Non-parametric Identification and Estimation of the Intergenerational Elasticity

Alejandro Puerta-Cuartas*

The intergenerational elasticity (IGE) has traditionally served as the primary measure of income persistence across generations. However, its estimation has long been hindered by the unobservability of lifetime income. We address this challenge by first establishing the nonparametric identification of the IGE, leveraging family characteristics and partial income data under standard missing-at-random assumptions. Building on this foundation, we derive a consistent and locally robust estimator using Neyman orthogonal moments that delivers valid inference. Our framework enables comparable IGE estimates across time and place, resolving long-standing challenges in identification and inference. Using the Panel Study of Income Dynamics, we estimate an IGE of 0.69 for the United States.

I. Introduction

The intergenerational elasticity (IGE) plays a central role in studying income persistence across generations. It underpins key applications, such as characterizing mobility at the national level (Solon, 1992; Mazumder, 2005), comparing mobility across countries (Björklund and Jäntti, 1997; Vosters and Nybom, 2017), and examining trends in intergenerational transmission (Aaronson and Mazumder, 2008). The IGE is defined as the slope coefficient from regressing the child's lifetime income on the parents' lifetime income and a constant. However, since lifetime income is rarely observed in practice, researchers typically rely on midlife income averages, which introduces measurement error and leads to downward-biased estimates (Nybom and Stuhler, 2017).

Recent methodological advances have made significant progress in addressing the attenuation bias. Early studies often relied on single-year measures of income or earnings, which were highly sensitive to transitory shocks and reporting errors (Becker and Tomes, 1986). To mitigate these issues, researchers began averaging income over multiple years (de Wolff and van Slijpe, 1973; Hauser et al., 1975; Freeman, 1978; Tsai, 1983). However, it was the seminal work of Solon (1992) that established income averaging over 3 to 5 years as a standard approach for measuring fathers' lifetime income. More recent studies have further improved reliability by adopting longer averaging periods, such as 10-year averages (Mazumder, 2005), or utilizing all available years for the child generation (Lee and Solon, 2009). More recently, Mello et al. (2024) leverage rich data on family characteristics to further reduce sensitivity to the age at which child income is measured.

* Department of Economics, Universidad Carlos III, Madrid, Spain, E-mail: alpuerta@eco.uc3m.es

Yet, a more fundamental question remains unaddressed: whether the IGE is identified with incomplete income data. Addressing this gap is essential for reliable estimation and meaningful comparisons across studies, as identification underpins consistency (Gabrielsen, 1978). Our analysis reveals that current methods, while empirically useful, do not identify the IGE, which explains the persistent challenges in achieving comparable estimates across applications. This insight motivates our formal analysis, which establishes the identification conditions necessary to deliver a consistent estimator of the intergenerational elasticity.

This paper shows non-parametric identification of the IGE in the presence of incomplete income data. To address the challenge of not observing lifetime income, we leverage family characteristics, which are commonly available in datasets used for mobility research. Under standard missing at random assumptions, we show that the IGE can be recovered from conditional means, including the conditional income profiles of parents and children, and the conditional covariance matrix of parental income. Non-parametric identification arises from using the structural definition of lifetime income as the sum of (log) annual earnings during working life, instead of the commonly assumed generalized error-in-variables model.

We construct a consistent and locally robust estimator for the intergenerational elasticity. Since identification depends on conditional expectations, bias or estimation error in their first-step estimation can distort IGE inference. To address this, we construct an orthogonal moment function following Chernozhukov et al. (2022). Specifically, we augment the moment identifying the intergenerational elasticity with its first-step influence function. This ensures that flexible estimation of the conditional expectations has no effect, locally, on the IGE estimate. The resulting estimator, obtained by minimizing the orthogonal moment via cross-fitting, delivers reliable and comparable estimates for analyzing intergenerational income persistence.

Asymptotic normality is established for the proposed estimator, enabling valid inference for the IGE. Our result directly follows from Theorem 9 in Chernozhukov et al. (2022), and explicitly accounts for uncertainty introduced by the first-stage estimation of nuisance parameters. Consequently, the resulting confidence intervals achieve correct asymptotic coverage.

To evaluate the performance of our proposed estimator in finite samples, we conduct a simulation study that mirrors key empirical features commonly found in intergenerational mobility data. We evaluate our locally robust estimator relative to three alternatives: (i) a naive plug-in estimator, uncorrected for first-step estimation errors, (ii) the standard IGE estimation approach, and (iii) the recently proposed life-cycle estimator (Mello et al., 2024). In contrast to the alternative estimators, which suffer from both bias and poor coverage, the locally robust estimator exhibits negligible bias which vanishes to zero as the sample size increases and maintains coverage rates close to nominal levels.

Applying our method to the United States using the core sample of the Panel Study of Income Dynamics (PSID), yields an IGE estimate of 0.69 (95% CI: 0.575, 0.804). In contrast, alternative estimators exhibit downward bias, producing estimates ranging of 0.38 to 0.51. A naive plug-in estimator delivers a slightly different point estimate of 0.6

and narrower confidence intervals, consistent with our simulation results and confirming its undercoverage. Our findings highlight the importance of identification, combined with local robustness, for studying income mobility through the lens of the intergenerational elasticity.

The contribution of this paper is two-fold. Methodologically, we develop a new framework that reconceptualizes IGE estimation as a missing data problem shifting focus from reducing attenuation bias to identification. Empirically, we deliver a consistent, asymptotically normal estimator that enables meaningful comparisons of intergenerational income persistence across time and place. An `R` package is currently under development to make the estimator readily accessible.

Our work engages with the literature's shift toward rank-based measures by addressing the methodological limitations that have sidelined the IGE. The recent shift toward rank-based measures (Mazumder, 2016) has been driven primarily by methodological concerns. Yet this trend has overlooked a crucial distinction: rank-rank correlations capture relative mobility, while the IGE reflects absolute mobility. As such, the selection of mobility measure should be guided by the research question, rather than implementation challenges. Our study addresses this gap by establishing valid inference for settings where the IGE is better suited, such as cross-country comparisons and analyses of mobility trends.

The remainder of the paper is as follows: Section II establishes the lack of identification under conventional procedures, characterizes the resulting biases, and discusses their empirical consequences. Section III shows non-parametric identification of the IGE, provides a locally robust estimator, and establishes asymptotic normality for the proposed estimator. Section IV reports simulation results, and Section V presents an empirical application of our estimator, measuring the IGE in the United States. Section VI concludes. Proofs are provided in the Appendix.

II. The Identification Problem of the IGE: Sources and Consequences

This section formalizes the intergenerational elasticity (IGE) as the parameter of interest and demonstrates that non-identification fundamentally limits its estimation. We show that the most widely used estimation procedure fails to identify the IGE, resulting in four sources of bias that vary with study design and income dynamics. We illustrate how these biases undermine comparability of estimates within the same country, cross-country comparisons, and analyses of temporal mobility trends. Finally, we show that the recently proposed life-cycle estimator by Mello et al. (2024), improves upon the traditional approach by eliminating two sources of bias in IGE estimation.

Our object of interest is the intergenerational elasticity, which captures the degree to which income differences between parents are associated with income differences among their children. It is defined as the slope coefficient from regressing the child's lifetime income $\left(Y_c^P\right)$ on father's lifetime income $\left(Y_f^P\right)$ and a constant:

$$Y_c^P = \alpha_0 + \beta_0 Y_f^P + u, \quad \mathbb{E}\left[u\left(1, Y_f^P\right)'\right] = 0, \tag{1}$$

$$Y_g^P := \sum_{t=1}^{T} Y_{gt}, \quad g \in \{c, f\}$$

where β_0 is the intergenerational elasticity, u is the component of the child's lifetime income uncorrelated to parental lifetime income, and lifetime income is defined as the sum of log annual income, where Y_{gt} represents log annual income for generation g in year t, with $g = c$ for children, and $g = f$ for fathers. This definition of lifetime income enables identification of the IGE under missing data and aligns with standard empirical practice. A detailed motivation and empirical justification for this choice are provided in Appendix A1.

While Mitnik and Grusky (2020) highlights important conceptual nuances in interpreting the intergenerational elasticity, our analysis adopts the conventional interpretation of IGE as capturing the persistence of income differences across generations. Accordingly, we interpret the IGE as the degree to which income differences between parents are associated with income differences among their children.

When lifetime income is observed for both generations, the IGE in equation (1) is identified as (see Appendix A2 for a detailed proof)

$$\beta_0 = \frac{\mathbb{E}\left[\left(Y_c^P - \mathbb{E}\left[Y_c^P\right]\right)\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)\right]}{\mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right]}. \tag{2}$$

However, complete lifetime income data is rarely available, forcing researchers to rely on snapshots of income over short age intervals (Mazumder, 2005).

Most empirical studies utilize longitudinal datasets such as the Panel Study of Income Dynamics (PSID), which contain only partial income trajectories for parents and children, along with additional individual and family characteristics. Formally, the observed data consists of a random independent and identically distributed (i.i.d) sample of $W = \left(\boldsymbol{Y}_c \odot \boldsymbol{D}_c, \boldsymbol{Y}_f \odot \boldsymbol{D}_f, \boldsymbol{D}_c, \boldsymbol{D}_f, \boldsymbol{X}\right)$, where $\boldsymbol{Y}_c$ and $\boldsymbol{Y}_f$ are T-dimensional random vectors with information on (log) annual child and parental income, respectively. $\odot$ is the element-wise product, meaning that each element Y_{gt} is observed if $D_{gt} = 1$ and not observed otherwise, for $g \in \{c, f\}$. The vector $\boldsymbol{X}$ contains observed characteristics for both generations.

In this setting, the IGE cannot be uniquely determined from the observed data, due to the fundamental unobservability of lifetime income. Accordingly, the IGE is not identified without further assumptions. This poses a significant challenge, as a parameter that is not identified cannot be consistently estimated (Gabrielsen, 1978).[1] While the literature has made valuable contributions by focusing on measuring lifetime income as accurately as possible, particularly to reduce attenuation bias from measurement error, it has overlooked the fundamental identification issue within a missing data framework.

[1] This follows from the fact that consistency implies identifiability; therefore, if A implies B, the negation of B implies the negation of A.

We now examine the identification and consistency properties of the customary approach to estimating the IGE, discussing its implications for comparability across studies and its broader impact on empirical research in intergenerational mobility.

The standard approach to estimate the IGE, which we label the mid-life income (MI) estimator, consists of a two-step approach. First, it proxies (average) lifetime income[2] as the average of T_f and T_c (log) annual income observations around mid-life for the fathers and the children, respectively.[3] In the second step, it regresses the estimated children's lifetime income on estimated parental lifetime income and a constant. A comprehensive discussion of the MI estimator's definition, theoretical underpinnings, assumptions and sources of bias can be found in Appendix A3.

Formally, the MI estimand is the slope coefficient in the projection:

$$\begin{aligned}\tilde{Y}_c^P &= \alpha^{MI} + \beta^{MI}\tilde{Y}_f^P + u^{MI}, \quad \mathbb{E}\left[u^{MI}\left(1, \tilde{Y}_f^P\right)'\right] = 0,\\ \tilde{Y}_g^P &\coloneqq \frac{1}{T_g}\sum_{j\in\mathcal{M}_g} Y_{gj}D_{gj}, \quad g \in \{c, f\},\end{aligned}$$

where $D_{gj} = 1$ when Y_{gj} is observed and zero otherwise, $\mathcal{M}_g$ is a set of pre-defined mid-life years for generation g,[4] and $T_g \coloneqq \sum_{j\in\mathcal{M}_g} D_{gj}$ is the number of years used for the average.

The following Theorem establishes that the MI estimand does not identify the IGE, even when imposing the strong identifying assumptions described in Appendix A3. The proof is provided in Appendix A4.

Theorem 1. *Given a random i.i.d sample of* $W = \left(\boldsymbol{Y}_c \odot D_c, \boldsymbol{Y}_f \odot D_f, D_c, D_f, \boldsymbol{X}\right)$ *the MI estimand does not identify the IGE under assumptions 1-MI and 2-MI. Instead, it is identified as:*

$$\beta^{MI} = \frac{\mathbb{E}\left[\left(Y_c^P - \mathbb{E}\left[Y_c^P\right]\right)\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)\right] + \frac{1}{T_c}\sum_t \mathbb{E}\left[Y_f^P v_{ct}\middle| D_{ct} = 1, t \in \mathcal{M}_c\right] \times p_c\left(t \in \mathcal{M}_c\right)}{\mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right] + \frac{1}{T_f^2}\sum_t\sum_j \mathbb{E}\left[v_{ft}v_{fj}\middle| D_{ft} = 1, D_{fj} = 1, \{t, j\} \in \mathcal{M}_f\right] \times p_f\left(\{t, j\} \in \mathcal{M}_f\right)}.$$

where v_{ct} and v_{ft} are children and parental age shocks to (log) annual income, $p_c\left(t \in \mathcal{M}_c\right)$ is the probability of observing child's income during mid-life in year t, and $p_f\left(\{t, j\} \in \mathcal{M}_f\right)$ is the probability of observing parental income during mid-life in years t and j.

The attenuation bias in IGE estimates is well-documented, and recent methodological advances have meaningfully reduced this bias. However, these advances overlook a more fundamental problem: the most widely used procedure fails to identify the IGE. This identification failure implies that MI-based estimates are estimating different population quantities across studies, making them inherently non-comparable even when identical

[2]Recall that it is equivalent to use lifetime income or average lifetime income to estimate the IGE.
[3]T_c is commonly set to 1 since measurement error in the dependent variable (usually) only affects efficiency.
[4]While some papers define parental mid-life according to their offspring's age (Chetty et al., 2014; Blanden et al., 2014), others use parental age (Björklund and Jäntti, 1997; Mazumder, 2005), so $\mathcal{M}_f$ can differ from $\mathcal{M}_c$.

estimation procedures are applied. Corollary 1.1 formalizes how this lack of identification gives rise to four distinct sources of bias that vary systematically with study design choices and underlying income dynamics. The proof is provided in Appendix A5.

Corollary 1.1. *Given a random i.i.d sample of* $W = \left(\boldsymbol{Y}_c \odot D_c, \boldsymbol{Y}_f \odot D_f, D_c, D_f, \boldsymbol{X}\right)$ *the MI estimator is inconsistent for* β_0 *under assumptions 1-MI and 2-MI. Specifically, its probability limit is given by:*

(3)

$$\hat{\beta}_n^{MI} \xrightarrow{p} \frac{\beta_0 \mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right] + \overbrace{\frac{1}{T_c}\sum_t \mathbb{E}\left[Y_f^P v_{ct} \middle| D_{ct} = 1, t \in \mathcal{M}_c\right]}^{(c)} \times \overbrace{p_c\left(t \in \mathcal{M}_c\right)}^{(d)}}{\mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right] + \underbrace{\frac{1}{T_f^2}\sum_t \sum_j \mathbb{E}\left[v_{ft} v_{fj} \middle| D_{ft} = 1, D_{fj} = 1, \{t, j\} \in \mathcal{M}_f\right]}_{(a)} \times \underbrace{p_f\left(\{t, j\} \in \mathcal{M}_f\right)}_{(b)}},$$

where v_{ct} *and* v_{ft} *are children and parental age shocks to (log) annual income,* $p_c\left(t \in \mathcal{M}_c\right)$ *is the probability of observing child's income during mid-life in year t, and* $p_f\left(\{t, j\} \in \mathcal{M}_f\right)$ *is the probability of observing parental income during mid-life in years t and j.*

Corollary 1.1 encapsulates the main sources of bias in estimating the IGE discussed in the literature. These include (a) the downward bias by measurement error (Solon, 1992; Mazumder, 2005), (b) the sensitivity of the IGE estimates to low, zero, and missing income (Couch and Lillard, 1998; Dahl and DeLeire, 2008; Chetty et al., 2014; Nybom and Stuhler, 2016), (c) the prediction error in the child's income being correlated with parental income (Nybom and Stuhler, 2016), and (d) the sensitivity to the number of years and the selected year(s) to measure children's income (Mello et al., 2024). If components (c) and (a) in equation (3) were equal to zero, the bias in β_n^{MI} would vanish, and it would converge in probability to β_0. However, there is evidence that (c) is different from zero (Mello et al., 2024), and (a) is always different from zero because average annual income is a noisy measure of lifetime income. A brief discussion of each component that hinders the consistent estimation of the IGE by $\hat{\beta}_n^{MI}$ can be found in Appendix A3.

This result contributes to the literature on bias in IGE estimation in two ways. First, as shown in Appendix A6, it generalizes prior findings in Solon (1992) and Nybom and Stuhler (2016), extending them to account for missing income data. Second, it provides a theoretical foundation for empirical findings about sample inclusion criteria (Couch and Lillard, 1998) and not observing complete income profiles (Heidrich, 2016) by showing how observation probabilities (terms (b) and (d)) influence the probability limit.

Corollary 1.1 explains the divergence in IGE estimates across studies: each of the four sources of bias varies systematically with study design choices and underlying income dynamics. Due to heterogeneity in institutional contexts and research protocols, these factors differ across datasets, regions, countries, and time, leading to non-comparable IGE estimates based on $\hat{\beta}_n^{MI}$.

A compelling example lies in the significant variation of recent IGE estimates for the

United States, which range from 0.35 to 0.65 (Mello et al., 2024). Even when components (a) and (c) in equation (3) are held constant across two studies, differing definitions of mid-life lead to changes in the estimand, suggesting that the studies are estimating fundamentally different objects.

The identification challenges in estimating the IGE may likewise affect analyses of mobility trends, as the four bias components in equation (3) can confound differences across cohorts. First, the relationship between children's income growth and family affluence may weaken or strengthen across generations. Second, age-related income shocks may vary systematically between cohorts. Third, the propensity scores, particularly the probability of observing child income (term (d)), can differ across cohorts due to survey design changes. For example, the PSID's transition from annual interviews (1968–1997) to biennial interviews thereafter mechanically reduces income observability for more recent cohorts, altering the MI estimand and potentially biasing trend comparisons. This structural limitation suggests that observed trends in intergenerational mobility over time may reflect differences in survey design and estimation bias rather than changes in income persistency.

These theoretical insights are supported empirically by Mello et al. (2024), who find that accounting for lifecycle effects leads to markedly different conclusions about how income mobility has evolved over time in Sweden. Estimates based on the MI estimator suggest that mobility declined sharply between the 1950s and 1970s cohorts. In contrast, the LC estimator yields remarkably stable mobility across these cohorts and a slight increase for those born in the 1980s.

Differences in estimated intergenerational elasticity across countries may reflect variation in bias magnitude rather than true differences in mobility. Just as in analyses of mobility trends, the four bias components in equation (3) can confound cross-country comparisons. These biases may arise from three primary sources: the persistence of transitory income shocks, captured by (a); the steepness of income growth for children from affluent families, represented by (c); and differing propensity scores ((b) and (d)), which may vary across countries even when mid-life income is defined consistently. This insight challenges the conventional interpretation of the Great Gatsby Curve, as IGE estimates are not directly comparable across countries, unlike other measures such as the Gini coefficient, which are based on the consistent estimates of a common estimand.

Understanding the empirical patterns behind sources of bias can significantly improve IGE measurement. Mello et al. (2024) evidence that children from affluent families exhibit faster income growth, even after controlling for observables. This causes the parental lifetime income to be correlated with the children's age shocks (component (c) in equation (3)). Consequently, measurement error in the child's lifetime income to not only affects efficiency but also introduce bias in the intergenerational elasticity estimate.

To address this concern, Mello et al. (2024) develop a life-cycle (LC) estimator that builds on two key innovations. First, it uses family background variables to purge the age-specific income prediction error (v_{ct}) of any correlation with parental income $\left(Y_f^P\right)$. Second, it constructs an estimate of children's lifetime income by: (i) predicting annual income at each age from family characteristics, and (ii) summing the predicted values in accordance with the lifetime income definition. A detailed exposition of the LC estimator, including its definition, theoretical underpinnings, and underlying assumptions, is provided in Appendix A7.

Formally, the LC estimand $\left(\beta^{LC}\right)$ is defined as the coefficient in the projection of the estimated lifetime income of the child on the mid-life average (log) income of the father:

$$\begin{aligned}\overline{Y}_c^P &= \alpha^{LC} + \beta^{LC}\tilde{Y}_f^P + u^{LC}, \quad \mathbb{E}\left[u^{LC}\tilde{Y}_f^P\right] = 0,\\ \overline{Y}_c^P &:= \sum_{t=1}^{T}\mathbb{E}\left[Y_{ct}|\boldsymbol{X}_{ct}\right], \quad \tilde{Y}_f^P := \frac{1}{T_f}\sum_{j\in\mathcal{M}_f} Y_{fj}D_{fj}.\end{aligned}$$

The following Corollary shows that the life-cycle estimator refines the MI approach by eliminating two sources of bias. This result relies on assumptions closely mirroring those of the MI estimand, supplemented by mild additional conditions (Assumptions 1-LC–4-LC in Appendix A7).

