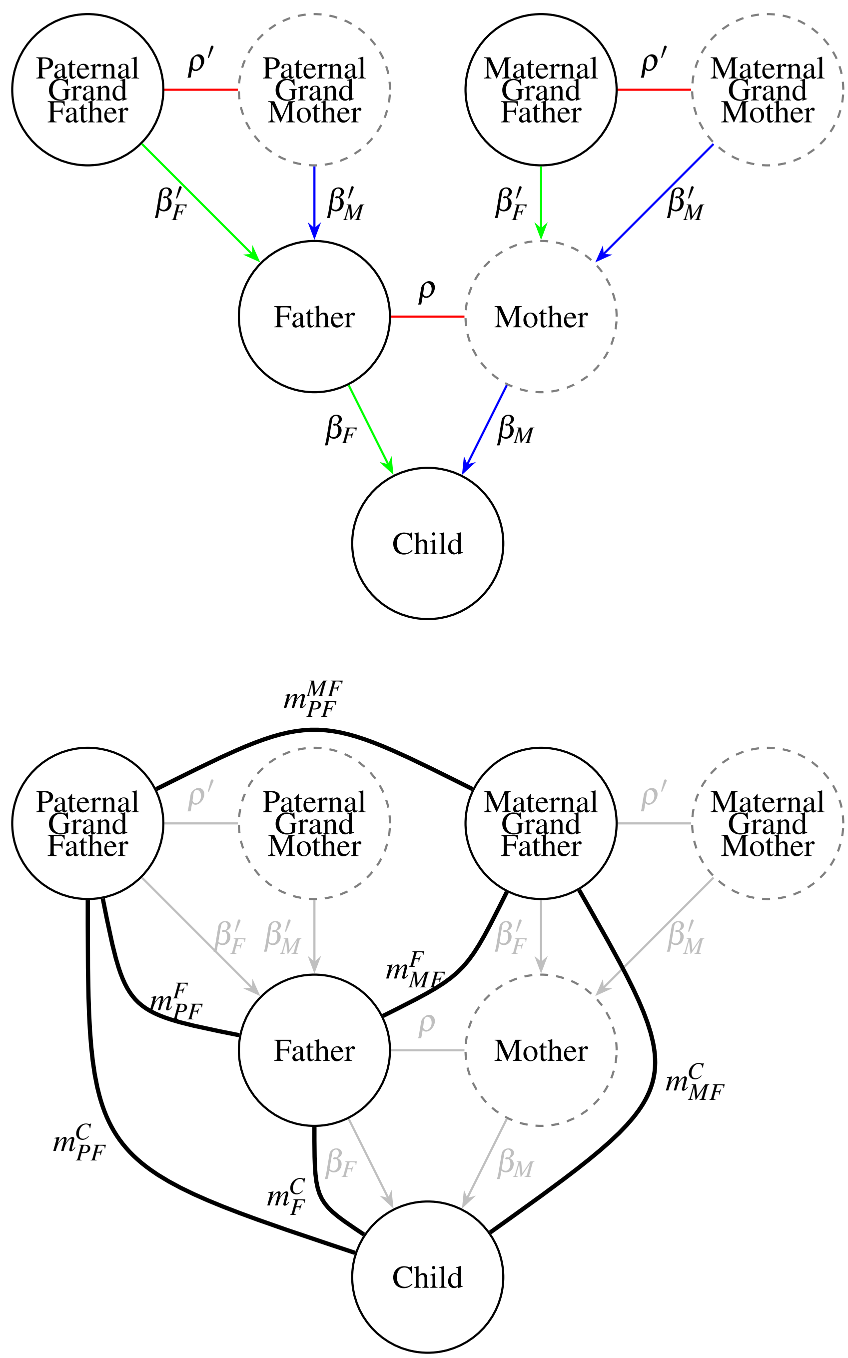
Throughout the paper we assume that the system formed by equations (4), (5), and (6) is well specified, i.e., that the error term is uncorrelated with the regressor in each equation. In particular, the exclusion restrictions are:

1. $\mathbb{E}\left[X_i^F \varepsilon_i^C\right] = 0$;
2. $\mathbb{E}\left[X_i^{PF} \varepsilon_i^C\right] = 0$;
3. $\mathbb{E}\left[X_i^{PF} \varepsilon_i^F\right] = 0$;
4. $\mathbb{E}\left[X_i^{PF} \varepsilon_i^M\right] = 0$;
5. $\mathbb{E}\left[X_i^{MF} \varepsilon_i^C\right] = 0$;
6. $\mathbb{E}\left[X_i^{MF} \varepsilon_i^F\right] = 0$;
7. $\mathbb{E}\left[X_i^{MF} \varepsilon_i^M\right] = 0$.

We have seven exclusion restrictions and they are analogous to exclusion restrictions in a time series that follows an $AR(1)$ process, i.e., $Y_t = \beta Y_{t-1} + \varepsilon_t$, where future shocks cannot affect the past: $\mathbb{E}\left[Y_{t-1}\varepsilon_t\right] = 0$. The child is at the end of the tree, so we cannot have any exclusion restriction multiplying X_i^C by any of the ε_i. The father is in the middle of the tree, so we can use one exclusion restriction for X_i^F: $\mathbb{E}\left[X_i^F \varepsilon_i^C\right] = 0$, i.e., $\mathbb{E}\left[Y_{t-1}\varepsilon_t\right] = 0$. X_i^F would be endogenous to all other error terms, except the one at the bottom of the tree, e.g., $\mathbb{E}\left[X_i^F \varepsilon_i^{PF}\right] \neq 0$, i.e., $\mathbb{E}\left[Y_t \varepsilon_{t-1}\right] \neq 0$. The grandfathers are at the top of the tree. X_i^{PF} and X_i^{MF} are exogenous to the error terms below them in the tree: two error terms in the middle of the tree, ε_i^F and ε_i^M ($\mathbb{E}\left[Y_{t-21}\varepsilon_{t-1}\right] = 0$) and the one at the end of the tree ε_i^C ($\mathbb{E}\left[Y_{t-2}\varepsilon_t\right] = 0$). These exclusion restrictions allow us to use the grandfathers as (non-linear) instruments to construct the relevant moments. The usual analysis of social mobility implicitly assumes that mothers play no role, or not a role independent from the father. This would be a particular case of our tree where mothers play no role, and there are only three relevant exclusion restrictions: $\mathbb{E}\left[X_i^F \varepsilon_i^C\right] = 0$ ($\mathbb{E}\left[Y_{t-1}\varepsilon_t\right] = 0$), $\mathbb{E}\left[X_i^{PF} \varepsilon_i^F\right] = 0$ ($\mathbb{E}\left[Y_{t-2}\varepsilon_{t-1}\right] = 0$), and $\mathbb{E}\left[X_i^{PF} \varepsilon_i^C\right] = 0$ ($\mathbb{E}\left[Y_{t-2}\varepsilon_t\right] = 0$). The extra exclusion restrictions coming from the maternal side allow us to identify all structural parameters.

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.

Figure 2: Structural Parameters and Empirical Relations



Notes: The horizontal lines (red) represent the degree of assortative matching; the vertical arrows represent the effects on mobility (green: father; blue: mother). The solid circles represent individuals (males) with observed outcomes while the dashed circles represent individuals (females) with unobserved outcomes.

$$\Sigma = Var \begin{bmatrix} X_i^{PF} \\ X_i^{PM} \\ X_i^{MF} \\ X_i^{MM} \end{bmatrix} = \begin{bmatrix} 1 & \rho' & \gamma_{PF}^{MF} & \gamma_{MM}^{PF} \\ \rho' & 1 & \gamma_{MF}^{PM} & \gamma_{PM}^{MM} \\ \gamma_{PF}^{MF} & \gamma_{MF}^{PM} & 1 & \rho' \\ \gamma_{MM}^{PF} & \gamma_{PM}^{MM} & \rho' & 1 \end{bmatrix}$$

where the nuisance parameters measure the four correlations among grandparent pairs, i.e., $\gamma_{PF}^{MF} \equiv \mathbb{E}\left[X_i^{PF} X_i^{MF}\right]$, $\gamma_{MF}^{PM} \equiv \mathbb{E}\left[X_i^{MF} X_i^{PM}\right]$, $\gamma_{PM}^{MM} \equiv \mathbb{E}\left[X_i^{PM} X_i^{MM}\right]$ and $\gamma_{MM}^{PF} \equiv \mathbb{E}\left[X_i^{MM} X_i^{PF}\right]$.

With this data-generating process, we can see in Figure 2 that there are only three end nodes in the tree: father, mother and child. Therefore, we need to use nothing more than equations (4), (5), and (6) to estimate the model. In other words, the data is originally generated by the outcomes 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 outcomes of the father and mother are realized. After those realizations, the transmission from the parents to the child occurs according to the causal relationships in equation (4), i.e., the outcome of the child is realized.

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 our interest as social scientists in social phenomena. β_F measures the effect of the father on the child. β_M measures the effect of the mother on the child. ρ measures assortative mating between the mother and the father.

Second, we define the empirical relationships in the data. In the standard model of social mobility we only observe two variables $\left(X_i^C, X_i^F\right)$ and are thus able to obtain one empirical relationship $m_F^C \equiv \mathbb{E}\left[X_i^C X_i^F\right]$. We have four variables available $\left(X_i^C, X_i^F, X_i^{PF}, X_i^{MF}\right)$, so we can get six empirical relationships, which correspond to m_F^C, m_{PF}^F, m_{MF}^C, m_{MF}^F, m_{PF}^C, and $m_{PF}^{MF} \equiv \gamma_{PF}^{MF}$ (Figure 2, bottom). Section 3 shows how we use these relationships to identify the structural parameters.

Third, we have the nuisance parameters $\left(\gamma_{PF}^{MF}, \gamma_{MF}^{PM}, \gamma_{PM}^{MM}, \gamma_{MM}^{PF}\right)$. We call them nuisance parameters because they are not our main object of interest here. With the exception of γ_{PF}^{MF}, 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 the most general way of thinking of family structure and assortative mating in our model. 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, e.g., arranged marriages.

In addition to the baseline tree, we also show identification results when the researcher has access to data relating to one maternal uncle. The equation determining the status of a maternal uncle (X_i^{MU}) is

$$X_i^{MU} = \beta_F X_i^{MF} + \beta_M X_i^{MM} + \varepsilon_i^{MU} \tag{7}$$

With this new equation, we have a new parameter $\eta \equiv \mathbb{E}\left[\varepsilon_i^M \varepsilon_i^{MU}\right]$ that measures household effects. Moreover, we can extend our baseline model to allow for parental effects to be different for sons and daughters. The system of equations in this case is:

$$\begin{array}{rcl} X_i^S & = & \beta_F^S X_i^F + \beta_M^S X_i^M + \varepsilon_i^S \\ X_i^F & = & \beta_F^S X_i^{PF} + \beta_M^S X_i^{PM} + \varepsilon_i^F \\ X_i^M & = & \beta_F^D X_i^{MF} + \beta_M^D X_i^{MM} + \varepsilon_i^M \end{array} \tag{8}$$

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. We now list some assumptions that are implicitly used in the literature.

Assumption 1 (Direct Mating)**.** $\gamma_{PF}^{MF} = \gamma_{MF}^{PM} = \gamma_{PM}^{MM} = \gamma_{MM}^{PF}$.

Assumption 2 (Mating on Observables)**.** $\mathbb{E}\left[X_i^F \varepsilon_i^M\right] = \mathbb{E}\left[X_i^{PF} \varepsilon_i^{PM}\right] = \mathbb{E}\left[X_i^{MF} \varepsilon_i^{MM}\right] = 0$.

Assumption 3 (Only Male Effects)**.** $\beta_M = 0$.

Assumption 4 (Gender Neutral)**.** $\beta_M = \beta_F$.

Assumption 5 (No Household Effects)**.** $\eta = 0$.

In empirical applications, one would sometimes impose some of them, but not all of them, and we discuss what assumptions are needed for identification depending on the data available to the researcher. It is worth emphasizing that these assumptions, or stronger ones, are implicitly used in the literature. We are just being intentionally explicit on what they are and how they relate to identification. A1 implies that mating is determined by individual characteristics of the spouses, not their parents. This is implicitly assumed in the literature and in our baseline case. Nonetheless, our framework allows for researchers to relax it, as we show in Proposition 4.

Assumption A2 imposes additional exclusion restrictions. We usually do not need to impose these restrictions to obtain mobility parameters, but sometimes we need it to estimate assortative mating. In the usual time series approach, where there is no tree, this would amount to $\mathbb{E}\left[Y_t \varepsilon_t\right] = 0$, which could not be true. With the tree structure, however, this assumption holds. The implication, again, is that the model is correctly specified and there is no mating on unobservables, i.e., X_i^F and X_i^M measure all the relevant characteristics for mating.

Assumption A3 assumes that only fathers affect their children, not mothers. Proposition 1 shows how the father effect would be overestimated if mothers have an effect on their children, but we assume they do not. Assumption A4 assumes that each parent has the same effect on their children. This is implicitly assumed in most of the labor economics literature when researchers have access to *household* income, rather than individual parental income. This assumption would be violated if mothers and fathers have different preferences on how to spend their income, or on how much time they spent with their children. Becker and Tomes (1979, 1986) assumes income is used to invest in the human capital of the child. Higher human capital children would then earn a higher income and will then invest a higher amount in the human capital of their own children. The income earned by the mother could be invested in the same way as the income earned by the father. If both parents have the same preferences on how to use that income, then $\beta_M = \beta_F$. If their preferences are different, even if the transformation of income into human capital of the child is the same, then $\beta_M \neq \beta_F$ (see Appendix E for details).[8] Proposition 3 and Corollaries 1 and 2 make this assumption. Finally, A5 assumes that there are not household effects, i.e., that the unexplained component of children's outcome is uncorrelated across siblings. Table 1 presents a summary of the various propositions, assumptions, and identified parameters.

3 Identification

In this section, we discuss point-identification of the structural parameters (β_F, β_M, ρ). Detailed proofs are shown in Appendix A. The goal of this section is twofold. First, it provides to the reader a cookbook: for each potential dataset that the reader may encounter, it discusses the most reasonable sets of assumptions to provide identification. Second, it allows the reader to understand the trade-offs between the sample generated and the parameters of the model. A sample with more relatives generates more empirical moments. This is good because it allows the identification of more parameters, but it may come at the expense of using a more selected sample.

In Subsection 3.1 we discuss identification using 2-generation trees. This is the simplest case regarding data requirements and generates only three empirical moments. In Subsection 3.2, we discuss identification using 3-generation records. In Subsection 3.3, we discuss identification using 3-generation records and one maternal uncle, allowing us to identify household effects. Finally, in Subsection 3.4 we allow mobility parameters to vary according to the child's gender, in addition to the parent's gender.

3.1 Identification Using 2-Generation Data

In this subsection, we present our baseline results. We provide the basic intuition of our approach and how we exploit all the variation in the data to get the most flexible estimates, given the data

[8] Studies using modern data usually have information on household income $(X_i^F + X_i^M)$ and run OLS estimation on $X_i^C = \beta\left(X_i^F + X_i^M\right) + \varepsilon_i$. If both parents are generating income this is equivalent to assuming $\beta = \beta_M = \beta_F$.

available. We assume that the econometrician has access to two generations of data prior to the child of interest data on male socioeconomic status $\left(X_i^F, X_i^{PF}, X_i^{MF}\right)$. With three variables, we can compute three empirical moments: m_{PF}^{MF}, m_{MF}^{F} and m_{PF}^{F}. We now show how these three empirical moments, together with the exclusion restrictions, generate three independent equations when Assumption A1 holds.

First, we use equation (5), multiply each term by X_i^{PF} and take expectations. Assuming $\mathbb{E}\left[X_i^{PF}\varepsilon_i^F\right]=0$ generates the first moment

$$\mathbb{E}\left[X_i^F X_i^{PF}\right]=\beta_F^{'}\mathbb{E}\left[X_i^{PF}X_i^{PF}\right]+\beta_M^{'}\mathbb{E}\left[X_i^{PF}X_i^{PM}\right]=\beta_F^{'}+\beta_M^{'}\rho^{\prime} \tag{9}$$

This moment is the analog of an OLS estimator where we multiply the estimating equation by one of the independent variables and take expectations. If we observe X_i^{PM}, we would multiply equation (5) by X_i^{PM} and take expectations. That moment, together with equation (9), would form a system with two equations and two unknowns $(\beta_F^{'},\beta_M^{'})$. $\rho^{\prime}=\mathbb{E}\left[X_i^{PF}X_i^{PM}\right]$ would be directly observable in the data. We do not observe X_i^{PM}, but we will create a system of three independent moments to identify $(\beta_F^{'},\beta_M^{'},\rho^{\prime})$.

We now use equation (5), multiply each term by X_i^{MF}, and take expectations. Assuming $\mathbb{E}\left[X_i^{MF}\varepsilon_i^F\right]=0$ and A1 generates the second moment

$$\mathbb{E}\left[X_i^F X_i^{MF}\right]=\beta_F^{'}\mathbb{E}\left[X_i^{PF}X_i^{MF}\right]+\beta_M^{'}\mathbb{E}\left[X_i^{PM}X_i^{MF}\right]=\mathbb{E}\left[X_i^{PF}X_i^{MF}\right]\left(\beta_F^{'}+\beta_M^{'}\right) \tag{10}$$

This moment is the analog of an instrumental variables (IV) estimator, where we multiply the estimating equation by a variable that is not in the estimating equation. In that sense we use X_i^{MF} as an instrument. Notice, however, that this instrument is not exogenous in our model. The usual approach to omitted variables is to find an instrument Z_i such that $\mathbb{E}\left[Z_i(\beta_M^{'}X_i^{PM}+\varepsilon_i^F)\right]=0$. Using that instrument, would generate the standard IV moment: $\mathbb{E}\left[X_i^F Z_i\right]=\beta_F^{'}\mathbb{E}\left[X_i^{PF}Z_i\right]$. The standard approach would produce an unbiased estimate for $\beta_F^{'}$ but would not be able to estimate $\beta_M^{'}$ or $\rho^{\prime}$. Our approach allows us to estimate all three parameters of interest, without having to find an exogenous instrument Z_i. Moreover, whereas the exogeneity of Z_i with respect to X_i^{PM} is assumed and untestable, our method directly computes the correlation between our instrument X_i^{MF} and the omitted variable X_i^{PM} through the tree structure.

Finally, we take equations (5) and (6) and multiply the left hand side in each equation, and the right hand side in each equation, and take expectations. Using the exclusion restrictions and Assumptions A1 and A2 we get

$$\rho=\mathbb{E}\left[X_i^{MF}X_i^{PF}\right]\left(\beta_F^{'}+\beta_M^{'}\right)^2 \tag{11}$$

Equation (9) is similar to an OLS equation, i.e., we use a variable that is included in the right hand side (X_i^{PF}) and multiply all terms by that variable. Because X_i^{PM} is unobserved, the second term would generate omitted variable bias. Equation (10) is similar to an IV equation, i.e., we use a variable that is not included in the right hand side (X_i^{MF}) and multiply all terms by that variable. Because X_i^{PM} is unobserved, the second term would generate omitted variable bias. What we need now is a third equation that relates these unobservable terms (ρ' and γ_{MF}^{PM}) to one another and to other observable moments or structural parameters. Equation (11) comes from our theoretical model and fulfills that role. We now have a system with three equations and seven unknowns: the four structural parameters $(\beta_F^{'}, \beta_M^{'}, \rho', \rho)$ and the three unknown nuisance parameters $(\gamma_{MF}^{PM}, \gamma_{PM}^{MM}, \gamma_{MM}^{PF})$. Proposition 1 below shows how one can identify $(\beta_F^{'}, \beta_M^{'}, \rho)$ making assumptions on the nuisance parameters, using information on 2-generation records alone.

Proposition 1. *Suppose X_i^F, X_i^{PF}, and X_i^{MF} are observed. If assumption A1 holds and $\rho' = \rho$, then $(\beta_F^{'}, \beta_M^{'}, \rho)$ is point identified.*

When assumption A1 holds $\rho' = \rho$, and we use equations (9), (10), and (11), we have a system with three unknowns and three independent equations. Solving this system we get

$$\rho = \frac{\left(m_{MF}^F\right)^2}{m_{PF}^{MF}} \tag{12}$$

$$\beta_F^{'} = \frac{m_{MF}^F}{m_{PF}^{MF}} - \frac{m_{MF}^F - m_{PF}^{MF} m_{PF}^F}{m_{PF}^{MF} - \left(m_{MF}^F\right)^2} \tag{13}$$

$$\beta_M^{'} = \frac{m_{MF}^F - m_{PF}^{MF} m_{PF}^F}{m_{PF}^{MF} - \left(m_{MF}^F\right)^2} \tag{14}$$

The seminal work by Chadwick and Solon (2002) uses a similar approach, but their framework is more restrictive in two respects. First, they impose Assumption A4. Second, they implicitly assume $m_{PF}^{MF} = m_{PF}^F m_{MF}^F$.[9] Both assumptions can be easily tested, and are rejected in our examples in Section 5. Indeed, if we substitute this restriction into equation (12) above, we get $\rho = m_{MF}^F / m_{PF}^F$, which is the ratio estimator used in Chadwick and Solon (2002). Thus, their estimator is a particular case of our framework under the assumption that $m_{PF}^{MF} = m_{PF}^F m_{MF}^F$.

Propositions 2 and 3 below restrict attention to the cases most commonly found in the literature: either women do not matter (Proposition 2, $\beta_M^{'} = 0$), or women matter as much as men (Proposition 3, $\beta_M^{'} = \beta_F^{'}$).

Proposition 2. *Suppose X_i^F, X_i^{PF}, and X_i^{MF} are observed. If assumption A3 holds, 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 use information on m_{PF}^{MF}, we can still estimate $(\beta_F^{'}, \rho)$ by writing

[9] This restriction comes from the particular variance decomposition used. The ratio of m_{PF}^{MF} over m_{MF}^F is equal to the effect of the parents on the child. But this effect is also equal to m_{PF}^F. Thus $m_{PF}^{MF} = m_{PF}^F m_{MF}^F$.

$\beta_F' = m_{PF}^F$ and $\rho = m_{MF}^F m_{PF}^F$. In this case, the formula for ρ is the product, not the ratio, of m_{MF}^F and m_{PF}^F. Proposition 3 below allow us to estimate ρ and ρ' independently by assuming $\beta_M' = \beta_F'$. This is a common assumption in the literature when the researcher only have data on aggregated household income.