Corollary 1.2. *Given a random i.i.d sample of $W = \left(\boldsymbol{Y}_c \odot D_c, \boldsymbol{Y}_f \odot D_f, D_c, D_f, \boldsymbol{X}\right)$, the life-cycle estimator eliminates two sources of the mid-life Income estimator's bias under Assumptions 1-LC-4-LC, since its probability limit is given by*

$$\hat{\beta}_n^{LC} \xrightarrow{p} \frac{\beta_0\mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right]}{\mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right] + \frac{1}{T_f^2}\sum_t\sum_j\mathbb{E}\left[v_{ft}v_{fj}\middle|D_{ft} = 1, D_{fj} = 1, \{t,j\}\in\mathcal{M}_f\right]\times p_f\left(\{t,j\}\in\mathcal{M}_f\right)},$$

where v_{ft} are parental age shocks to (log) annual income and $p_f\left(\{t,j\}\in\mathcal{M}_f\right)$ is the probability of observing parental income during mid-life in years t and j.

Corollary 1.2 establishes that the LC estimator is robust to the age at which child income is measured. Mello et al. (2024) provide empirical evidence that their approach reduces sensitivity to the age at which child income is measured. Our result extends this finding by formalizing a stronger property: complete insensitivity. This robustness stems from projecting children's income onto family characteristics, which eliminates the correlation between age shocks and parental lifetime income. As a result, the probability of observing a child's income during midlife also disappears, since both terms interact multiplicatively in equation (3). With these source of bias removed, the only remaining threat to consistency in the LC estimator arises from measurement error in parental lifetime income.

The life-cycle estimator constitutes a substantial advancement beyond the mid-life estimator by mitigating two key sources of bias. Nonetheless, because inconsistency precludes identification, this exposes a fundamental limitation: reliable and comparable estimates across studies cannot be assured without formal identification. Therefore, we now turn our attention to the essential problem of IGE identification, setting the stage for consistent estimation.

III. Identification, Estimation, and Inference for the IGE with Incomplete Data

This section shows nonparametric identification of the IGE in the presence of incomplete income data leveraging the availability of family characteristics. We then build upon this foundation to construct a consistent and locally robust estimator based on Neyman orthogonal moments. Finally, we establish asymptotic normality for the proposed estimator, delivering valid inference for the intergenerational elasticity.

A. Non-parametric Identification

We establish the non-parametric identification of the IGE in the presence of incomplete income data, drawing on the framework of identification with observational data under unconfoundedness (Rosenbaum and Rubin, 1983). Under standard missing at random assumptions, we show that the IGE can be recovered from conditional means, including the conditional income profiles of parents and children, and the conditional covariance of parental income. Instead of imposing a parametric structure relating unobserved lifetime income to annual income we leverage the definition of lifetime income.

The fundamental challenge in identifying the IGE lies in its dependence on unobserved lifetime incomes. To address this, the literature has traditionally relied on the generalized error-in-variables (GEIV) model (Haider and Solon, 2006), which relates unobserved (average) lifetime income to observed (log) annual income through the parametric relationship

$$Y_{gt} = \lambda_t Y_g^P + v_{gt}, \quad \mathbb{E}\left[v_{gt} Y_g^P\right] = 0, \quad g \in \{c, f\}, \quad t = 1, \ldots, T, \tag{4}$$

where Y_g^P denotes (average) lifetime income, λ_t reflects life-cycle variations in income persistence, and v_{gt} represents an age-specific shock. The GEIV model provides a useful framework by linking unobserved lifetime income to observed annual income. However, it has several limitations. To begin with, when used for identification, the validity of the result relies on the correct specification of the model. In addition, it introduces T unknown parameters, λ_t, and $2T$ additional unobserved components, v_{gt}, which make necessary imposing additional orthogonality assumptions unlikely to hold in practice.

In constrast, we leverage the definition of lifetime income

$$Y_g^P = \sum_{t=1}^{T} Y_{gt}, \quad g \in \{c, f\},$$

which deterministically maps unobserved lifetime income to partially observed (log) annual income. This approach underpins the non-parametric nature of our identification result, enabling us to recover the intergenerational elasticity without restrictive functional form assumptions.

While the IGE fundamentally depends on unobserved lifetime incomes for both generations, the definition of lifetime income allows us to reformulate the target parameter in terms of partially observed (log) annual incomes:

$$\beta_0 = \frac{\mathbb{E}\left[\left(Y_c^P - \mathbb{E}\left[Y_c^P\right]\right)\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)\right]}{\mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right]} = \frac{\sum_{t=1}^T \sum_{j=1}^T \mathbb{E}\left[\left(Y_{ct} - \mathbb{E}\left[Y_c^P\right]\right)\left(Y_{fj} - \mathbb{E}\left[Y_f^P\right]\right)\right]}{\sum_{t=1}^T \sum_{j=1}^T \mathbb{E}\left[\left(Y_{ft} - \mathbb{E}\left[Y_f^P\right]\right)\left(Y_{fj} - \mathbb{E}\left[Y_f^P\right]\right)\right]}. \tag{5}$$

Although this is a crucial step, it is not sufficient for identification. To bridge this gap, we leverage observable characteristics by decomposing (log) annual income as

$$Y_{gt} = \mathbb{E}\left[Y_{gt} \mid \boldsymbol{X}_{gt}\right] + \epsilon_{gt}, \quad \mathbb{E}\left[\epsilon_{gt} \mid \boldsymbol{X}_{gt}\right] = 0, \quad g \in \{c, f\}, \quad t = 1, \ldots, T, \tag{6}$$

where $\boldsymbol{X}_{gt}$ are the elements in the observed characteristics $\boldsymbol{X}$ relevant for predicting (log) annual income of generation g at time t.[5]

To establish identification of the IGE, we proceed in two steps. First, we substitute the income decomposition into equation (5) and impose conditional mean independence and orthogonality assumptions involving observables and prediction errors, thereby eliminating dependence on unobserved components. Second, we impose standard missing-at-random assumptions to recover the necessary conditional moments from the available data. For the income profiles, the MAR assumption enables identification of the conditional mean:

$$\mathbb{E}[Y_{gt} \mid \boldsymbol{X}_{gt}] = \mathbb{E}[Y_{gt} \mid \boldsymbol{X}_{gt}, D_{gt} = 1],$$

where D_{gt} indicates income observability. In a similar way, we are able to identify the conditional second moments arising in the denominator of (5):

$$\mathbb{E}[Y_{ft}Y_{fj} \mid \boldsymbol{X}_{ftj}] = \mathbb{E}[Y_{ft}Y_{fj} \mid \boldsymbol{X}_{ftj}, D_{ft} = 1, D_{fj} = 1],$$

where $\boldsymbol{X}_{ftj}$ comprises the elements in the observed characteristics $\boldsymbol{X}$ relevant for predicting the covariance between parental incomes at ages t and j, and $\boldsymbol{X}_{ftj}$ is defined such that $\boldsymbol{X}_{ft} \subset \boldsymbol{X}_{ftj}$ for $t, j = 1, ..., T$. With these foundations in place, we now formally state our complete set of identifying assumptions.

Assumption 1-NP. *(Conditional Mean Independence and Orthogonality)*

i. The observable characteristics satisfy:

1. $\mathbb{E}\left[Y_{ct} \mid \boldsymbol{X}_{ct}, \boldsymbol{X}_{cj}, \boldsymbol{X}_{fj}\right] = \mathbb{E}\left[Y_{ct} \mid \boldsymbol{X}_{ct}\right]$ *for* $t, j = 1, ...T$,
2. $\mathbb{E}\left[Y_{ft} \mid \boldsymbol{X}_{ft}, \boldsymbol{X}_{ftj}, \boldsymbol{X}_{cj}\right] = \mathbb{E}\left[Y_{ft} \mid \boldsymbol{X}_{ft}\right]$ *for* $\boldsymbol{X}_{fj} \subset \boldsymbol{X}_{ftj}$, $t, j = 1, ...T$.

[5] Although the prediction error ϵ_{ct} in equation (6) can be interpreted as an age shock, it differs conceptually from the age shock v_{ct} in the GEIV model of equation (4). Specifically, ϵ_{ct} captures the component of (log) annual income that is not explained by observed parental and own characteristics, that is, the residual from a predictive model based on observables. In contrast, v_{ct} reflects transitory deviations from an individual's lifetime income and arises within a latent factor structure that distinguishes between the permanent and transitory components of income.

ii. The prediction errors of the children's annual income are uncorrelated to parental lifetime income

$$\mathbb{E}\left[\epsilon_{ct} Y_f^P\right] = 0, \quad t = 1, ..., T,$$
$$\epsilon_{ct} \coloneqq Y_{ct} - \mathbb{E}\left[Y_{ct} \mid \boldsymbol{X}_{ct}\right].$$

iii. Parental income prediction errors are uncorrelated for $|t - j| > h$

$$\mathbb{E}\left[\epsilon_{ft}\epsilon_{fj}\right] = 0, \quad \epsilon_{ft} \coloneqq Y_{ft} - \mathbb{E}\left[Y_{ft} \mid \boldsymbol{X}_{ft}\right].$$

The first condition establishes that $\boldsymbol{X}_{gt}$ contains all relevant predictors for annual income for generation g at time t, implying the remaining information in $\boldsymbol{X}$ provides no additional explanatory power. In Section V we illustrate that the specification of the characteristics predictive of income profiles and parental income covariance, namely, $\boldsymbol{X}_{ct}$, $\boldsymbol{X}_{ft}$, and $\boldsymbol{X}_{ftj}$,can be designed to satisfy Assumption 1-NP.*i* by construction.

Children's age shocks being correlated with parental lifetime income constitutes a source of bias of the MI estimator (component (c) in equation (3)). One of the empirical patterns driving this dependence stems for children from affluent families exhibiting faster income growth, even after controlling for observables (Mello et al., 2024). The life-cycle estimator addresses this by projecting children's annual income into the space of observables. In particular, by including in X_{ct} the interaction between average parental (log) annual income observations around mid-life $\left(\tilde{Y}_f^P = \frac{1}{T_f}\sum_{j\in\mathcal{M}_f} Y_{fj}D_{fj}\right)$ and children's age at time t, the prediction errors of children's income $\left(\epsilon_{gt} = Y_{gt} - \mathbb{E}\left[Y_{gt} \mid \boldsymbol{X}_{gt}\right]\right)$ become uncorrelated with parental lifetime income Y_f^P. Accordingly, Assumption 1-NP.*ii* imposes that children prediction errors are orthogonal to parental lifetime income, once we have controlled for the relevant family characteristics. In section III.D we provide a formal test for Assumption 1-NP.*ii*, as well as for Assumption 1-NP.*iii*.

The requirement of Assumption 1-NP.*iii* arises from the fundamental mismatch between the complete income profiles required by equation (5) and the income snapshots typically available in practice. Specifically, joint observation of parental incomes (Y_{ft}, Y_{fj}) (i.e., $D_{ft} = 1, D_{fj} = 1$) occurs only for relatively close time periods, such as incomes observed between ages 25 and 35 for a given individual. Consequently, income pairs for distant periods ($|t - j| > h$) are systematically absent in available data. Assumption 1-NP.*iii* addresses this empirical constraint by imposing that conditional on family characteristics $\boldsymbol{X}_{ftj}$, parental income shocks (prediction errors ϵ_{ft} and ϵ_{fj}) are uncorrelated for periods separated by more than h years. The availability of rich family characteristics $\boldsymbol{X}$ makes this assumption empirically plausible, as it allows us to account for the persistent components of intertemporal dependence.

Income autocorrelation captures two distinct sources: a permanent component driven by family characteristics (e.g., wealth, neighborhood quality, and race), and a transitory component, driven by short-term shocks (e.g., unemployment spells, economic crises, or health events). Crucially, while the influence of transitory shocks decays as the time gap $(t - j)$ widens, the effect of family background characteristics remains over time. Assumption 1-NP.*iii* states that parental annual income from periods more than h years in the past influences current income solely through observed characteristics. This specification serves dual purposes: it realistically captures the (conditional) short-memory of transitory shocks while accommodating the limitations inherent in available longitudinal datasets.

The following assumption formalizes some necessary conditions for identifying the intergenerational elasticity using partial income data and family characteristics. First, it requires that income realizations, for both generations and across nearby ages for fathers, are independent of their observability conditional on family characteristics. This ensures that survey attrition or non-reporting is not systematically associated with unobserved income determinants, ruling out selection bias. Second, it imposes an overlap condition guaranteeing sufficient data coverage across individuals and age windows, preventing estimates from being driven by specific reporting patterns or missing subpopulations. Together, these conditions prevent two key threats to validity: estimates being distorted either by systematic missingness (e.g., concentrated among low-income families) or by over-reliance on narrow age clusters. When satisfied, they ensure that inference is driven by income dynamics rather than data availability.

Assumption 2-NP. *(Missing At Random)*

i. The missingness of children's annual income Y_{ct} is as good as random once we control for $\boldsymbol{X}_{ct}$

$$Y_{ct} \perp D_{ct} \mid \boldsymbol{X}_{ct}, \quad t = 1, ..., T.$$

ii. Given family characteristics, there is both missing and non-missing children incomes for every age

$$0 < p\left(D_{ct} = 1 \mid \boldsymbol{X}_{ct}\right) < 1 \quad a.s, \quad t = 1, ..., T.$$

iii. The missingness of parental annual income pairs $\left(Y_{ft}, Y_{fj}\right)$ is as good as random once we control for $\boldsymbol{X}_{ftj}$

$$\left(Y_{ft}, Y_{fj}\right) \perp \left(D_{ft}, D_{fj}\right) \mid \boldsymbol{X}_{ftj}, \quad \textit{for all } t - j > h > 0,$$

where $\boldsymbol{X}_{ftj}$ are the family characteristics predictive of parental income covariance between years t and j, and $\boldsymbol{X}_{ftj} \coloneqq \boldsymbol{X}_{ft}$ for $j = t$.

iv. Given family characteristics, there is both missing and non-missing parental incomes for every age and its neighboring ages

$$0 < p\left(D_{ft} = 1, D_{fj} = 1 \mid \boldsymbol{X}_{ftj}\right) < 1 \quad a.s., \quad \textit{for all } t - j > h > 0.$$

According to equation (5), the IGE depends on two distinct components: the covariance between parent and child income, and the covariance within parental income. As a result, identification requirements differ across generations. For children, unconfoundedness needs only apply to single income observations, as the IGE exploits contemporaneous parent-child pairs. For fathers, stronger conditions on income tuples are needed to capture the temporal structure of their income process. This explains why Assumptions 2-NP.*iii* and 2-NP.*iv* impose ignorability and boundedness requirements on these joint income realizations.

The assumption that income missingness in the PSID is missing at random is supported by empirical evidence. Fitzgerald et al. (1998) finds that attrition in the PSID is highly selective, primarily affecting lower socioeconomic individuals and those with unstable earnings, marriage, and migration histories, but these factors explain little of the overall attrition, and regression-to-the-mean effects mitigate selection bias. This conclusion is reinforced by Lillard and Panis (1998), who find that ignoring attrition induces only very mild biases in household income models.

Fitzgerald (2011) examines attrition in intergenerational models of health, education, and earnings, finding that sibling correlations in outcomes are marginally higher among individuals who remain in the panel longer, though the differences are not statistically significant. Models of intergenerational links with covariates show negligible attrition bias for females. In contrast, the evidence for males is mixed but generally weak, suggesting that conditioning on observables largely mitigates selective attrition. The study finds little evidence of attrition bias, though analyses of educational and earnings outcomes for men appear to benefit from conditioning on observables.

Schoeni and Wiemers (2015) show that applying sample weights reduces differences in intergenerational income elasticity estimates between the full sample, the attriting sample, and the non-attriting sample, rendering these differences statistically insignificant. Their findings highlight that attrition, particularly higher among lower-income individuals, is influenced by the correlation between child and parental income outcomes, emphasizing the importance of incorporating both parental and child characteristics in analyses of intergenerational mobility.

Taken together, the literature suggests that the MAR assumption for income missingness in the PSID is empirically plausible, provided analyses carefully account for relevant observables. To address the concerns raised by Schoeni and Wiemers (2015), particularly the influence of the correlation between parental and child income outcomes on observability, our analysis incorporates both parental and child characteristics in the conditioning set, thereby strengthening the plausibility of the MAR assumption in our analysis.

The following Theorem establishes the nonparametric identification of the intergenerational elasticity in the presence of incomplete income data and family characteristics. This fundamental result ensures that estimates derived from the identification result are comparable across studies, providing a building block to analyze intergenerational mobility under valid inference.

Theorem 2. *Given a random i.i.d sample of* $W = \left(\boldsymbol{Y}_c \odot D_c, \boldsymbol{Y}_f \odot D_f, D_c, D_f, \boldsymbol{X}\right)$, *the*

IGE is non-parametrically identified under assumptions 1-NP and 2-NP:

$$\beta_0 = \frac{\mathbb{E}\left[\sum_{t=1}^{T}\left(\mu_{ct}\left(\boldsymbol{X}_{ct},1\right)-\mu_c^P\right)\sum_{j=1}^{T}\left(\mu_{fj}\left(\boldsymbol{X}_{fj},1\right)-\mu_f^P\right)\right]}{\mathbb{E}\left[\sum_{|t-j|\le h}\sigma_{tj}\left(\boldsymbol{X}_{ftj},1,1\right)+\sum_{|t-j|>h}\left(\mu_{ft}\left(\boldsymbol{X}_{ft},1\right)-\mu_f^P\right)\left(\mu_{fj}\left(\boldsymbol{X}_{fj},1\right)-\mu_f^P\right)\right]}, \tag{7}$$

where the conditional expectation of (log) annual income for generation $g \in \{c, f\}$ *at time* t *is given by* $\mu_{gt}(\boldsymbol{X}_{gt}, 1) := \mathbb{E}\left[Y_{gt} \mid \boldsymbol{X}_{gt}, D_{gt} = 1\right]$*; the average lifetime income for generation is given by* $\mu_g^P := \mathbb{E}\left[\sum_{t=1}^{T}\mu_{gt}(\boldsymbol{X}_{gt}, 1)\right]$*; and the conditional covariance of parental (log) annual income between periods* t *and* j *is given by* $\sigma_{tj}\left(\boldsymbol{X}_{ftj}, 1, 1\right) := \mathbb{E}\left[\left(Y_{ft} - \mu_f^P\right)\left(Y_{fj} - \mu_f^P\right) \mid \boldsymbol{X}_{ftj}, D_{ft} = 1, D_{fj} = 1\right]$.

Theorem 2 establishes the identification of the intergenerational elasticity in the presence of incomplete income data. To the best of our knowledge, the only existing identification result in this framework is that of (An et al., 2022) who non-parametrically identify the mobility function relating children's to parents lifetime income. While their more general framework nests the linear IGE as a special case, since they leave the relationship of parental and child incomes unspecified, our approach offers three important advantages. First, we relax their classical errors-in-variables assumption (equation (4) with $\lambda_t = 1$) for two measurement periods, by exploiting the definition of lifetime income. Second, we relax the assumption that transitory shocks to children's income are uncorrelated with parental lifetime income and parental transitory shocks. In contrast, we assume that the prediction error of the children's annual income is uncorrelated to parental lifetime income conditional on family characteristics (Assumption 1-NP).[6] Finally, our framework explicitly addresses the missing data structure inherent in real-world income observations, while incorporating all available information on both income dynamics and family characteristics.

Our identification result provides two valuable contributions to the study of intergenerational mobility. First, it resolves persistent methodological challenges by establishing sufficient conditions for identifying the intergenerational elasticity from incomplete income observations and family characteristics. Second, and more importantly, it provides the theoretical foundation for constructing a consistent estimator, enabling researchers to obtain valid and comparable estimates of the intergenerational elasticity. We now proceed to derive this estimator formally.

B. Locally Robust Estimation of the IGE

To estimate the intergenerational elasticity, we adopt a GMM approach building upon Theorem 2. To express this identifying result as a moment condition, we first define the

[6] This accounts for the empirically documented correlation between parental income and children's income growth rates.

following conditional expectations:

$$
\begin{aligned}
\mu_{gt}\left(F_\tau\right)(z) &:= \mathbb{E}_\tau\left[Y_{gt}|Z_{gt}=z\right], \quad \boldsymbol{Z}_t := \left(\boldsymbol{X}_{gt}, D_{gt}\right), \quad g\in\{c,f\}, \quad t=1,...,T,\\
\sigma_{tj}\left(F_\tau\right)(z) &:= \mathbb{E}_\tau\left[\left(Y_{ft}-\mu_f^P\right)\left(Y_{fj}-\mu_f^P\right)|Z_{tj}=z\right], \quad \boldsymbol{Z}_{tj} := \left(\boldsymbol{X}_{ftj}, D_{ft}, D_{fj}\right), \quad t,j=1,...,T,\\
\mu_g^{1,T}\left(F_\tau\right) &:= \left(\mu_{1,t}\left(F_\tau\right)(z),...,\mu_{gt}\left(F_\tau\right)(z)\right), \quad g\in\{c,f\},\\
\sigma^{t,1,T}\left(F_\tau\right) &:= \left(\sigma_{t,1}\left(F_\tau\right)(z),...,\sigma_{t,T}\left(F_\tau\right)(z)\right), \quad t=1,...,T\\
\sigma^{1,T,1,T}\left(F_\tau\right) &:= \left(\sigma^{1,1,T}\left(F_\tau\right)(z),...,\sigma^{T,1,T}\left(F_\tau\right)(z)\right),\\
\gamma\left(F_\tau\right) &:= \left(\mu_c^{1,T}\left(F_\tau\right), \gamma^{f,1,T}\left(F_\tau\right), \gamma^{f,1,T,1,T}\left(F_\tau\right)\right),
\end{aligned}
$$

where $\mathbb{E}_\tau$ denotes the expectation under $F_\tau=(1-\tau)F_0+\tau H$, and F_0 is the unknown cumulative distribution function of W. Thus, equation (7), which identifies our parameter of interest β_0, can be rewritten as

$$
\begin{aligned}
&\mathbb{E}\left[g_1\left(W,\gamma_0,\beta_0,\mu_c^P,\mu_f^P\right)\right]=0,\\
&\quad g_1\left(W,\gamma_0,\beta_0,\mu_c^P,\mu_f^P\right)=\beta_0\sum_{|t-j|\le h}\sigma_{tj}\left(F_0\right)\left(\boldsymbol{X}_{ftj},1,1\right)\\
&\qquad +\beta_0\sum_{|t-j|>h}\left(\mu_{ft}\left(F_0\right)\left(\boldsymbol{X}_{ft},1\right)-\mu_f^P\right)\left(\mu_{fj}\left(F_0\right)\left(\boldsymbol{X}_{fj},1\right)-\mu_f^P\right)\\
&\qquad -\sum_{t=1}^T\left(\mu_{ct}\left(F_0\right)\left(\boldsymbol{X}_{ct},1\right)-\mu_c^P\right)\sum_{j=1}^T\left(\mu_{fj}\left(F_0\right)\left(\boldsymbol{X}_{fj},1\right)-\mu_f^P\right),
\end{aligned}
\tag{8}
$$

where $\mathbb{E}[\cdot]$ is the expectation under the true distribution of W (F_0) and $\gamma_0:=\gamma\left(F_0\right)$ is the probability limit under F_0 of a first step estimator $\hat{\boldsymbol{\gamma}}$. Notice that β_0 also depends on the mean of children and parental income $\left(\mu_c^P,\mu_f^P\right)$. According to equation (B2) these two parameters are identified by

$$
\mu_g^P=\mathbb{E}\left[\sum_{t=1}^T\mu_{gt}\left(F_0\right)\left(\boldsymbol{X}_{gt},1\right)\right] \quad g\in\{c,f\},
$$

so, by defining the augmented parameter of interest $\theta_0:=\left(\beta_0,\mu_c^P,\mu_f^P\right)$, the moment identifying μ_c^P can be expressed as

$$
\begin{aligned}
&\mathbb{E}\left[g_2\left(W,\gamma\left(F_0\right),\theta\right)\right]=0,\\
&\quad g_2\left(W,\gamma\left(F_0\right),\theta\right)=\sum_{t=1}^T\mu_{ct}\left(F_0\right)\left(\boldsymbol{X}_{ct},1\right)-\mu_c^P,
\end{aligned}
$$

and analogously for μ_f^P

$$\mathbb{E}\left[g_3\left(W, \gamma\left(F_0\right), \theta\right)\right] = 0,$$
$$g_3\left(W, \gamma\left(F_0\right), \theta\right) = \sum_{t=1}^{T} \mu_{ft}\left(F_0\right)\left(\boldsymbol{X}_{ft}, 1\right) - \mu_f^P.$$

Accordingly, the identifying moment is given by

$$g\left(W, \gamma\left(F_\tau\right), \theta\right) = \begin{pmatrix} g_1\left(W, \gamma\left(F_\tau\right), \theta\right) \\ g_2\left(W, \gamma\left(F_\tau\right), \theta\right) \\ g_3\left(W, \gamma\left(F_\tau\right), \theta\right) \end{pmatrix}.$$

The challenge in accurately estimating θ arises from the fact that the plug-in moment

$$\hat{g}(\theta) = \frac{1}{n}\sum_{i=1}^{n} g(W_i, \hat{\gamma}, \theta)$$

is sensitive to errors in $\hat{\gamma}$. That is, it inherits first-order bias from estimation errors in the preliminary nuisance parameter estimates $\hat{\gamma}$. To illustrate this sensitivity, consider a local perturbation of γ_0 given by $\gamma_\tau := \gamma_0 + \tau \cdot h$. A Taylor expansion reveals how these perturbations affect the moment condition:

$$\mathbb{E}[g(W, \gamma_\tau, \theta)] = \mathbb{E}[g(W, \gamma_0, \theta)] + \tau \cdot \frac{d}{d\tau}\mathbb{E}[g(W, \gamma_\tau, \theta)]\Bigg\|_{\tau=0} + o(\tau),$$

where the second component corresponds to the Gateaux derivative, which captures the sensitivity of the identifying moment condition to changes in $\boldsymbol{\gamma}$ under general misspecification.

Equivalently, the Gateaux derivative can be represented using the first-step influence function (FSIF), denoted by ϕ:

$$\frac{d}{d\tau}\mathbb{E}[g(W, \gamma_\tau, \theta)]\Bigg\|_{\tau=0} = \mathbb{E}[\phi(W, \gamma_0, \alpha_0, \theta)],$$

where α_0 is the Riesz representer associated with the functional of interest (γ).[7]

Chernozhukov et al. (2022) provide a general procedure to construct orthogonal moment functions[8] for GMM, where first steps have no effect, locally, on average moment

[7]The parameter β_0 is identified through equation (8), allowing us to express our target estimand as $\beta_0 = \mathbb{E}\left[g\left(W, \gamma_0, \mu_c^P, \mu_f^P\right)\right]$. By rearranging terms, we can equivalently express it as

$$\beta_0 = \mathbb{E}\left[\gamma_0 \alpha_0\right], \quad \text{for all possible } \gamma_0,$$

where α_0 is known as the Riesz representer of the functional γ_0.

[8]Orthogonal moment functions are those whose identifying moment conditions have zero derivatives with respect to

functions. In particular, the authors show that an orthogonal (locally robust) moment function ψ can be constructed by augmenting the identifying moment function g with its FSIF:

$$\psi\left(W,\gamma,\alpha,\theta\right)=g\left(W,\gamma,\theta\right)+\phi\left(W,\gamma,\alpha,\theta\right).$$

To understand the local robustness property, consider the Taylor expansion of the orthogonal moment function:

$$\begin{aligned}\mathbb{E}[\psi(W,\gamma_\tau,\alpha_0,\theta)]&=\mathbb{E}[g(W,\gamma_0,\theta)]-\tau\cdot\frac{d}{d\tau}\mathbb{E}[g(W,\gamma_\tau,\theta)]\Bigg\|_{\tau=0}+\tau\cdot\frac{d}{d\tau}\mathbb{E}[g(W,\gamma_\tau,\theta)]\Bigg\|_{\tau=0}+o(\tau)\\&=\mathbb{E}[g(W,\gamma_0,\theta)]+o(t)\end{aligned}$$

where the first-order terms cancel exactly by construction. Thus, estimation of θ becomes locally robust to errors in $\hat{\gamma}$.

In Appendix B1, we derive the FSIF for our identifying moment condition and establish that the locally robust moment takes the form:

$$\begin{aligned}(9)\quad &\psi\left(W,\gamma,\alpha,\theta\right)=\left(\psi_1\left(W,\gamma,\alpha_1,\theta\right),\psi_2\left(W,\gamma,\alpha_2,\theta\right),\psi_3\left(W,\gamma,\alpha_3,\theta\right)\right),\\
&\psi_1\left(W,\gamma,\alpha,\theta\right)=\beta\sum_{|t-j|\le h}\sigma_{tj}\left(\boldsymbol{X}_{ftj},1,1\right)+\beta\sum_{|t-j|>h}\left(\mu_{ft}\left(\boldsymbol{X}_{ft},1\right)-\mu_f^P\right)\left(\mu_{fj}\left(\boldsymbol{X}_{fj},1\right)-\mu_f^P\right)\\
&\quad-\sum_{t=1}^{T}\left(\mu_{ct}\left(\boldsymbol{X}_{ct},1\right)-\mu_c^P\right)\sum_{j=1}^{T}\left(\mu_{fj}\left(\boldsymbol{X}_{fj},1\right)-\mu_f^P\right)\\
&\quad+\beta\sum_{|t-j|\le h}\frac{D_{ft}D_{fj}}{p\left(D_{ft}=1,D_{fj}=1|\boldsymbol{X}_{ftj}\right)}\left(\left(Y_{ft}-\mu_f^P\right)\left(Y_{fj}-\mu_f^P\right)-\sigma_{tj}\left(\boldsymbol{X}_{ftj},1,1\right)\right)\\
&\quad+\beta\sum_{|t-j|>h}\left(\mu_{fj}\left(\boldsymbol{X}_{fj},1\right)-\mu_f^P\right)\frac{D_{ft}}{p\left(D_{ft}=1|\boldsymbol{X}_{ft}\right)}\left(Y_{ft}-\mu_{ft}\left(\boldsymbol{X}_{ft},1\right)\right)\\
&\quad+\beta\sum_{|t-j|>h}\left(\mu_{ft}\left(\boldsymbol{X}_{ft},1\right)-\mu_f^P\right)\frac{D_{fj}}{p\left(D_{fj}=1|\boldsymbol{X}_{fj}\right)}\left(Y_{fj}-\mu_{fj}\left(\boldsymbol{X}_{fj},1\right)\right)\\
&\quad-\sum_{t=1}^{T}\sum_{t=j}^{T}\left(\mu_{ct}\left(\boldsymbol{X}_{ct},1\right)-\mu_c^P\right)\frac{D_{fj}}{p\left(D_{fj}=1|\boldsymbol{X}_{fj}\right)}\left(Y_{fj}-\mu_{fj}\left(\boldsymbol{X}_{fj},1\right)\right)\\
&\quad-\sum_{t=1}^{T}\sum_{t=j}^{T}\left(\mu_{fj}\left(\boldsymbol{X}_{fj},1\right)-\mu_f^P\right)\frac{D_{ct}}{p\left(D_{ct}=1|\boldsymbol{X}_{ct}\right)}\left(Y_{ct}-\mu_{ct}\left(\boldsymbol{X}_{ct},1\right)\right).\end{aligned}$$

$$\psi_2\left(W,\gamma,\alpha,\theta\right)=\sum_{t=1}^{T}\mu_{ct}\left(\boldsymbol{X}_{ct},1\right)-\mu_c^P+\sum_{t=1}^{T}\frac{D_{ct}}{p\left(D_{ct}=1|\boldsymbol{X}_{ct}\right)}\left(Y_{ct}-\mu_{ct}\left(\boldsymbol{X}_{ct},1\right)\right)$$

the first-step estimates.

$$\psi_3\left(W,\gamma,\alpha,\theta\right)=\sum_{t=1}^{T}\mu_{ft}\left(\boldsymbol{X}_{ft},1\right)-\mu_f^P+\sum_{t=1}^{T}\frac{D_{ct}}{p\left(D_{ft}=1|\boldsymbol{X}_{ft}\right)}\left(Y_{ft}-\mu_{ft}\left(\boldsymbol{X}_{ft},1\right)\right).$$

To construct a debiased machine learning estimator for the IGE, we use the orthogonal moment condition specified in equation (9) combined with cross-fitting to ensure robustness and mitigate overfitting. Following Semenova et al. (2023), cross-fitting in settings with dependence should be performed at the level of independent sampling units, in our case, families, rather than individual child–father pairs. Accordingly, let $f\in\{1,\ldots,n_f\}$ index families, with $\mathcal{P}_f$ denoting the set of all child–father pairs in family f. We partition the set of family indices into L mutually exclusive and exhaustive folds $\{\mathcal{F}_\ell\}_{\ell=1}^{L}$. For each fold $\ell=1,\ldots,L$, the nuisance parameters $\hat{\gamma}^{(\ell)}$ and $\hat{\alpha}^{(\ell)}$ are estimated using only data from families not in $\mathcal{F}_\ell$, thereby preserving independence between the samples used for first-stage estimation and those used for evaluation.

The debiased moment function is then computed as

$$\hat{\psi}(\theta)=\frac{1}{n}\sum_{\ell=1}^{L}\sum_{f\in\mathcal{F}_\ell}\sum_{i\in\mathcal{P}_f}\sum_{(t,j)\in\mathcal{J}_i}\hat{\psi}_{i,tj}^{(\ell)},\quad \hat{\psi}_{i,tj}^{(\ell)}:=g(W_{i,tj},\hat{\gamma}^{(\ell)},\theta)+\phi(W_{i,tj},\hat{\gamma}^{(\ell)},\hat{\alpha}^{(\ell)},\theta),$$

where $\mathcal{J}_i$ denotes the set of all tuples (t,j) observed for child–father pair i. Since the system is exactly identified, there is no need to compute fold-specific $\hat{\theta}^{(\ell)}$. The locally robust estimator of the IGE is thus obtained by solving

$$\hat{\theta}_n^{LR}=\arg\min_{\theta\in\Theta\subset\mathbb{R}^3}\hat{\psi}(\theta)'\hat{\Upsilon}\hat{\psi}(\theta).$$

where $\hat{\Upsilon}$ is a positive semi-definite weighting matrix, and Θ denotes the set of parameter values. This objective function incorporates orthogonal moments and cross-fitting. While the influence function corrects for conditional mean prediction errors, cross-fitting eliminates overfitting in nuisance parameter estimation. Furthermore, by grouping folds at the family level, our approach aligns with the principle of leaving out dependent "neighbor" units in panel settings (Semenova et al., 2023), ensuring that dependence within families does not bias the orthogonalization step. We now examine the large-sample behavior of $\hat{\theta}_n^{LR}$, to provide rigorous justification for its empirical implementation.

C. *Asymptotic Theory and Inference*

Asymptotic Properties of the Locally Robust Estimator. — In this section we show consistency and asymptotic normality of our proposed estimator. Consistency follows from standard M-estimation theory, adapted to the locally robust framework of Chernozhukov et al. (2022). While their main asymptotic results assume consistency, Theorem A3 provides primitive conditions under which it holds. The following Lemma adapts these conditions to our setting. Appendix B4 describes and discusses the required additional assumptions to establish consistency.

Lemma 1 (Consistency of the Locally Robust Estimator)**.** *Let $\hat{\theta}_n^{LR}$ be the solution to the cross-fitted orthogonal moment condition:*

$$\hat{\theta}_n^{LR} = \arg\min_{\theta\in\Theta\subset\mathbb{R}^3} \hat{\psi}(\theta)'\hat{\Upsilon}\hat{\psi}(\theta), \quad \hat{\psi}(\theta) = \frac{1}{n}\sum_{\ell=1}^{L}\sum_{f\in\mathcal{F}_\ell}\sum_{i\in\mathcal{P}_f}\sum_{(t,j)\in\mathcal{J}_i}\hat{\psi}_{i,tj}^{(\ell)},$$
$$\hat{\psi}_{i,tj}^{(\ell)} := g(W_{i,tj},\hat{\gamma}^{(\ell)},\theta) + \phi(W_{i,tj},\hat{\gamma}^{(\ell)},\hat{\alpha}^{(\ell)},\theta),,$$

where $\hat{\Upsilon}$ is a positive semi-definite weighting matrix. Then $\hat{\theta}_n^{LR} \xrightarrow{p} \theta_0$, by Theorem A3 in Chernozhukov et al. (2022), provided Assumptions 1-NP, 2-NP and C-NP hold.

Lemma 1 shows that under mild regularity conditions $\hat{\theta}_n^{LR} = \left(\hat{\beta}_n^{LR}, \hat{\mu}_{c,n}^{P}, \hat{\mu}_{F,n}^{P}\right)$ converges in probability to the true parameter θ_0. The consistency of the locally robust estimator guarantees that, under the specified conditions, the estimated intergenerational elasticity $\hat{\beta}_n^{LR}$ converges to the true value β_0 as the sample size increases. This ensures that the estimator remains stable even when machine learning methods are used to estimate nuisance parameters. As a result, the estimates of the intergenerational elasticity are both reliable and comparable across different studies.

Under the regularity conditions described in Appendix B4, we establish the asymptotic normality of our proposed estimator, that explicitly accounts for uncertainty from the first-stage estimation of the nuisance parameters. This yields confidence intervals with valid coverage, a crucial requirement for drawing meaningful conclusions about intergenerational mobility patterns.

The following Lemma formalizes the validity of inference for the estimator $\hat{\theta}_n^{LR}$, even when nuisance components are estimated using high-dimensional or nonparametric methods. This robustness is achieved through the use of orthogonal moment conditions, which ensure that estimation errors in the first stage enter the moment function only at second order. As a result, standard $\sqrt{n}$ asymptotic normality can be established under relatively weak conditions. Crucially, cross-fitting plays a central role in mitigating own-observation bias and avoids the need for stringent entropy or Donsker-type conditions,[9] which are not known to hold for many machine learning first steps. Together, these features allow us to leverage flexible first-stage methods while maintaining valid inference.

Lemma 2 (Asymptotic Normality of the Locally Robust Estimator)**.** *Under Assumptions 1-NP-5-LR and 7-LR, $\hat{\theta}_n^{LR} \xrightarrow{p} \theta_0$, and non-singularity of $G'\Upsilon G$, the asymptotic normality*

[9]Donsker conditions require function classes to be sufficiently "simple" (typically via entropy bounds or VC dimension constraints) to ensure uniform convergence of empirical processes to Gaussian limits. Flexible machine learning methods (e.g., random forests, neural networks) might violate these conditions, as their complexity increases with sample size, potentially leading to non-Gaussian limits or invalid inference. Cross-fitting addresses this by decoupling the estimation of nuisance functions and moment evaluation, using independent samples to prevent overfitting-induced correlations. It also relaxes uniformity requirements, ensuring convergence of the empirical process only at the estimated nuisance functions rather than over the entire function class.

of the estimator $\hat{\theta}_n^{LR}$ directly follows from Theorem 9 of Chernozhukov et al. (2022). Specifically, we have:

$$\sqrt{n}(\hat{\theta}_n^{LR} - \theta_0) \xrightarrow{d} \mathcal{N}(0, V),$$

where$V = (G'\Upsilon G)^{-1}$, $G = \mathbb{E}[\partial_\theta g(W, \gamma, \alpha, \theta)]$, and $\hat{\Upsilon}$ is the estimated efficient weighting matrix defined as $\hat{\Upsilon} = \hat{\Psi}^{-1}$ for $\hat{\Psi} = \frac{1}{n}\sum_{\ell=1}^{L}\sum_{i\in\mathcal{I}_\ell}\sum_{(tj)\in\mathcal{J}_i}\hat{\psi}_{i,tj}^{(\ell)}\hat{\psi}_{i,tj}^{(\ell)\prime}$. In addition, if Assumption 6-LR holds, then $\hat{V} \xrightarrow{p} V$.

Lemma 2 completes the theoretical framework by integrating our three contributions: (i) the nonparametric identification of the intergenerational elasticity in the presence of incomplete income data; (ii) a consistent, locally robust estimator that corrects for first-step prediction errors; and (iii) valid inference that accounts for uncertainty from the first-stage estimation of nuisance parameters. Appendix B4 characterizes the asymptotic variance V associated with this result. Together, these advances provide a theoretically grounded toolkit for studying income persistence through the lens of the intergenerational elasticity.

Our framework complements rather than replaces rank-based approaches, as rank-rank correlation and intergenerational elasticity capture fundamentally different aspects of mobility. The Rank-Rank correlation tracks positional changes in the income distribution by assessing how children from disadvantaged backgrounds fare compared to their peers within a given income distribution. In contrast, the IGE measures the persistence of absolute economic advantages by directly comparing children's economic status to that of their parents.