Proposition 3. *Suppose X_i^F, X_i^{PF}, and X_i^{MF} are observed. If assumptions A1 and A4 hold, then (β_F', ρ', ρ) is point identified.*

3.2 Identification Using 3-Generation Data

We now discuss what parameters can be identified with 3-generation data. We now combine equations (4), (5), and (6) with the seven exclusion restrictions above to generate six moments.[10] With the exclusion restrictions above we can generate the following moments.

Using equation (4) and $\mathbb{E}\left[X_i^F \varepsilon_i^C\right] = 0$ we get

$$\begin{aligned} m_F^C &\equiv \mathbb{E}\left[X_i^C X_i^F\right] \\ &= \beta_F \mathbb{E}\left[X_i^F X_i^F\right] + \beta_M \mathbb{E}\left[X_i^M X_i^F\right] \\ &= \beta_F + \rho \beta_M \end{aligned} \tag{15}$$

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

$$\begin{aligned} m_{PF}^F &\equiv \mathbb{E}\left[X_i^F X_i^{PF}\right] \\ &= \beta_F' \mathbb{E}\left[X_i^{PF} X_i^{PF}\right] + \beta_M' \mathbb{E}\left[X_i^{PM} X_i^{PF}\right] \\ &= \beta_F' + \rho' \beta_M' \end{aligned} \tag{16}$$

Using equations (4) and (6) and $\mathbb{E}\left[X_i^{MF} \varepsilon_i^M\right] = 0$ and $\mathbb{E}\left[X_i^{MF} \varepsilon_i^C\right] = 0$ we get

$$\begin{aligned} m_{MF}^C &\equiv \mathbb{E}\left[X_i^{MF} X_i^C\right] = \beta_F \mathbb{E}\left[X_i^{MF} X_i^F\right] + \beta_M \mathbb{E}\left[X_i^{MF} X_i^M\right] \\ &= \beta_F \mathbb{E}\left[X_i^{MF} X_i^F\right] + \beta_M \left(\beta_F' \mathbb{E}\left[X_i^{MF} X_i^{MF}\right] + \beta_M' \mathbb{E}\left[X_i^{MF} X_i^{MM}\right]\right) \\ &= \beta_F m_{MF}^F + \beta_M \left(\beta_F' + \rho' \beta_M'\right) \end{aligned} \tag{17}$$

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

$$\begin{aligned} m_{MF}^F &\equiv \mathbb{E}\left[X_i^F X_i^{MF}\right] \\ &= \beta_F' \mathbb{E}\left[X_i^{PF} X_i^{MF}\right] + \beta_M' \mathbb{E}\left[X_i^{PM} X_i^{MF}\right] \\ &= \beta_F' \gamma_{PF}^{MF} + \beta_M' \gamma_{MF}^{PM} \end{aligned} \tag{18}$$

[10] 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 X_i^F, and then take expectations and use $\mathbb{E}\left[X_i^F \varepsilon_i^C\right] = 0$ to get equation (15). See Appendix A for details.

Using equation (4) and $\mathbb{E}\left[X_i^{PF}\varepsilon_i^C\right]=0$ and $\mathbb{E}\left[X_i^{PF}\varepsilon_i^M\right]=0$ we get

$$\begin{aligned} m_{PF}^C &\equiv \mathbb{E}\left[X_i^{PF}X_i^C\right] \\ &= \beta_F\mathbb{E}\left[X_i^{PF}X_i^F\right]+\beta_M\mathbb{E}\left[X_i^{PF}X_i^M\right] \\ &= \beta_F\mathbb{E}\left[X_i^{PF}X_i^F\right]+\beta_M\left(\beta_F^{'}\mathbb{E}\left[X_i^{PF}X_i^{MF}\right]+\beta_M^{'}\mathbb{E}\left[X_i^{PF}X_i^{MM}\right]\right) \\ &= \beta_F m_{PF}^F+\beta_M\left(\beta_F^{'}\gamma_{PF}^{MF}+\beta_M^{'}\gamma_{MM}^{PF}\right) \end{aligned} \tag{19}$$

Using equations (5) and (6) we get

$$\mathbb{E}\left[X_i^F X_i^M\right]=\mathbb{E}\left[\left(\beta_F^{'}X_i^{PF}+\beta_M^{'}X_i^{PM}\right)\left(\beta_F^{'}X_i^{MF}+\beta_M^{'}X_i^{MM}\right)\right]$$

and solving we get

$$\rho=\left(\beta_F^{'}\right)^2\gamma_{PF}^{MF}+\left(\beta_M^{'}\right)^2\gamma_{PM}^{MM}+\beta_F^{'}\beta_M^{'}\left(\gamma_{MF}^{PM}+\gamma_{MM}^{PF}\right) \tag{20}$$

Our estimator can be seen as a generalization of both IV in cross-sectional data and time series data.[11] Equation (15) is as a standard equation with omitted variables in cross-sectional data where X_i^C is the dependent variable, X_i^F is the regressor and X_i^M is the omitted variable. The standard solution in such a case would be to use an instrumental variable that is correlated with the regressor X_i^F but not with the omitted variable X_i^M. What we do here instead is to use a variable that is correlated with the omitted variable X_i^M, and estimate that correlation. For example, equation (17) is equivalent to using X_i^{MF} as an instrument in equation (4). However, instead of the usual assumption that the instrument is uncorrelated with the omitted variable, e.g., $\mathbb{E}\left[X_i^{MF}X_i^M\right]=0$, our model indicates that this correlation is a function of the structural parameters, e.g., $\mathbb{E}\left[X_i^{MF}X_i^M\right]=\beta_F^{'}+\rho'\beta_M^{'}$. 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, instead of assuming that it is zero.

Mathematically, all six moments, i.e., equations (15), (16), (17), (18), (19), and (20), are generated using exclusion restrictions and the data. However, there is a qualitative difference between the first three and the last three moments. In the first three moments, i.e., equations (15), (16), and (17), the exclusion restrictions come from using the error term of a person and the status of their father, e.g., ε_i^C and X_i^F; ε_i^F and X_i^{PF}; and ε_i^M and X_i^{MF}. These moments are analogous to a time series model that follows an $AR(1)$ process. In that case the model is $Y_t=\beta Y_{t-1}+\varepsilon_t$ and the exclusion restriction is $\mathbb{E}\left[Y_{t-1}\varepsilon_t\right]=0$. When $\beta_M^{'}=0$ our model is identical to that $AR(1)$ process. This is the implicit assumption in Becker and Tomes (1979, 1986) and most of the literature after that. In that sense, time series is a particular case of econometric trees where only the patrilineal effects matter. 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

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

time series, where we also use the matrilineal *lags* in the data, e.g., X_i^{MF}.

There are no nuisance parameters in equations (15), (16), and (17). 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. Notice that, assuming $\rho = \rho'$ does not provide a new independent equation, because it would make equations (15) and (16) collinear. When the parameters are constant across two generations, i.e., $(\beta_F, \beta_M, \rho) = (\beta_F^{'}, \beta_M^{'}, \rho')$, equations (18), (19), and (20) contain the nuisance parameters $\left(\gamma_{MF}^{PM}, \gamma_{PM}^{MM}, \gamma_{MM}^{PF}\right)$, as well as the parameters relating to the child'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 the nuisance parameters.[12]

Combining equations (16) and (17) we can express β_M as a function of β_F and empirical moments

$$\beta_M m_{PF}^F = m_{MF}^C - \beta_F m_{MF}^F \tag{21}$$

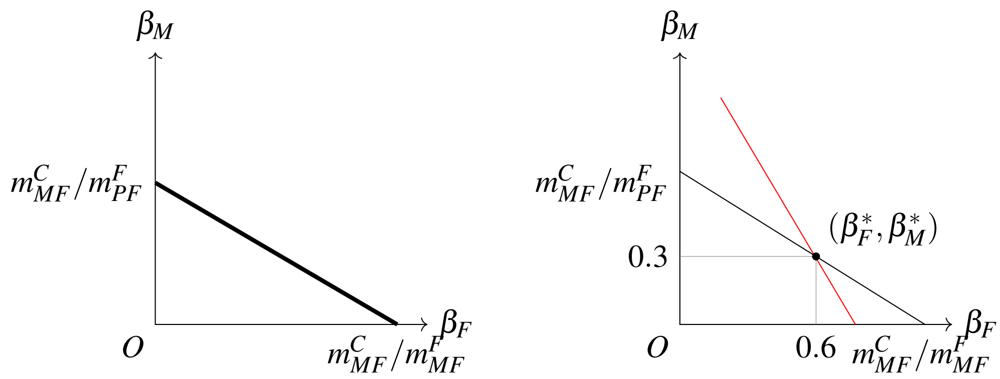
Thus, we can use two identifying equations to get an equation that is a linear combination of β_F and β_M alone. Figure 3 (left) shows an equation that depends only on β_F and β_M and data, by combining equations (16) and (17). Notice that all the other moments depend on ρ, ρ', and the nuisance parameters. Thus, without further assumptions, this line is our identified set for β_F and β_M. We now show that we can get point identification on $(\beta_F, \beta_M, \rho', \rho)$ when we only make a mild assumption on the nuisance parameters, relaxing assumptions A1. This would be useful for settings where there are arranged marriages, or the status of the parents of the groom or bride have a direct effect on mating, beyond the status of the groom and bride. When the econometrician has access to 3-generation data on male socioeconomic status $\left(X_i^C, X_i^F, X_i^{PF}, X_i^{MF}\right)$, we have four variables, we can compute six empirical moments: $\left(m_{PF}^{MF}, m_F^C, m_{PF}^F, m_{MF}^C, m_{MF}^F, m_{PF}^C\right)$ (see Figure 2). Thus, we only need one restriction on the nuisance parameters to get point identification. To illustrate this point, we assume that the correlation among grandparents across genders is equal to the product of the standard deviations, i.e., $\gamma_{MF}^{PM} = \gamma_{MM}^{PF}$. This is, in our opinion, the weakest assumption on the nuisance parameters. Proposition 4 shows how this assumption generates point identification.

Proposition 4. *Suppose X_i^C, X_i^F, X_i^{PF} and X_i^{MF} are observed. If we assume $\gamma_{MF}^{PM} = \gamma_{MM}^{PF}$, then $(\beta_F, \beta_M, \rho', \rho)$ is point identified.*

Combining equations (16) and (17), 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 3. Assuming $\gamma_{MF}^{PM} = \gamma_{MM}^{PF}$, and combining equations (18) and (19), we get an independent equation that depends only on β_F and

[12] The intuition also works conversely: if we can identify $\left(\gamma_{MF}^{PM}, \gamma_{PM}^{MM}, \gamma_{MM}^{PF}\right)$, then we can identify (β_F, β_M, ρ). This intuition is the basis for the proofs of the propositions below.

Figure 3: Set Identification and Point Identification



Notes: In this example, we use $\beta_F = 0.3$, $\beta_M = 0.6$, and $\rho = 0.5$. Simulating data using this set of parameteres generates a set of correlations that we insert in our estimating moments. Panel A. The equation comes from using equations (16) and (17). The equation is $\beta_M = \frac{m^C_{MF}}{m^F_{PF}} - \frac{m^F_{MF}}{m^F_{PF}}\beta_F = 0.62 - 0.59\beta_F$. Panel B. Black line is the same as in panel A. The second equation (red line) comes from using equations (18) and (19) and the assumption $\gamma^{PM}_{MF} = \gamma^{PF}_{MM}$. The resulting equation is $\beta_M = \frac{m^C_{PF}}{m^F_{MF}} - \frac{m^F_{PF}}{m^F_{MF}}\beta_F = 1.32 - 1.71\beta_F$.

β_M and data. This equation is the red line on the right panel on Figure 3. Thus, using these four moments and the restriction on the nuisance parameters, we can identify β_F and β_M. Identification on ρ' and ρ follows using the other moments.[13]

Proposition 5 shows identification results when we do not observe X^{PF}_i, but assume A1 and that parameters are constant across generations. Observing only X^C_i, X^F_i, and X^{PF}_i will not produce identification of all structural parameters, even with the same assumptions as Proposition 5. We can see this by noting that the moment generated by $\mathbb{E}\left[X^C_i X^F_i\right]$ is colinear with equation (9). This demonstrates the importance of the male relatives on the mother side. Maternal grandfathers and maternal uncles can be used as instruments to obtain identification. Paternal grandfathers (or paternal uncles) alone would not help in identifying all parameters.

Proposition 5. *Suppose X^C_i, X^F_i, and X^{MF}_i are observed. If assumption A1 holds and $\rho' = \rho$, then (β_F, β_M, ρ) is point identified.*

Summarizing, in Proposition 1, we use Assumption A1 but let the mobility parameters vary across generations. In Proposition 4 we relax Assumption A1, but impose that the mobility parameters are constant across two consecutive generations. Proposition 5 relaxes the data requirement in Proposition 4, but requires similar assumptions as Proposition 1. Appendix A.5 discusses several intermediate cases when we make assumptions weaker than A1 about the nuisance parameters,

[13]Proposition 4 implicitly identifies all the nuisance parameters, too. With $\gamma^{PM}_{MF} = \gamma^{PF}_{MM}$, this means that the six empirical moments produce a vector of six parameters $\left(\beta_F, \beta_M, \rho', \rho, \gamma^{PM}_{MF}, \gamma^{MM}_{PM}\right)$.

and we impose some restrictions on mobility parameters across time, without imposing they are constant.

3.3 Identification Using Maternal Uncles

In this section, we discuss identification when we have information on one maternal uncle (X_i^{MU}). In addition to the six moments generated using $\left(X_i^C, X_i^F, X_i^{MF}, X_i^{PF}\right)$, we can get additional moments if we have information on one maternal uncle. Proposition 6 below shows that we can get identification of all structural parameters, including household effects η, without any additional assumptions. Typically one needs to observe the status of two siblings in order to estimate household 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 X_i^C, X_i^F, X_i^{MU}, and X_i^{MF} are observed. Then $(\beta_F, \beta_M, \rho', \rho, \eta)$ is point identified.*

Proposition 6 uses only four variables and generates six equations and six empirical moments $\left(m_F^C, m_{MF}^C, m_{MF}^F, m_{MU}^F, m_{MU}^{MF}, m_{MU}^C\right)$. We use six equations, where there are no nuisance parameters and estimate five structural parameters $(\beta_F, \beta_M, \rho', \rho, \eta)$. 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 η, 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. Of course, requiring information on a maternal uncle would restrict the sample to families where the mother had a brother.

Proposition 7. *Suppose X_i^C, X_i^F, X_i^{MU} and X_i^{PF} are observed. Then $(\beta_F, \beta_M, \rho', \rho, \eta)$ is point identified.*

Proposition 7 shows that we can identify all the parameters of interest using X_i^{PF} instead of X_i^{MF}. 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 $\left(X_i^C, X_i^F, X_i^{MU}\right)$, when we assume there are no household effects. 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 or any equation involving grandparents.

Proposition 8. *Suppose X_i^C, X_i^F, X_i^{MU} are observed. If assumption A5 holds and $\rho' = \rho$, then (β_F, β_M, ρ) is point identified.*

3.4 Gendered Effects

In this subsection, we present identification results when we allow for gendered effects according to equation (8). Proposition 9 shows that we can identify all six parameters of interest in the model with gendered effects when Assumption 1 holds. This is a somewhat surprising result. In the previous propositions we were able to identify the effect of the mother on her child, without observing her outcome X_i^M. Proposition 9 goes one step further and shows that we are able to identify the effect of mothers on daughters β_M^D without observing the outcome of either one of them (X_i^{MM} and X_i^M).

Proposition 9. *Suppose X_i^S, X_i^F, X_i^{PF}, and X_i^{MF} are observed. If assumption A1 holds, then $\left(\beta_F^S, \beta_M^S, \beta_F^D, \beta_M^D, \rho', \rho\right)$ is point identified.*

4 Simulation Results

We now present the structural model estimates using simulated data, using the estimators developed in Section 3. 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 a weighting matrix. We simulate outcomes for the grandparents using a multivariate normal distribution with variance-covariance matrix Σ as a function of $\left(\gamma_{PF}^{MF}, \gamma_{MF}^{PM}, \gamma_{PM}^{MM}, \gamma_{MM}^{PF}; \rho'\right)$. We also simulate idiosyncratic normal random shocks for the father, mother, and child $\left(e_i^F, e_i^M, e_i^C\right)$ with zero mean. We use the structural parameters to create the appropriate standard deviation for the shocks, e.g., $\sigma_S = \sqrt{1 - \beta_F^2 - \beta_M^2 - 2\rho'\beta_F\beta_M}$. We then use equations (4), (5), and (6) to forward simulate the outcome of the father, mother, and child.[14]

For the analysis we compute both the standard deviation (across simulations) of the estimated parameters and the median of the estimated standard errors calculated by the GMM asymptotic variance formula. In each analysis we run 1,000 simulations, and we vary the sample size n.[15]

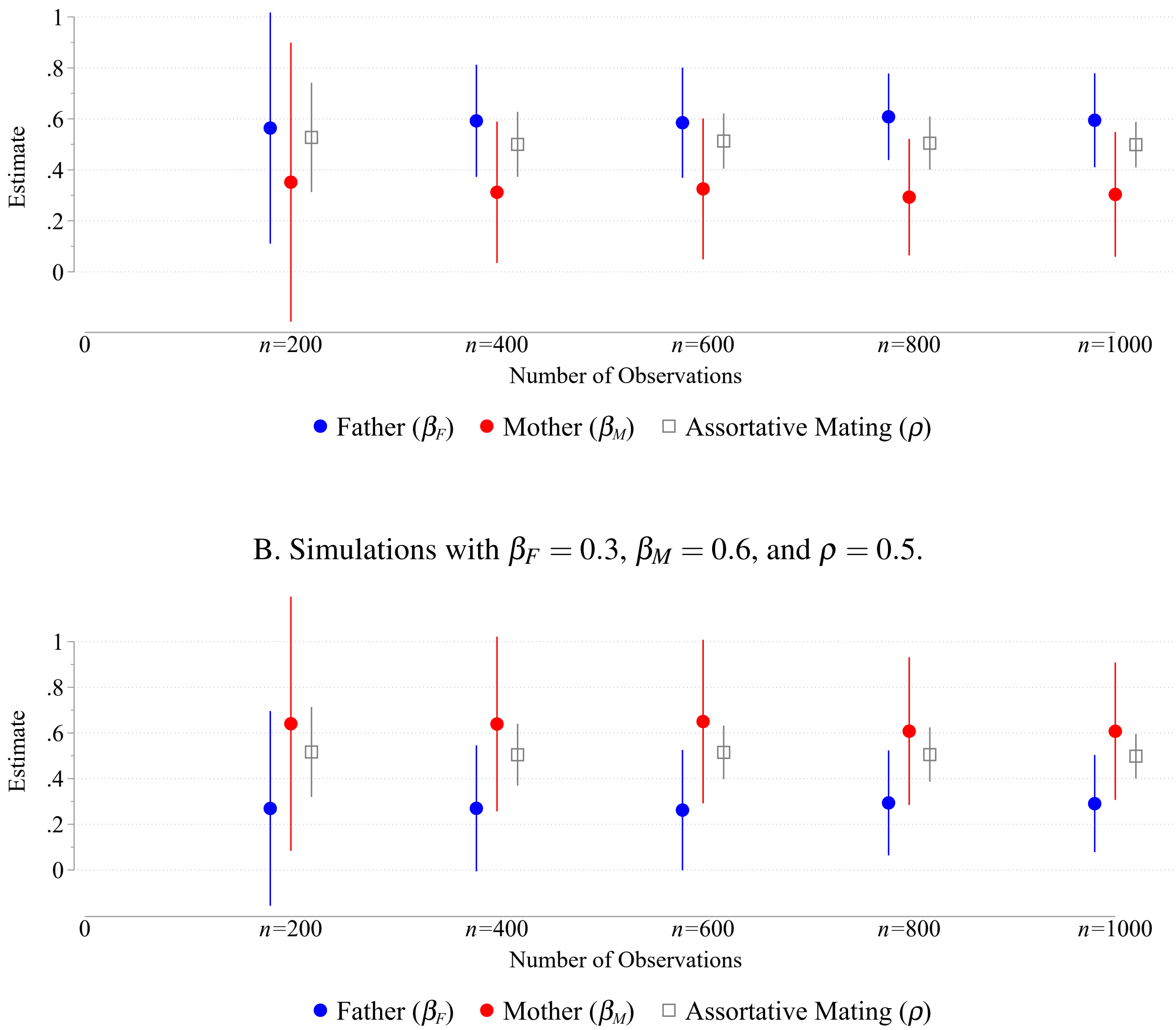
Figure 4 shows the estimation results using our estimator for Proposition 1 in simulated samples of size ranging 200-1000. The sample parameters are $\beta_F = 0.6$, $\beta_M = 0.3$, and $\rho = 0.5$ for panel A and $\beta_F = 0.3$, $\beta_M = 0.6$, and $\rho = 0.5$ for panel B. In panel A, we can see how the point estimates are always very close to the true value. At $n = 400$ all parameters are statistically different than zero, and β_F is statistically different than β_M. In panel B, the effect of the mother is larger than that of the father. That makes it more difficult for the estimator to estimate the errors more precisely,

[14] The simulation for the analysis with the maternal uncle is similar. We simulate the maternal uncle's outcome X_i^{MU} with a correlation among siblings of ρ_{MU} and get $\eta = \sigma_S^2 \rho_{MU}$.

[15] In Appendix B we also show results using both the standard deviation (SD) of the estimated parameter across simulations and the median of the estimated standard errors (SE) calculated by the GMM asymptotic variance formula.

Figure 4: Simulation Results from Proposition 1.

A. Simulations with $\beta_F = 0.6$, $\beta_M = 0.3$, and $\rho = 0.5$.



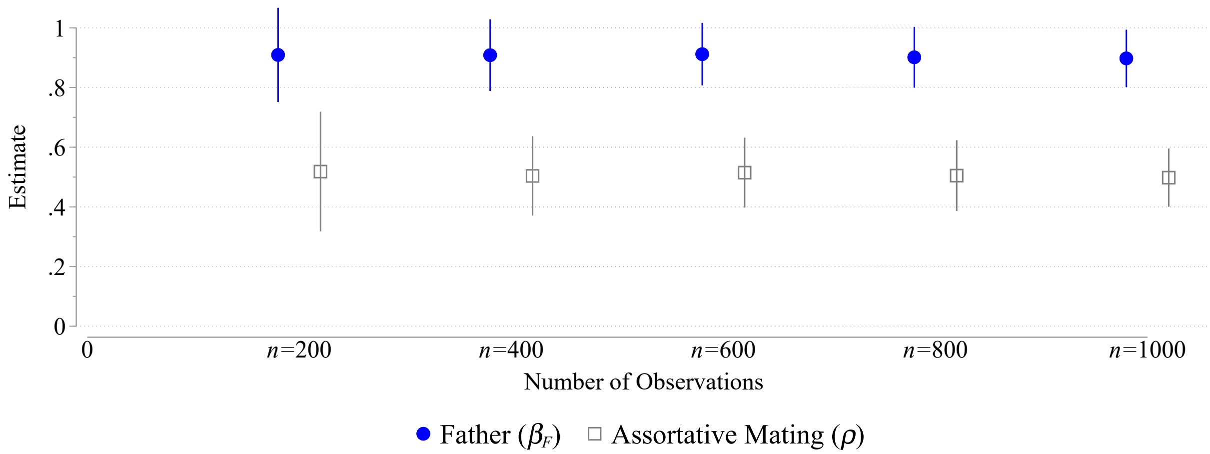
Notes: Estimated parameters and confidence intervals at $p = 0.05$ using Proposition 1.

due to the way each parameter enters the moment. Even in this case, all parameters are statistically different than zero at $n = 400$, but we need to go to $n = 600$ to see β_F statistically different than β_M.