This conceptual distinction highlights that the choice of mobility measure should be guided by the specific research question, rather than by methodological concerns. For instance, fixing ranks to the national distribution allows for more meaningful comparisons across cities (Mazumder, 2016). However, this standardization inherently masks variations in income inequality across time or place (Mello et al., 2024), making the IGE more suitable for analyses such as cross-country comparisons or examining trends in mobility.

In practice, methodological considerations have outweighed conceptual considerations in selecting the mobility measure. Following the influential work of Chetty et al. (2014), there has been a noticeable shift in the literature toward rank-based measures (Mazumder, 2016). This trend is driven by legitimate concerns regarding the nonlinear relationship between log child income and log parent income, as well as the high sensitivity of IGE estimates to the treatment of zero and very small income observations. Recent theoretical advances by Chetverikov and Wilhelm (2023) further strengthen the rank-based approach by establishing valid inference for rank-rank regressions when lifetime income is observed. However, extending this framework for the setting of incomplete income data remains an open question.

Our study bridges this gap by establishing valid inference for contexts where the IGE is more appropriate. It addresses one of the key limitations that have led the literature to favor rank-based measures. Specifically, our estimator incorporates propensity scores and imputes lifecycle income profiles. The first feature directly addresses missing data, while the second reduces the influence of individual income observations on the estimation of lifetime income. Together, these elements mitigate the high sensitivity to the treatment of small or zero income values.

While asymptotic guarantees provide essential theoretical foundations, empirical researchers ultimately depend on methods that perform reliably in realistic sample sizes. Having established the asymptotic properties and theoretical advantages of our estimator, we now turn to evaluating its finite-sample performance, which is the critical determinant of its practical applicability.

D. Tests for Identification Assumptions

This section develops formal hypothesis tests for Assumptions 1-NP.*ii* and 1-NP.*iii*. In contrast, no formal tests are provided for the remaining assumptions. Specifically, as illustrated In Section V the specification of the characteristics predictive of income profiles and parental income covariance, namely, $\boldsymbol{X}_{ct}$, $\boldsymbol{X}_{ft}$, and $\boldsymbol{X}_{ftj}$, can be designed to satisfy Assumption 1-NP.*i* by construction. The MAR conditions in Assumptions 2-NP.*i* and 2-NP.*iii* are not directly testable from the observed data, as they involve unobserved missingness mechanisms. Nevertheless, as discussed above, the literature suggests that the MAR assumption for income missingness in the PSID is empirically plausible, provided that analyses carefully account for relevant observables. Consistent with the findings in Schoeni and Wiemers (2015), we include both child and father characteristics in the conditioning set, thereby strengthening the plausibility of the MAR assumption in our analysis. Finally, the boundedness condition on the propensity score in Assumptions 2-NP.*ii* and 2-NP.*iv* can be assessed informally through visual inspection.

We start by considering a test for the orthogonality between children's

(10)
$$H_0 : \mathbb{E}\left[\epsilon_{ct} Y_f^P\right] = 0, \quad \forall t = 1, \ldots, T \quad \text{vs} \quad H_1 : \mathbb{E}\left[\epsilon_{ct} Y_f^P\right] \neq 0 \text{ for some } t \in \{1, \ldots, T\},$$

where $\epsilon_{ct} := Y_{ct} - \mathbb{E}\left[Y_{ct} \mid \boldsymbol{X}_{ct}\right]$ denotes the children's income prediction errors at time t and Y_f^P represents parental lifetime income. The main challenge in testing this hypothesis is that both random variables are unobserved, and their machine learning estimation introduces regularization and model selection bias when testing H_0. To address these issues, we propose a three stages procedure. First, we establish identification of the object of interest $\theta_{cft} := \mathbb{E}\left[\epsilon_{ct} Y_f^P\right]$. Second, we construct a locally robust estimator $\hat{\theta}_t$. Finally, we provide a Wald test based on $\hat{\boldsymbol{\theta}} = (\theta_1, \ldots, \theta_{cft})$.

In Appendix B5 we show that a locally robust Wald test for H_0 in (10) is given by

$$W_{cf,n} = n \hat{\theta}_{cf,n} \hat{V}_{cf,n}^{-1} \hat{\theta}_{cf,n},$$

where $\hat{\theta}_{cf,n}$ is the argument solving the cross-fitted locally robust moment

$$\psi_{cf}\left(W, \gamma_{cf}, \theta_{cf}\right) = \left(\psi_{cf1}\left(W, \gamma_{cf1}, \theta_{cf1}\right), \cdots, \psi_{cfT}\left(W, \gamma_{cfT}, \theta_{cfT}\right)\right),$$

$$
\begin{aligned}
\psi_{cft}\left(W, \gamma_{cft}, \theta_{cft}\right) = & \sum_{j=1}^{T} \mu_{cftj}\left(\boldsymbol{X}_{ct}, \boldsymbol{X}_{fj}, 1, 1\right) - \mu_{ct}\left(\boldsymbol{X}_{ct}, 1\right) \mu_{fj}\left(\boldsymbol{X}_{fj}, 1\right) - \theta_{cft} \\
& - \sum_{j=1}^{T} \mu_{fj}\left(\boldsymbol{X}_{fj}, 1\right) \frac{D_{ct}}{p\left(D_{ct} = 1 | \boldsymbol{X}_{ct}\right)} \left(Y_{ct} - \mu_{ct}\left(\boldsymbol{X}_{ct}, 1\right)\right) \\
& - \mu_{ct}\left(\boldsymbol{X}_{ct}, 1\right) \sum_{j=1}^{T} \frac{D_{fj}}{p\left(D_{fj} = 1 | \boldsymbol{X}_{fj}\right)} \left(Y_{fj} - \mu_{fj}\left(\boldsymbol{X}_{fj}, 1\right)\right) \\
& + \sum_{j=1}^{T} \frac{D_{ct} D_{fj}}{p\left(D_{ct} = 1, D_{fj} = 1 | \boldsymbol{X}_{ct}, \boldsymbol{X}_{fj}\right)} \left(Y_{ct} Y_{fj} - \mu_{cftj}\left(\boldsymbol{X}_{ct}, \boldsymbol{X}_{fj}, 1, 1\right)\right),
\end{aligned}
$$

and $\hat{V}_{cf,n}$ is a consistent estimator of the asymptotic variance of $\hat{\theta}_{cf,n}$.

The following Theorem establishes the asymptotic properties of this locally robust Wald test.

Theorem 3. *(Size, Consistency, and Local Power of the Locally Robust Wald Test I) Under Assumptions 1-NP', 2-NP', 3-LR-7-LR and C-NP, the asymptotic properties of the locally robust Wald statistic*

$$
W_{cf,n} \;=\; n\, \hat{\theta}_{cf,n}\, \hat{V}_{cf,n}^{-1}\, \hat{\theta}_{cf,n},
$$

are given by the following statements:

1) *(*Asymptotic size*) Under $H_0 : \theta_{cf0} = \mathbf{0}$,*

$$
W_{cf,n} \Rightarrow d\chi^2_T \quad \textit{and} \quad \lim_{n\to\infty} \Pr\left(W_{cf,n} > \chi^2_{T,1-\alpha}\right) = \alpha.
$$

2) *(*Consistency under fixed alternatives*) For any fixed alternative with $\theta_{cf0} \neq \mathbf{0}$,*

$$
\lim_{n\to\infty} \Pr(W_{cf,n} > \chi^2_{T,1-\alpha}) = 1.
$$

3) *(*Local alternatives*) Under $H_{1n} : \theta_{cf0} = \delta / \sqrt{n}$ with fixed $\delta \in \mathbb{R}^T$,*

$$
W_{cf,n} \Rightarrow d\chi^2_T(\lambda), \qquad \lambda = \delta' V_{cf}^{-1} \delta,
$$

so the limiting power is

$$
\lim_{n\to\infty} \Pr(W_{cf,n} > \chi^2_{T,1-\alpha}) \;=\; 1 - F_{\chi^2_T(\lambda)}(\chi^2_{T,1-\alpha}) \;>\; \alpha \quad \textit{whenever } \delta \neq \mathbf{0}.
$$

To test that the parental income prediction errors are uncorrelated for $|t - j| > h$, we propose the test

(11)
$H_0 : \mathbb{E}[\epsilon_{ft}\,\epsilon_{fj}] = 0 \quad \text{for all } |t-j| = h+1, \quad \text{vs} \quad H_1 : \mathbb{E}[\epsilon_{ft}\,\epsilon_{fj}] \neq 0 \quad \text{for some } |t-j| = h+1.$

In Appendix B5 we show that a locally robust Wald test for H_0 in (11) is given by

$$W_{fh,n} = n\hat{\theta}_{fh,n}\hat{V}_{fh,n}^{-1}\hat{\theta}_{fh,n}$$

where $\hat{\theta}_{fh,n}$ is the argument solving the cross-fitted orthogonal moment

$$\begin{aligned}
\psi_{cf}\left(W,\gamma_{fh},\theta_{fh}\right) &= \left(\psi_{fh1}\left(W,\gamma_{fh1},\theta_{fh1}\right),\cdots,\psi_{fhT}\left(W,\gamma_{fhT},\theta_{fhT}\right)\right),\\
\psi_{ftj}\left(W,\gamma_{ftj},\theta_{ftj}\right) &= \mu_{ftj}\left(\boldsymbol{X}_{ftj},1,1\right) - \mu_{ft}\left(\boldsymbol{X}_{fj},1\right)\mu_{fj}\left(\boldsymbol{X}_{fj},1\right) - \theta_{ftj}\\
&\quad + \frac{D_{ft}D_{fj}}{p\left(D_{ft}=1,D_{fj}=1|\boldsymbol{X}_{ftj}\right)}\left(Y_{ft}Y_{fj} - \mu_{ftj}\left(\boldsymbol{X}_{ftj},1,1\right)\right)\\
&\quad - \mu_{fj}\left(\boldsymbol{X}_{fj},1\right)\frac{D_{ft}}{p\left(D_{ft}=1|\boldsymbol{X}_{ft}\right)}\left(Y_{ft} - \mu_{ft}\left(\boldsymbol{X}_{ft},1\right)\right)\\
&\quad - \mu_{ft}\left(\boldsymbol{X}_{ft},1\right)\frac{D_{fj}}{p\left(D_{fj}=1|\boldsymbol{X}_{fj}\right)}\left(Y_{fj} - \mu_{fj}\left(\boldsymbol{X}_{fj},1\right)\right).
\end{aligned}$$

The following Corollary establishes the asymptotic properties of this locally robust Wald test.

Corollary 3.1. *(Size, Consistency, and Local Power of the Locally Robust Wald Test II) Under Assumptions 1-NP", 2-NP", 3-LR-7-LR, C-NP, and 8-NP", the asymptotic properties of the locally robust Wald statistic*

$$W_{cf,n} \;=\; W_{fh,n} = n\hat{\theta}_{fh,n}\hat{V}_{fh,n}^{-1}\hat{\theta}_{fh,n}$$

are given by the following statements:

1) (Asymptotic size) *Under $H_0 : \theta_{fh0} = \mathbf{0}$,*

$$W_{fh,n} \Rightarrow d\chi^2_{K_\theta} \quad \textit{and} \quad \lim_{n\to\infty}\Pr\left(W_{fh,n} > \chi^2_{K_\theta,1-\alpha}\right) = \alpha,$$

where K_θ is the dimension of θ_{fh}

2) (Consistency under fixed alternatives) *For any fixed alternative with $\theta_{fh0} \neq \mathbf{0}$,*

$$\lim_{n\to\infty}\Pr(W_{fh,n} > \chi^2_{K_\theta,1-\alpha}) = 1.$$

*3) (*Local alternatives*) Under* $H_{1n} : \theta_{fh0} = \delta/\sqrt{n}$ *with fixed* $\delta \in \mathbb{R}^{K_\theta}$,

$$W_{fh,n} \Rightarrow d\chi^2_{K_\theta}(\lambda), \qquad \lambda = \delta' V_{fh}^{-1}\delta,$$

so the limiting power is

$$\lim_{n\to\infty} \Pr(W_{fh,n} > \chi^2_{K_\theta,1-\alpha}) \;=\; 1 - F_{\chi^2_{K_\theta}(\lambda)}(\chi^2_{K_\theta,1-\alpha}) \;>\; \alpha \quad \textit{whenever } \delta \neq \mathbf{0}.$$

IV. Simulations

This section evaluates the finite-sample properties of our proposal via Monte Carlo simulations. The simulation designs capture key characteristics of observed intergenerational income data. We assess its performance in terms of bias and coverage, comparing it to the plug-in machine learning approach, mid-life income estimator, and lifecycle estimator. The results show that our locally robust estimator exhibits negligible bias, which converges to zero as the sample size increases, and coverage rates closely aligned with nominal values, outperforming the benchmarks across different empirically relevant scenarios.

We consider the following D.G.P for generation g at age t:

$$Y_{gt} = \gamma_{0,g} + \gamma_{1,g}X_{1,g} + \gamma_{2,g}X_{2,g} + \gamma_{3,g}t + \gamma_{4,g}t^2 + \gamma_{5,g}X_{1,g}t + \epsilon_{gt}, \quad t = 1,...,T, \quad g \in \{c,f\},$$
$$\epsilon_{gt} \sim \mathcal{N}\left(0, \sigma^2_\epsilon\right)$$
$$\begin{pmatrix} X_{j,c} \\ X_{j,f} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \sigma_j \\ \sigma_j & 1 \end{pmatrix}\right), \quad j = 1,2.$$

Using the definition of lifetime income:

$$Y^P_g = \gamma_{0,g} + \gamma_{1,g}X_{1,g} + \gamma_{2,g}X_{2,g} + \gamma_{3,g}\bar{t} + \gamma_{4,g}\bar{t^2} + \gamma_{5,g}X_{1,g}\bar{t} + \bar{\epsilon}_g, \qquad g \in \{c,f\},$$
$$\bar{t} = \sum_{t=1}^{T} t, \quad \bar{t^2} = \sum_{t=1}^{T} t^2, \quad \bar{\epsilon}_g = \sum_{t=1}^{T} \epsilon_{gt}, \tag{12}$$

so the covariance between lifetime incomes is given by

(13)
$$\begin{aligned}\mathrm{Cov}\left(Y^P_c, Y^P_f\right) &= \mathrm{Cov}\left(\gamma_{1,c}X_{1,c}, \gamma_{1,f}X_{1,f}\right) + \mathrm{Cov}\left(\gamma_{2,c}X_{2,c}, \gamma_{2,f}X_{2,f}\right) + \mathrm{Cov}\left(\gamma_{5,c}X_{1,c}\bar{t}, \gamma_{5,f}X_{1,f}\bar{t}\right) \\ &\quad + \mathrm{Cov}\left(\gamma_{1,c}X_{1,c}, \gamma_{5,f}X_{1,f}\bar{t}\right) + \mathrm{Cov}\left(\gamma_{5,c}X_{1,c}\bar{t}, \gamma_{1,f}X_{1,f}\right) \\ &= \gamma_{1,c}\gamma_{1,f}\sigma_1 + \gamma_{2,c}\gamma_{2,f}\sigma_2 + \gamma_{5,c}\gamma_{5,f}\bar{t^2}\sigma_1 + \gamma_{1,c}\gamma_{5,f}\bar{t}\sigma_1 + \gamma_{5,c}\gamma_{1,f}\bar{t}\sigma_1,\end{aligned}$$

where we have used that the covariates come from a bivariate normal distribution with zero mean and correlation σ_j among generations.

According to equation (12), the variance of parental income corresponds to

$$\mathrm{Var}(Y_f^P) = \gamma_{1,f}^2 + \gamma_{2,f}^2 + \gamma_{5,f}^2 \bar{t^2} + 2\gamma_{1,f}\gamma_{5,f}\bar{t} + \sigma_\epsilon^2. \tag{14}$$

Finally, by plugging equations (13) and (14) into (2) yields

$$\begin{aligned}\beta_0 &= \frac{\mathbb{E}\left[\left(Y_c^P - \mathbb{E}\left[Y_c^P\right]\right)\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)\right]}{\mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right]} \\ &= \frac{\sigma_1\left(\gamma_{1,c}\gamma_{1,f} + \gamma_{5,c}\gamma_{5,f}\bar{t^2} + \left(\gamma_{1,c}\gamma_{5,f} + \gamma_{5,c}\gamma_{1,f}\right)\bar{t}\right) + \gamma_{2,c}\gamma_{2,f}\sigma_2}{\gamma_{1,f}^2 + \gamma_{2,f}^2 + \gamma_{5,f}^2\bar{t^2} + 2\gamma_{1,f}\gamma_{5,f}\bar{t} + \sigma_\epsilon^2}. \end{aligned} \tag{15}$$

Setting the parameter values to

$$\begin{array}{llll} \gamma_{0,c} = 8.5, & \gamma_{0,f} = 5, & \gamma_{1,c} = 0.275, & \gamma_{1,f} = 0.4, \\ \gamma_{2,c} = 0.2, & \gamma_{2,f} = 0.25, & \gamma_{3,c} = 0.4, & \gamma_{3,f} = 0.5, \\ \gamma_{4,c} = -0.005, & \gamma_{4,f} = -0.0045, & \gamma_{5,c} = 0.01, & \gamma_{5,f} = 0.015, \\ \sigma_1 = 0.75, & \sigma_2 = 0.75, & \sigma_\epsilon = 1, & t = 20, \ldots, 60, \end{array}$$

yields $T = 41$, and $\beta_0 = 0.50$ according to equation (15).

We consider the setting above for sample sizes $n = 100$, 500, 1000, and 2000, where we randomly draw 20%, 35%, and 50% of each sample by drawing income snapshots from parents and children as follows. First, for each individual i, we draw a contiguous observation period length from a right-censored Poisson distribution:

$$\ell_i = \min\left(\max(OW_i^P, 2), 41\right), \quad OW_i^P \sim \mathrm{Pois}(\lambda),$$

where $\lambda \approx 10$, 20, andor 31 years for 20%, 35%, and 50% coverage respectively. Then, the observation window begins at a random age:

$$a_i \sim \mathcal{U}(20,\ 60 - \ell_i + 1)$$

ensuring complete coverage within the 20-60 age range. Thus, only incomes satisfying $t \in [a_i, a_i + \ell_i)$ are observed, with other years being missing (completely at random), mimicking common data limitations in mobility studies. This creates contiguous observation blocks that mimic real-world data limitations where income histories armay only be observed during certain life periods, mimicking realistic administrative or survey-based data constraints. The sampling is performed separately for children and parents.

We assess the performance of our Locally Robust (LR) estimator by examining its bias and coverage properties relative to three alternative approaches: (1) the plug-in machine learning estimator, (2) the mid-life income estimator, and (3) the life-cycle estimator. We estimate income profiles for both generations using XGBoost Regression, which also allows us to compute the conditional covariance of parental income. Propensity scores are estimated via logistic regression. The core difference between the locally robust (LR) and plug-in machine learning estimators lies in their moment conditions: the LR estimator uses a Neyman-orthogonal moment that incorporates the influence function of the first-stage estimates, while the plug-in estimator relies on the uncorrected identifying moment. For the mid-life income estimator, fathers' lifetime income is proxied by averaging earnings from ages 30 to 40, and children's income is based on a single mid-life earnings draw. In contrast, the life-cycle estimator uses the same paternal income proxy but estimates children's lifetime income as the sum of predicted earnings over the life cycle from XGBoost Regression. Estimation proceeds in two steps: hyperparameter tuning using 5-fold cross-validation, followed by cross-fitting to prevent overfitting. In all simulations, we use 500 Monte Carlo replications.

Table 1 presents the finite-sample performance of four estimators for the intergenerational elasticity, evaluated through bias and coverage rates across 500 Monte Carlo replications. The true IGE is 0.5, with a nominal coverage rate of 0.95. The analysis spans three sample sizes ($n = 100, 500, 1000, 2000$) and three observation probabilities ($\kappa = 0.20, 0.35, 0.50$).

Table 1—: Bias and coverage of different estimators for the IGE for different sample sizes and observation probability.

	Locally Robust		Plug-in Machine Learning		Life-cycle		Mid-life	
n	Bias	Coverage	Bias	Coverage	Bias	Coverage	Bias	Coverage
$\kappa = 0.20$								
100	-0.00506	0.91	-0.07358	0.56	-0.19255	0.29	-0.19890	0.86
500	-0.00347	0.95	-0.03991	0.43	-0.17193	0.00	-0.18762	0.43
1000	-0.00129	0.92	-0.02981	0.41	-0.17010	0.00	-0.18462	0.17
2000	-0.00132	0.93	-0.03414	0.11	-0.17311	0.00	-0.18699	0.01
$\kappa = 0.35$								
100	-0.00685	0.91	-0.05702	0.66	-0.17529	0.17	-0.18318	0.75
500	-0.00314	0.94	-0.02800	0.66	-0.15818	0.00	-0.17540	0.17
1000	-0.00230	0.93	-0.03077	0.35	-0.15885	0.00	-0.17099	0.01
2000	0.00109	0.92	-0.01376	0.67	-0.15142	0.00	-0.16894	0.00
$\kappa = 0.50$								
100	-0.00317	0.91	-0.04778	0.72	-0.16108	0.12	-0.15682	0.73
500	0.00561	0.93	-0.01547	0.84	-0.14557	0.00	-0.15302	0.12
1000	0.00460	0.93	-0.01928	0.67	-0.14741	0.00	-0.15481	0.01
2000	0.00306	0.92	-0.00949	0.82	-0.13861	0.00	-0.15465	0.00

Results based on 500 Monte Carlo replications with true IGE equal to 0.5 and the nominal coverage is 0.95.