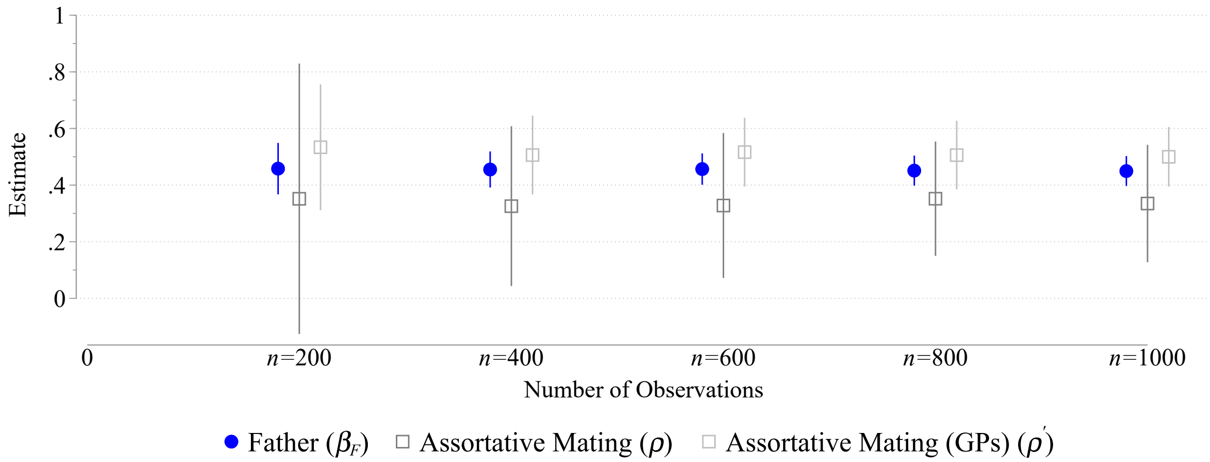
Figure 5 shows the estimation results using our estimator for Propositions 2 and 3 in simulated samples of size ranging 200-1000. Notice that the model is intentionally mispecified here. The sample parameters are $\beta_F = 0.6$, $\beta_M = 0.3$, and $\rho = 0.5$ for both panels, but the estimator assumes $\beta_M = 0$ in panel A (Proposition 2) and $\beta_M = \beta_F$ in panel B (Proposition 3). Unsurprisingly, β_F is estimated at 0.9 in panel A (the sum of the effects of both parents), and at 0.45 in panel B

Figure 5: Simulation Results from Propositions 2 and 3

A. Simulations with $\beta_F = 0.3$, $\beta_M = 0.6$, and $\rho = 0.5$, assuming $\beta_M = 0$



B. Simulations with $\beta_F = 0.3$, $\beta_M = 0.6$, and $\rho = \rho' = 0.5$, assuming $\beta_M = \beta_F$



Notes: Estimated parameters and confidence intervals at $p = 0.05$ using Propositions 2 and 3.

(the average of the effects of both parents). Regarding assortative mating, in both cases, ρ' is consistently estimated and with very small error even at small samples. Notice, however, that in panel B, ρ is inconsistently (and imprecisely) estimated and hovering below 0.4.[16]

Figure 6 shows the estimation results using maternal uncles as non-linear instruments, for Propositions 6, 7, and 8 in simulated samples of size ranging 200-1000. In panels A and B, we see that the results are very similar when we use a maternal or a paternal grandfather as non-linear instruments. In both cases all parameters but η are significantly different than zero, even at

[16] For results about Propositions 4 and 5 see Appendix B.

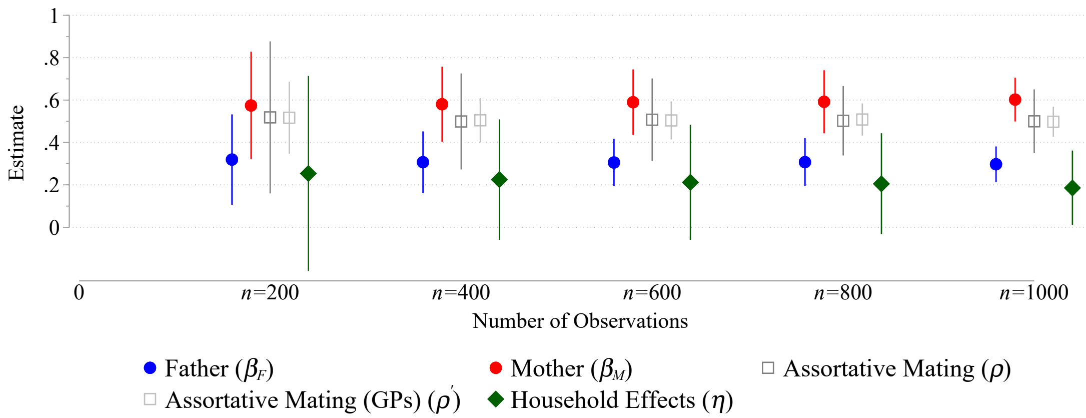
$n = 200$, and the effects of the father and the mother are significantly different at $n = 400$. The parameter measuring household effects η has the largest confidence intervals, and is only statistically significant at $n = 1,000$. Panel C shows the estimates when we mispecify the model and impose $\eta = 0$. In this case, assortative mating is consistently estimated, but the effect of the father is underestimated and the effect of the mother is overestimated. In other words, by imposing no correlation with the maternal uncle, we are over-attributing effects to the mother.

Figure 7 shows the estimation results using maternal uncles as non-linear instruments, for Proposition 9 in simulated samples of size ranging 1,000-5,000. Panels A and B show results with different parameters. In Panel A, fathers have a larger effect. In Panel B, mothers have a larger effect. In both panels we see gendered effects, i.e., the effects are stronger when the parent and the child have the same gender. We use larger samples here because some parameters do not show statistical significance at lower sample sizes. The parameters that have slower convergence are those related to the daughters: β_F^D and β_M^D. β_F^D is statistically different than zero at $n = 3,000$ in Panel A, but even at $n = 5,000$ it is not statistically significant in Panel B. Notice, however, that all parameters are always very close to the true values for all sample sizes. Looking at Figure 8, we can see why β_F^D and β_M^D have larger confidence intervals. In the tree, we can see that whereas the other mobility parameters appear twice: in the arrows leading to the child and to the father; β_F^D and β_M^D only appear once, in the arrows leading to the mother. In the six moments that we use to identify all parameters, β_F^D and β_M^D always appear multiplying other parameters (non-linearly) but β_F^S and β_M^S sometimes appear linearly (see Appendix A.4 for details).

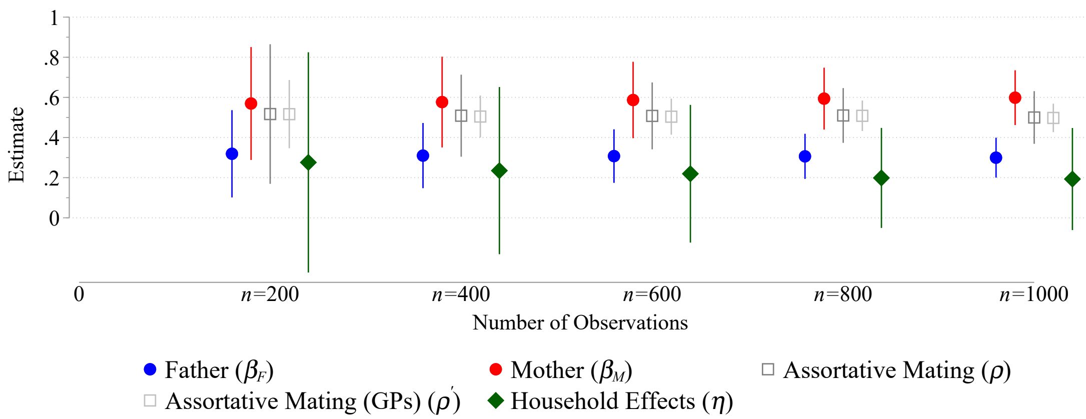
In summary, our framework allows us to estimate the effects of women (and effects on women) without observing outcome variables for them. The array of results that we provide will allow the researcher to estimate the parameters of interest under different sets of assumptions and available data. In addition to the list of econometric results summarized in Table 1 we validate the results using simulated data under different specifications, and use our GMM estimators to estimate structural parameters using these samples. In most cases the estimates converge to the true parameters at very small sample sizes. Indeed, when the model is well specified the estimates are never statistically different than the true values, even with $n = 200$, and are usually statistically different than zero at $n = 400$. This suggests that our results could be applied to many other settings when the researcher have access to datasets with small sample sizes. Moreover, our propositions allow for a variety of assumptions regarding the structural parameters, the nuisance parameters, and household effects. But how do we choose among these assumptions? One answer is that the choice is context-dependent. Another is that we can sometimes rely on ancillary information on outcomes other than the outcome that is the primary focus.

Figure 6: Simulation Results from Propositions 6, 7 and 8

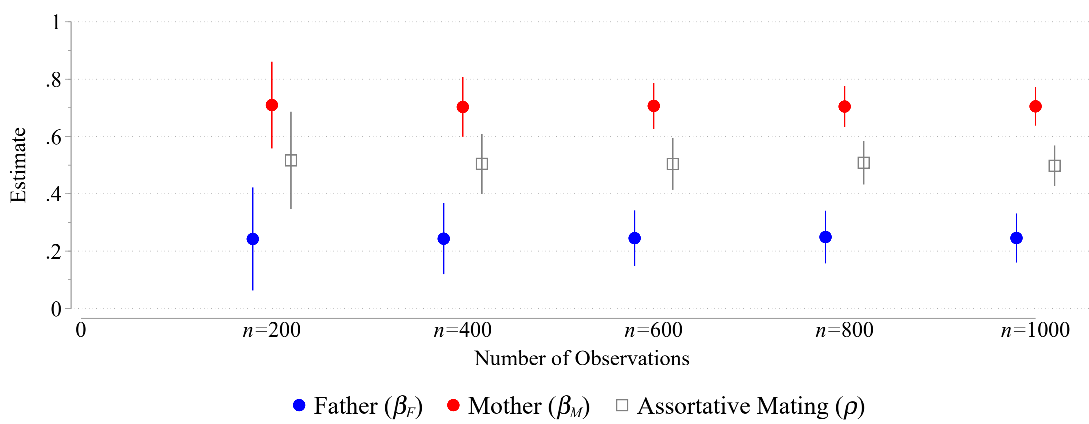
A. Simulations with $\beta_F = 0.3$, $\beta_M = 0.6$, $\rho = \rho' = 0.5$, and $\eta = 0.185$, using X_i^{MF}



B. Simulations with $\beta_F = 0.3$, $\beta_M = 0.6$, $\rho = \rho' = 0.5$, and $\eta = 0.185$, using X_i^{PF}



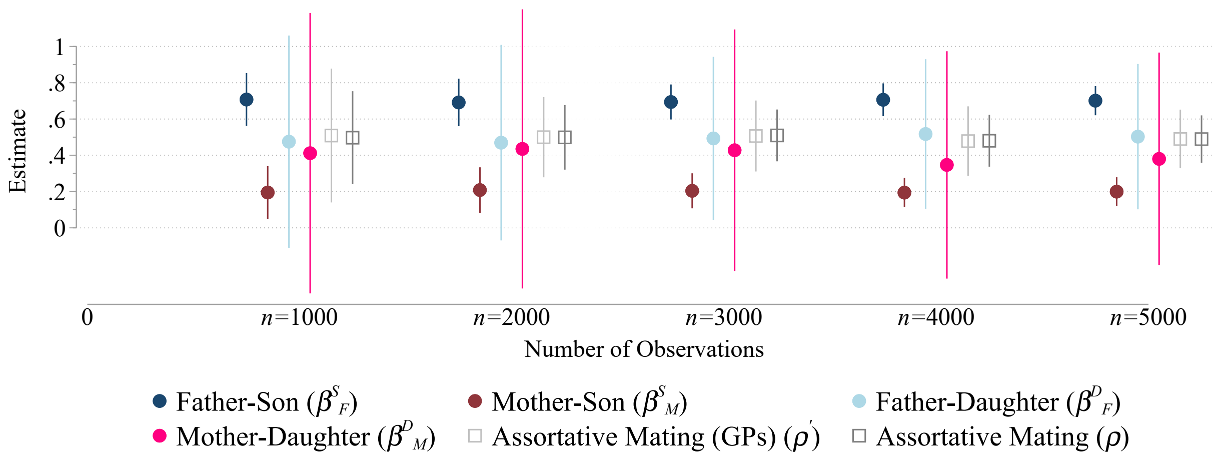
C. Simulations with $\beta_F = 0.3$, $\beta_M = 0.6$, $\rho = \rho' = 0.5$, and $\eta = 0.185$, assuming A5.



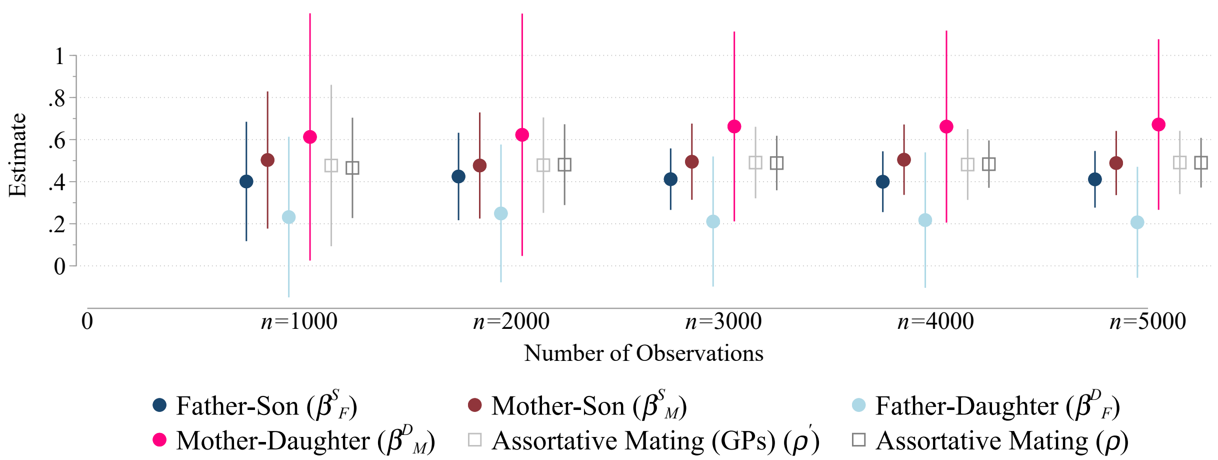
Notes: Estimated parameters and confidence intervals at $p = 0.05$ using Propositions 6, 7, and 8. Assuming $\rho_{MU} = 0.5$ we get $\eta = \sigma_S^2 \rho_{MU} = (1 - \beta_F^2 - \beta_M^2 - 2\rho' \beta_F \beta_M)\rho_{MU} = 0.185$.

Figure 7: Simulation Results from Proposition 9

A. Simulations with $\beta_F^S = 0.7$, $\beta_M^S = 0.2$, $\beta_F^D = 0.5$, $\beta_M^D = 0.4$, and $\rho = \rho' = 0.5$



B. Simulations with $\beta_F^S = 0.4$, $\beta_M^S = 0.5$, $\beta_F^D = 0.2$, $\beta_M^D = 0.7$, and $\rho = \rho' = 0.5$



Notes: Estimated parameters and confidence intervals at $p = 0.05$ using Proposition 9.

5 Empirical Application

5.1 Data

We now employ our econometric framework to assess intergenerational mobility in the U.S. from 1870 to 1940. This requires linkages of families across two (parent-child) or three (grandparent-parent-child) generations, with an outcome measure that is defined and measured consistently across a span of 70 years. We use the 100% U.S. Census of Population manuscript files 1870-

1880 and 1900-1940 as the basis for our linkages (Ruggles et al., 2024a), and follow recent practice in historical studies of intergenerational mobility to generate income measures for our linked individuals.[17]

5.1.1 Family Tree Construction

To construct family trees, we use the Buckles et al. (2023a) linkages of individuals in the censuses. This collection is the largest, most comprehensive, and most accurate set of census linkages ever generated. It relies on both algorithmic linkage and the Family Search collection of genealogies. We rely on three family trees: (1) 2-generation trees which contain a married couple (F, M), his parents (PF, PM), and her parents (MF, MM); (2) 3-generation trees which contain an adult male child (S), their parents (F and M), both of their paternal grandparents (PF, PM), and both of their maternal grandparents (MF, MM); and (3) The 3-generation trees but additionally with a maternal uncle (MU).

To start, we take the 1880 Census and every individual who is listed as a male child in a household with their parents present.[18] The starting population is 11,851,814 individuals. Of these, 6,683,159 (56.4%) can be linked to themselves 30 years later, in the 1910 Census. Of these individuals, 4,987,881 (74.6%) are married in 1910 with a spouse present. We then use the wife's age to look for her as a child, and hence find the father-in-law of the adult male F originally observed in 1880, in the Census in which she would have been a child less than 10 years old. This step results in 1,745,280 separate trees (35.0% of the brides being able to be traced to the father-in-law).

The second set of trees is constructed by starting with the 2-generation trees, taking the 2,686,082 children (S) of the 1,745,280 fathers and mothers, and searching for them in the 1940 Census. We find 1,991,768 of these (74.4% of the total), which completes the 3-generation trees. Finally, we take the maternal grandfathers (that is, the fathers-in-law in the two generation trees), identify their male children, and search for them 30 years later. This generates the 3-generation-with-uncle trees, of which we have 2,704,830.[19]

Ward (2023) shows that measurement error can substantially bias estimates of intergenerational mobility. To account for this, for each matched individual, we look for them one Census earlier and/or later, depending on the availability of the Census. We require that an individual have at least two observations with a valid occupation in these Censuses, in which case we take the two

[17] The manuscript schedules of the 1890 census were destroyed, so there is no 100% file for this census.

[18] These are restricted-access data available from the IPUMS project at the Minnesota Population Center. By child, we mean that that they are recorded as a child of the head-of-household. They need not be a minor, though in practice the vast majority are. We assess the sensitivity of our results to starting our family trees from 1880 in Appendix C.

[19] Adding maternal uncles can make the number of trees either smaller or larger. In particular, if two brothers each have two sons, that makes four possible trees, as a tree in this case is defined by a son-uncle dyad.

observations closest to age 40. We assess the sensitivity of our results to this in Appendix C.

5.1.2 Occupational Incomes

The outcome most often examined in studies of intergenerational mobility is income. The U.S. Census did not begin to collect data on the income of respondents until 1940. The U.S. Census did, however, report age, sex, race, location, and occupation for all respondents from 1850 forward. Several recent studies have followed a common practice to generate pre-1940 income measures: predicting incomes for pre-1940 censuses based on the relationship between 1940 income and a set of characteristics observed in the census, where the latter are available in 1940 and before (Abramitzky et al., 2021; Collins and Wanamaker, 2022; Jácome et al., 2025). The relationship we estimate and use for prediction is described by the regression

$$\ln W_i = \sum_{s=1}^{S} \delta_s ST_s + \sum_{a=1}^{A} (\delta_a AGE_a + \gamma_a AGE_a^2) + \sum_{r=1,i=1}^{R,I} \delta_{r,i} REG_r \times CLA_i + \eta_{REG} + \eta_{CLA} + e_i$$

where W is income in 1940 and ST, AGE, REG, and CLA are, respectively, fixed effects for state of residence, age, fixed effects for Census region of residence, and fixed effects for one-digit occupational categories (derived from the three-digit numeric code assigned using the 1950 Census coding scheme which is in turn a standardization of the reported occupational title; e.g. "blacksmith"=501).[20]

There are three adjustments we make to income, again in line with past practice, before estimating this relationship:

1. The 1940 Census reports only wage and salary income, so self-employment income is excluded. Self-employment income is, however, reported in 1950 in addition to wage and salary income. For all occupations except farmers, we calculate the 1950 ratio of self-employment income to wage and salary income for each state and three-digit occupational category. We then use these ratios to calculate each respondent's total 1940 income (wage and salary income + self-employment income). For occupations with no wage and salary income in 1940 (e.g. proprietors such as grocers), we estimate their total income in 1940 as the 1950 self-employment income in their state/occupation cell, deflated to 1940 prices.

2. Farmers' incomes are complicated not just by the fact that they are generally self-employed and therefore do not have 1940 wage and salary income but also by their receipt of income in-kind from their farms (e.g. food, fuel, lodging) and by the nature of their income as the net product of a business enterprise with a variety of inputs and outputs. To account for these

[20]This regression-based adjustment approach has been shown to yield more accurate measures of income for the analysis of intergenerational mobility than other approaches (Saavedra and Twinam, 2020).

complications, we follow Abramitzky et al. (2021) and Collins and Wanamaker (2022) in assuming that the relationship between farmers' incomes and the incomes of farm laborers is relatively stable. We calculate the ratio of farmers' 1950 total income to farm laborers' 1950 wage and salary income by state and county and use this ratio together with average 1940 farm laborers' wage and salary income by state and county to assign farmers a 1940 income.[21]

3. Two occupations (farm laborer, and farm manager) received a significant part of their income as in-kind payments (e.g. food, lodging, garden plots). The USDA estimated the total value of these perquisites by region for both 1925 (Folsom, 1931) and 1945 (Reagan, 1946). We used the employment-weighted mid-point of these regional figures to scale up the 1940 wage and salary incomes of farm laborers and farm managers and added this to their 1940 total income calculated in (1) above.

5.2 Empirical Results

In Table 2, we show the results from the estimation of Proposition 1. Recall that this estimation requires data from a male individual, his father, and his father-in-law. The transmission from the maternal side is considerably stronger than that from the paternal side.[22] Table 2 also shows the results from the estimation of Propositions 2 and 3. The results are identified from 2-generation data, but imposing stronger assumptions on the parameters than those usually found in the literature. Proposition 2 assumes that there is no maternal effect ($\beta_M = 0$). Thus, all the mobility effect is attributed to the father, resulting in an estimate for the paternal effect that is the sum of the two effects using Proposition 1. Proposition 3 assumes that maternal and paternal effects are identical ($\beta_F = \beta_M$), resulting in an estimate for each of the effects that is the average of the two effects using Proposition 1. The results for Propositions 1 through 3 yield an assortative matching measure for the parents that is neither perfect nor random ($\rho = 0.339$). This means that excluding mothers

[21] In making this adjustment, both Abramitzky et al. (2021) and Collins and Wanamaker (2022) used the the 5% sample from the 1960 Census to calculate this ratio. The 5% sample from the 1950 Census provided too few observations as income in that year was a "sample line" question asked of only 20% of respondents, a substantial fraction of whom were under age 18. We instead use the recently released 100% 1950 Census file as the source for the ratio of farm laborer income to farmer income, and are able to do this both for each of 3,100 state/county cells rather than states or regions as they used and for a census year (1940) closer to our own linked data.

[22] These results for β_F and β_M are in line with the structural estimations in Eriksson et al. (2023) using Massachusetts marriage registers, although we find stronger assortative mating. One possible reason for the difference is that mating takes place within local geographies, and so measured relative to the national distribution, spouses are more alike than measured relative to local distributions. We formally test the restriction used in the literature that $m_{PF}^{MF} = m_{PF}^{F} m_{MF}^{F}$. We use a Fisher's z-transformation. The value of the Z-statistic is 261.35 and the p-value is smaller than 10^{-11}. We strongly reject that this restriction is valid in our sample.

Table 1: Summary of Identification Results

Proposition	Data	Literature	Assumptions	Parameters
Identification using 2-generation data				
Proposition 1	(F, PF, MF)	A1, A2	$\rho = \rho'$	(β_F, β_M, ρ)
Proposition 2	(F, PF, MF)	A2, A3		(β_F, ρ)
Proposition 3	(F, PF, MF)	A1, A2, A4		(β_F, ρ', ρ)
Identification using 3-generation data				
Proposition 4	(S, F, PF, MF)	A2	$\gamma_{MF}^{PM} = \gamma_{MM}^{PF}$, $(\beta_F, \beta_M) = (\beta_F^{'}, \beta_M^{'})$	$(\beta_F, \beta_M, \rho', \rho)$
Proposition 5	(S, F, MF)	A1, A2	$\Theta = \Theta'$	(β_F, β_M, ρ)
Identification from maternal uncles				
Proposition 6	(S, F, MU, MF)		$(\beta_F, \beta_M) = (\beta_F^{'}, \beta_M^{'})$	$(\beta_F, \beta_M, \rho', \rho, \eta)$
Proposition 7	(S, F, MU, PF)		$(\beta_F, \beta_M) = (\beta_F^{'}, \beta_M^{'})$	$(\beta_F, \beta_M, \rho', \rho, \eta)$
Proposition 8	(S, F, MU)	A5	$\Theta = \Theta'$, $\eta = 0$	(β_F, β_M, ρ)
Identification allowing heterogeneous effects by gender				
Proposition 9	(S, F, PF, MF)	A1, A2		$(\beta_F^S, \beta_M^S, \beta_F^D, \beta_M^D, \rho', \rho)$

Notes: $\Theta \equiv (\beta_F, \beta_M, \rho)$ refers to the full vector of parameters. Assumption A2 in Proposition 3 is only needed to identify ρ.

from the analysis has two consequences: the estimator for the father's effect provides neither a consistent estimator of the combined effect of both parents nor the actual effect of the father alone.

The first column uses only one assumption: that the four nuisance parameters measuring the correlations across the two sets of grandparents are identical. The measured mother's effect in this case is more than three times the measured father's effect, and both are precisely estimated. If we are interested in understanding causal pathways from parental income to child's income, this finding provides strong evidence that it is likely much more than just the family's material resources that matter in producing outcomes in the next generation – even if the mother is not working outside the home, she contributes something to the child's earning capacity that is much more valuable than what the father's actual income provides.

The second column assumes that there is no "mother effect" and only the father matters. The third column assumes both that the nuisance parameters are identical and that the impact on the child's income is identical for both parents. These are all plausible assumptions, but some might be appropriate in some contexts but not others. The key insight the analysis provides is how the estimates vary dramatically with the particular assumptions underlying them.[23]

In Table 3, we turn to Propositions 4 through 9. Again, varying the assumptions results in different estimates. With 3-generation data (Propositions 4 and 5 in the first and second columns), going from assuming only that the within-marriage correlation is the same for both sets of grandparents to again assuming that both the two within-marriage male-female correlations and the two cross-marriage male-female correlations are identical reduces the gap between the paternal and maternal effects. Moving to the use of maternal uncles (Propositions 6 through 8 in the third, fourth, and fifth columns) which imposes the least restrictive assumptions (none in the first two cases and only $\eta = 0$ so no household fixed effects in the third) yields parental effects broadly similar to those in Proposition 5 (second column). Finally, when we allow the effect to differ by the sex of the child (Proposition 9, last column), the mother's impact is greater than the father's for both sons and daughters and the mother's effect on the son is substantially greater than her effect on the daughter.

This pattern suggests that whatever it is that mothers convey in income-generating capacity is provided in greater quantity to both sons and daughters than what the father provides to them. At the same time, if what is provided to both sons and daughters is provided in equal quantity to each, the son derives greater benefit. A mother's provision of human capital in the home simultaneously to both sons and daughters that later in their lives has a larger return in the labor market than in home production would fit this pattern of coefficients.

We are not prepared to say which of the specifications in Tables 2 and 3 we prefer. This exercise is undertaken not to find the single correct measure of intergenerational mobility, but rather to demonstrate how different sets of assumptions generate different sets of results. Some assumptions will be more appropriate in some contexts than in others. What we can unambiguously conclude, however, is that the omission of mothers from this sort of analysis leaves out a potentially substantial part of the process of the intergenerational transmission of (dis)advantage.

6 Discussion

We now discuss the implications of our model and the problems that may arise when a model is mispecified in this context. Subsection 6.1 discusses the main implicit assumptions in the literature and the problem that arises when these models are mispecified but our model is correct. Subsection 6.2 discusses potential extensions to our model, and how researchers could address whether the extended models are correct and our model is mispecified.

[23] With the estimated parameter vector and the corresponding variance-covariance matrix in the first column, we compute a Wald statistic to test the hypothesis $H_0 : \beta_F = \beta_M$. Under the null hypothesis, this statistic follows a chi-square distribution with one degree of freedom. The value of the Wald statistic is $2,834.27$ and the null hypothesis is rejected with a p-value smaller than 10^{-10}.

Table 2: Empirical Results for Propositions 1, 2, and 3

Parameter	Prop. 1	Prop. 2	Prop. 3
β_F	0.194	0.868	0.427
	(0.003)	(0.002)	(0.001)
β_M	0.674		
	(0.005)		
ρ	0.339	0.339	0.339
	(0.002)	(0.002)	(0.001)
N	817,962	817,962	817,962
MSE	3.534e-16	7.538e-16	6.179e-05

Notes: Empirical results from the estimation of Propositions 1, 2, and 3, for 1880 cohort. A cohort is defined as any male child in the Census for the specified year; hence, there can be some overlap in the individuals in these trees.

6.1 Model Mispecification in the Literature

We want to emphasize the importance of being explicit about the assumptions that the researcher is making. If the goal of the researcher is to describe correlations in the data, and how they change over time and space, then the use of raw correlations is appropriate. However, if the researcher wants to asses causality or to test a particular model, empirical correlations commonly used as "proxies" in the literature could create biased estimates. Moreover, these biases may change over time, making claims about trends invalid (Olivetti et al., 2024). We show below how many common, and usually unremarked upon, assumptions made in the literature would produce unbiased estimates only in extreme cases, such as $\beta_M = 0$ and $\rho = 1$.

Only Male Effects ($\beta_M = 0$) This is the most common case in the literature. As explained in Section 2, researchers would use the empirical correlation between father and child (m_S^F) and interpret it as the effect of the father on the child (β_F). However, in our model, this is only true when mothers do not contribute to their children's outcome so that $\beta_M = 0$. The bias here (*bias* $\equiv \mathbb{E}\left[X_i^F X_i^C\right] - \beta_F = \rho\beta_M$) is increasing on the degree of assortment ρ and in the effect of mothers β_M.

Gender Neutral ($\beta_M = \beta_F$) This is another common assumption in labor economics, where the sum of the incomes of both parents is used which assigns the same weight to the mother's and father's contribution to the child's outcome (Jácome et al., 2025; Chetty et al., 2014). This

Table 3: Empirical Results from Propositions 4-9

Param.	Prop. 4	Prop. 5	Prop. 6	Prop. 7	Prop. 8	Param.	Prop. 9
β_F	0.303	0.189	0.199	0.195	0.175	β_F^S	0.200
	(0.020)	(0.004)	(0.002)	(0.003)	(0.004)		(0.002)
						β_F^D	0.105
							(0.004)
β_M	0.312	0.547	0.451	0.479	0.514	β_M^S	0.691
	(0.020)	(0.008)	(0.004)	(0.006)	(0.013)		(0.004)
						β_M^D	0.440
							(0.007)
ρ	0.215	0.332	0.381	0.381	0.381	ρ	0.246
	(0.049)	(0.002)	(0.001)	(0.001)	(0.001)		(0.002)
$\rho^{'}$	0.579		0.551	0.644		$\rho^{'}$	0.410
	(0.027)		(0.005)	(0.005)			(0.003)
η			0.051	0.000			
			(0.002)	(0.004)			
N	575,170	575,170	586,931	586,931	586,931		575,170
MSE	4.381e-14	9.163e-16	2.819e-16	6.050e-04	1.357e-16		2.909e-16

Notes: Results from the estimation of Propositions 4, 5, 6, 7, 8, and 9, for 1880 cohort. A cohort is defined as any male child in the census for the specified year; hence, there can be some overlap in the individuals in these trees.

assumption is motivated by only having household outcome data, but not individual outcomes for the father and the mother. Our Proposition 3 shows the consequences of this assumption. The estimated β is usually between the true values of β_M and β_F. These maternal and paternal contributions need not be equal, nor constant over time. For example, Brandén et al. (2024), using modern data from both the US and Sweden, find an increasing contribution of mothers to overall persistence over recent decades. They find a larger contribution of fathers, attributable in part to the greater variation of the income of fathers.

Perfect Assortment ($\rho = 1$) This assumption is usually combined with the previous one ($\beta_M = \beta_F$). The researcher implicitly assumes that the status of a wife is fully determined by the status of her husband ($\rho = 1$). In that case, one can use the correlation between the husband (father) and the wife's father (maternal grandfather) $\mathbb{E}\left[X_i^F X_i^{MF}\right]$ to measure the effect of the wife's father on her ($\beta_F^C \neq \beta_F^D$). Propositions 9 shows the results when the researcher assumes that the effect of each parent is different depending on the gender of the child, but without assuming perfect assortment. $\rho = 1$ is a very strong assumption.

Proxies for assortative mating Many papers use $\mathbb{E}\left[X_i^{PF} X_i^{MF}\right]$ or $\mathbb{E}\left[X_i^F X_i^{MF}\right]$ as a proxy for ρ. These proxies are biased estimates of ρ and the bias may or may not change over time. This approach is common in cases where researchers can link a man to his father-in-law, using marriage records or similar sources. Our moments in Proposition 1 state

$$\rho = \mathbb{E}\left[X_i^{PF} X_i^{MF}\right](\beta_F + \beta_M)^2$$

$$\rho = \mathbb{E}\left[X_i^F X_i^{MF}\right](\beta_F + \beta_M)$$

$\mathbb{E}\left[X_i^{PF} X_i^{MF}\right]$ and $\mathbb{E}\left[X_i^F X_i^{MF}\right]$ would only be a proxy for ρ when $\beta_F + \beta_M = 1$. In the typical case in which $\beta_F + \beta_M < 1$, we have $\rho < \mathbb{E}\left[X_i^{PF} X_i^{MF}\right]$ and $\rho < \mathbb{E}\left[X_i^F X_i^{MF}\right]$; i.e., both proxies would overestimate ρ, and in more mobile societies the bias for using proxies is larger. Moreover, when using the most common proxy $\mathbb{E}\left[X_i^{PF} X_i^{MF}\right]$, the bias grows quadratically in $1 - (\beta_F + \beta_M)$.

6.2 Model Mispecification in Our Empirical Analysis

We now discuss the sensitivity of our results to our assumptions and how they could be extended to other settings.[24] To this point, we have used the simplest model that accounts for the effects of fathers and mothers on their children. This econometric model is derived from a model similar to Becker and Tomes (1979, 1986) but having two parents, that bargain à la Nash (see Appendix E). In certain settings, the researcher may want to use a more general model. We could generalize our model to allow for non-linear parental effects, or interaction terms, as

$$X_i^C = \beta_F X_i^F + \beta_M X_i^M + \zeta(X_i^F X_i^M) + \varepsilon_i^S \tag{22}$$

where ζ measures the interaction effect (matching surplus). Edwards and Roff (2016) argue that the matching surplus has an added effect on the child. Again, since this would require more than three parameters, we would need to use a three generation tree, or maternal uncles.

We could also generalize our model to allow for direct grandparent effects. The evidence for grandparent effects is mixed. Long and Ferrie (2007) show evidence of direct grandparent effects in the US and England. Braun and Stuhler (2018) provide evidence from intergenerational correlations of educational status consistent with our model. In a variety of German samples, in regressions of educational status on that of relatives, the coefficient on grandparents falls dramatically when mothers are included directly, leading them to caution against a causal interpretation

[24] Appendix D.1 provides more details on comparative statistics and measurement error.

of grandparent effects. In a horse race of different models, they find that latent variable models perform best in out-of-sample tests. Bratsberg et al. (2023) similarly find that the inclusion of grandparent measures does not substantially change estimates of a production function for education and occupational rank of earnings of children when mothers' characteristics are directly measurable. Our 3-generation tree allow us to identify up to six parameters. Thus, we could estimate two assortment parameters (ρ and ρ'), two first-generation mobility parameters (β_F and β_M), and two second-generation mobility parameters (ξ_{GF} and ξ_{GM}) according to following equation

$$X_i^C = \beta_F X_i^F + \beta_M X_i^M + \xi_{GF}(X_i^{PF} + X_i^{MF}) + \xi_{GM}(X_i^{PM} + X_i^{MM}) + \varepsilon_i^S \tag{23}$$

Identifying this would require symmetry assumptions similar to Assumption A1.

Finally, the researcher could be concerned about that the measure of socioeconomic status is measured with error.[25] This could be because there is an unobservable component that is inheritable and affects the status of the child (see Appendix E.3 for details)

$$e_i^C = \lambda_F e_i^F + \lambda_M e_i^M + \nu_i$$

where λ_F measures the inheritability from the father and λ_M measures the inheritability from the mother. In that case, our estimating equation becomes

$$X_i^C = (\lambda_F + \beta_F) X_i^F + (\beta_M + \lambda_M) X_i^M - \lambda_F \beta_F X_i^{PF} - \lambda_F \beta_M X_i^{PM} - \lambda_M \beta_F X_i^{MF} - \lambda_M \beta_M X_i^{MM} + \nu_i \tag{24}$$

There are six variables in the right hand side (two parents and four grandparents), but there are only four parameters: the mobility coefficients β_F and β_M and the heritability coefficients λ_F and λ_M. Therefore, we could use a 3-generation tree, which generates six moments, to identify these four parameters and the two assortment parameters ρ and ρ'.

The variables in the right hand side of equations (23) and (24) are the same. However, equation (23) predicts that the coefficients on the grandparents would be weakly positive, whereas equation (24) predicts that they would be negative and small. This generalizes the discussion in the monoparental case as to whether the status transmission declines more slowly than geometrically (Solon, 2014).

More generally, one could use recent results to assess the validity of the instruments, whether the model is well-specified, and the sensitivity of the estimates to each of the empirical moments (Andrews et al., 2017, 2020; Bonhomme and Weidner, 2022).[26] The goal of this paper is not to present the definitive estimator of social mobility, but rather, to provide a series of results that any researcher could apply depending on the restrictions on available data and what assumptions

[25] Appendix D.2 discusses how some of our estimators are immune to classical measurement error and how we could assess the direction and magnitude of biases in this case.

[26] Appendix D.1 computes the derivatives of each estimator in Proposition 1 with respect to each empirical moment, and their signs.

about the parameters they are willing to make. Moreover, as mentioned above, the simple model presented here could be extended to allow for other effects using the techniques that we developed.

7 Conclusions

We have provided a systematic and coherent framework for the analysis of intergenerational mobility. That framework takes seriously the ways in which family formation, social norms, and biology shape how (dis)advantage is transmitted across generations in a pattern of influence that fans out backward from the current generation. In doing so, we show how women can matter for social mobility, both through their direct effect on their children and indirectly through the correlation in status between spouses. Our framework could be easily extended in a number of ways, two of which we discuss briefly here:

1. In the third moment of our main specification (equation (11)) 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 (outcome). We do not actually 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 effects η we estimate using maternal uncles.

2. The estimators here can be easily extended by adding covariates or dummy variables or applying our estimators to subpopulations to better understand the mechanisms underlying the transmission of status. How are the estimates different for children whose fathers die when they were young? or their mothers? How about the children of divorced parents? Does it matter whether the divorce happened when the son/daughter was a child? These and many other questions can be analyzed by looking at subpopulations (or adding dummy variables) to study status transmission under different circumstances.

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 demonstrate how strong effects on mobility from the maternal side of the family line can be uncovered even without directly observing female labor market outcomes. In fact, in our new analysis of linked U.S. data from 1870 to 1940, the effects of mothers on their children's outcomes are consistently larger than those of fathers. We also show the importance of accounting for underlying patterns of assortative mating that shape links across generations. And we identify a number of previous studies that embody assumptions that are captured by our framework, potentially crucial, and in many cases directly testable.

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 intra-household bargaining. Our approach could be applied to different regions and different population subgroups and of course to other countries and periods. The goal of this article is not to provide the last word on social mobility, or even the role of women in social mobility. Rather, we provide a set of tools for researchers to use that will be applied to a wide range of settings thereby helping 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.

Alesina, Alberto, Sebastian Hohmann, Stelios Michalopoulos, and Elias Papaioannou, "Intergenerational mobility in Africa," *Econometrica*, 2021, *89* (1), 1–35.

Althoff, Lukas, Harriet Brookes Gray, and Hugo Reichardt, "The Missing Link(s): Women and Intergenerational Mobility," 2024.

Andrews, Isaiah, Matthew Gentzkow, and Jesse Shapiro, "Measuring the Sensitivity of Parameter Estimates to Estimation Moments," *Quarterly Journal of Economics*, 2017, *132* (4), 1553–1592.

_, _, and _ , "On the Informativeness of Descriptive Statistics for Structural Estimates," *Econometrica*, 2020, *88* (6), 2231–2258.

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*, 1991, *58* (2), 277–297.

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), S1–S39.

Blume, Lawrence, Neil A. Cholli, Steven N. Durlauf, and Aleksandra Lukina, "Immobility As Memory: Some New Approaches to Characterizing Intergenerational Persistence via Markov Chains," *NBER Working Paper*, 2024, (33166).

Bonhomme, Stephane and Martin Weidner, "Minimizing Sensitivity to Model Misspecification," *Quantitative Economics*, 2022, *13* (3), 907–954.

Brandén, Gunnar, Martin Nybom, and Kelly Vosters, "Like mother, like child? The rise of women's intergenerational income persistence in Sweden and the United States," *Journal of Labor Economics*, 2024.

Bratsberg, Bernt, Simen Markussen, Oddbjørn Raaum, Knut Røed, and Ole Røgeberg, "Trends in Assortative Mating and Offspring Outcomes," *The Economic Journal*, 2023, *133*, 928–950.

Braun, Sebastian Till and Jan Stuhler, "The transmission of inequality across multiple generations: testing recent theories with evidence from Germany," *The Economic Journal*, 2018, *128*, 576–611.

Buckles, Kasey, Adrian Haws, Joseph Price, and Haley EB Wilbert, "Breakthroughs in Historical Record Linking Using Genealogy Data: The Census Tree Project," Technical Report 31671, National Bureau of Economic Research 2023.

_ , Joseph Price, Zachary Ward, and Haley EB Wilbert, "Family Trees and Falling Apples: Historical Intergenerational Mobility Estimates for Women and Men," Technical Report 31918, National Bureau of Economic Research 2023.

Card, David, Ciprian Domnisoru, and Lowell Taylor, "The intergenerational transmission of human capital: Evidence from the golden age of upward mobility," *Journal of Labor Economics*, 2022, *40* (1), 39–95.

Carneiro, Pedro, Italo Lopez Garcia, Kjell G. Salvanes, and Emma Tominey, "Intergenerational mobility and the timing of parental income," *Journal of Political Economy*, 2021, *129* (3), 757–788.

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

Chetty, Raj, John N. Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan, "Income segregation and intergenerational mobility across colleges in the United States," *The Quarterly Journal of Economics*, 2020, *135* (3), 1567–1633.

_ , Nathaniel Hendren, Patrick Kline, and Emmanuel Saez, "Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States," *Quarterly Journal of Economics*, 2014, (129.4), 1553–1623.

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).

Collado, M. Dolores, Ignacio Ortuño-Ortín, and Jan Stuhler, "Estimating Intergenerational and Assortative Processes in Extended Family Data," *Review of Economic Studies*, 2023, *90* (3), 1195–1227.

Collins, William J and Marianne H Wanamaker, "African American intergenerational economic mobility since 1880," *American Economic Journal: Applied Economics*, 2022, *14* (3), 84–117.

Curtis, Matthew, "The her in inheritance: how marriage matching has always mattered, Quebec 1800-1970," *working paper*, 2022.

Daruich, Diego and Julian Kozlowski, "Explaining intergenerational mobility: The role of fertility and family transfers," *Review of Economic Dynamics*, 2020, *36*, 220–245.

Davis, Jonathan M.V. and Bhashkar Mazumder, "The decline in intergenerational mobility after 1980," *Review of Economics and Statistics*, 2024, pp. 1–47.

Edwards, Ryan and Jennifer Roff, "What mom and dad's match means for junior: Marital sorting and child outcomes," *Labour Economics*, 2016, *40*, 43–56.

Eriksson, Katherine, Gregory T Niemesh, Myera Rashid, and Jacqueline Craig, "Marriage and the Intergenerational Mobility of Women: Evidence from Marriage Certificates 1850-1920," 2023.