The locally robust estimator demonstrates superior performance, with bias decreasing as sample size increases (e.g., from -0.0051 at $n = 100$ to -0.0015 at $n = 2000$ for $\kappa = 0.20$). This aligns with expected $\sqrt{n}$-consistency, reflecting its robustness to sample size variations. Coverage rates remain close to the nominal 0.95, ranging from 0.91 to 0.95 across all scenarios, with minor undercoverage at smaller sample sizes ($n = 100$). Notably, both bias and coverage are largely insensitive to changes in κ, indicating stability across varying observation probabilities.

In contrast, the plug-in machine learning estimator exhibits substantially higher bias in absolute terms (e.g., -0.0736 at $n = 100$ vs. -0.0095 at $n = 2000$ for $\kappa = 0.50$). Its coverage rates are consistently below the nominal 0.95, improving from 0.56 to 0.82 as sample size increases for $\kappa = 0.50$, but remaining inadequate. This poor performance underscores the limitations of the plug-in approach, particularly in smaller samples or lower observation probabilities, justifying the preference for the locally robust estimator.

The life-cycle and mid-life estimators show significant and persistent bias across all sample sizes and κ values (e.g., life-cycle bias ranges from -0.19 to -0.14, mid-life from -0.199 to -0.154. Their coverage rates are notably poor, deteriorating to 0 for larger samples ($n = 2000$) due to miss-centered confidence intervals. As sample size increases, reduced sampling variability narrows these intervals, but uncorrected bias causes them to systematically miss the true IGE.

Caution is warranted when interpreting the results for the LC versus MI estimator. The comparable performance of the life-cycle and mid-life estimators arises from the design of the data-generating process (DGP), which may not fully capture real-world income dynamics. The life-cycle estimator, designed to account for empirical income process patterns, may exhibit understated bias reduction in this simulation due to the DGP's simplified structure. In practical applications, where income processes are more complex, the life-cycle estimator would potentially outperform the mid-life estimator.

Overall, the locally robust estimator emerges as the most reliable, offering low bias and near-nominal coverage across all settings. The plug-in machine learning estimator, while improving with larger samples, remains inferior due to higher bias and poor coverage. The life-cycle and mid-life estimators are consistently outperformed, highlighting the importance of consistent and locally robust estimation in analyzing the IGE.

V. Consistent Estimation of the Intergenerational Elasticity in the United States

In this section, we implement our locally robust estimator to measure the intergenerational elasticity (IGE) of income in the United States. Our analysis employs the Panel Study of Income Dynamics (PSID), the world's longest-running longitudinal household survey. Launched in 1968 with a nationally representative sample of 5,000 U.S. families (over 18,000 individuals), the PSID has continuously tracked these families and their descendants, collecting rich data on income, wealth, employment, education, health, and other socioeconomic outcomes.

Our sample consists of 1138 child–father pairs, for children born between 1951 and 1960. Following Lee and Solon (2009), we use the PSID core sample, corresponding to the Survey Research Center component, and defined the income measure as family income, which allows us to include both male and female children. We exclude individuals with only zero or missing income values. All dollar values are adjusted to 1968 dollars using the CPI. To address negative and zero-income cases, we bottom-code family income at the first percentile, which affects 0.12% of the observations in the raw PSID data. For our analysis, we consider the lifetime span from ages 20 to 60.

The family characteristics in our analysis are drawn from the rich data provided by the PSID and are organized into several domains. Education is measured by years of schooling completed and whether the household head received additional training beyond standard school or college. Regional location follows the PSID's classification into Northeast, North Central, South, or West. Family structure includes the birth order of the children, the father's age at first birth, and we leverage the PSID's intergenerational mapping to incorporate the number of offspring per father. Assets are captured through indicators of housing and business ownership. Demographics include race (classified as White or Non-White),[10] sex of the children (given our focus on fathers), religion, and age at the time of interview. While the PSID offers a broader set of variables, we focus on these selected characteristics to ensure consistency and availability across survey waves.

Based on this available data, we construct the characteristics predictive of income profiles and parental income covariance, namely, $\boldsymbol{X}_{ct}$, $\boldsymbol{X}_{ft}$, and $\boldsymbol{X}_{ftj}$. This specification must account for three key requirements: handling missing data in observables, incorporating the dynamics of the income process, and satisfying Assumptions 1-NP and 2-NP. To address the first, we summarize variables over the lifetime span (ages 20–60) using averages for time-varying characteristics (excluding education), modes for religion and region, and the maximum value for education.

To accurately model yearly income as a function of covariates, it is essential to incorporate the dynamics of the income process and capture relevant empirical patterns. To this end, we closely follow the covariate specification in Mello et al. (2024). Specifically, we include a quartic polynomial in age to account for concavities and nonlinearities in income profiles. For the child generation, we incorporate a noisy proxy for parental permanent income, defined as the three-year average of log family income when the child was aged 15–17.[11] Additionally, we include interactions of this noisy proxy and parental education with a quadratic polynomial in age to capture the greater variability in income growth at younger ages and the typically faster income growth observed among children from high-income families.

Our covariate specification is designed to satisfy Assumption 1-NP.*i* by construction, while also making the remaining assumptions plausible in practice, although not guaranteed to hold. To satisfy Assumption 1-NP.*i*, we merge $\boldsymbol{X}_{ft}$ and $\boldsymbol{X}_{cj}$ such that their non-overlapping components $\left(\boldsymbol{X}_{ft} \cap \boldsymbol{X}_{cj}\right)^c = \{age_{ft}, age_{cj}\}$, are both deterministic. In doing so, we also include in $\boldsymbol{X}_{ft}$ the noisy measure of parental permanent income along with its interaction with age. This time-invariant measure serves as a relevant predictor in the presence of missing income data, helping to compensate for the absence of income leads and lags. Moreover, it enhances the first-step estimation of income profiles (and parental income covariance), which is fundamentally a prediction task. Finally, to construct $\boldsymbol{X}_{ftj}$, we merge $\boldsymbol{X}_{ft}$ and $\boldsymbol{X}_{fj}$, ensuring $\left(\boldsymbol{X}_{ft} \cap \boldsymbol{X}_{ftj}\right)^c = \{age_{fj}\}$. The current specification of the covariates ensures that Assumption 1-NP.*i* is satisfied by construction; that is, the

[10] To ensure consistency across survey waves, we classify respondents only as "White" or "Non-White." The earliest waves included White, Black, Puerto Rican or Mexican, and Other, while later waves incorporated additional categories such as American Indian or Alaska Native, Asian, and Native Hawaiian or Pacific Islander.

[11] If family income data during this period is insufficient, we use the closest available three-year window to ages 15–17.

covariates used to predict children's income satisfy:

$$\mathbb{E}\left[Y_{ct} \mid \boldsymbol{X}_{ct}, \boldsymbol{X}_{cj}, \boldsymbol{X}_{fj}\right] = \mathbb{E}\left[Y_{ct} \mid \boldsymbol{X}_{ct}\right] \quad \text{for } t, j = 1, \ldots, T,$$

and those used to predict fathers' income satisfy:

$$\mathbb{E}\left[Y_{ft} \mid \boldsymbol{X}_{ft}, \boldsymbol{X}_{ftj}, \boldsymbol{X}_{cj}\right] = \mathbb{E}\left[Y_{ft} \mid \boldsymbol{X}_{ft}\right] \quad \text{for } \boldsymbol{X}_{fj} \subset \boldsymbol{X}_{ftj}, \quad t, j = 1, ...T.$$

This follows from the complements of the intersections, $\left(\boldsymbol{X}_{ft} \cap \boldsymbol{X}_{ft}\right)^c = \{age_{ft}, age_{cj}\}$ and $\left(\boldsymbol{X}_{ft} \cap \boldsymbol{X}_{ftj}\right)^c = \{age_{fj}\}$, consisting solely of age, which is deterministic and thus do not add stochastic variation beyond what is captured by $\boldsymbol{X}_{ct}$ and $\boldsymbol{X}_{ft}$.

Our covariate specification further enhances the plausibility of the remaining assumptions in practice. The inclusion of an interaction between parental permanent income and age, for example, strengthens the orthogonality condition between children's income prediction errors and parental lifetime income required by Assumption 1-NP.*ii*. In Section III.D, we develop formal tests to empirically evaluate Assumptions 1-NP.*ii* and 1-NP.*iii*.

The dimensionality of the covariate set reflects the trade-off between the plausibility of the missing-at-random assumption and the boundedness of the propensity scores in Assumption 2-NP. While increasing the dimension of $\boldsymbol{X}_{gt}$, can make the conditional independence $Y_{ct} \perp D_{ct} | \boldsymbol{X}_{ct}$ more plausible, it may reduce the likelihood that the propensity score remains bounded away from zero. Nonetheless, it is not merely the dimensionality, but rather the informativeness of the covariates that determines whether missingness is conditionally at random. In other words, we aim to control for the relevant features such that, conditional on them, the missingness of annual income occurs conditionally at random. While the MAR conditions in Assumptions 2-NP.*i* and 2-NP.*iii* cannot be directly tested from observed data, the boundedness conditions in Assumptions 2-NP.*ii* and 2-NP.*iv* can be assessed informally through visual inspection.

Table 2 provides summary statistics for socioeconomic characteristics of children and their fathers revealing important intergenerational patterns in socioeconomic characteristics. Fathers exhibit higher mean logged annual income (9.39 vs. 9.18) and home ownership rates (88% vs. 67%), while children show greater educational attainment (mean 13.7 vs. 12.2 years) and additional training participation (19% vs. 12%). Both generations share identical rates of business ownership (16% for father) and white composition (91%), though fathers report higher religious affiliation (92% vs. 83%). Half of the children are female, reflecting a balanced gender distribution. The median birth order indicates that most families in the data have two or more children, with relatively few only children. The mean and median values for the proxy of lifetime income closely match those of annual income, but with less variation and a narrower range, suggesting that especially low incomes tend to occur outside of midlife. Most fathers had their first child around age 25. Regional distributions show similar patterns across generations, with children slightly more concentrated in the South (30% vs. 27%) and fathers slightly more in the Northeast (23% vs. 22%).

The income measures exhibit tighter dispersion for fathers, with smaller standard de-

Table 2—: Summary Statistics for Children and Fathers

	Children					Fathers				
Variable	Mean	Median	SD	Min	Max	Mean	Median	SD	Min	Max
Annual Income	9.18	9.27	0.90	-1.42	13.8	9.39	9.46	0.79	-1.42	12.6
Proxy of Lifetime Income	-	-	-	-	-	9.38	9.42	0.53	7.30	11.3
Education Level	13.70	14.00	2.61	0.00	17.0	12.20	12.00	3.20	0.00	17.0
House Ownership	0.67	0.76	0.30	0.00	1.00	0.88	1.00	0.27	0.00	1.00
Business Ownership	0.16	0.05	0.23	0.00	1.00	0.16	0.00	0.28	0.00	1.00
Additional Training	0.19	0.00	0.39	0.00	1.00	0.12	0.00	0.33	0.00	1.00
Religion	0.83	1.00	0.38	0.00	1.00	0.92	1.00	0.28	0.00	1.00
White	0.91	1.00	0.28	0.00	1.00	0.91	1.00	0.28	0.00	1.00
Sex	0.50	1.00	0.50	0.00	1.00	-	-	-	-	-
Birth Order	2.19	2.00	1.26	1.00	8.00	-	-	-	-	-
Northeast Region	0.22	0.00	0.42	0.00	1.00	0.23	0.00	0.42	0.00	1.00
South Region	0.30	0.00	0.46	0.00	1.00	0.27	0.00	0.45	0.00	1.00
West Region	0.19	0.00	0.39	0.00	1.00	0.17	0.00	0.38	0.00	1.00
Age at First Child	-	-	-	-	-	26.30	25.00	5.08	16.00	44.0

viations (0.79 vs. 0.90 for annual income). Educational attainment shows greater variability among fathers (SD 3.20 vs. 2.61), potentially reflecting cohort differences in educational access. Fathers tend to have higher ownership rates, with a median home ownership of 1 compared to 76% for children. The regional distributions are remarkably consistent across generations, with all regional variables showing similar summary statistics. After outlining these descriptive results, we shift to the core investigation of intergenerational income persistence.

Table 3 presents the estimation results of intergenerational elasticity in the United States from the PSID core sample using alternative estimators. The estimation procedure employs XGBoost Regression to model income profiles for both generations and estimate conditional covariances of parental income, while logistic regression estimates the propensity scores. The key distinction between the LR and plug-in machine learning estimators lies in their moment conditions: the LR approach utilizes the orthogonal moment that accounts for the first-step influence function, whereas the plug-in version relies solely on the identifying moment without such correction. For the mid-life income estimator, fathers' lifetime income is proxied using a three-year average of log family income when the child was aged 15–17.[12] , while children's income is represented by a single random draw from their midlife earnings (years 25–33, following Solon (1992)). The life-cycle estimator maintains the same paternal income measure but measures children's lifetime income by summing predicted earnings profiles from OLS Regression. Our estimation procedure involves two key steps: hyperparameter tuning via 5-fold cross-validation followed by cross-fitting (using 10 folds) to mitigate overfitting.

We find that the locally robust estimator yields an IGE of 0.69 with a 95% confidence interval of (0.575, 0.804). This result aligns closely with the findings in Mazumder (2016), which suggest an IGE for family income in the U.S. likely exceeding 0.6, indi-

[12]If family income data during this period is insufficient, we use the closest available three-year window to ages 15–17.

cating relatively low intergenerational mobility. Additionally, our result falls within the range of 0.55 to 0.74 reported by Mitnik et al. (2015) using a non-parametric approach for traditional IGE.

Table 3—: Estimation Results of Intergenerational Elasticity in the United States Using Alternative Estimators

Locally Robust	**Naive ML**	**Life-cycle**	**Mid-life Income**
0.690	0.596	0.508	0.378
(0.575, 0.804)	(0.502, 0.690)	(0.475, 0.541)	(0.273, 0.484)

Sample size consists of 1138 child-father pairs drawn from the PSID core sample (Survey Research Center component), 25,929 child observations, and 16,033 father observations. 95% confidence intervals clustered at the family level are reported in parenthesis.

The Naive ML estimator yields an IGE of 0.596 (95% CI: (0.502, 0.690)), considerably lower than the Locally Robust estimate. As expected from our simulations (Table 1), the plug-in estimator suffers from finite-sample bias, which—together with estimation error, results in undercoverage. The differences between the locally robust and plug-in ML estimates and confidence intervals highlight the importance of employing locally robust moment conditions to ensure valid inference for the IGE. These differences in both point estimates and confidence intervals illustrate how conventional machine learning approaches, while useful for prediction, may require robustness corrections for proper statistical inference.

Our results reveal important differences in estimator performance: while the Life-Cycle (LC) estimator underestimates the IGE, it remains closer to the Locally Robust (LR) benchmark, whereas the Mid-life Income (MI) estimator exhibits substantial downward bias. Specifically, the LC estimator yields an IGE of 0.508 (95% CI: 0.475, 0.541). In contrast, the MI estimator produces a markedly lower IGE of 0.378 (95% CI: 0.273, 0.484), significantly underestimating the true intergenerational elasticity. These findings are in line with Corollaries 1.1 and 1.2, which explain the downward bias in both estimators while demonstrating the LC estimator improvement upon the MI approach. Moreover, as discussed in Section IV, a key advantage of the LC estimator lies in its ability to explicitly account for the empirical patterns driving the bias in IGE estimation. This explains why the LC estimator outperforms the MI estimator in practice.

Our results underscore the key advantages of the locally robust (LR) estimator. The contrast between the LR estimate and those from alternative approaches—0.51 for the life-cycle (LC) estimator and 0.38 for the mid-life income (MI) estimator—highlights the importance of our method for consistent estimation. Additionally, the comparison with the naive machine learning (ML) estimate of 0.60 further motivates the construction of a locally robust moment over the plug-in approach, particularly with regard to coverage. Taken together, these findings highlight the importance of identification, combined with local robustness, for studying income mobility through the lens of the intergenerational elasticity.

VI. Conclusions

The primary challenge in estimating the intergenerational elasticity (IGE) arises from the unavailability of complete income profiles. Consequently, researchers have relied on midlife income averages, which introduces measurement error, leading to downward-biased estimates. While recent methodological advances have mitigated some of this attenuation bias, they overlook a more fundamental issue: the absence of formal identification of the IGE when income data is incomplete. This is not merely a technical concern; without proper identification, no consistent estimator exists (Gabrielsen, 1978), which undermines both the reliability and comparability of IGE estimates.

This paper addresses this issue by providing valid inference for the intergenerational elasticity through three key contributions: (i) the identification of the IGE in the presence of incomplete income data; (ii) the development of a consistent, locally robust estimator that corrects for first-step prediction errors; and (iii) valid inference that accounts for uncertainty from the first-stage estimation of nuisance parameters.

First, we establish nonparametric identification by leveraging family characteristics under standard missing at random assumptions. Moving beyond the conventional generalized error-in-variables model, we instead exploit the structural definition of lifetime income as the sum of annual earnings during working life. This approach allows us to recover the intergenerational elasticity from conditional moments of parental and child incomes.

Second, we develop a consistent and locally robust estimator by constructing an orthogonal moment function. This ensures that the machine learning estimation of nuisance parameters, such as conditional means, has no local effect on the IGE estimate. Finally, we establish the estimator's asymptotic normality.

Our framework enables comparable IGE estimates across time and place in the presence of incomplete income data. Importantly, our study complements rather than replaces rank-based measures by enabling valid inference for contexts where the IGE is more appropriate, such as cross-country comparisons or analyses of absolute mobility trends. By addressing key methodological challenges, our approach establishes a robust foundation for studying income persistence, enhancing the reliability and interpretability of mobility research across diverse economic settings.

Our simulation analysis illustrates the superior finite-sample performance of the locally robust estimator, which exhibits negligible bias and near-nominal coverage rates across different scenarios, outperforming alternative approaches that exhibit both higher bias and poor coverage. Using the PSID core sample, our approach yields an intergenerational elasticity of 0.69 for the United States, delivering a reliable and comparable measure of income persistence across generations.

Our study highlights two important directions for future research. First, our identification strategy requires longitudinal data that are often unavailable in developing countries - where understanding income persistence is most relevant. Future work should establish alternative identification results for data-scarce environments.

Second, the shift in the literature toward rank-based measures has been partly motivated by concerns about the nonlinear relationship between log child income and log parent income. Accordingly, future research should study identification and locally robust estimation for a nonlinear version of the intergenerational elasticity. One promising direction involves estimating the regression of child lifetime income on parent lifetime income in levels. A quantile-specific elasticity can then be constructed by multiplying the marginal effect at each point in the parental income distribution by the ratio of average child to parent income at that quantile. This would provide a richer, distributional perspective on income persistence and allow researchers to quantify how mobility varies across the income ladder patterns while avoiding the limitations of log-linear specifications.

References

Aaronson, D. and Mazumder, B. (2008). Intergenerational economic mobility in the united states, 1940 to 2000. *Journal of Human Resources*, 43(1):139–172.

An, Y., Wang, L., and Xiao, R. (2022). A nonparametric nonclassical measurement error approach to estimating intergenerational mobility elasticities. *Journal of Business & Economic Statistics*, 40(1):169–185.

Becker, G. S. and Tomes, N. (1986). Human capital and the rise and fall of families. *Journal of labor economics*, 4(3, Part 2):S1–S39.

Björklund, A. and Jäntti, M. (1997). Intergenerational income mobility in sweden compared to the united states. *The American Economic Review*, 87(5):1009–1018.

Blanden, J., Haveman, R., Smeeding, T., and Wilson, K. (2014). Intergenerational mobility in the u nited s tates and g reat b ritain: A comparative study of parent–child pathways. *Review of Income and Wealth*, 60(3):425–449.

Böhlmark, A. and Lindquist, M. J. (2006). Life-cycle variations in the association between current and lifetime income: replication and extension for sweden. *Journal of Labor Economics*, 24(4):879–896.