Eshaghnia, Sadegh, James J. Heckman, and Rasmus Landersø, "The Impact of the Level and Timing of Parental Resources on Child Development and Intergenerational Mobility," *Journal of Labor Economics*, 2025.

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.

Folsom, Josiah C, "Perquisites and wages of hired farm laborers," Technical Report 213, USDA Technical Bulletin 1931.

Goldin, Claudia, Lawrence F. Katz, and Ilyana Kuziemko, "The homecoming of American college women: The reversal of the college gender gap," *Journal of Economic perspectives*, 2006, *20* (4), 133–156.

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 twentieth century," *Journal of Political Economy*, 2025, *133* (1), 306–354.

Keller, Wolfgang and Carol H Shiue, "Mobility through Marriage over Multiple Generations: Evidence from Imperial China," 2024.

Lochner, Lance, "Intergenerational Transmission," in "The New Palgrave Dictionar of Economics," London: Palgrave, 2016, pp. 1–7.

_ and Youngmin Park, "Earnings dynamics and intergenerational transmission of skill," *Journal of Econometrics*, 2024, *243* (1-2), 105255.

Long, Jason and Joseph Ferrie, "The path to convergence: intergenerational occupational mobility in Britain and the US in three eras," *The Economic Journal*, 2007, *117*, C61–C71.

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

Manser, Marilyn and Murray Brown, "Marriage and Household Decision-Making: A Bargaining Analysis," *International Economic Review*, 1980, *21* (1), 31–44.

Mazumder, Bhashkar, "Estimating the Intergenerational Elasticity and Rank Association in the US: Overcoming the Current Limitations of Tax Data," *Federal Reserve Bank of Chicago Working Paper*, 2015, pp. 1–56.

Mogstad, Magne and Gaute Torsvik, "Family background, neighborhoods, and intergenerational mobility," in "Handbook of the Economics of the Family," Vol. 1 2023, pp. 327–387.

Nybom, Martin and Jan Stuhler, "Interpreting trends in intergenerational mobility," *Journal of Political Economy*, 2024, *132* (8), 2531–2570.

Olivetti, Claudia, M. Daniele Paserman, Laura Salisbury, and E. Anna Weber, "The Emergence of the Socioeconomic Gradient in Women's Marriage Outcomes," *Review of Economics and Statistics*, 2024.

Raaum, Oddbjørn, Bernt Bratsberg, Knut Røed, Eva Österbacka, Tor Eriksson, Markus Jäntti, and Robin A. Naylor, "Marital sorting, household labor supply, and intergenerational earnings mobility across countries," *The BE Journal of Economic Analysis & Policy*, 2008, *7* (2).

Reagan, Barbara B, "Perquisites Furnished Hired Farm Workers: United States and Major Regions, 1945," *USDA Surveys of Wages and Wage Rates in Agriculture*, 1946, (18).

Ruggles, Steven, Matt A Nelson, Matthew Sobek, Catherine A Fitch, Ronald Goeken, J. David Hacker, Evan Roberts, and J. Robert Warren, "IPUMS Ancestry Full Count Data: Version 4.0 [dataset]," 2024.

_ , Sarah Flood, Matthew Sobek, Daniel Backman, Annie Chen, Grace Cooper, Stephanie Richards, Renae Rodgers, and Megan Schouweiler, "IPUMS USA: Version 15.0 [dataset]," 2024.

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

Shepherd-Banigan, Megan, Anirban Basu, Janice F. Bell, Cathryn Booth-LaForce, and Jeffrey R. Harris, "Is maternal income in childhood associated with adolescent health and behavioral outcomes?," *Journal of family issues*, 2019, *40* (7), 911–928.

Solon, Gary, "Theoretical models of inequality transmission across multiple generations," *Research in Social Stratification and Mobility*, 2014, *35*, 13–18.

Song, Xi, Catherine G. Massey, Karen A. Rolf, Joseph P. Ferrie, Jonathan L. Rothbaum, and Yu Xie, "Long-term Decline in Intergenerational Mobility in the United States since the 1850s," *Proceedings of the National Academy of Sciences*, 2020, *117* (1), 251–58.

Ward, Zachary, "Intergenerational Mobility in American History: Accounting for Race and Measurement Error," *American Economic Review*, 2023, *113* (12), 3213–3248.

A Econometrics Appendix

A.1 Identification Using 2-Generation Data (proofs)

Proof of Proposition 1.

Proof. We use equations (9), (10), and (11), and $\gamma_{PF}^{MF} \equiv m_{PF}^{MF}$. With equation (10), we can directly get the sum of the mobility effects as a function of the data $\left(\beta_F' + \beta_M'\right) = m_{MF}^{F}/m_{PF}^{MF}$. Using this and equation (11), we get

$$\rho = \left(m_{MF}^{F}\right)^2 / m_{PF}^{MF}.$$

We can now solve for β_F' as a function of β_M' and data using the second equation, i.e, $\beta_F' = \frac{m_{MF}^{F}}{m_{PF}^{MF}} - \beta_M'$. Plugging in this and the formula for ρ in the first equation we get

$$m_{PF}^{F} = \frac{m_{MF}^{F}}{m_{PF}^{MF}} - \beta_M' + \frac{\left(m_{MF}^{F}\right)^2}{m_{PF}^{MF}}\beta_M'$$

rearranging we get

$$\beta_M' = -\frac{m_{PF}^{MF} m_{PF}^{F} - m_{MF}^{F}}{m_{PF}^{MF} - \left(m_{MF}^{F}\right)^2}$$

$$\beta_F' = \frac{m_{MF}^{F}}{m_{PF}^{MF}} + \frac{m_{PF}^{MF} m_{PF}^{F} - m_{MF}^{F}}{m_{PF}^{MF} - \left(m_{MF}^{F}\right)^2}$$

□

Proof of Proposition 2.

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

$$m_{PF}^{F} = \beta_F'; \qquad m_{MF}^{F} = \beta_F' m_{PF}^{MF}; \qquad \rho = \left(\beta_F'\right)^2 m_{PF}^{MF}$$

This is a system with two unknowns and three equations. Thus, it is overidentified. Similar to Proposition 1, we can use the second and third moment and we have $\rho = \left(m_{MF}^{F}\right)^2 / m_{PF}^{MF}$. Alternatively, we can substitute the first two moments into the third moment and get $\rho = m_{PF}^{F} m_{MF}^{F}$. The second moment then gives us $\beta_F' = m_{MF}^{F}/m_{PF}^{MF}$. The system is overidentified, and the first equation implies $\beta_F' = m_{PF}^{F}$. If in the data we have $m_{PF}^{F} \neq m_{MF}^{F}/m_{PF}^{MF}$. We should reject the assumption $\beta_M' = 0$. □

Proof of Proposition 3.

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

$$m_{PF}^{F} = \beta_F'\left(1+\rho'\right); \qquad m_{MF}^{F} = 2\beta_F' m_{PF}^{MF}; \qquad \rho = \left(2\beta_F'\right)^2 m_{PF}^{MF}$$

This is a system with three unknowns $\left(\beta_F', \rho', \rho\right)$ and three independent equations, so it is identified. Similar to Proposition 1, we can use the second and third moment and we have $\rho =$

$\left(m_{MF}^{F}\right)^{2}/m_{PF}^{MF}$. From the second equation we get $\beta_F^{'} = m_{MF}^{F}/2m_{PF}^{MF}$. We can then plug this in the first equation to get $\rho' = m_{PF}^{F}/\beta_F^{'} - 1 = 2m_{PF}^{MF}m_{PF}^{F}/m_{MF}^{F} - 1$. □

A.2 Identification Using 3-generation Data (proofs)

Proof of Proposition 4.

Proof. We now use three moments to get identification

- $\mathbb{E}\left[X_i^{PF}\varepsilon_i^{C}\right] = 0$ and $\mathbb{E}\left[X_i^{PF}\varepsilon_i^{M}\right] = 0$. Multiplying equations (4) and (6) by X_i^{PF} and then substituting the definition of X_i^{M}, and taking expectations, we get

$$m_{PF}^{C} - \beta_F m_{PF}^{F} = \beta_M\left(\beta_F\gamma_{PF}^{MF} + \beta_M\gamma_{MF}^{PM}\right) \tag{25}$$

- $\mathbb{E}\left[X_i^{MF}\varepsilon_i^{F}\right] = 0$. Multiplying the second equation by X_i^{MF} and taking expectations and using $\gamma_{MF}^{PM} = \gamma_{MM}^{PF}$, we get

$$m_{MF}^{F} = \beta_F\gamma_{PF}^{MF} + \beta_M\gamma_{MF}^{PM} \tag{26}$$

By imposing $\gamma_{MF}^{PM} = \gamma_{MM}^{PF}$, we can now identify the model. First, we can take equation (26) and insert it in equation (25). This way we get

$$m_{PF}^{C} - \beta_F m_{PF}^{F} = \beta_M m_{MF}^{F} \tag{27}$$

Now we have four equations ((15), (16), (17) and (27)) and four unknowns $(\beta_F, \beta_M, \rho', \rho)$. Notice that equations (16) and (17) combined create an equation on β_F and β_M only. This together with equation (27) above, creates a system with two equations and two unknowns β_F and β_M. We can then use equation (15) to solve for ρ and equation (16), to solve for ρ'. The solution to this system is

$$\begin{aligned}
\beta_F &= \frac{m_{PF}^{F}m_{PF}^{C} - m_{MF}^{C}m_{MF}^{F}}{\left(m_{PF}^{F}\right)^{2} - \left(m_{MF}^{F}\right)^{2}} \\
\beta_M &= \frac{m_{PF}^{F}m_{MF}^{C} - m_{PF}^{C}m_{MF}^{F}}{\left(m_{PF}^{F}\right)^{2} - \left(m_{MF}^{F}\right)^{2}} \\
\rho &= \frac{m_{F}^{C}\left(\left(m_{PF}^{F}\right)^{2} - \left(m_{MF}^{F}\right)^{2}\right) - \left(m_{PF}^{F}m_{PF}^{C} - m_{MF}^{C}m_{MF}^{F}\right)}{m_{PF}^{F}m_{MF}^{C} - m_{PF}^{C}m_{MF}^{F}} \\
\rho' &= \frac{m_{PF}^{F}\left(\left(m_{PF}^{F}\right)^{2} - \left(m_{MF}^{F}\right)^{2}\right) - \left(m_{PF}^{F}m_{PF}^{C} - m_{MF}^{C}m_{MF}^{F}\right)}{m_{PF}^{F}m_{MF}^{C} - m_{PF}^{C}m_{MF}^{F}}
\end{aligned}$$

□

A.3 Identification Using Maternal Uncles (proofs)

Using similar assumptions as above we can get the following moments using X_i^{MU}

$$m_{MU}^{C} = \beta_F m_{MU}^{F} + \beta_M\left(\beta_F^{2} + \beta_M^{2} + 2\rho'\beta_F\beta_M + \eta\right) \tag{28}$$

$$m_{MU}^{F} = \rho' \tag{29}$$

$$m_{MU}^{MF} = \beta_F + \rho'\beta_M \tag{30}$$

$$m_{MU}^{PF} = \beta_F\gamma_{PF}^{MF} + \beta_M\gamma_{MM}^{PF} \tag{31}$$

where $\eta \equiv \mathbb{E}\left[\varepsilon_i^{M}\varepsilon_i^{MU}\right]$ measures household effects.

Proof of Proposition 5.

Proof. First, we impose $\gamma_{PF}^{MF} = \gamma_{MF}^{PM} = \gamma_{PM}^{MM} = \gamma_{MM}^{PF}$ and $\rho = \rho'$ in the system of equations above and we get

$$m_F^C = \beta_F + \rho\beta_M$$

$$m_{MF}^C = \beta_F m_{MF}^F + \beta_M (\beta_F + \rho\beta_M)$$

$$m_{MF}^F = (\beta_F + \beta_M)\gamma_{PF}^{MF}$$

$$\rho = (\beta_F + \beta_M)^2 \gamma_{PF}^{MF}$$

Notice that now, because we do not observe X_i^{PF}, we do not observe m_{PF}^F, m_{PF}^C and γ_{PF}^{MF}. Nonetheless, we can take the last two equations and write $\rho = m_{MF}^F (\beta_F + \beta_M)$. 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. □

Proof of Proposition 6.

Proof. The moments that do not use X_i^{PF} are equations (15), (17) and (18). Equation (18) includes two nuisance parameters that are unobserved here $\left(\gamma_{PF}^{MF}, \gamma_{MF}^{PM}\right)$. 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 (15) and (17), we can use equations (28), (29) and (30). 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 (29) into equation (15). We can also substitute $\beta_F + \rho'\beta_M$ from equation (30) into the equation (17). Equations (15) and (16) become

$$m_F^C = \beta_F + \beta_M m_{MU}^F$$

$$m_{MF}^C = \beta_F m_{MF}^F + \beta_M m_{MU}^{MF}$$

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 (30) and solve for ρ'. We can use equation (28) to identify the household fixed effects η. □

Proof of Proposition 7.

Proof. The moments that do not use X_i^{MF} are equations (15), (16) and (19). Equation (17) includes two nuisance parameters that are unobserved here $\left(\gamma_{PF}^{MF}, \gamma_{MM}^{PF}\right)$. However, this would not be a problem as we see below. In addition to equations (15), (16) and (19), we can use equations (28), (29) and (31). These six equations form a system of five independent equations, when we substitute equation (31) into the equation (19). 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 (29) into equation (15). We can also substitute $\left(\beta_F \gamma_{PF}^{MF} + \beta_M \gamma_{MM}^{PF}\right)$ from equation (31) into the equation (19). Equations (15) and (19) become

$$m_F^C = \beta_F + \beta_M m_{MU}^F$$

$$m^C_{PF} = \beta_F m^F_{PF} + \beta_M m^{PF}_{MU}$$

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 (16) and solve for ρ'. We can use equation (28) to identify the household fixed effects η. □

Proof of Proposition 8.

Proof. We can generate three moments (28), (29) and (30). Assuming $\rho' = \rho$ and $\eta = 0$ we get

$$m^C_F = \beta_F + \rho \beta_M$$

$$m^F_{MU} = \rho$$

$$m^C_{MU} = \beta_F m^F_{MU} + \beta_M \left(\beta_F^2 + \beta_M^2 + 2\rho\beta_F\beta_M\right)$$

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.4 Gendered Effects (proofs)

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

$$m^F_{MU} = \beta^S_F + \rho \beta^S_M \tag{32}$$

$$m^F_{PF} = \beta^S_F + \rho' \beta^S_M \tag{33}$$

$$m^S_{MF} = \beta^S_F m^F_{MF} + \beta^S_M \left(\beta^D_F + \rho' \beta^D_M\right) \tag{34}$$

$$m^F_{MF} = \beta^S_F \gamma^{MF}_{PF} + \beta^S_M \gamma^{PM}_{MF} \tag{35}$$

$$m^S_{PF} = \beta^S_F m^F_{PF} + \beta^S_M \left(\beta^D_F \gamma^{MF}_{PF} + \beta^D_M \gamma^{PF}_{MM}\right) \tag{36}$$

$$\rho = \beta^S_F \beta^D_F \gamma^{MF}_{PF} + \beta^D_F \beta^S_M \gamma^{PM}_{MF} + \beta^S_F \beta^D_M \gamma^{MM}_{PF} + \beta^S_M \beta^D_M \gamma^{MM}_{PM} \tag{37}$$

The system above shows six equations with six structural parameters $\left(\beta^S_F, \beta^S_M, \beta^D_F, \beta^D_M, \rho', \rho\right)$, and three nuisance parameters $\left(\gamma^{PM}_{MF}, \gamma^{MM}_{PM}, \gamma^{PF}_{MM}\right)$. 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.

Proof of Proposition 9.

Proof. First, we impose $\gamma^{MF}_{PF} = \gamma^{PM}_{MF} = \gamma^{MM}_{PM} = \gamma^{PF}_{MM}$ in the system of equations above. Notice that γ^{MF}_{PF} is observable. Second, take the system of six equations above and notice that: ρ only appears in equations (32) and (37); and ρ' only appears in equations (33) and (34). We can take equation (32), solve for ρ and substitute it in equation (37) and take equation (33), solve for ρ' and substitute in equation (34). We get

$$\frac{m_F^S - \beta_F^S}{\beta_M^S} = \left(\beta_F^S + \beta_M^S\right)\left(\beta_F^D + \beta_M^D\right)\gamma_{PF}^{MF}$$

$$m_{MF}^S = \beta_F^S m_{MF}^F + \beta_M^S \beta_F^D + \left(m_{PF}^F - \beta_F^S\right)\beta_M^D$$

The two equations above, together with equations (35) and (36) 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 (32) to solve for ρ and equation (33) to solve for ρ'. □

A.5 Identification of Generational Effects

In the baseline model in Section 2, 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, but impose restrictions on the nuisance parameters. The system of equations that we are considering is the following

$$\begin{array}{rcl} X_i^C & = & \beta_F X_i^F + \beta_M X_i^M + \varepsilon_i^C \\ X_i^F & = & \beta_F' X_i^{PF} + \beta_M' X_i^{PM} + \varepsilon_i^F \\ X_i^M & = & \beta_F' X_i^{MF} + \beta_M' X_i^{MM} + \varepsilon_i^M \end{array} \tag{38}$$

where β_F and β_M are the effects of the father and mother in the second generation, respectively and β_F' and β_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 10. *Suppose X_i^C, X_i^F, X_i^{PF} and X_i^{MF} are observed. If we assume $\gamma_{MF}^{PM} = \gamma_{MM}^{PF}$, then (β_F, β_M, ρ) and β' (with $\beta' \equiv \beta_F' + \beta_M'$) is point identified. However, ρ', β_F' and β_M' are not point identified.*

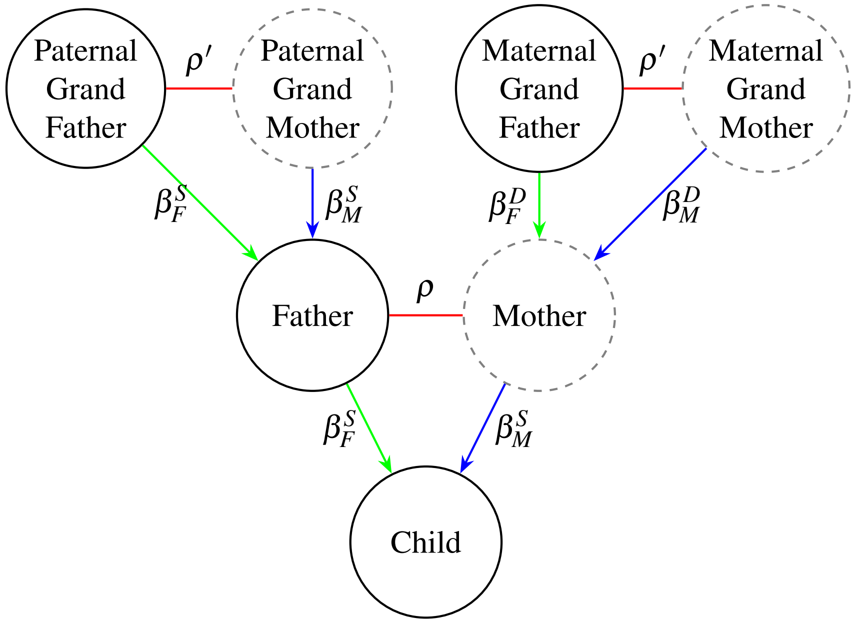
Proposition 10 shows that imposing assumptions on the nuisance parameters, but no assumptions on the structural parameters, is not enough to get point identification here. Nonetheless, identifying $(\beta_F, \beta_M, \rho, \beta')$ could be of interest in many settings. For example, the econometrician might be willing to assume that the coefficients for father and mother for the first generations are equal to each other, i.e., $\beta_F' = \beta_M'$. Corollary 1 below shows that, with this extra assumption, all parameters are identified.

Corollary 1. *Suppose X_i^C, X_i^F, X_i^{PF} and X_i^{MF} are observed. If we assume $\gamma_{PF}^{MF} = \gamma_{MF}^{PM} = \gamma_{MM}^{PF}$ and $\beta_F' = \beta_M'$, then $(\beta_F, \beta_M, \beta_F', \rho', \rho)$ is point identified.*

Proposition 11 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 10.

Proposition 11. *Suppose X_i^C, X_i^F, X_i^{PF} and X_i^{MF} are observed. If we assume $\gamma_{MF}^{PM} = \gamma_{MM}^{PF} = 0$ and $\gamma_{PF}^{MF} = \gamma_{PM}^{MM}$, then $(\beta_F, \beta_M, \beta_F', \beta_M', \rho', \rho)$ is point identified.*

Figure 8: Family Trees for Gendered 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 outcomes while the dashed circles represent individuals (females) with unobserved outcomes.

A reasonable assumption that one can make is that the degree of assortative mating is constant across generations, i.e., $\rho' = \rho$. Proposition 12 shows that this assumption provides point identification in all the other structural parameters.

Notice that this result contrasts with the negative results shown at the beginning of this section. In the simple model, assuming $\rho' = \rho$ did not add any identification to our model, but here, it provides an independent equation, with respect to the results in proposition 10. 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 also that 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., $\beta_F \neq \beta_F^{'}$ and $\beta_M \neq \beta_M^{'}$. Thus, assuming that the mating parameters are constant, i.e., $\rho' = \rho$, does add more identification power here.

Proposition 12. *Suppose X_i^C, X_i^F, X_i^{PF} and X_i^{MF} are observed. If we assume $\gamma_{PF}^{MF} = \gamma_{MF}^{PM} = \gamma_{MM}^{PF}$ and $\rho' = \rho$, then $(\beta_F, \beta_M, \beta_F^{'}, \beta_M^{'}, \rho)$, is point identified.*

Propositions 10 and 11 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 11, the six equations are independent and thus we can get point identification in all six parameters of interest. In Proposition 10, however, the equations are not independent and we get point identifications in all parameters but $\beta_F^{'}$ and $\beta_M^{'}$. We get identification in their sum $\beta^{'}$ and we end up with a system of over-identifying restrictions. Proposition 13 below extends this intuition and shows how we can also get point identification on $(\beta_F, \beta_M, \rho', \rho)$, with a weaker assumption on the nuisance parameters: $\gamma_{MF}^{PM} = \gamma_{MM}^{PF} = (\gamma_{PF}^{MF}\gamma_{PM}^{MM})^{0.5}$. The downside is that now we cannot point identify the mobility parameters in the older generation $\beta_F^{'}$ and $\beta_M^{'}$, or the nuisance parameter γ_{PM}^{MM}, but we can identify $\tilde{\beta}^{'} \equiv \beta_F^{'}(\gamma_{PF}^{MF})^{0.5} + \beta_F^{'}(\gamma_{PM}^{MM})^{0.5}$.

Proposition 13. *Suppose X_i^C, X_i^F, X_i^{PF} and X_i^{MF} are observed. If we assume $\gamma_{MF}^{PM} = \gamma_{MM}^{PF} = \sqrt{\gamma_{PF}^{MF}\gamma_{PM}^{MM}}$, then (β_F, β_M, ρ) and $\tilde{\beta}^{'}$ (with $\tilde{\beta}^{'} \equiv \beta_F^{'}\sqrt{\gamma_{PF}^{MF}} + \beta_M^{'}\sqrt{\gamma_{PM}^{MM}}$) is point identified. However, $\beta_F^{'}$, $\beta_M^{'}$, and ρ' are not point identified.*

Proposition 13 presents a negative result when we use the assumption $\gamma_{MF}^{PM} = \gamma_{MM}^{PF} = \sqrt{\gamma_{PF}^{MF}\gamma_{PM}^{MM}}$. Corollary 1 above shows how the negative result in Proposition 10 can be overcome by a simple restriction on the structural parameters such as $\beta_F^{'} = \beta_M^{'}$. Corollary 2 below shows how the same assumption does not solve the identification issues from Proposition 13.

Corollary 2. *Suppose X_i^C, X_i^F, X_i^{PF} and X_i^{MF} are observed. If we assume $\gamma_{MF}^{PM} = \gamma_{MM}^{PF} = \sqrt{\gamma_{PF}^{MF}\gamma_{PM}^{MM}}$ and $\beta_F^{'} = \beta_M^{'}$, (β_F, β_M, ρ) and $\tilde{\beta}^{'}$ (with $\tilde{\beta}^{'} \equiv \beta_F^{'}\left(\sqrt{\gamma_{PF}^{MF}} + \sqrt{\gamma_{PM}^{MM}}\right)$) is point identified. However, $\beta_F^{'}$, $\beta_M^{'}$, and ρ' are not point identified.*

A.6 Generational Effects (proofs)

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

$$m_F^C = \beta_F + \rho\beta_M \tag{39}$$

$$m_{PF}^F = \beta_F^{'} + \rho'\beta_M^{'} \tag{40}$$

$$m_{MF}^C = \beta_F m_{MF}^F + \beta_M\left(\beta_F^{'} + \rho'\beta_M^{'}\right) \tag{41}$$

$$m_{MF}^F = \beta_F^{'}\gamma_{PF}^{MF} + \beta_M^{'}\gamma_{MF}^{PM} \tag{42}$$

$$m_{PF}^C = \beta_F m_{PF}^F + \beta_M\left(\beta_F^{'}\gamma_{PF}^{MF} + \beta_M^{'}\gamma_{MM}^{PF}\right) \tag{43}$$

$$\rho = \left(\beta_F^{'}\right)^2\gamma_{PF}^{MF} + \left(\beta_M^{'}\right)^2\gamma_{PM}^{MM} + \beta_F^{'}\beta_M^{'}\left(\gamma_{MF}^{PM} + \gamma_{MM}^{PF}\right) \tag{44}$$

Proof of Proposition 10.

Proof. First, we impose $\gamma_{MF}^{PM} = \gamma_{MM}^{PF}$ in the system of equations above. Notice that γ_{PF}^{MF} is observable. We can substitute equation (40) into equation (41) and substitute equation (42) into equation (43). We have now a system with two equations and two unknowns (β_F, β_M).

$$m_{MF}^C = \beta_F m_{MF}^F + \beta_M m_{PF}^F$$

$$m_{PF}^C = \beta_F m_{PF}^F + \beta_M m_{MF}^F$$

We can go to equation (39) and identify ρ. This is as far as we can get. In equations (42), (43) and (44) we can only identify $\beta^{'}$ but not each component. In equations (40) and (41) we can only identify $\left(\beta_F^{'} + \rho'\beta_M^{'}\right)$. Thus we have two independent equations, say (40) and (42), for three unknowns ρ', $\beta_F^{'}$ and $\beta_M^{'}$. □

Proof of Corollary 1.

Proof. We can follow the same steps as in Proposition 10 to get identification on (β_F, β_M, ρ). Unlike before, we have now an extra assumption $\beta_F^{'} = \beta_M^{'}$. Moreover, we are imposing $\gamma_{PF}^{MF} = \gamma_{MF}^{PM}$ now. We can use equation (42) and get $m_{MF}^F = 2\beta_F^{'}\gamma_{PF}^{MF}$ to identify $\beta_F^{'}$. We can then use equation (40) to identify ρ'. □

Proof of Proposition 11.

Proof. First, we impose $\gamma_{MF}^{PM} = \gamma_{MM}^{PF} = 0$ and $\gamma_{PF}^{MF} = \gamma_{PM}^{MM}$ in the system of equations above. Notice that γ_{PF}^{MF} is observable. Second, take the system of six equations above and notice that: ρ only appears in equations (39) and (44); and ρ' only appears in equations (40) and (41). We can take equation (39), solve for ρ and substitute it in equation (44) and take equation (40), solve for ρ' and substitute in equation (41). We get

$$\frac{m_F^C - \beta_F}{\beta_M} = \left(\left(\beta_F^{'}\right)^2 + \left(\beta_M^{'}\right)^2\right)\gamma_{PF}^{MF}$$

$$m_{MF}^C = \beta_F m_{MF}^F + \beta_M m_{PF}^F$$

With the assumption here, equation (42) identifies $\beta_F^{'}$ directly, i.e, $\beta_F^{'} = m_{MF}^F/\gamma_{PF}^{MF}$. If we substitute this in equation (43), we get

$$m_{PF}^C = \beta_F m_{PF}^F + \beta_M m_{MF}^F$$

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 $\beta_F^{'}$, β_F and β_M, together with the first equation above, we can identify $\beta_M^{'}$. Finally, we can use the mobility parameters $\beta_F^{'}$, $\beta_M^{'}$, β_F and β_M and using equations (39) and (40), we get the mating parameters ρ and ρ'. □

Proof of Proposition 12.

Proof. First, we impose $\gamma_{PF}^{MF} = \gamma_{MF}^{PM} = \gamma_{MM}^{PF}$ in the system of equations above. Notice that γ_{PF}^{MF} is observable. Second, take the system of six equations above and notice that ρ' only appears in equations (40) and (41). We can take equation (40), solve for ρ' and substitute in equation (41). We get

$$m_{MF}^C = \beta_F m_{MF}^F + \beta_M m_{PF}^F$$

We can take equation (42) and substitute in equation (43) to get

$$m_{PF}^C = \beta_F m_{PF}^F + \beta_M m_{MF}^F$$

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 3 (right). Now we can just use equation (39) to solve for ρ.

$$m_F^C = \beta_F + \rho\beta_M$$

Now imposing $\rho = \rho'$ means that we can use equations (40) and (42) for form the system below.

$$m_{PF}^F = \beta_F^{'} + \rho\beta_M^{'}$$

$$m_{MF}^F = \left(\beta_F^{'} + \beta_M^{'}\right)\gamma_{PF}^{MF}$$

The previous two equations form a system of two equations and two unknowns $\left(\beta_F^{'}, \beta_M^{'}\right)$. □

Proof of Proposition 13.

Proof. We can substitute equation (40) into equation (41) and get

$$m_{MF}^{C} = \beta_F m_{MF}^{F} + \beta_M m_{PF}^{F}$$

Using $\gamma_{MF}^{PM} = \gamma_{MM}^{PF}$ we can substitute equation (42) into equation (43) and get

$$m_{PF}^{C} = \beta_F m_{PF}^{F} + \beta_M m_{MF}^{F}$$

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 (39) we can solve for ρ. If we would write $\rho = \rho'$ we can write equation (40) as

$$m_{PF}^{F} = \beta_F^{'} + \rho \beta_M^{'}$$

This gives us one equation to identify $\beta_F^{'}$ and $\beta_M^{'}$. Notice, however, that $\beta_F^{'}$ and $\beta_M^{'}$ appear in all other equations as $\tilde{\beta}' \equiv \beta_F^{'}\sqrt{\gamma_{PF}^{MF}} + \beta_M^{'}\sqrt{\gamma_{PM}^{MM}}$.[27] This means we have only one independent equation to estimate $\tilde{\beta}'$ and equation (40) that relates $\beta_F^{'}$ and $\beta_M^{'}$. Without further assumptions on γ_{PM}^{MM} we do not get point identification. □

Proof of Corollary 2.

Proof. First, we impose $\gamma_{MF}^{PM} = \gamma_{MM}^{PF} = \sqrt{\gamma_{PF}^{MF}\gamma_{PM}^{MM}}$ and $\beta_F^{'} = \beta_M^{'}$ in the system of equations above. Notice that γ_{PF}^{MF} is observable. Second, take the system of six equations above and notice that: ρ only appears in equations (39) and (44); and ρ' only appears in equations (40) and (41). We can take equation (39), solve for ρ and substitute it in equation (44) and take equation (40), solve for ρ' and substitute in equation (41). We get

$$\frac{m_F^C - \beta_F}{\beta_M} = \left(\beta_F^{'}\right)^2 \left(\sqrt{\gamma_{PF}^{MF}} + \sqrt{\gamma_{PM}^{MM}}\right)^2 = \frac{\left(\beta_F^{'}\right)^2}{\gamma_{PF}^{MF}} \left(\gamma_{PF}^{MF} + \sqrt{\gamma_{PF}^{MF}\gamma_{PM}^{MM}}\right)^2$$

$$m_{MF}^{C} = \beta_F m_{MF}^{F} + \beta_M m_{PF}^{F}$$

The previous two equations, together with equations (42) and (43), below, form a system with four equations and four unknowns.

$$m_{MF}^{F} = \beta_F^{'}\left(\gamma_{PF}^{MF} + \sqrt{\gamma_{PF}^{MF}\gamma_{PM}^{MM}}\right)$$

$$m_{PF}^{C} = \beta_F m_{PF}^{F} + \beta_M \beta_F^{'}\left(\gamma_{PF}^{MF} + \sqrt{\gamma_{PF}^{MF}\gamma_{PM}^{MM}}\right)$$

We can take the last two equations and get $m_{PF}^{C} = \beta_F m_{PF}^{F} + \beta_M m_{MF}^{F}$. This together with the second equation above ($m_{MF}^{C} = \beta_F m_{MF}^{F} + \beta_M m_{PF}^{F}$) 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 (39), we can identify ρ.

[27] Equation (42), dividing both sides by $\sqrt{\gamma_{PF}^{MF}}$, can be written as $\frac{m_{MF}^{F}}{\sqrt{\gamma_{PF}^{MF}}} = \tilde{\beta}'$ and similarly for equation (43).

Notice that $\beta_F^{'}$ appears three times in the equations above, but each time appears in the same form $\beta_F^{'}\left(\gamma_{PF}^{MF}+\sqrt{\gamma_{PF}^{MF}\gamma_{PM}^{MM}}\right)$ thus, we only have one independent equation for $\beta_F^{'}$ and γ_{PM}^{MM}. The other time that $\beta_F^{'}$ appears is in equations (11) and 12. In both cases it appears together with ρ' as $\beta(1+\rho')$. Thus, we have two independent equations to estimate three parameters $\left(\beta_F^{'},\rho',\gamma_{PM}^{MM}\right)$. Thus, without further assumptions, $\beta_F^{'}$, ρ', and γ_{PM}^{MM}, are not point identified. □

B Extra Simulation Results

Figure 9 shows the estimation results using our estimator for Propositions 4 and 5 in simulated samples of size ranging 200-1000. Unlike in other results, Panel A shows that the point estimates are not at the true value for small samples sizes. Even with $n = 1,000$ the point estimates seem a bit off. This highlights how Assumption A1 helps identification by making some moments simpler and, thus, making some structural parameter linear in those moments. Without using Assumption A1, even if it holds in reality as in this simulated case, makes the estimates much less precise. We would recommend the researcher use results that use Assumption A1, unless she has a specific model in mind of arranged marriages, or settings where the parents of the groom and the bride may have a direct and asymmetric effect on mating. In contrast to Proposition 4 in Panel A, Proposition 5 in Panel B uses no information on the paternal grandfather X_i^{PF} but imposes Assumption A1. We can see here how the point estimates are right at the true values and statistically significant for very small sample sizes ($n = 400$). This, again, highlights the importance of Assumption A1 and the little importance that males in the patrilineal side such as the paternal grandfather has for identification and estimation.

Table 4 shows results from Proposition 1, similar to those used in Figure 4. We report three magnitudes for each parameter and sample size: 1) the mean of the estimated parameters across simulations; 2) the standard deviation (SD) of the estimated parameter across simulations; and 3) the median of the estimated standard errors (SE) calculated by the GMM asymptotic variance formula. We can see how the mean of the parameters converges very quickly to the true value for all parameters. Moreover, both the SD and the SE are small, and quickly become smaller for all parameters. This is particularly true for ρ where we get statistical significance even at $n = 200$.

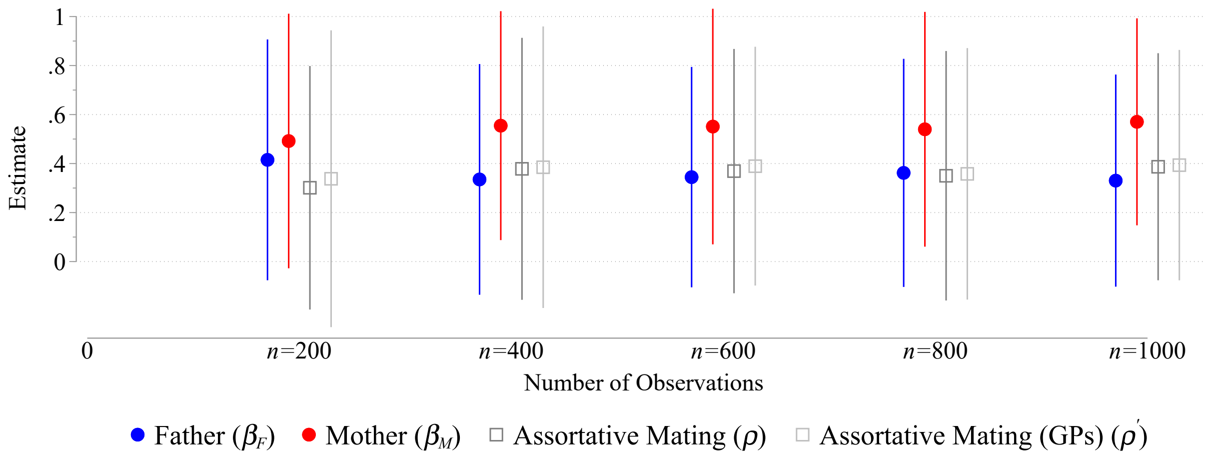
C Extra Empirical Results

We now show other empirical results as a robustness check to our main empirical estimates. Tables 5 and 6 show results when we use only one measure of income for each individual. In other words, we observe each individual only once in adulthood. Due to measurement error, the pairwise correlations in this sample are lower than the pairwise correlations when observing each individual twice. The effects on the empirical estimates could go up, down, or not be affected. On the other hand, there is sample attrition when requiring that each individual is observed in two different census records. In that sense, the sample here are larger and less selected.

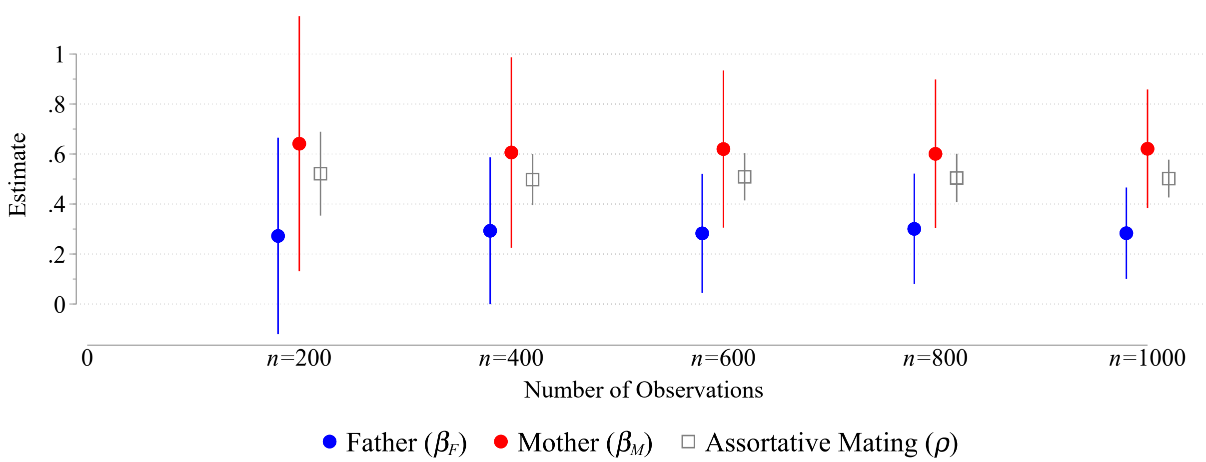
As a second robustness check, we compute income measures for the 1870 cohort, similarly to the ones for the 1880 cohort. Tables 7 and 8 display these results. The results are similar than those

Figure 9: Simulation Results from Propositions 4 and 5

A. Simulations with $\beta_F = 0.3$, $\beta_M = 0.6$, and $\rho = \rho' = 0.5$. Proposition 4



B. Simulations with $\beta_F = 0.3$, $\beta_M = 0.6$, and $\rho = 0.5$. Proposition 5



Notes: Estimated parameters and confidence intervals at $p = 0.05$ using Propositions 4 and 5.

for the 1880 cohort. The effect of the mother and the degree of assortative mating is larger for the 1870 cohort.

D Model Mispecification Analysis

D.1 Comparative Statistics

In this section, we show a comparative statistics analysis for Proposition 1. Matrix J below shows the derivative of each estimator for each structural parameter (β_F, β_M, ρ) over each empirical moment $(x_F^{MF}, x_F^{PF}, x_{PF}^{MF})$. We restrict attention to the case with $0 < \beta_F, \beta_M, \rho < 1$. Without further

Table 4: Simulation Results from Proposition 1

Sample Size	200	400	600	800	1000
	0.269	0.270	0.262	0.294	0.291
$\beta_F = 0.3$	(0.218)	(0.141)	(0.135)	(0.117)	(0.109)
	0.267	*0.161*	*0.142*	*0.111*	*0.097*
	0.640	0.639	0.650	0.608	0.608
$\beta_F = 0.6$	(0.284)	(0.195)	(0.183)	(0.165)	(0.154)
	0.342	*0.228*	*0.193*	*0.153*	*0.136*
	0.517	0.505	0.515	0.505	0.498
$\rho = 0.5$	(0.101)	(0.069)	(0.060)	(0.061)	(0.050)
	0.112	*0.078*	*0.063*	*0.055*	*0.048*

*Not*es: We report three magnitudes for each parameter and sample size: 1) the mean of the estimated parameters across simulations; 2) the standard deviation (SD) of the estimated parameter across simulations, in parenthesis; and 3) the median of the estimated standard errors (SE) calculated by the GMM asymptotic variance formula, in italics.

Table 5: Empirical Results for Propositions 1, 2, and 3 (one measure)

Parameter	Prop. 1	Prop. 2	Prop. 3
β_F	0.140	0.874	0.382
	(0.003)	(0.003)	(0.001)
β_M	0.734		
	(0.006)		
ρ	0.287	0.287	0.287
	(0.002)	(0.002)	(0.001)
N	2,552,748	2,552,748	2,552,748
MSE	2.710e-16	6.633e-16	0.0028

Notes: Empirical results from the estimation of Propositions 1, 2, and 3, for 1880 cohort, using one measure. A cohort is defined as any male child in the Census for the specified year; hence, there can be some overlap in the individuals in these trees.

restrictions, three of the nine derivatives have an ambiguous sign.

$$J = \begin{pmatrix} \frac{\partial \rho}{\partial x_F^{MF}} & \frac{\partial \rho}{\partial x_F^{PF}} & \frac{\partial \rho}{\partial x_{PF}^{MF}} \\ \frac{\partial \beta_F}{\partial x_F^{MF}} & \frac{\partial \beta_F}{\partial x_F^{PF}} & \frac{\partial \beta_F}{\partial x_{PF}^{MF}} \\ \frac{\partial \beta_M}{\partial x_F^{MF}} & \frac{\partial \beta_M}{\partial x_F^{PF}} & \frac{\partial \beta_M}{\partial x_{PF}^{MF}} \end{pmatrix} = \begin{pmatrix} \frac{2x_F^{MF}}{x_F^{PF}} & -\frac{(x_F^{MF})^2}{(x_F^{PF})^2} & 0 \\ \frac{1}{x_F^{PF}} + \frac{2x_F^{MF} \cdot x_F^{PF} \cdot x_{PF}^{MF} - (x_F^{MF})^2 - x_F^{PF}}{(x_F^{PF} - (x_F^{MF})^2)^2} & \frac{x_F^{MF}(1 - x_{PF}^{MF} \cdot x_F^{MF})}{(x_F^{PF} - (x_F^{MF})^2)^2} - \frac{x_F^{MF}}{(x_F^{PF})^2} & \frac{x_F^{PF}(x_F^{PF} - (x_F^{MF})^2)}{(x_F^{PF} - (x_F^{MF})^2)^2} \\ \frac{x_F^{PF} - 2x_F^{MF} \cdot x_F^{PF} \cdot x_{PF}^{MF} + (x_F^{MF})^2}{(x_F^{PF} - (x_F^{MF})^2)^2} & \frac{x_F^{MF}(x_{PF}^{MF} \cdot x_F^{MF} - 1)}{(x_F^{PF} - (x_F^{MF})^2)^2} & \frac{x_F^{PF}((x_F^{MF})^2 - x_F^{PF})}{(x_F^{PF} - (x_F^{MF})^2)^2} \end{pmatrix}$$

Table 6: Empirical Results from Propositions 4-9 (one measure)

Param.	Prop. 4	Prop. 5	Prop. 6	Prop. 7	Prop. 8	Param.	Prop. 9
β_F	0.227	0.128	0.147	0.158	0.143	β_F^S	0.147
	(0.023)	(0.003)	(0.001)	(0.002)	(0.016)		(0.002)
						β_F^D	0.077
							(0.002)
β_M	0.338	0.591	0.468	0.456	0.478	β_M^S	0.733
	(0.023)	(0.007)	(0.004)	(0.005)	(0.058)		(0.003)
						β_M^D	0.431
							(0.005)
ρ	0.165	0.262	0.287	0.287	0.287	ρ	0.185
	(0.057)	(0.002)	(0.001)	(0.001)	(0.001)		(0.001)
ρ'	0.476		0.406	0.537		ρ'	0.329
	(0.035)		(0.003)	(0.004)			(0.002)
η			0.000	0.000			
			(0.002)	(0.003)			
N	2,800,897	2,800,897	507,885	507,885	507,885		2,800,897
MSE	9.816e-07	1.770e-16	5.741e-05	9.493e-04	8.877e-17		2.986e-16

Notes: Results from the estimation of Propositions 4, 5, 6, 7, 8, and 9, for 1880 cohort, using one measure. A cohort is defined as any male child in the census for the specified year; hence, there can be some overlap in the individuals in these trees.