Chernozhukov, V., Escanciano, J. C., Ichimura, H., Newey, W. K., and Robins, J. M. (2022). Locally robust semiparametric estimation. *Econometrica*, 90(4):1501–1535.

Chetty, R., Hendren, N., Kline, P., and Saez, E. (2014). Where is the land of opportunity? the geography of intergenerational mobility in the united states. *The Quarterly Journal of Economics*, 129(4):1553–1623.

Chetverikov, D. and Wilhelm, D. (2023). Inference for rank-rank regressions. *arXiv preprint arXiv:2310.15512*.

Couch, K. A. and Lillard, D. R. (1998). Sample selection rules and the intergenerational correlation of earnings. *Labour Economics*, 5(3):313–329.

Dahl, M. W. and DeLeire, T. (2008). *The association between children's earnings and fathers' lifetime earnings: estimates using administrative data*. University of Wisconsin-Madison, Institute for Research on Poverty Madison.

de Wolff, P. and van Slijpe, A. R. (1973). The relation between income, intelligence, education and social background. *European Economic Review*, 4(3):235–264.

Fitzgerald, J. (2011). Attrition in models of intergenerational links in health and economic status in the psid. *The BE Journal of Economic Analysis & Policy*, 11(3):1–61.

Fitzgerald, J., Gottschalk, P., and Moffitt, R. A. (1998). An analysis of sample attrition in panel data: The michigan panel study of income dynamics.

Francesconi, M. and Nicoletti, C. (2006). Intergenerational mobility and sample selection in short panels. *Journal of Applied Econometrics*, 21(8):1265–1293.

Freeman, R. B. (1978). Black economic progress after 1964: who has gained and why? Technical report, National Bureau of Economic Research.

Friedman, M. (1957). The permanent income hypothesis. In *A theory of the consumption function*, pages 20–37. Princeton University Press.

Gabrielsen, A. (1978). Consistency and identifiability. *Journal of Econometrics*, 8(2):261–263.

Haider, S. and Solon, G. (2006). Life-cycle variation in the association between current and lifetime earnings. *American economic review*, 96(4):1308–1320.

Hauser, R. M., Sewell, W. H., and Lutterman, K. G. (1975). Socioeconomic background, ability, and achievement.

Hausman, J. (2001). Mismeasured variables in econometric analysis: problems from the right and problems from the left. *Journal of Economic perspectives*, 15(4):57–67.

Heidrich, S. (2016). A study of the missing data problem for intergenerational mobility using simulations. Technical Report 930, Umeå University, Department of Economics.

Lee, C.-I. and Solon, G. (2009). Trends in intergenerational income mobility. *The review of economics and statistics*, 91(4):766–772.

Lillard, L. A. and Panis, C. W. (1998). Panel attrition from the panel study of income dynamics: Household income, marital status, and mortality. *Journal of Human Resources*, pages 437–457.

Lubotsky, D. and Wittenberg, M. (2006). Interpretation of regressions with multiple proxies. *The Review of Economics and Statistics*, 88(3):549–562.

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

Mazumder, B. (2016). Estimating the intergenerational elasticity and rank association in the united states: Overcoming the current limitations of tax data. In *Inequality: Causes and consequences*, pages 83–129. Emerald group publishing limited.

Mello, U., Nybom, M., and Stuhler, J. (2024). A lifecycle estimator of intergenerational income mobility. Technical report, Working Paper.

Mitnik, P., Bryant, V., Weber, M., and Grusky, D. B. (2015). New estimates of intergenerational mobility using administrative data. *Statistics of Income Division working paper, Internal Revenue Service*.

Mitnik, P. A. and Grusky, D. B. (2020). The intergenerational elasticity of what? the case for redefining the workhorse measure of economic mobility. *Sociological Methodology*, 50(1):47–95.

Nybom, M. and Stuhler, J. (2016). Heterogeneous income profiles and lifecycle bias in intergenerational mobility estimation. *Journal of Human Resources*, 51(1):239–268.

Nybom, M. and Stuhler, J. (2017). Biases in standard measures of intergenerational income dependence. *Journal of Human Resources*, 52(3):800–825.

Rosenbaum, P. R. and Rubin, D. B. (1983). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society: Series B (Methodological)*, 45(2):212–218.

Schoeni, R. F. and Wiemers, E. E. (2015). The implications of selective attrition for estimates of intergenerational elasticity of family income. *The Journal of Economic Inequality*, 13(3):351–372.

Semenova, V., Goldman, M., Chernozhukov, V., and Taddy, M. (2023). Inference on heterogeneous treatment effects in high-dimensional dynamic panels under weak dependence. *Quantitative Economics*, 14(2):471–510.

Solon, G. (1992). Intergenerational income mobility in the united states. *The American Economic Review*, pages 393–408.

Stuhler, J. et al. (2018). A review of intergenerational mobility and its drivers. *Publications Office of the European Union, Luxembourg*.

Tsai, S.-L. (1983). *Sex differences in the process of stratification*. The University of Wisconsin-Madison.

Vosters, K. and Nybom, M. (2017). Intergenerational persistence in latent socioeconomic status: evidence from sweden and the united states. *Journal of Labor Economics*, 35(3):869–901.

APPENDIX A

A1. Definition of Lifetime Income

We define lifetime income as the sum of log annual income:

$$Y_g^P := \sum_{t=1}^{T} Y_{gt}, \quad g \in \{c, f\},$$

where Y_{gt} represents the log annual income for generation g (children or fathers) at time t. This section motivates this definition, justifies its alignment with standard practices, and illustrates its robustness for estimating the intergenerational elasticity (IGE).

The primary motivation for considering lifetime income as the sum of log annual income is to enable the identification of the IGE. Considering an alternative definition of lifetime income as $Y_g^{PL} := \log\left(\sum_t e^{Y_{gt}}\right)$, where $e^{Y_{gt}}$ denotes absolute annual income, complicates identification of the IGE. This stems from the nonlinearity introduced by the logarithm and summation prevents the expectation from being expressed as a simple sum of individual expectations due to Jensen's inequality:

$$\mathbb{E}\left[Y_g^{PL}\right] = \mathbb{E}\left[\log\left(\sum_t e^{Y_{gt}}\right)\right] \neq \mathbb{E}\left[Y_g^P\right] = \mathbb{E}\left[\sum_t Y_{gt}\right] = \sum_t \mathbb{E}\left[Y_{gt}\right].$$

The left-hand side requires taking the expectation over the full (unobserved) joint distribution of the vector $\left(Y_{g1}, \ldots, Y_{gT}\right)$, which captures all dependencies across time periods. In contrast, the right-hand side depends only on the marginal distributions of each Y_{gt}, which can be partially observed and estimated from available data. This distinction is critical under missing data scenarios, where annual incomes are not fully observed for all individuals or periods. Our identification strategy for the IGE relies on the linearity of $Y_g^P = \sum_t Y_{gt}$, which allows us to interchange the summation and expectation operators. Under standard missing at random (MAR) and conditional mean independence assumptions, we can recover $\mathbb{E}\left[Y_g^P\right]$ using observed conditional means:

$$\begin{aligned}
\mathbb{E}\left[Y_g^P\right] &= \mathbb{E}\left[\sum_{t=1}^{T} Y_{gt}\right] \\
&= \sum_{t=1}^{T} \mathbb{E}\left[Y_{gt}\right] \\
&= \sum_{t=1}^{T} \mathbb{E}\left[\mathbb{E}\left[Y_{gt} \mid \boldsymbol{X}_{gt}\right]\right]
\end{aligned}$$

$$= \sum_{t=1}^{T} \mathbb{E}\left[\mathbb{E}\left[Y_{gt} \mid \boldsymbol{X}_{gt}, D_{gt} = 1\right]\right]$$
$$= \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{E}\left[Y_{gt} \mid \boldsymbol{X}_{gt}, D_{gt} = 1\right]\right],$$

where $\boldsymbol{X}_{gt}$ represents family characteristics predictive of annual income for generation g at time t, and D_{gt} is an indicator equal to 1 if Y_{gt} is observed and 0 otherwise. The MAR assumption ensures that $\mathbb{E}\left[Y_{gt} \mid \boldsymbol{X}gt\right] = \mathbb{E}\left[Ygt \mid \boldsymbol{X}gt, Dgt = 1\right]$, allowing us to impute missing values using observed data.

For Y_g^{PL}, however, no such straightforward decomposition exists. The expectation $\mathbb{E}\left[\log\left(\sum_t e^{Y_{gt}}\right)\right]$ cannot be reduced to a sum of conditional expectations of individual Y_{gt}. Instead, it would require knowledge (or estimation) of the entire joint distribution of $\left(Y_{g1}, \ldots, Y_{gT}\right)$ to account for correlations and higher-order moments across periods. Under missing data, this would necessitate stronger assumptions about the joint distribution and potentially complex imputation methods for the full vector of incomes, rather than period-by-period conditional means. This makes identification infeasible with our strategy, which exploits only marginal conditional expectations and covariances from partially observed data.

Having motivated our definition of lifetime income, we now provide three empirical justifications for it. First, it is consistent with the common practice of measuring parental lifetime income as the average of log annual earnings (Solon, 1992; Lee and Solon, 2009; Mello et al., 2024). For the purpose of estimating the IGE, using the sum or the average of (log) income is equivalent, although the estimand for the constant differs due to the constant scaling factor $(1/T)$ across observations.[13]

Second, when measuring children's lifetime income, it is common to use only one observation, as measurement error in the dependent variable (usually) only affects efficiency. In this case, the distinction between specifications collapses: with $T = 1$, both $\log(\sum_t e^{Y_{gt}})$ and $\sum_t Y_{gt}$ reduce to the same quantity.

Third, our definition yields virtually identical results to the log-sum specification (Y_c^{PL}) when applied to the life-cycle estimator of Mello et al. (2024). The authors define children's lifetime income as the log of the sum of absolute incomes. Their life-cycle estimator involves first predicting log annual incomes for each individual of the child generation and converting these predictions into absolute incomes. These absolute incomes are subsequently summed to form lifetime income, after which the IGE is estimated by regressing the child's log lifetime income on the parents' average log family income. To rationalize this similarity, consider the average log annual income , $\overline{Y}_g := \frac{1}{T} \sum_t Y_{gt}$. The

[13]This equivalence arises because the scaling component $(1/T)$ is constant across observations.

log-sum can be expressed as:

$$\log\left(\sum_{t=1}^{T} e^{Y_{gt}}\right) = \log\left(e^{\overline{Y}_g} \sum_{t=1}^{T} e^{Y_{gt}-\overline{Y}_g}\right)$$
$$= \overline{Y}_g + \log\left(\sum_{t=1}^{T} e^{Y_{gt}-\overline{Y}_g}\right).$$

When annual log income deviates minimally from the average ($Y_{gt} - \overline{Y}_g \approx 0$), then $e^{Y_{gt}-\overline{Y}_g} \approx 1$ for all t, leading to:

$$\log\left(\sum_{t=1}^{T} e^{Y_{gt}}\right) \approx \overline{Y}_g + \log(T).$$

This approximation explains the similarity in IGE estimates, as the difference between the two definitions is approximately a constant, $\log(T)$, which does not affect the IGE slope.

A2. Identification of the IGE When Lifetime Income Is Observed

To obtain a closed-form expression for the intergenerational elasticity, we depart from the population regression of parental lifetime income on children's lifetime income:

$$Y_c^P = \alpha_0 + \beta_0 Y_f^P + u, \quad \mathbb{E}\left[u\left(1, Y_f^P\right)'\right] = 0,$$

and exploit the orthogonality conditions as follows:

$$\begin{aligned}
\mathbb{E}\left[Y_c^P\right] &= \alpha_0 + \beta_0 \mathbb{E}\left[Y_f^P\right] + \mathbb{E}\left[u\right] \\
Y_c^P - \mathbb{E}\left[Y_c^P\right] &= \alpha_0 - \alpha_0 + \beta_0\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right) + u - \mathbb{E}\left[u\right] \\
Y_c^P - \mathbb{E}\left[Y_c^P\right] &= \beta_0\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right) + u \\
\left(Y_c^P - \mathbb{E}\left[Y_c^P\right]\right)\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right) &= \beta_0\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2 + u\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right) \\
\mathbb{E}\left[\left(Y_c^P - \mathbb{E}\left[Y_c^P\right]\right)\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)\right] &= \beta_0 \mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right] + \mathbb{E}\left[u\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)\right] \\
\beta_0 &= \frac{\mathbb{E}\left[\left(Y_c^P - \mathbb{E}\left[Y_c^P\right]\right)\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)\right]}{\mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right]}.
\end{aligned}$$

A3. The Mid-life Income Estimator

The standard approach to estimating intergenerational elasticity, which we label the mid-life income (MI) estimator, proxies lifetime income by averaging (log) annual income snapshots around mid-life, primarily for fathers, though also applicable to children. This strategy is motivated by the errors-in-variables (EIV) framework, building on the permanent income hypothesis of Friedman (1957), which posits that observed income depends on a permanent (lifetime) and a transitory component. By averaging multiple years of income, the MI estimator isolates the permanent component, reducing the impact of transitory fluctuations on intergenerational elasticity estimates.

Following the seminal work of Solon (1992), this approach has become standard for measuring fathers' lifetime income, with researchers using a simple average of (log) yearly income. Solon's key contribution was showing that averaging multiple income snapshots reduces attenuation bias, with the bias decreasing as the number of periods averaged increases. However, as noted by Becker and Tomes (1986), earlier studies (de Wolff and van Slijpe, 1973; Hauser et al., 1975; Freeman, 1978; Tsai, 1983) had already employed income averaging to mitigate response errors and transitory components, laying the groundwork for this practice.

Using mid-life observations to proxy permanent income is rationalized by the generalized error-in-variables (GEIV) model (Haider and Solon, 2006)

$$Y_{gt} = \lambda_t Y_g^P + v_{gt}, \quad \mathbb{E}\left[v_{gt} Y_g^P\right] = 0, \quad g \in c, f, \quad t = 1, \dots, T, \tag{A1}$$

where Y_g^P is lifetime income, λ_t captures varying persistence over the life cycle, and v_{gt} is an age-specific shock. As suggested by equation (A1), annual income at younger and older ages is a noisier measure of lifetime income compared to mid-life income, as lifetime income is less persistent during these periods (indicated by a smaller λ_t). This phenomenon, known as life-cycle bias, can be mitigated by measuring income during mid-life, when persistence approaches one (Haider and Solon, 2006). Extensive evidence supports this approach, demonstrating that mid-life income yields more accurate estimates of lifetime income for fathers (e.g., Böhlmark and Lindquist (2006); Nybom and Stuhler (2017)).

For children's lifetime income, standard practice uses a single mid-life observation, as measurement error in the dependent variable (usually) only affects efficiency, while error in the independent variable causes attenuation bias (Hausman, 2001).

Formally, the MI estimand $\left(\beta^{MI}\right)$ corresponds to the slope coefficient of the projection of the average child's (log) income during mid-life $\left(\tilde{Y}_c^P\right)$ on the average parental (log) income during mid-life $\left(\tilde{Y}_f^P\right)$:

$$\begin{aligned} \tilde{Y}_c^P &= \alpha^{MI} + \beta^{MI} \tilde{Y}_f^P + u^{MI}, \quad \mathbb{E}\left[u^{MI}\left(1, \tilde{Y}_f^P\right)'\right] = 0, \\ \tilde{Y}_g^P &:= \frac{1}{T_g} \sum_{j \in \mathcal{M}_g} Y_{gj} D_{gj}, \quad g \in \{c, f\}, \end{aligned} \tag{A2}$$

where $D_{gj} = 1$ when Y_{gj} is observed and zero otherwise,[14] $\mathcal{M}_g$ is a set of pre-defined mid-life years for generation g,[15] and $T_g := \sum_{j \in \mathcal{M}_g} D_{gj}$ is the number of years used for the average. Following the same procedure as in equation (2) leads to the closed-form expression for the MI estimand

$$\beta^{MI} = \frac{\mathbb{E}\left[\left(\tilde{Y}_c^P - \mathbb{E}\left[\tilde{Y}_c^P\right]\right)\left(\tilde{Y}_f^P - \mathbb{E}\left[\tilde{Y}_f^P\right]\right)\right]}{\mathbb{E}\left[\left(\tilde{Y}_f^P - \mathbb{E}\left[\tilde{Y}_f^P\right]\right)^2\right]} := \frac{\mathbb{E}\left[\tilde{y}_c^P \tilde{y}_f^P\right]}{\mathbb{E}\left[\left(\tilde{y}_f^P\right)^2\right]}, \tag{A3}$$

where low-case letters denote the random variables in deviations from their population mean. Thus, the corresponding MI estimator is given by

$$\hat{\beta}_n^{MI} = \frac{\mathbb{E}_n\left[\left(\tilde{Y}_c^P - \mathbb{E}_n\left[\tilde{Y}_c^P\right]\right)\left(\tilde{Y}_f^P - \mathbb{E}_n\left[\tilde{Y}_f^P\right]\right)\right]}{\mathbb{E}_n\left[\left(\tilde{Y}_f^P - \mathbb{E}_n\left[\tilde{Y}_f^P\right]\right)^2\right]}, \tag{A4}$$

where $\mathbb{E}_n[X] := \frac{1}{n}\sum_{i=1}^n X_i$ is the empirical expectation operator.

To study identification of the MI estimand, we now state its underlying assumptions. The first one begins by imposing that the GEIV model of equation (A1), which motivates the MI estimator, is correctly specified and assumes that persistency of lifetime income equals 1 during mid-life. This assumption serves the crucial function of mapping unobserved lifetime income to (partially) observed (log) annual income.

Assumption 1-MI. *(Annual Income Process) The relationship between annual and lifetime income is governed by*

$$\begin{aligned} Y_{gt} &= \lambda_t Y_g^P + v_{gt}, \quad \mathbb{E}\left[v_{gt} Y_g^P\right] = 0, \quad g \in \{c, f\}, \quad t = 1, ..., T, \\ \lambda_t &= 1, \forall t \in \mathcal{M}_g, \quad g \in \{c, f\}. \end{aligned}$$

where λ_t captures that the persistence of lifetime income may vary over the life-cycle period, and v_{gt} is an age shock.

While empirical evidence suggests that λ_t approaches one during mid-life (Haider and Solon, 2006; Nybom and Stuhler, 2016), the assumption that λ_t equals one in this period is unlikely to hold. This motivates the use of optimally weighted income measures (Lubotsky and Wittenberg, 2006), which provide more accurate estimates than simple

[14] While $D_{gj} = 1$ is formally defined as indicating when Y_{gj} is observed, in practice, it also implicitly requires that Y_{gj} is used for estimation. This distinction arises because empirical studies often sample parental income selectively—for example, by focusing on log annual income during midlife (e.g., ages 30–50) to reduce lifecycle bias or measurement error. Thus, even if income is observed in other years, it may be excluded from estimation due to sampling design. This refinement clarifies that D_{gj} reflects both data availability and inclusion criteria, ensuring consistency with standard empirical approaches.

[15] While some papers define parental mid-life according to their offspring's age (Chetty et al., 2014; Blanden et al., 2014), others use parental age (Björklund and Jäntti, 1997; Mazumder, 2005), so $\mathcal{M}_f$ can differ from $\mathcal{M}_c$.

averages. Crucially, even if $\lambda_t = 1$ during mid-life, the MI estimator remains inconsistent for the IGE(Nybom and Stuhler, 2016).

The model in equation (A2), together with Assumption 1-MI, involves a random i.i.d sample of $\tilde{W} = \left(\boldsymbol{Y}_c \odot D_c, \boldsymbol{Y}_f \odot D_f, \boldsymbol{D}_c, \boldsymbol{D}_f\right)$,[16] unobserved components including: (i) the time-varying shocks v_{gt} for both children and fathers, indexed by $g \in c, f$ and $t = 1, \ldots, T$; (ii) the lifetime income Y_g^P of both generations $g \in c, f$; and (iii) the error term u^{MI} associated with the MI estimator. The model parameters consist of the parameter of interest β^{MI}, and the nuisance parameter α^{MI}.[17] To evaluate whether β^{MI} identifies β_0, we now introduce zero conditional mean restrictions involving the observed and unobserved components in this setting.