Given $\rho = \frac{(x_F^{MF})^2}{x_F^{PF}}$ and $0 < \rho < 1$:

$$0 < \frac{(x_F^{MF})^2}{x_F^{PF}} < 1$$
$$\Rightarrow 0 < (x_F^{MF})^2 < x_F^{PF}$$

This gives us our first key relationship: $(x_F^{MF})^2 < x_F^{PF}$, which means the denominator $x_F^{PF} - (x_F^{MF})^2$ is positive.

From $0 < \beta_M < 1$

Table 7: Empirical Results for Propositions 1, 2, and 3 (1870 cohort)

Parameter	Prop. 1	Prop. 2	Prop. 3
β_F	0.163	0.900	0.448
	(0.006)	(0.004)	(0.002)
β_M	0.738		
	(0.010)		
ρ	0.385	0.385	0.385
	(0.003)	(0.003)	(0.002)
N	254,341	254,341	254,341
MSE	3.821e-16	8.668e-16	6.106e-06

Notes: Empirical results from the estimation of Propositions 1, 2, and 3, for 1870 cohort, using two measures. A cohort is defined as any male child in the Census for the specified year; hence, there can be some overlap in the individuals in these trees.

Given $\beta_M = \frac{x_F^{MF} - x_F^{PF} \cdot x_{PF}^{MF}}{x_F^{PF} - (x_F^{MF})^2}$ and $0 < \beta_M < 1$:

$$0 < \frac{x_F^{MF} - x_F^{PF} \cdot x_{PF}^{MF}}{x_F^{PF} - (x_F^{MF})^2} < 1$$

Since we know from our previous constraint that $x_F^{PF} - (x_F^{MF})^2 > 0$, we can multiply all terms by this positive denominator:

$$0 < x_F^{MF} - x_F^{PF} \cdot x_{PF}^{MF} < x_F^{PF} - (x_F^{MF})^2$$
$$\Rightarrow 0 < x_F^{MF} - x_F^{PF} \cdot x_{PF}^{MF}$$
$$\Rightarrow x_F^{PF} \cdot x_{PF}^{MF} < x_F^{MF}$$

For the upper bound:

$$x_F^{MF} - x_F^{PF} \cdot x_{PF}^{MF} < x_F^{PF} - (x_F^{MF})^2$$
$$\Rightarrow x_F^{MF} + (x_F^{MF})^2 < x_F^{PF} + x_F^{PF} \cdot x_{PF}^{MF}$$
$$\Rightarrow x_F^{MF}(1 + x_F^{MF}) < x_F^{PF}(1 + x_{PF}^{MF})$$

From $0 < \beta_F < 1$

Given $\beta_F = \frac{x_F^{MF}}{x_F^{PF}} - \frac{x_F^{MF} - x_F^{PF} \cdot x_{PF}^{MF}}{x_F^{PF} - (x_F^{MF})^2}$ and $0 < \beta_F < 1$:

Since $\beta_M = \frac{x_F^{MF} - x_F^{PF} \cdot x_{PF}^{MF}}{x_F^{PF} - (x_F^{MF})^2}$, we can rewrite β_F as:

$$\beta_F = \frac{x_F^{MF}}{x_F^{PF}} - \beta_M$$

Table 8: Empirical Results from Propositions 4-9 (1870 cohort)

Param.	Prop. 4	Prop. 5	Prop. 6	Prop. 7	Prop. 8	Param.	Prop. 9
β_F	0.102	0.160	0.193	0.207	0.184	β_F^S	0.187
	(0.042)	(0.005)	(0.002)	(0.003)	(0.005)		(0.002)
						β_F^D	0.112
							(0.003)
β_M	0.505	0.591	0.461	0.442	0.480	β_M^S	0.741
	(0.042)	(0.009)	(0.004)	(0.007)	(0.014)		(0.004)
						β_M^D	0.410
							(0.006)
ρ	0.526	0.351	0.386	0.386	0.386	ρ	0.244
	(0.040)	(0.002)	(0.001)	(0.001)	(0.001)		(0.002)
ρ'	0.758		0.539	0.673		ρ'	0.401
	(0.020)		(0.004)	(0.006)			(0.002)
η			0.000	0.000			
			(0.002)	(0.003)			
N	684,663	684,663	626,490	626,490	626,490		684,663
MSE	9.297e-14	2.739e-15	5.244e-05	9.612e-04	1.382e-16		2.814e-16

Notes: Results from the estimation of Propositions 4, 5, 6, 7, 8, and 9, for 1870 cohort, using two measures. A cohort is defined as any male child in the census for the specified year; hence, there can be some overlap in the individuals in these trees.

With $0 < \beta_F < 1$:

$$0 < \frac{x_F^{MF}}{x_F^{PF}} - \beta_M < 1$$

$$\Rightarrow \beta_M < \frac{x_F^{MF}}{x_F^{PF}} < 1 + \beta_M$$

Since $\beta_M < 1$, we know:

$$\frac{x_F^{MF}}{x_F^{PF}} < 2$$

We now analyze the partial derivatives of ρ, β_F, and β_M with respect to each of the three variables.

- Derivatives of ρ
  - With respect to x_F^{MF}

$$\frac{\partial \rho}{\partial x_F^{MF}} = \frac{\partial}{\partial x_F^{MF}} \left(\frac{(x_F^{MF})^2}{x_F^{PF}} \right) = \frac{2x_F^{MF}}{x_F^{PF}}$$

Sign: POSITIVE
Since $x_F^{MF} > 0$ and $x_F^{PF} > 0$, we have $\frac{\partial \rho}{\partial x_F^{MF}} > 0$.

  - With respect to x_F^{PF}

$$\frac{\partial \rho}{\partial x_F^{PF}} = \frac{\partial}{\partial x_F^{PF}} \left(\frac{(x_F^{MF})^2}{x_F^{PF}} \right) = -\frac{(x_F^{MF})^2}{(x_F^{PF})^2}$$

Sign: NEGATIVE
Since $x_F^{MF} > 0$ and $x_F^{PF} > 0$, we have $\frac{\partial \rho}{\partial x_F^{PF}} < 0$.

  - With respect to x_{PF}^{MF}

$$\frac{\partial \rho}{\partial x_{PF}^{MF}} = \frac{\partial}{\partial x_{PF}^{MF}} \left(\frac{(x_F^{MF})^2}{x_F^{PF}} \right) = 0$$

Sign: ZERO
Since ρ does not depend on x_{PF}^{MF}, we have $\frac{\partial \rho}{\partial x_{PF}^{MF}} = 0$.

- Derivatives of β_F
  - With respect to x_F^{MF}

$$\frac{\partial \beta_F}{\partial x_F^{MF}} = \frac{\partial}{\partial x_F^{MF}} \left(\frac{x_F^{MF}}{x_F^{PF}} - \frac{x_F^{MF} - x_F^{PF} \cdot x_{PF}^{MF}}{x_F^{PF} - (x_F^{MF})^2} \right) = \frac{1}{x_F^{PF}} + \frac{2x_F^{MF} \cdot x_F^{PF} \cdot x_{PF}^{MF} - (x_F^{MF})^2 - x_F^{PF}}{(x_F^{PF} - (x_F^{MF})^2)^2}$$

Sign: INDETERMINATE
The first term $\frac{1}{x_F^{PF}} > 0$ is clearly positive.
For the second term, the denominator $(x_F^{PF} - (x_F^{MF})^2)^2 > 0$ is positive.
The numerator $2x_F^{MF} \cdot x_F^{PF} \cdot x_{PF}^{MF} - (x_F^{MF})^2 - x_F^{PF}$ has no clear sign based on our constraints.
Therefore, the overall sign of $\frac{\partial \beta_F}{\partial x_F^{MF}}$ cannot be determined without additional information.

– With respect to x_F^{PF}

$$\frac{\partial \beta_F}{\partial x_F^{PF}} = \frac{\partial}{\partial x_F^{PF}} \left(\frac{x_F^{MF}}{x_F^{PF}} - \frac{x_F^{MF} - x_F^{PF} \cdot x_{PF}^{MF}}{x_F^{PF} - (x_F^{MF})^2} \right)$$
$$= -\frac{x_F^{MF}}{(x_F^{PF})^2} + \frac{x_F^{MF}(1 - x_{PF}^{MF} \cdot x_F^{MF})}{(x_F^{PF} - (x_F^{MF})^2)^2}$$

Sign: INDETERMINATE
The first term $-\frac{x_F^{MF}}{(x_F^{PF})^2} < 0$ is negative.
For the second term, the denominator $(x_F^{PF} - (x_F^{MF})^2)^2 > 0$ is positive.
For the numerator $x_F^{MF}(1 - x_{PF}^{MF} \cdot x_F^{MF})$, since $x_{PF}^{MF} \cdot x_F^{MF} < 1$ (as both $x_{PF}^{MF}, x_F^{MF} < 1$), we have $x_F^{MF}(1 - x_{PF}^{MF} \cdot x_F^{MF}) > 0$.
This means the first term is negative and the second term is positive.
The overall sign depends on which term has greater magnitude.
Therefore, the sign of $\frac{\partial \beta_F}{\partial x_F^{PF}}$ cannot be determined without additional information.

– With respect to x_{PF}^{MF}

$$\frac{\partial \beta_F}{\partial x_{PF}^{MF}} = \frac{\partial}{\partial x_{PF}^{MF}} \left(\frac{x_F^{MF}}{x_F^{PF}} - \frac{x_F^{MF} - x_F^{PF} \cdot x_{PF}^{MF}}{x_F^{PF} - (x_F^{MF})^2} \right)$$
$$= \frac{x_F^{PF}}{x_F^{PF} - (x_F^{MF})^2}$$
$$= \frac{x_F^{PF}}{x_F^{PF} - (x_F^{MF})^2} \cdot \frac{x_F^{PF} - (x_F^{MF})^2}{x_F^{PF} - (x_F^{MF})^2}$$
$$= \frac{x_F^{PF}(x_F^{PF} - (x_F^{MF})^2)}{(x_F^{PF} - (x_F^{MF})^2)^2}$$

Sign: POSITIVE
Since $x_F^{PF} > (x_F^{MF})^2$ (from our constraints), the numerator $x_F^{PF}(x_F^{PF} - (x_F^{MF})^2) > 0$ is positive.
The denominator $(x_F^{PF} - (x_F^{MF})^2)^2 > 0$ is positive.
Therefore, $\frac{\partial \beta_F}{\partial x_{PF}^{MF}} > 0$ is positive.

- Derivatives of β_M

– With respect to x_F^{MF}

$$\frac{\partial \beta_M}{\partial x_F^{MF}} = \frac{\partial}{\partial x_F^{MF}} \left(\frac{x_F^{MF} - x_F^{PF} \cdot x_{PF}^{MF}}{x_F^{PF} - (x_F^{MF})^2} \right)$$
$$= \frac{(x_F^{PF} - (x_F^{MF})^2) - (x_F^{MF} - x_F^{PF} \cdot x_{PF}^{MF})(2x_F^{MF})}{(x_F^{PF} - (x_F^{MF})^2)^2}$$
$$= \frac{x_F^{PF} - (x_F^{MF})^2 - 2x_F^{MF}(x_F^{MF} - x_F^{PF} \cdot x_{PF}^{MF})}{(x_F^{PF} - (x_F^{MF})^2)^2}$$
$$= \frac{x_F^{PF} - (x_F^{MF})^2 - 2(x_F^{MF})^2 + 2x_F^{MF} \cdot x_F^{PF} \cdot x_{PF}^{MF}}{(x_F^{PF} - (x_F^{MF})^2)^2}$$
$$= \frac{x_F^{PF} - (x_F^{MF})^2 - 2(x_F^{MF})^2 + 2x_F^{MF} \cdot x_F^{PF} \cdot x_{PF}^{MF}}{(x_F^{PF} - (x_F^{MF})^2)^2}$$
$$= \frac{x_F^{PF} - 2(x_F^{MF})^2 - (x_F^{MF})^2 + 2x_F^{MF} \cdot x_F^{PF} \cdot x_{PF}^{MF}}{(x_F^{PF} - (x_F^{MF})^2)^2}$$
$$= \frac{x_F^{PF} - 3(x_F^{MF})^2 + 2x_F^{MF} \cdot x_F^{PF} \cdot x_{PF}^{MF}}{(x_F^{PF} - (x_F^{MF})^2)^2}$$

Sign: INDETERMINATE
The denominator $(x_F^{PF} - (x_F^{MF})^2)^2 > 0$ is positive.
For the numerator, the term $x_F^{PF} - 3(x_F^{MF})^2$ could be positive or negative, and the term $2x_F^{MF} \cdot x_F^{PF} \cdot x_{PF}^{MF}$ is positive.
Without additional constraints on the relative magnitudes of these terms, the sign of $\frac{\partial \beta_M}{\partial x_F^{MF}}$ cannot be determined.

– With respect to x_F^{PF}

$$\frac{\partial \beta_M}{\partial x_F^{PF}} = \frac{\partial}{\partial x_F^{PF}} \left(\frac{x_F^{MF} - x_F^{PF} \cdot x_{PF}^{MF}}{x_F^{PF} - (x_F^{MF})^2} \right)$$
$$= \frac{-x_{PF}^{MF} \cdot (x_F^{PF} - (x_F^{MF})^2) - (x_F^{MF} - x_F^{PF} \cdot x_{PF}^{MF}) \cdot 1}{(x_F^{PF} - (x_F^{MF})^2)^2}$$
$$= \frac{-x_{PF}^{MF} \cdot x_F^{PF} + x_{PF}^{MF} \cdot (x_F^{MF})^2 - x_F^{MF} + x_F^{PF} \cdot x_{PF}^{MF}}{(x_F^{PF} - (x_F^{MF})^2)^2}$$
$$= \frac{-x_{PF}^{MF} \cdot x_F^{PF} + x_{PF}^{MF} \cdot (x_F^{MF})^2 - x_F^{MF} + x_F^{PF} \cdot x_{PF}^{MF}}{(x_F^{PF} - (x_F^{MF})^2)^2}$$
$$= \frac{x_{PF}^{MF} \cdot (x_F^{MF})^2 - x_F^{MF}}{(x_F^{PF} - (x_F^{MF})^2)^2}$$
$$= \frac{x_F^{MF}(x_{PF}^{MF} \cdot x_F^{MF} - 1)}{(x_F^{PF} - (x_F^{MF})^2)^2}$$

Sign: NEGATIVE
The denominator $(x_F^{PF} - (x_F^{MF})^2)^2 > 0$ is positive.
For the numerator, since $x_{PF}^{MF} < 1$ and $x_F^{MF} < 1$, we have $x_{PF}^{MF} \cdot x_F^{MF} < 1$.
Therefore, the term $(x_{PF}^{MF} \cdot x_F^{MF} - 1) < 0$ is negative.

Since $x_F^{MF} > 0$, the numerator $x_F^{MF}(x_{PF}^{MF} \cdot x_F^{MF} - 1) < 0$ is negative.
Therefore, $\frac{\partial \beta_M}{\partial x_F^{PF}} < 0$ is negative.

– With respect to x_{PF}^{MF}

$$\begin{aligned}
\frac{\partial \beta_M}{\partial x_{PF}^{MF}} &= \frac{\partial}{\partial x_{PF}^{MF}} \left(\frac{x_F^{MF} - x_F^{PF} \cdot x_{PF}^{MF}}{x_F^{PF} - (x_F^{MF})^2} \right) \\
&= \frac{-x_F^{PF}}{x_F^{PF} - (x_F^{MF})^2} \\
&= -\frac{x_F^{PF}}{x_F^{PF} - (x_F^{MF})^2} \cdot \frac{x_F^{PF} - (x_F^{MF})^2}{x_F^{PF} - (x_F^{MF})^2} \\
&= -\frac{x_F^{PF}(x_F^{PF} - (x_F^{MF})^2)}{(x_F^{PF} - (x_F^{MF})^2)^2} \\
&= \frac{x_F^{PF}((x_F^{MF})^2 - x_F^{PF})}{(x_F^{PF} - (x_F^{MF})^2)^2}
\end{aligned}$$

Sign: NEGATIVE
Since $x_F^{PF} > (x_F^{MF})^2$ (from our constraints), the term $((x_F^{MF})^2 - x_F^{PF}) < 0$ is negative.
The denominator $(x_F^{PF} - (x_F^{MF})^2)^2 > 0$ is positive.
Therefore, $\frac{\partial \beta_M}{\partial x_{PF}^{MF}} < 0$ is negative.

We can arrange all these derivatives in a 3×3 matrix form, where rows correspond to the functions (ρ, β_F, β_M) and columns correspond to the variables (x_F^{MF}, x_F^{PF}, x_{PF}^{MF}). In summary, the constraints $0 < \beta_F, \beta_M, \rho < 1$, allow us to determine the signs of six out of the nine partial derivatives. The most critical constraint is $x_F^{PF} > (x_F^{MF})^2$, which derives directly from $0 < \rho < 1$, and ensures that denominators are positive and helps establish the signs of several derivatives. For the three derivatives with indeterminate signs, additional constraints or specific parameter values would be needed to determine their signs conclusively. Matrix J below shows the derivatives of each estimator (structural parameter) as a function of each empirical moment

$$J = \begin{pmatrix} \frac{\partial \rho}{\partial x_F^{MF}} & \frac{\partial \rho}{\partial x_F^{PF}} & \frac{\partial \rho}{\partial x_{PF}^{MF}} \\ \frac{\partial \beta_F}{\partial x_F^{MF}} & \frac{\partial \beta_F}{\partial x_F^{PF}} & \frac{\partial \beta_F}{\partial x_{PF}^{MF}} \\ \frac{\partial \beta_M}{\partial x_F^{MF}} & \frac{\partial \beta_M}{\partial x_F^{PF}} & \frac{\partial \beta_M}{\partial x_{PF}^{MF}} \end{pmatrix}; sign(J) = \begin{pmatrix} + & - & 0 \\ ? & ? & + \\ ? & - & - \end{pmatrix}$$

D.2 Measurement Error

In studies of intergenerational mobility, measurement error and attenuation bias could be a concern. Social status is transmitted across generations, but it is hard to measure precisely. Researchers usually have access to some variable, such as income, that is only imperfectly correlated with status. Therefore, the correlation in outcomes between father and child is lower than the correlation of their status. The difference (bias) between these two correlations would be a function of how correlated the outcome used is to status. Historical data very rarely have information on income, but usually contain information on occupation. Our approach here is to improve the usual estimates for income in a given occupation by allowing variation across time and space, and, more importantly, providing better estimates for farmer's income. Song et al. (2020) take a different approach and create a measure of normalized literacy by occupation. Instead of looking at the literacy of each individual, they look at the fraction of individuals that are literate and have the same occupation, and then assign that index to all individuals with the same occupation. Ward (2023) takes a similar approach but computes indexes not only by occupation, but by occupation, race, and state. The goal of these alternative measures of status is to reduce the attenuation bias in the mobility estimates. We now discuss whether some estimators could be *immune* to attenuation bias.

Curtis (2022) uses the ratio estimator in Chadwick and Solon (2002) to measure assortative mating, i.e., $\rho' = m^F_{MF}/m^F_{PF}$, and notes that this estimator is not subject to attenuation bias if the bias θ when computing the correlation in the numerator is the same as the bias when computing the correlation in the denominator. Recent papers have used this estimator to compute the degree of marital assortment (Clark and Cummins, 2022; Clark et al., 2022; Clark, 2023). Let $\widehat{m^F_{PF}}$ and $\widehat{m^F_{MF}}$ be the income correlation between X^F_i and X^{PF}_i, and X^F_i and X^{MF}_i, respectively; and let m^F_{PF} and m^F_{MF} be the status correlations. Then, $\widehat{\rho'} \equiv \frac{\widehat{m^F_{MF}}}{\widehat{m^F_{PF}}} = \frac{\theta m^F_{MF}}{\theta m^F_{PF}} = \frac{m^F_{MF}}{m^F_{PF}} \equiv \rho'$. In other words, the ratio estimator is *immune* to attenuation bias.[28] In general, any estimator would be *immune* to attenuation bias if

1. The bias θ in any pairwise correlation is the same.
2. The degree of the correlations in the numerator is the same as that in the denominator.

To illustrate this point, we can look at our mobility estimators in Proposition 4 (see Appendix A for details). Following the notation above we can write.

$$\widehat{\beta_F} \equiv \frac{\widehat{m^F_{PF}}\widehat{m^C_{PF}} - \widehat{m^C_{MF}}\widehat{m^F_{MF}}}{\widehat{m^F_{PF}}^2 - \widehat{m^F_{MF}}^2} = \frac{\theta m^F_{PF}\theta m^C_{PF} - \theta m^C_{MF}\theta m^F_{MF}}{(\theta m^F_{PF})^2 - (\theta m^F_{MF})^2} = \frac{m^F_{PF}m^C_{PF} - m^C_{MF}m^F_{MF}}{(m^F_{PF})^2 - (m^F_{MF})^2} \equiv \beta_F$$

$$\widehat{\beta_M} \equiv \frac{\widehat{m^F_{PF}}\widehat{m^C_{MF}} - \widehat{m^C_{PF}}\widehat{m^F_{MF}}}{\widehat{m^F_{PF}}^2 - \widehat{m^F_{MF}}^2} = \frac{\theta m^F_{PF}\theta m^C_{MF} - \theta m^C_{PF}\theta m^F_{MF}}{(\theta m^F_{PF})^2 - (\theta m^F_{MF})^2} = \frac{m^F_{PF}m^C_{MF} - m^C_{PF}m^F_{MF}}{(m^F_{PF})^2 - (m^F_{MF})^2} \equiv \beta_M$$

In other words, our mobility estimators in Proposition 4 are *immune* to attenuation bias.[29] If the researcher is very concerned about attenuation bias in their sample, but it is not very concerned about changes in mobility over time, Proposition 4 would be the right choice for her.

D.3 Occupational Scores

One particular type of measurement error common in the literature is generated by using occupational scores. In historical data, it is common to have information on occupation, but not on income. The solution is to impute income based on occupation, and maybe other characteristics

[28] We have emphasized that there are inherent trade-offs between the assumptions imposed and the parameters that can be estimated. The ratio estimator is one point on that set of trade-offs that is particularly useful when particular data requirements are encountered. For example, the ratio estimator is particularly useful in settings, like the PSID, where one cannot construct trees, but can independently estimate husband/father correlations and husband/father-in-law correlations.

[29] Notice that this property need not apply to all our estimators.

such as state of residency, year, and race. We perform a sensitivity exercise to assess how this type of measurement error could affect our estimates. We replicate the simulation exercise as explained in Appendix B. Once the simulated data is generated, using our underlying structure, we fix the number of quantiles q and compute the median income of each quantile. These median incomes are the reference points, or income scores. We then assign each individual the income score that is closest to their simulated income. In practice, this is a dataset that introduces non-classical measurement error and would induce lower pairwise correlations. It is a more extreme version of non-classical measurement than what we would see in the data. If we consider each quantile as an occupation, our simulation exercise assumes that there is no income overlap between occupations.

Figure 10 displays simulation results using Proposition 1 for different quantile sizes. The measurement error produces attenuation bias in all pairwise correlations. This attenuation bias has an a priori ambiguous effect on the structural estimates. From formula 12 we see that the power in the numerator is 2 and the power in the denominator is 1. If the degree of attenuation bias is the same in both correlations, then due to attenuation bias, using Proposition 1 would underestimate ρ. With a small number of quantiles ($q = 10$) we see that β_M is over-estimated and β_F and ρ are underestimated. With an intermediate number of quantiles ($q = 50$), the bias has vanished and the estimates and not statistically different from the true values. For reference, in the 1900 US Decennial census there are 264 occupations reported. In summary, we learn several things from this exercise. First, unsurprisingly, using occupational scores instead of actual incomes would underestimate all pairwise correlations, due to attenuation bias. Second, the effect on each structural parameters is ambiguous, but we can use comparative statics using the formulas for our estimators to predict whether they would be overestimated or underestimated. The simulation results confirm the analytical results. Third, the biases disappear at relatively low quantile sizes ($q = 50$). This suggest that the type of measurement error induce by using occupational scores would not have an effect on our estimators.

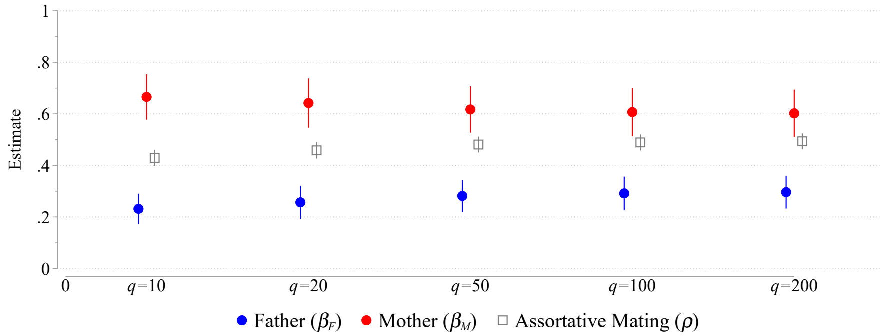
Figure 11 displays simulation results using Proposition 4 for different quantile sizes. As shown above in Appendix D.2, we would expect our mobility estimates β_F and β_M to be immune to measurement error. Indeed, this is what we see. Even at very small quantile sizes ($q = 10$ and $q = 20$) we see the estimates for β_F and β_M be not statistically different from the true values. The estimates for assortative mating ρ and ρ', however, remain far from the true values until we reach intermediate quantile sizes ($q = 50$ and $q = 100$). This exercise corroborates the results in Figure 10 using Proposition 1. In addition to that, it we learn that the intuition above regarding estimates being immune to classical measurement error, extends to the type of non-classical measurement error study here, which is commonly found in the literature.

E Economic Model

In this section, we show a simple economic model in the spirit of Becker and Tomes (1979) when there are two parents, instead of one. Each parent care about their own consumption and about investing in a public good: their child. Unlike in Becker and Tomes (1979), where the investment decision by a single parent is a simple trade off between utility today (consumption) and utility in the future (investment in my child), the problem here is complicated by the bargaining between the parents. We first present the problem and the notation in subsection E.1, and then solve the bargaining problem à la Nash in subsection E.2.

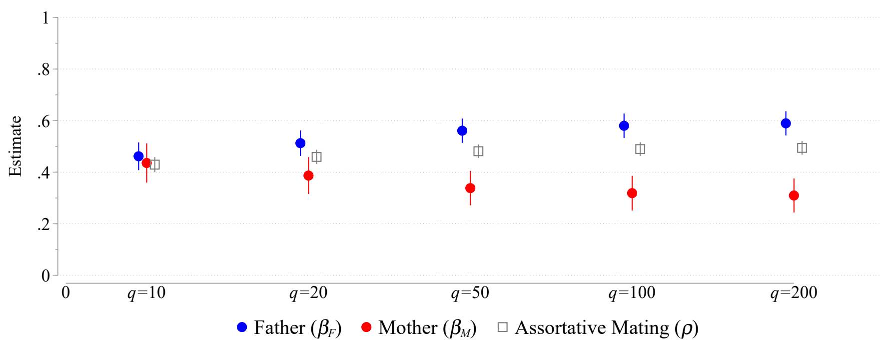
Figure 10: Simulation Results from Proposition 1 with aggregated data.

A. Simulations with $\beta_F = 0.3$, $\beta_M = 0.6$, and $\rho = \rho' = 0.5$.



Correlation	True	g=10	g=20	g=50	g=100	g=200
$\mathbb{E}\left[X_i^F X_i^M\right]$	0.5	0.4759	0.4994	0.5081	0.5106	0.5112
$\mathbb{E}\left[X_i^F X_i^{PF}\right]$	0.6	0.5566	0.5838	0.5958	0.5987	0.6000
$\mathbb{E}\left[X_i^F X_i^{MF}\right]$	0.56	0.5244	0.5520	0.5607	0.5623	0.5621
$\mathbb{E}\left[X_i^{PF} X_i^{MF}\right]$	0.62	0.5812	0.6071	0.6153	0.6189	0.6209

B. Simulations with $\beta_F = 0.6$, $\beta_M = 0.3$, and $\rho = \rho' = 0.5$.



Correlation	True	g=10	g=20	g=50	g=100	g=200
$\mathbb{E}\left[X_i^F X_i^M\right]$	0.5	0.4763	0.5003	0.5098	0.5128	0.5131
$\mathbb{E}\left[X_i^F X_i^{PF}\right]$	0.75	0.6993	0.7306	0.7467	0.7500	0.7516
$\mathbb{E}\left[X_i^F X_i^{MF}\right]$	0.56	0.5291	0.5512	0.5600	0.5616	0.5617
$\mathbb{E}\left[X_i^{PF} X_i^{MF}\right]$	0.62	0.5819	0.6089	0.6188	0.6207	0.6217

*Not*es: Estimated parameters and confidence intervals at $p = 0.05$ using Proposition 1, and data generated aggregating at the median of the nearest quantile. We use a sample size of $n = 10,000$ and $1,000$ simulations.

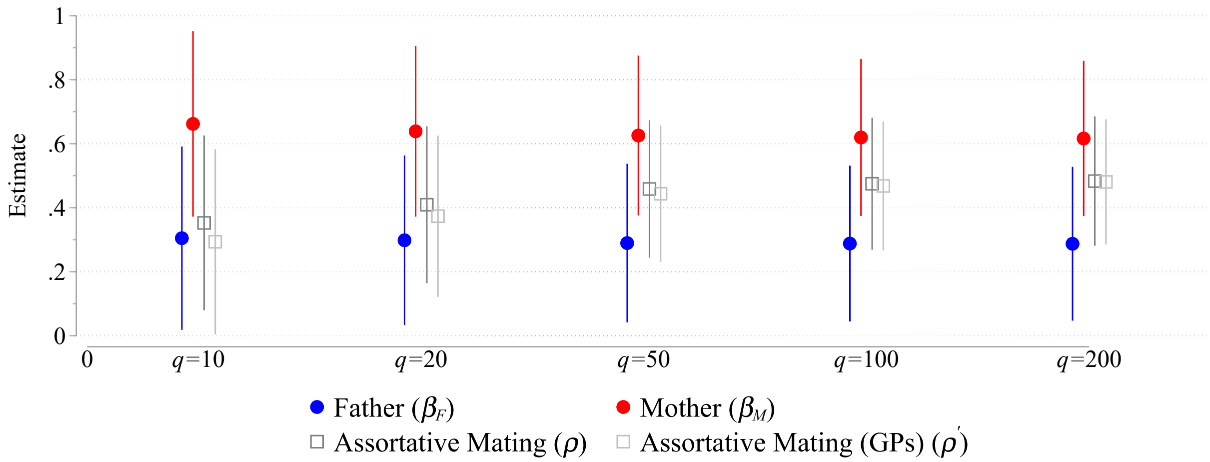
E.1 Becker and Tomes (1979) with mother and father

Here we reproduce a simple version of the model in Becker and Tomes (1979), as written in Solon (2014), but extending the analysis to two parents. We modify the formulas slightly by removing the logs to adapt to our utility function. Each parent $j \in \{F, M\}$ must allocate their lifetime income X_i^j between the parent's own consumption C_i^j and investment in the child's human capital I_i^j:

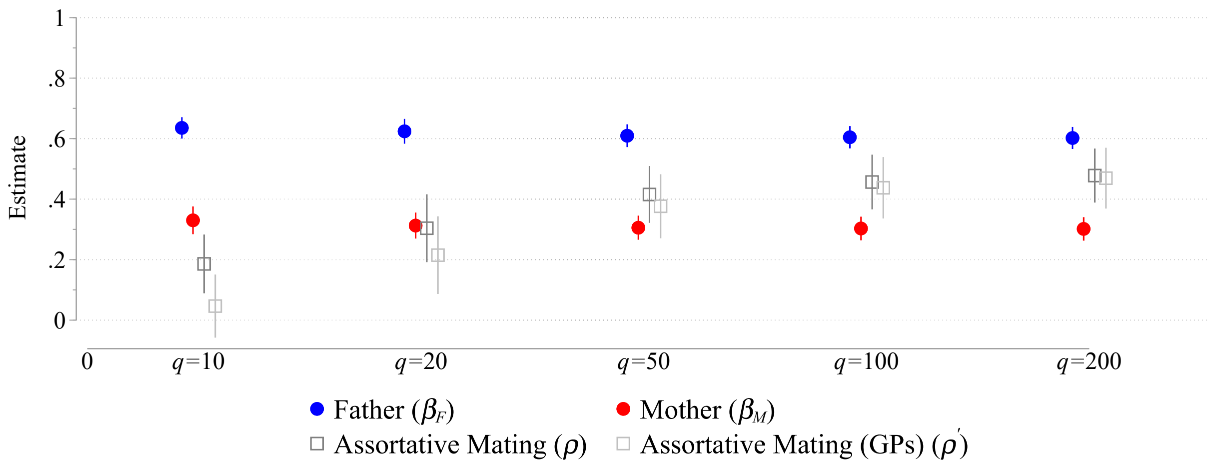
$$X_i^F = C_i^F + I_i^F \tag{45}$$

Figure 11: Simulation Results from Proposition 4 with aggregated data.

A. Simulations with $\beta_F = 0.3$, $\beta_M = 0.6$, and $\rho = \rho' = 0.5$.



B. Simulations with $\beta_F = 0.6$, $\beta_M = 0.3$, and $\rho = \rho' = 0.5$.



Notes: Estimated parameters and confidence intervals at $p = 0.05$ using Proposition 4, and data generated aggregating at the median of the nearest quantile. We use a sample size of $n = 10,000$ and $1,000$ simulations.

$$X_i^M = C_i^M + I_i^M \tag{46}$$

The technology translating total investment $I_i \equiv I_i^F + I_i^M$ into the child's human capital h_i is

$$h_i = \theta I_i + \nu_i \tag{47}$$

where $\theta > 0$ represents a positive marginal product for human capital investment, and ν_t denotes the human capital endowment of the child. The child life-time income X_i^C is determined by the semi-log earnings function

$$X_i^C = \mu + p h_i \tag{48}$$

where p is the earning return to human capital. Substituting equation (47) into equation (48) yields

$$X_i^C = \mu + \gamma I_i + p\nu_i \tag{49}$$

where $\gamma = \theta p$ is the elasticity of the child's income with respect to investment in the child's human capital. The parents divide their income income X_i^j between the parent's own consumption C_i^j and investment in the child's human capital I_i^j, so as to maximize the following utility function

$$U_i^F = (1-\alpha_F)C_i^F + \alpha_F X_i^C \tag{50}$$

$$U_i^M = (1-\alpha_M)C_i^M + \alpha_M X_i^C \tag{51}$$

The altruism parameters α_F and α_M, which lie between 0 and 1, measure the parent's taste for investing in their child's human capital relative to their own consumption. This utility function can be rewritten as

$$U_i^F = (1-\alpha_F)\left(X_i^F - I_i^F\right) + \alpha_F\mu + \alpha_F\gamma\left(I_i^F + I_i^M\right) + \alpha_F p\nu_i \tag{52}$$

$$U_i^M = (1-\alpha_M)\left(X_i^M - I_i^M\right) + \alpha_M\mu + \alpha_M\gamma\left(I_i^F + I_i^M\right) + \alpha_M p\nu_i \tag{53}$$

which expresses the objective functions in terms of the choice variables I_i^F and I_i^M. In the next subsection we derive the solution of this model when parents bargain à la Nash.

E.2 Bargaining with a Public Good

Here we follow the simple setting of intra-household bargaining in Manser and Brown (1980). We first define the elements needed for the analysis and their formulation when utility preferences are hedonic. In particular the utility functions are

$$U^F\left(I_i^F, I_i^M\right) = (1-\alpha_F)\left(X_i^F - I_i^F\right) + \alpha_F\gamma\left(I_i^F + I_i^M\right) + \alpha_F\mu + \alpha_F p\nu_i \tag{54}$$

$$U^M\left(I_i^F, I_i^M\right) = (1-\alpha_M)\left(X_i^M - I_i^M\right) + \alpha_M\gamma\left(I_i^F + I_i^M\right) + \alpha_M\mu + \alpha_M p\nu_i \tag{55}$$

We now compute the optimal choice for each of the parents if they were in autarky. In this case, each parent will invest all their income in their child's education and the solution is:[30]

$$I_i^j = X_i^j$$

$$C_i^j = X_i^j - I_i^j = 0$$

We can use this autarky solution to compute the threat points for each of the parents

$$V_i^F = \alpha_F\gamma\left(X_i^F\right) + \alpha_F\mu + \alpha_F p\nu_i \tag{56}$$

$$V_i^M = \alpha_M\gamma\left(X_i^M\right) + \alpha_M\mu + \alpha_M p\nu_i \tag{57}$$

The Nash bargaining solution objective function is

$$\max_{I_i^F, I_i^M} N = \left[\left(U^F\left(I_i^F, I_i^M\right) - V_i^F\right)\left(U^M\left(I_i^F, I_i^M\right) - V_i^M\right)\right]$$

We define $\Delta^F\left(I_i^F, I_i^M\right) \equiv U^F\left(I_i^F, I_i^M\right) - V_i^F$. We can simplify and get

$$\Delta^F\left(I_i^F, I_i^M\right) = (1-\alpha_F(1+\gamma))\left(X_i^F - I_i^F\right) + \alpha_F\gamma I_i^M \tag{58}$$

[30]With the hedonic preferences, in autarky, they would invest all their income if $(1-\alpha_i) < \alpha_i\gamma + \alpha_i\mu + \alpha_i p e_t$. Notice that this is a threat point because in autarky, one parent does not enjoy the benefits invested in the children by the other parent, since they are not their children.

The expression for $\Delta^M\left(I_i^F, I_i^M\right)$ is analogous. Notice that $\Delta^F\left(I_i^F, I_i^M\right)$ depends on both I_i^F and I_i^M. Moreover, the expression in equation (58) depends on the ability ν_i and the market term p, since $\gamma = \theta p$. We can write the objective function now as

$$\max_{I_i^F, I_i^M} N = \left[\left((1-\alpha_F(1+\gamma))\left(X_i^F - I_i^F\right) + \alpha_F \gamma I_i^M\right)\left((1-\alpha_M(1+\gamma))\left(X_i^M - I_i^M\right) + \alpha_M \gamma I_i^F\right)\right]$$

This can be rewritten as

$$\max_{I_i^F, I_i^M} N = \left[A \cdot I_i^M \cdot I_i^F + B\left(I_i^M\right)^2 + C\left(I_i^F\right)^2 + D \cdot I_i^F + E \cdot I_i^M + F\right]$$

where $A = (1-\alpha_F(1+\gamma))(1-\alpha_M(1+\gamma))$,
$B = -\alpha_F \gamma (1-\alpha_M(1+\gamma))$,
$C = -\alpha_M \gamma (1-\alpha_F(1+\gamma))$,
$D = \alpha_M \gamma (1-\alpha_F(1+\gamma)) X_{t-1}^F - (1-\alpha_F(1+\gamma))(1-\alpha_M(1+\gamma)) X_{t-1}^M$,
$E = \alpha_F \gamma (1-\alpha_M(1+\gamma)) X_{t-1}^M - (1-\alpha_M(1+\gamma))(1-\alpha_F(1+\gamma)) X_{t-1}^F$,
$F = (1-\alpha_F(1+\gamma)) X_{t-1}^F (1-\alpha_M(1+\gamma)) X_{t-1}^M$.

The FOC are

$$\frac{\partial N}{\partial I_i^F} = A I_i^M + 2C I_i^F + D = 0$$

$$\frac{\partial N}{\partial I_i^M} = A I_i^F + 2B I_i^M + E = 0$$

The solution to this system is

$$I_i^F = \frac{2BD - AE}{A^2 - 4BC}$$

$$I_i^M = \frac{2CE - AD}{A^2 - 4BC}$$

Notice that the only terms that contains the parents incomes X_i^F and X_i^M, are D, E, and F. F is a constant and does not appear in the FOC. The terms D and E appear in the numerator in both expressions, and they appear additively. Therefore, we can write the solution of this problem in the form

$$I_i^F = a_F X_i^F + b_F X_i^M + c_F \tag{59}$$

$$I_i^M = a_M X_i^F + b_M X_i^M + c_M \tag{60}$$

where a_j, b_j, and c_j are constants that are known functions of the original parameters $(\alpha_F, \alpha_M, \gamma)$ and the endowments $\left(X_i^F, X_i^M\right)$. The total investment in the child is then $I_i \equiv I_i^F + I_i^M = aX_i^F + bX_i^M + c$.[31] With that investment, using equation (49), we get the equation for the income of the child

$$X_t^C = a\gamma X_i^F + b\gamma X_i^M + \mu + c\gamma + p\nu_i \tag{61}$$

Equation (61) corresponds to equation (2) with $\beta_F = a\gamma$ and $\beta_M = b\gamma$, and where we normalized the variables so that there is no constant term in the equation. In summary, the model proposed here, under Nash bargaining shows that the investment in the child's education, and thus the child's income in the next period, is a linear combination of the income of each parent in the previous

[31] With $d_F = a_F + a_M$, $b = b_F + b_M$ and $c = c_F + c_M$.

generation.

E.3 Inheritability

Becker and Tomes (1979); Solon (2014) shows, for the monoparental case, that when there is an unobservable inherited component, we can write the income of the child as a linear function of the income of the father and the income of the paternal grandfather. In that case, the coefficient on the paternal grandfather will be negative and small. We now extend our model to allow for unobserved inheritability of traits (Lochner, 2016). This is represented by equation (62), with two extra parameters: λ_F is the heritability coefficient from the father and λ_M is the heritability coefficient from the mother

$$e_i^C = \lambda_F e_i^F + \lambda_M e_i^M + \nu_i \tag{62}$$

This is just an extension of our previous model where the unobservable component in the equation of the child e_i^C is not exogenous, but rather it is a linear combination function of the unobservable components in the equations of the father and the mother. The new equations for the child and parents are then

$$X_i^C = \beta_F X_i^F + \beta_M X_i^M + e_i^C \tag{63}$$

$$X_i^F = \beta_F X_i^{PF} + \beta_M X_i^{PM} + e_i^F \tag{64}$$

$$X_i^M = \beta_F X_i^{MF} + \beta_M X_i^{MM} + e_i^M \tag{65}$$

We now take equation (64) and multiply it by λ_F, and we take equation (65) and multiply it by λ_M. We then take equation (63) and subtract, equation (64) (multiplied by λ_F) and equation (65) (multiplied by λ_M), and we get

$$X_i^C - \lambda_F X_i^F - \lambda_M X_i^M = \beta_F X_i^F + \beta_M X_i^M + e_i^C - \lambda_F \left(\beta_F X_i^{PF} + \beta_M X_i^{PM} + e_i^F\right) - \lambda_M \left(\beta_F X_i^{MF} + \beta_M X_i^{MM} + e_i^M\right)$$

$$X_i^C = (\lambda_F + \beta_F) X_i^F + (\beta_M + \lambda_M) X_i^M - \lambda_F \beta_F X_i^{PF} - \lambda_F \beta_M X_i^{PM} - \lambda_M \beta_F X_i^{MF} - \lambda_M \beta_M X_i^{MM} + \nu_i \tag{66}$$

where $\nu_i = e_i^C - \lambda_F e_i^F - \lambda_M e_i^M$. The income of the child is now a function of the incomes of the parents and the grandparents. The coefficients on the parents are the sum of the mobility coefficients β_F and β_M and the heritability coefficients λ_F and λ_M. The coefficients on the grandparents are are the products of the mobility coefficients β_F and β_M and the heritability coefficients λ_F and λ_M. Notice that there are six variables in the right hand side (two parents and four grandparents), but there are only four parameters: the mobility coefficients β_F and β_M and the heritability coefficients λ_F and λ_M. Therefore, we could use a 3-generation tree, which generates six moments, to identify these four parameters and the two assortment parameters ρ and ρ'. This model is a natural extension to the monoparental model with inheritable characteristics and our baseline model with two parents, but no unobservable inherited characteristics. First, if there is no inheritability of unobserved characteristic, i.e., $\lambda_F = \lambda_M = 0$, then we are back to our baseline model. Equation (66) becomes

$$X_i^C = \beta_F X_i^F + \beta_M X_i^M + \nu_i$$

which is our original equation. Second, if there are no maternal effects, i.e., $\beta_M = \lambda_M = 0$. Equation (66) becomes

$$X_i^C = (\lambda_F + \beta_F) X_i^F - \lambda_F \beta_F X_i^{PF} + \nu_i$$

which is the equation in the literature with monoparental households with unobserved inheritability Becker and Tomes (1979); Solon (2014).