Assumption 2-MI. *(Conditional Mean Independence) The following conditional mean restrictions hold*

$$
\begin{aligned}
\mathbb{E}\left[v_{ct}v_{fj}\middle|D_{ct}, D_{fj}\right] &= 0, \quad t \in \mathcal{M}_c, \quad j \in \mathcal{M}_f,\\
\mathbb{E}\left[v_{fj}Y_c^P\middle|D_{fj}\right] &= 0, \quad j \in \mathcal{M}_f,\\
\mathbb{E}\left[v_{ft}Y_f^P|D_{ft}, D_{fj}\right] &= 0, \quad tj \in \mathcal{M}_f,\\
\mathbb{E}\left[v_{gj}|D_{gj}\right] &= 0, \quad g \in \{c, f\}, \quad j \in \mathcal{M}_j.
\end{aligned}
$$

Assumption 2-MI imposes a set of orthogonality conditions involving age shocks, missingness indicators, and unobserved lifetime income. The first condition states that, conditional on the missingness indicators for child and parent income, mid-life age shocks to children and parents are mean independent. The second condition requires that parental age shocks during mid-life are mean independent of the child's lifetime income, conditional on the missingness status of parental income. The third condition assumes that, given the missingness indicators for a tuple of years of parental income, age shocks are mean independent of the parent's lifetime income.[18] Lastly, the fourth condition states that observed age shocks have zero mean, conditional on the missingness status of the corresponding annual income observation.

Previous work (e.g., Couch and Lillard (1998); Mazumder (2005); Heidrich (2016)) has shown that the MI estimator does not perform well. Furthermore, there is an extensive literature discussing its sources of bias (e.g., Solon (1992); Mazumder (2005); Nybom and Stuhler (2016)). We now briefly examine the sources of bias that prevent the MI estimator from being consistent. As shown in Corollary 1.1, the probability limit of

[16]Because the MI estimator does not incorporate family characteristics in its estimation procedure, we abstract from their observation in our analysis.

[17]While $\{\lambda_t\}_{t=1}^T$ would typically be nuisance parameters in an unrestricted model, our framework does not classify them as such, as we impose the restriction $\lambda_t = 1$ during mid-life.

[18]By the law of iterated expectations, this condition implies that age shocks are uncorrelated with lifetime income for the parental generation, as already assumed in Assumption 1-MI, but the reverse does not necessarily hold.

$\hat{\beta}_n^{MI}$ takes the form:

(A5)

$$\hat{\beta}_n^{MI} \xrightarrow{p} \frac{\beta_0 \mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right] + \overbrace{\frac{1}{T_c}\sum_t \mathbb{E}\left[Y_f^P v_{ct} \middle| D_{ct} = 1, t \in \mathcal{M}_c\right]}^{(c)} \times \overbrace{p_c\left(t \in \mathcal{M}_c\right)}^{(d)}}{\mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right] + \underbrace{\frac{1}{T_f^2}\sum_t \sum_j \mathbb{E}\left[v_{ft} v_{fj} \middle| D_{ft} = 1, D_{fj} = 1, \{t, j\} \in \mathcal{M}_f\right]}_{(a)} \times \underbrace{p_f\left(\{t, j\} \in \mathcal{M}_f\right)}_{(b)}}.$$

The downward bias by measurement error highlighted by Solon (1992) and Mazumder (2005) corresponds to component (a) in equation (A5). Consider rewriting (a) as

(A6)

$$\frac{1}{T_f^2}\sum_t \mathbb{E}\left[v_{ft}^2 \middle| D_{ft} = 1, D_{fj} = 1, \{t, j\} \in \mathcal{M}_f\right] + \frac{2}{T_f^2}\sum_t \sum_{j \neq t} \mathbb{E}\left[v_{ft} v_{fj} \middle| D_{ft} = 1, D_{fj} = 1, \{t, j\} \in \mathcal{M}_f\right],$$

where the first component is the variance of the transitory income component, causing the attenuation bias shown is Solon (1992), while the second term comprises the autoregressive nature of the transitory component illustrated in Mazumder (2005). This term rationalizes why even the 10-year average is not enough for the attenuation bias to vanish due to the transitory component of income being highly serially correlated (Mazumder, 2005). As T_f grows (more years are used for the average), the first component in the last display might vanish. However, the second one does not, because the number of covariances in the second term is $T_f^2 - T_f$. Consequently, the second term in the last display will not disappear when the transitory income component is highly serially correlated.

Component (b) captures the sensitivity of the IGE estimates to low, zero, and missing income documented in the literature. There is extensive evidence that IGE estimates are not robust to how extreme and missing incomes are treated (Couch and Lillard, 1998; Dahl and DeLeire, 2008; Chetty et al., 2014; Nybom and Stuhler, 2016). When observations with zero and low income are dropped, the probability of observing a given tuple of years for parents changes. Moreover, not observing lifetime income also affects the probability of observing a given tuple of years. If we were to observe lifetime income, components (d) and (b) in equation (A5) would be equal to 1, and would not induce bias.

The sensitivity of the IGE to sample inclusion rules is also comprised in component (d). Couch and Lillard (1998) show with empirical evidence that the MI estimator is sensitive to different sample inclusion rules. As noted by Francesconi and Nicoletti (2006), studies usually restrict their analysis to children from specific birth cohorts. The upper bound for the birth cohort is required to ensure that children's socioeconomic status is observed as long as possible so their observed status is a reliable measure of long-run lifetime status. Imposing such a restriction mechanically affects the probability of observing income in a given year, which corresponds to component (b) in equation (A5).

Even when using the same data, changes in the definition of mid-life alter the estimand in equation (A3), further limiting comparability. The transitory component of children's income depending on parental income is encompassed by (c) in equation (A5). Nybom and Stuhler (2016) highlight that children from affluent families might experience faster income growth, which would cause the steepness of the income trajectory to depend on parental lifetime income. Thus, the correlation between age shocks to children's (log) annual income and parental lifetime income induces bias in estimating the IGE.

Component (d) in equation (A5) captures that the estimate of the IGE depends both on the number of years used to measure children's income and the selected year(s). Mello et al. (2024) provide evidence that using $\hat{\beta}_n^{MI}$ to estimate the IGE is sensitive to the span of ages where the child generation is observed and the number of income observations available for each individual. The first finding is captured by the pre-defined mid-life years ($\mathcal{M}_c$) used to proxy children's lifetime income. The second one, by the cardinality of $\mathcal{M}_c$ affecting the magnitude of component (d).

As shown in equation (A27), if we were to drop the assumption that $\lambda_t = 1$ during mid-life, our inconsistency result would also capture the life-cycle bias in estimating the IGE. As previously mentioned, imposing $\lambda_t = 1$ allows us to obtain the closed-form solution in equation (A5). However, relaxing this assumption allows our inconsistency result to capture another source of bias discussed in the literature.

A4. Proof of Theorem 1

We aim to characterize the population quantity identified by the MI estimand, defined as

$$\beta^{MI} = \frac{\mathbb{E}\left[\left(\tilde{Y}_c^P - \mathbb{E}\left[\tilde{Y}_c^P\right]\right)\left(\tilde{Y}_f^P - \mathbb{E}\left[\tilde{Y}_f^P\right]\right)\right]}{\mathbb{E}\left[\left(\tilde{Y}_f^P - \mathbb{E}\left[\tilde{Y}_f^P\right]\right)^2\right]} := \frac{\mathbb{E}\left[\tilde{y}_c^P \tilde{y}_f^P\right]}{\mathbb{E}\left[\left(\tilde{y}_f^P\right)^2\right]},$$

where low-case letters denote the random variables in deviations from their population mean. To this end, we derive closed-form expressions for the numerator and denominator under Assumptions 1-MI and 2-MI. We start by analyzing the denominator:

$$\begin{aligned}
\mathbb{E}\left[\tilde{y}_c^P \tilde{y}_f^P\right] &= \mathbb{E}\left[\tilde{Y}_c^P \tilde{Y}_f^P\right] - \mathbb{E}\left[\tilde{Y}_c^P\right]\mathbb{E}\left[\tilde{Y}_f^P\right] \\
&= \mathbb{E}\left[\left(\frac{1}{T_c}\sum_{t\in\mathcal{M}_c} Y_{ct}D_{ct}\right)\left(\frac{1}{T_f}\sum_{j\in\mathcal{M}_f} Y_{fj}D_{fj}\right)\right] - \mathbb{E}\left[\tilde{Y}_c^P\right]\mathbb{E}\left[\tilde{Y}_f^P\right] \\
&= \frac{1}{T_cT_f}\sum_{t\in\mathcal{M}_c}\sum_{j\in\mathcal{M}_f} \mathbb{E}\left[Y_{ct}Y_{fj}D_{ct}D_{fj}\right] - \mathbb{E}\left[\tilde{Y}_c^P\right]\mathbb{E}\left[\tilde{Y}_f^P\right] \\
&= \frac{1}{T_cT_f}\sum_{t\in\mathcal{M}_c}\sum_{j\in\mathcal{M}_f} \mathbb{E}\left[\left(\lambda_t Y_c^P + v_{ct}\right)\left(\lambda_j Y_f^P + v_{fj}\right)D_{ct}D_{fj}\right] - \mathbb{E}\left[\tilde{Y}_c^P\right]\mathbb{E}\left[\tilde{Y}_f^P\right]
\end{aligned}$$

$$= \frac{1}{T_cT_f}\mathbb{E}\left[Y_c^PY_f^P\sum_{t\in\mathcal{M}_c}\lambda_tD_{ct}\sum_{j\in\mathcal{M}_f}\lambda_jD_{fj}\right] + \frac{1}{T_cT_f}\mathbb{E}\left[Y_f^P\sum_{t\in\mathcal{M}_c}v_{ct}D_{ct}\sum_{j\in\mathcal{M}_f}\lambda_jD_{fj}\right]$$

$$+ \frac{1}{T_cT_f}\sum_{t\in\mathcal{M}_c}\lambda_t\sum_{j\in\mathcal{M}_f}\mathbb{E}\left[Y_c^Pv_{fj}D_{ct}D_{fj}\right] + \frac{1}{T_cT_f}\sum_{t\in\mathcal{M}_c}\sum_{j\in\mathcal{M}_f}\mathbb{E}\left[v_{ct}v_{fj}D_{ct}D_{fj}\right]$$

$$- \mathbb{E}\left[\tilde{Y}_c^P\right]\mathbb{E}\left[\tilde{Y}_f^P\right], \tag{A7}$$

where we have used the definition of average (log) income during mid-life as given in equation (A2) in the second equality, and the fourth equality follows by Assumption 1-MI.

The first term in equation (A7) can be expressed as

$$\begin{aligned}
\frac{1}{T_cT_f}\mathbb{E}\left[Y_c^PY_f^P\sum_{t\in\mathcal{M}_c}\lambda_tD_{ct}\sum_{j\in\mathcal{M}_f}\lambda_jD_{fj}\right] &= \frac{1}{T_cT_f}\mathbb{E}\left[Y_c^PY_f^P\sum_{t\in\mathcal{M}_c}D_{ct}\sum_{j\in\mathcal{M}_f}D_{fj}\right]\\
&= \frac{1}{T_cT_f}\mathbb{E}\left[Y_c^PY_f^P\right]\times T_cT_f\\
&= \mathbb{E}\left[Y_c^PY_f^P\right],
\end{aligned} \tag{A8}$$

where the first equality follows from assuming $\lambda_t = 1, \forall t \in \mathcal{M}_g$ for $g \in \{c, f\}$ in Assumption 1-MI, and the second one by the definition of $T_g := \sum_{j\in\mathcal{M}_g} D_{gj}$. As regards the second term in equation (A7), it can be simplified as follows:

$$\begin{aligned}
\frac{1}{T_cT_f}\mathbb{E}\left[Y_f^P\sum_{t\in\mathcal{M}_c}v_{ct}D_{ct}\sum_{j\in\mathcal{M}_f}\lambda_jD_{fj}\right] &= \frac{1}{T_cT_f}\mathbb{E}\left[Y_f^P\sum_{t\in\mathcal{M}_c}v_{ct}D_{ct}\sum_{j\in\mathcal{M}_f}D_{fj}\right]\\
&= \frac{1}{T_cT_f}\mathbb{E}\left[Y_f^P\sum_{t\in\mathcal{M}_c}v_{ct}D_{ct}\right]T_f\\
&= \frac{1}{T_c}\sum_{t\in\mathcal{M}_c}\mathbb{E}\left[Y_f^Pv_{ct}D_{ct}\right]\\
&= \frac{1}{T_c}\sum_{t}\mathbb{E}\left[Y_f^Pv_{ct}\middle|D_{ct}=1, t\in\mathcal{M}_c\right]\times p\left(D_{ct}=1\middle|t\in\mathcal{M}_c\right)\\
&:= \frac{1}{T_c}\sum_{t}\mathbb{E}\left[Y_f^Pv_{ct}\middle|D_{ct}=1, t\in\mathcal{M}_c\right]\times p_c\left(t\in\mathcal{M}_c\right),
\end{aligned} \tag{A9}$$

where the fourth equality follows by the law of total probability.

The third term in equation (A7) equals zero by the law of iterated expectations (LIE) and Assumption 2-MI

$$\mathbb{E}\left[Y_c^P v_{fj} D_{ct} D_{fj}\right] = \mathbb{E}\left[\mathbb{E}\left[Y_c^P v_{fj} \middle| D_{ct}, D_{fj}\right] D_{ct} D_{fj}\right] = \mathbb{E}\left[\mathbb{E}\left[Y_c^P v_{fj} \middle| D_{fj}\right] D_{ct} D_{fj}\right] = 0. \tag{A10}$$

Similarly, the fourth term in equation (A7) also equals zero by Assumption 2-MI and LIE

$$\mathbb{E}\left[v_{ct} v_{fj} D_{ct} D_{fj}\right] = \mathbb{E}\left[\mathbb{E}\left[v_{ct} v_{fj} \middle| D_{ct}, D_{fj}\right] D_{ct} D_{fj}\right] = \mathbb{E}\left[\mathbb{E}\left[v_{ct} v_{fj} \middle| D_{ct}, D_{fj}\right] D_{ct} D_{fj}\right] = 0. \tag{A11}$$

Finally, for the last term in equation (A7) we have

$$\begin{aligned}
\mathbb{E}\left[\tilde{Y}_c^P\right]\mathbb{E}\left[\tilde{Y}_f^P\right] &= \mathbb{E}\left[\frac{1}{T_c}\sum_{t\in\mathcal{M}_c} Y_{ct} D_{ct}\right]\mathbb{E}\left[\frac{1}{T_f}\sum_{j\in\mathcal{M}_f} Y_{fj} D_{fj}\right] \\
&= \mathbb{E}\left[\frac{1}{T_c}\sum_{t\in\mathcal{M}_c}\left(\lambda_t Y_c^P + v_{ct}\right) D_{ct}\right]\mathbb{E}\left[\frac{1}{T_f}\sum_{j\in\mathcal{M}_f}\left(\lambda_j Y_f^P + v_{fj}\right) D_{fj}\right] \\
&= \mathbb{E}\left[Y_c^P \frac{1}{T_c}\sum_{t\in\mathcal{M}_c} D_{ct} + \frac{1}{T_c}\sum_{t\in\mathcal{M}_c} v_{ct} D_{ct}\right]\mathbb{E}\left[Y_f^P \frac{1}{T_f}\sum_{j\in\mathcal{M}_j} D_{fj} + \frac{1}{T_f}\sum_{j\in\mathcal{M}_f} v_{fj} D_{fj}\right] \\
&= \left(\mathbb{E}\left[Y_c^P\right] + \frac{1}{T_c}\sum_{t\in\mathcal{M}_c}\mathbb{E}\left[v_{ct} D_{ct}\right]\right)\left(\mathbb{E}\left[Y_f^P\right] + \frac{1}{T_f}\sum_{j\in\mathcal{M}_f}\mathbb{E}\left[v_{fj} D_{fj}\right]\right) \\
&= \left(\mathbb{E}\left[Y_c^P\right] + \frac{1}{T_c}\sum_{t\in\mathcal{M}_c}\mathbb{E}\left[\mathbb{E}\left[v_{ct}|D_{ct}\right] D_{ct}\right]\right) \\
&\quad\times\left(\mathbb{E}\left[Y_f^P\right] + \frac{1}{T_f}\sum_{j\in\mathcal{M}_f}\mathbb{E}\left[\mathbb{E}\left[v_{fj}|D_{fj}\right] D_{fj}\right]\right) \\
&= \mathbb{E}\left[Y_c^P\right]\mathbb{E}\left[Y_f^P\right], && \text{(A12)}
\end{aligned}$$

where the last equality follows by Assumption 1-MI.

By plugging equations (A9), (A8), (A10), (A11), and (A12) into equation (A7), we have

$$\begin{aligned}
\mathbb{E}\left[\tilde{y}_c^P \tilde{y}_f^P\right] &= \mathbb{E}\left[\left(Y_c^P - \mathbb{E}\left[Y_c^P\right]\right)\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)\right] \\
&\quad + \frac{1}{T_c}\sum_t \mathbb{E}\left[Y_f^P v_{ct} \middle| D_{ct} = 1, t \in \mathcal{M}_c\right] \times p_c\left(t \in \mathcal{M}_c\right). && \text{(A13)}
\end{aligned}$$

We now analyze the denominator in equation (A3)

$$
\begin{aligned}
\mathbb{E}\left[\left(\tilde{y}_f^P\right)^2\right] &= \mathbb{E}\left[\left(\frac{1}{T_f}\sum_{t\in\mathcal{M}_f} Y_{ft}D_{ft}\right)\left(\frac{1}{T_f}\sum_{j\in\mathcal{M}_f} Y_{fj}D_{fj}\right)\right] - \mathbb{E}\left[\left(\tilde{Y}_f^P\right)^2\right] \\
&= \frac{1}{T_f^2}\sum_{t\in\mathcal{M}_f}\sum_{j\in\mathcal{M}} \mathbb{E}\left[Y_{ft}Y_{fj}D_{ct}D_{fj}\right] - \mathbb{E}\left[\left(\tilde{Y}_f^P\right)^2\right] \\
&= \frac{1}{T_f^2}\sum_{t\in\mathcal{M}_f}\sum_{j\in\mathcal{M}_f} \mathbb{E}\left[\left(\lambda_t Y_f^P + v_{ft}\right)\left(\lambda_j Y_f^P + v_{fj}\right)D_{ft}D_{fj}\right] - \mathbb{E}\left[\left(\tilde{Y}_f^P\right)^2\right] \\
&= \frac{1}{T_f^2}\mathbb{E}\left[\left(Y_f^P\right)^2\sum_{t\in\mathcal{M}_f}\lambda_t D_{ft}\sum_{j\in\mathcal{M}_f}\lambda_j D_{fj}\right] + \frac{1}{T_f^2}\sum_{t\in\mathcal{M}_f}\sum_{j\in\mathcal{M}_f}\lambda_j \mathbb{E}\left[v_{ft}Y_f^P D_{ft}D_{fj}\right] \\
&+ \frac{1}{T_f^2}\sum_{t\in\mathcal{M}_f}\lambda_t\sum_{j\in\mathcal{M}_f}\mathbb{E}\left[Y_f^P v_{fj}D_{ft}D_{fj}\right] + \frac{1}{T_f^2}\sum_{t\in\mathcal{M}_f}\sum_{j\in\mathcal{M}_f}\mathbb{E}\left[v_{ft}v_{fj}D_{ft}D_{fj}\right] \\
&- \mathbb{E}\left[\left(\tilde{Y}_f^P\right)^2\right]. \qquad \text{(A14)}
\end{aligned}
$$

The first term of equation (A14) can be expressed as

$$
\begin{aligned}
\frac{1}{T_f^2}\mathbb{E}\left[\left(Y_f^P\right)^2\sum_{t\in\mathcal{M}_f}\lambda_t D_{ft}\sum_{j\in\mathcal{M}_f}\lambda_j D_{fj}\right] &= \frac{1}{T_f^2}\mathbb{E}\left[\left(Y_f^P\right)^2\sum_{t\in\mathcal{M}_f}D_{ft}\sum_{j\in\mathcal{M}_f}D_{fj}\right] \\
&= \frac{1}{T_f^2}\mathbb{E}\left[\left(Y_f^P\right)^2\right]T_f^2 \\
&= \mathbb{E}\left[\left(Y_f^P\right)^2\right] \qquad \text{(A15)}
\end{aligned}
$$

The second and third terms in the last display equal zero by Assumption 1-MI, since

$$
\mathbb{E}\left[v_{ft}Y_f^P D_{ft}D_{fj}\right] = \mathbb{E}\left[\mathbb{E}\left[v_{ft}Y_f^P | D_{ft}, D_{fj}\right]D_{ft}D_{fj}\right] = 0, \qquad \text{(A16)}
$$

and the fourth term

$$
\begin{aligned}
\frac{1}{T_f^2}\sum_{t\in\mathcal{M}_f}\sum_{j\in\mathcal{M}_f}\mathbb{E}\left[v_{ft}v_{fj}D_{ft}D_{fj}\right] &= \frac{1}{T_f^2}\sum_t\sum_j \mathbb{E}\left[v_{ft}v_{fj}\middle|D_{ft}=1, D_{fj}=1, \{t,j\}\in\mathcal{M}_f\right] \\
&\times p\left(D_{ft}=1, D_{fj}=1\middle|\{t,j\}\in\mathcal{M}_f\right) \\
&:= \frac{1}{T_f^2}\sum_t\sum_j \mathbb{E}\left[v_{ft}v_{fj}\middle|D_{ft}=1, D_{fj}=1, \{t,j\}\in\mathcal{M}_f\right] \\
&\times p_f\left(\{t,j\}\in\mathcal{M}_f\right). \qquad \text{(A17)}
\end{aligned}
$$

By the same arguments as those of equation (A12), the last term in equation (A14) boils down to

$$\mathbb{E}\left[\left(\tilde{Y}_f^P\right)^2\right] = \mathbb{E}\left[\left(Y_f^P\right)^2\right] \tag{A18}$$

By plugging equations (A15), (A16), (A17), and (A18) into equation (A14), we have

$$\begin{aligned}\mathbb{E}\left[\left(\tilde{y}_f^P\right)^2\right] &= \mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right] \\ &+ \frac{1}{T_f^2}\sum_t\sum_j \mathbb{E}\left[v_{ft}v_{fj}\middle| D_{ft}=1, D_{fj}=1, \{t,j\}\in\mathcal{M}_f\right]\times p_f\left(\{t,j\}\in\mathcal{M}_f\right).\end{aligned} \tag{A19}$$

Finally, by plugging equations (A13) and (A19) into (A3), we get

$$\beta^{MI} = \frac{\mathbb{E}\left[\left(Y_c^P - \mathbb{E}\left[Y_c^P\right]\right)\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)\right] + \frac{1}{T_c}\sum_t \mathbb{E}\left[Y_f^P v_{ct}\middle| D_{ct}=1, t\in\mathcal{M}_c\right]\times p_c\left(t\in\mathcal{M}_c\right)}{\mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right] + \frac{1}{T_f^2}\sum_t\sum_j \mathbb{E}\left[v_{ft}v_{fj}\middle| D_{ft}=1, D_{fj}=1, \{t,j\}\in\mathcal{M}_f\right]\times p_f\left(\{t,j\}\in\mathcal{M}_f\right)} .Q.E.D.$$

A5. Proof of Corollary 1.1

To characterize the limit in probability of the MI estimator, defined as

$$\hat{\beta}_n^{MI} = \frac{\mathbb{E}_n\left[\left(\tilde{Y}_c^P - \mathbb{E}_n\left[\tilde{Y}_c^P\right]\right)\left(\tilde{Y}_f^P - \mathbb{E}_n\left[\tilde{Y}_f^P\right]\right)\right]}{\mathbb{E}_n\left[\left(\tilde{Y}_f^P - \mathbb{E}_n\left[\tilde{Y}_f^P\right]\right)^2\right]}.$$

The numerator converges in probability to

$$\begin{aligned}\mathbb{E}_n\left[\tilde{y}_c^P\tilde{y}_f^P\right] &\xrightarrow{p} \mathbb{E}\left[\left(\tilde{Y}_c^P - \mathbb{E}\left[\tilde{Y}_c^P\right]\right)\left(\tilde{Y}_f^P - \mathbb{E}\left[\tilde{Y}_f^P\right]\right)\right] \\ &= \mathbb{E}\left[\left(Y_c^P - \mathbb{E}\left[Y_c^P\right]\right)\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)\right] \\ &+ \frac{1}{T_c}\sum_t \mathbb{E}\left[Y_f^P v_{ct}\middle| D_{ct}=1, t\in\mathcal{M}_c\right]\times p_c\left(t\in\mathcal{M}_c\right) \\ &= \mathbb{E}\left[\left(\beta_0\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right) + u\right)\left(\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)\right)\right] \\ &+ \frac{1}{T_c}\sum_t \mathbb{E}\left[Y_f^P v_{ct}\middle| D_{ct}=1, t\in\mathcal{M}_c\right]\times p_c\left(t\in\mathcal{M}_c\right) \\ &= \beta_0\mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right] + \frac{1}{T_c}\sum_t \mathbb{E}\left[Y_f^P v_{ct}\middle| D_{ct}=1, t\in\mathcal{M}_c\right]\times p_c\left(t\in\mathcal{M}_c\right)\end{aligned} \tag{A20}$$

where the convergence in probability follows by the Law of Large Numbers (LLN), the first equality from equation (A13), and the second and third by equation (1). As regards the denominator in equation (A4), it converges in probability to

$$
\begin{aligned}
\mathbb{E}_n\left[\left(\tilde{y}_f^P\right)^2\right] &\xrightarrow{p} \mathbb{E}\left[\left(\tilde{Y}_f^P - \mathbb{E}\left[\tilde{Y}_f^P\right]\right)^2\right] \\
&= \mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right] \\
&+ \frac{1}{T_f^2}\sum_t\sum_j \mathbb{E}\left[v_{ft}v_{fj}\middle|D_{ft}=1, D_{fj}=1, \{t,j\}\in\mathcal{M}_f\right]\times p_f\left(\{t,j\}\in\mathcal{M}_f\right)
\end{aligned}
\tag{A21}
$$

where the convergence in probability follows by the LLN, and the equality from equation (A19). Thus, by plugging equations (A20) and (A21) into (A3) and applying the Continuous Mapping Theorem (CMT) yields

$$
\hat{\beta}_n^{MI} \xrightarrow{p} \frac{\beta_0\mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right] + \frac{1}{T_c}\sum_t \mathbb{E}\left[Y_f^P v_{ct}\middle|D_{ct}=1, t\in\mathcal{M}_c\right]\times p_c\left(t\in\mathcal{M}_c\right)}{\mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right] + \frac{1}{T_f^2}\sum_t\sum_j \mathbb{E}\left[v_{ft}v_{fj}\middle|D_{ft}=1, D_{fj}=1, \{t,j\}\in\mathcal{M}_f\right]\times p_f\left(\{t,j\}\in\mathcal{M}_f\right)}. Q.E.D.
\tag{A22}
$$

A6. Equivalence with Previous Inconsistency Results

Corollary 1.1 encompasses previous formalizations of bias in estimating the IGE. In particular, we first show that the inconsistency result in Solon (1992) is a particular case of Corollary 1.1 under an additional assumption. We then show that the two inconsistency results in Nybom and Stuhler (2016) are particular cases of Corollary 1.1 when variants of Assumption 1-MI are considered, and Assumption 2-MI is relaxed.

We now show that equation (3) in Corollary 1.1 simplifies to the inconsistency result in (Solon, 1992, p. 400), if we further assume that parental lifetime income is uncorrelated to child age shocks, for the observed years during mid-life. In particular, consider assuming

Assumption 3-S.

$$
\begin{aligned}
\mathbb{E}\left[Y_f^P v_{ct}|D_{ct}=1, t\in\mathcal{M}_c\right] &= 0 \\
D_{ft} &= 1 \quad \forall t\in\mathcal{M}_f,
\end{aligned}
$$

so that

$$
\mathbb{E}\left[v_{ft}v_{fj}\middle|D_{ft}=1, D_{fj}=1, \{t,j\}\in\mathcal{M}_f\right]\times p_f\left(\{t,j\}\in\mathcal{M}_f\right) = \mathbb{E}\left[v_{ft}v_{fj}\right]\times 1.
$$

Then, under Assumptions 1-MI, 2-MI,and 3-S equation (A22) (which corresponds to equation (3) in Corollary 1.1) boils down to

$$\text{(A23)} \qquad \hat{\beta}_n^{MI} \xrightarrow{p} \frac{\beta_0 \mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right]}{\mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right] + \frac{1}{T_f^2}\sum_t \sum_j \mathbb{E}\left[v_{ft}v_{fj}\right]},$$

which is the result in Solon (1992). The shape of the second term in the denominator of the last display depends on the assumptions of the transitory shock to parental income. For instance, if we assume that is white noise, it boils down to V_v^2/T_f. Conversely, if we assume it follows a MA(1) it becomes $\left(V_v^2/T_f\right) \times \left[1 + 2\theta\left(T_f - 1\right)/T_f\right]$, where θ denotes the first-order autocorrelation. Finally, if we assume a stationary AR(1) process the second term in the denominator of the last display becomes $\left(V_v^2/T_f\right) \times \left[1 + 2\theta\left\{T_f - (1-\theta_f^T)/(1-\theta)\right\}/\left(T_f[1-\theta]\right)\right]$ (see footnote 17 in Solon (1992)). We now turn to contrasting Corollary 1.1 with more recent results.

There are two inconsistency results in Nybom and Stuhler (2016). Both of them impose the additional assumption that income is measured in a single year during mid-life. However, while the first result assumes an error-in-variables model, the second one assumes a generalized error-in-variables model. By formalizing these assumptions, we show that Corollary 1.1 encompasses both of these inconsistency results.

We first consider the inconsistency result in equation (2) in Nybom and Stuhler (2016). For this purpose, consider the following alternative to Assumption 1-MI

Assumption 1-NS. *(Annual Income Process) The relationship between annual and lifetime income is governed by*

$$\begin{aligned} &Y_{gt} = \lambda_t Y_g^P + v_{gt}, \quad \mathbb{E}\left[v_{gt}\right] = 0, \quad g \in \{c, f\}, \quad t = 1, ..., T, \\ &\lambda_t = 1, \forall t \in \mathcal{M}_g, \quad g \in \{c, f\}. \end{aligned}$$

Moreover, we further assume

$$\begin{aligned} &\mathcal{M}_g = T_g, \quad g \in \{c, f\} \\ &D_{gt} = 1 \quad \forall t = T_g, \quad g \in \{c, f\}. \end{aligned}$$

That is, we assume that parental and child's income are measured in a given year so that every child and parent is observed in that year. Moreover, we assume that the age shock to annual income has zero mean, and relax the assumption that transitory income shocks are uncorrelated to parental lifetime income. Furthermore, in their result, they authors relax the conditional mean restrictions in Assumption 2-MI.

Since we are only using one observation for both parents and children, we have $\tilde{Y}_c^P =$

Y_{ct} and $\tilde{Y}_f^P = Y_{fj}$, so that

$$
\begin{aligned}
\mathbb{E}\left[\tilde{y}_c^P \tilde{y}_f^P\right] &= \mathbb{E}\left[Y_{ct} Y_{fj}\right] - \mathbb{E}\left[Y_{ct}\right] \mathbb{E}\left[Y_{fj}\right] \\
&= \mathbb{E}\left[\left(Y_c^P + v_{ct}\right)\left(Y_f^P + v_{fj}\right)\right] - \mathbb{E}\left[\left(Y_c^P + v_{ct}\right)\right] \mathbb{E}\left[\left(Y_f^P + v_{fj}\right)\right] \\
&= \mathbb{E}\left[Y_c^P Y_f^P\right] + \mathbb{E}\left[Y_f^P v_{ct}\right] + \mathbb{E}\left[Y_c^P v_{fj}\right] + \mathbb{E}\left[v_{ct} v_{fj}\right] - \mathbb{E}\left[Y_c^P\right] \mathbb{E}\left[Y_f^P\right] \\
&= \mathbb{E}\left[\left(Y_c^P - \mathbb{E}\left[Y_c^P\right]\right)\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)\right] + \mathbb{E}\left[Y_f^P v_{ct}\right] + \mathbb{E}\left[Y_c^P v_{fj}\right] + \mathbb{E}\left[v_{ct} v_{fj}\right] \\
&= \beta_0 \mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right] + \mathbb{E}\left[Y_f^P v_{ct}\right] + \mathbb{E}\left[Y_c^P v_{fj}\right] + \mathbb{E}\left[v_{ct} v_{fj}\right],
\end{aligned} \tag{A24}
$$

and

$$
\begin{aligned}
\mathbb{E}\left[\left(\tilde{y}_f^P\right)^2\right] &= \mathbb{E}\left[\left(Y_{ft}\right)^2\right] - \mathbb{E}\left[Y_{ft}\right]^2 \\
&= \mathbb{E}\left[\left(Y_f^P + v_{ft}\right)^2\right] - \mathbb{E}\left[\left(Y_f^P + v_{ft}\right)\right]^2 \\
&= \mathbb{E}\left[\left(Y_f^P\right)^2\right] + \mathbb{E}\left[\left(v_{ft}\right)^2\right] + 2\mathbb{E}\left[Y_f^P v_{ft}\right] - \mathbb{E}\left[Y_f^P\right]^2 \\
&= \mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right] + \mathbb{E}\left[v_{ft}^2\right] + 2\mathbb{E}\left[Y_f^P v_{ft}\right]
\end{aligned} \tag{A25}
$$

Then, by equation (A24)

$$
\begin{aligned}
\mathbb{E}_n\left[\tilde{y}_c^P \tilde{y}_f^P\right] &\xrightarrow{p} \mathbb{E}\left[\tilde{y}_c^P \tilde{y}_f^P\right] \\
&= \beta_0 \mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right] + \mathbb{E}\left[Y_f^P v_{ct}\right] + \mathbb{E}\left[Y_c^P v_{fj}\right] + \mathbb{E}\left[v_{ct} v_{fj}\right],
\end{aligned}
$$

and by equation (A25)

$$
\begin{aligned}
\mathbb{E}_n\left[\left(\tilde{y}_f^P\right)^2\right] &\xrightarrow{p} \mathbb{E}\left[\left(\tilde{y}_f^P\right)^2\right] \\
&= \mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right] + \mathbb{E}\left[v_{ft}^2\right] + 2\mathbb{E}\left[Y_f^P v_{ft}\right],
\end{aligned}
$$

so that, under Assumption 1-NS, equation (A22) boils down to

$$
\hat{\beta}_n^{MI} \xrightarrow{p} \frac{\beta_0 \mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right] + \mathbb{E}\left[Y_f^P v_{ct}\right] + \mathbb{E}\left[Y_c^P v_{fj}\right] + \mathbb{E}\left[v_{ct} v_{fj}\right]}{\mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right] + \mathbb{E}\left[v_{ft}^2\right] + 2\mathbb{E}\left[Y_f^P v_{ft}\right]}, \tag{A26}
$$

which is equation (2) in Nybom and Stuhler (2016).

We now turn to the second inconsistency result, which in contrast to Assumption 1-MI, assumes

Assumption 1'-NS. *(Annual Income Process) The relationship between annual and lifetime income is governed by*

$$Y_{gt} = \lambda_t Y_g^P + v_{gt}, \quad \mathbb{E}\left[v_{gt}\left(1, Y_g^P\right)'\right] = 0, \quad g \in \{c, f\}, \quad t = 1, ..., T.$$

where λ_t captures that the persistence of lifetime income may vary over the life-cycle period, and v_{gt} is an age shock, uncorrelated by construction with Y_g^P. Moreover, we further assume $\mathcal{M}_c = t_c$, $\mathcal{M}_f = t_f$, $D_{ct} = 1$ for $t = t_c$, and $D_{ft} = 1$ for $t = t_f$.

That is, similar to Assumption 1-MI, we consider the linear projection of Y_g^P on Y_{gt} so that v_{gt} is uncorrelated to Y_g^P by construction. Moreover, we relax the assumption of $\lambda_t = 1, \forall t \in \mathcal{M}_g, g \in \{c, f\}$ in Assumption 1-MI. Thus, we have that

$$\begin{aligned}
\mathbb{E}\left[\tilde{y}_c^P \tilde{y}_f^P\right] &= \mathbb{E}\left[Y_{ct} Y_{fj}\right] - \mathbb{E}\left[Y_{ct}\right] \mathbb{E}\left[Y_{fj}\right] \\
&= \mathbb{E}\left[\left(\lambda_t Y_c^P + v_{ct}\right)\left(\lambda_j Y_f^P + v_{fj}\right)\right] - \mathbb{E}\left[\left(\lambda_t Y_c^P + v_{ct}\right)\right] \mathbb{E}\left[\left(\lambda_j Y_f^P + v_{fj}\right)\right] \\
&= \lambda_t \lambda_j \mathbb{E}\left[Y_c^P Y_f^P\right] + \lambda_j \mathbb{E}\left[Y_f^P v_{ct}\right] + \lambda_t \mathbb{E}\left[Y_c^P v_{fj}\right] + \mathbb{E}\left[v_{ct} v_{fj}\right] - \lambda_t \mathbb{E}\left[Y_c^P\right] \lambda_j \mathbb{E}\left[Y_f^P\right] \\
&= \lambda_t \lambda_j \mathbb{E}\left[\left(Y_c^P - \mathbb{E}\left[Y_c^P\right]\right)\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)\right] + \lambda_j \mathbb{E}\left[Y_f^P v_{ct}\right] + \lambda_t \mathbb{E}\left[Y_c^P v_{fj}\right] + \mathbb{E}\left[v_{ct} v_{fj}\right] \\
&= \lambda_t \lambda_j \mathbb{E}\left[\left(\beta_0\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right) + u\right)\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)\right] \\
&+ \lambda_j \mathbb{E}\left[Y_f^P v_{ct}\right] + \lambda_t \mathbb{E}\left[Y_c^P v_{fj}\right] + \mathbb{E}\left[v_{ct} v_{fj}\right] \\
&= \beta_0 \lambda_t \lambda_j \mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right] + \lambda_j \mathbb{E}\left[Y_f^P v_{ct}\right] + \lambda_t \mathbb{E}\left[Y_c^P v_{fj}\right] + \mathbb{E}\left[v_{ct} v_{fj}\right],
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}\left[\left(\tilde{y}_f^P\right)^2\right] &= \mathbb{E}\left[\left(\lambda_j Y_f^P + v_{fj}\right)^2\right] - \mathbb{E}\left[\left(\lambda_j Y_f^P + v_{fj}\right)\right]^2 \\
&= \lambda_j^2 \mathbb{E}\left[\left(Y_f^P\right)^2\right] + 2\lambda_j \mathbb{E}\left[Y_f^P v_{fj}\right] + \mathbb{E}\left[v_{fj}^2\right] - \lambda_j^2 \mathbb{E}\left[Y_f^P\right]^2 \\
&= \lambda_j^2 \mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right] + \mathbb{E}\left[v_{fj}^2\right]
\end{aligned}$$

so that

$$\begin{aligned}
\mathbb{E}_n\left[\tilde{Y}_c^P \tilde{Y}_f^P\right] &\xrightarrow{p} \mathbb{E}\left[\tilde{Y}_c^P \tilde{Y}_f^P\right] \\
&= \beta_0 \lambda_t \lambda_j \mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right] + \lambda_j \mathbb{E}\left[Y_f^P v_{ct}\right] + \lambda_t \mathbb{E}\left[Y_c^P v_{fj}\right] + \mathbb{E}\left[v_{ct} v_{fj}\right],
\end{aligned}$$

and

$$\mathbb{E}_n\left[\left(\tilde{Y}_f^P\right)^2\right] \xrightarrow{p} \mathbb{E}\left[\left(\tilde{Y}_f^P\right)^2\right]$$
$$= \lambda_j^2\mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right] + \mathbb{E}\left[v_{fj}^2\right].$$

Thus, under Assumption 1'-NS we have that equation (A22) becomes

$$\hat{\beta}_n^{MI} \xrightarrow{p} \frac{\beta_0\lambda_t\lambda_j\mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right] + \lambda_j\mathbb{E}\left[Y_f^P v_{ct}\right] + \lambda_t\mathbb{E}\left[Y_c^P v_{fj}\right] + \mathbb{E}\left[v_{ct}v_{fj}\right]}{\lambda_j^2\mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right] + \mathbb{E}\left[v_{fj}^2\right]}, \tag{A27}$$

which is equation (6) in Nybom and Stuhler (2016).

A7. *The Life-cycle Estimator*

The life-cycle (LC) estimator, introduced by Mello et al. (2024), addresses one source of bias of the MI estimator: the age shocks in children's income to be correlated with parental lifetime income. Affluent families systematically exhibit faster income growth even after controlling for observables, causing age shocks (prediction errors) in children's lifetime income to correlate with parental income (Y_f^P). This correlation not only reduces efficiency but also biases conventional intergenerational elasticity estimates. Consequently, measurement error in the child's lifetime income not only affects efficiency but also introduces bias in the intergenerational elasticity estimate.

The LC estimator tackles this issue by introducing two key innovations. First, it leverages the availability of family characteristics to decorrelate the prediction error (age shock) in child income (captured by v_{ct}) with parental lifetime income $\left(Y_f^P\right)$. Second, it estimates children's lifetime income by: (i) predicting their annual income at each age using family characteristics, and (ii) summing these predicted values (converted into absolute income) according to the lifetime income definition. As a result, the estimator eliminates bias components (c) and (d) in equation (3), providing more accurate estimates of the intergenerational elasticity.

In line with the discussion in Appendix A1 regarding lifetime income definitions, we adopt the sum of log annual incomes ($Y_g^P = \sum_{t=1}^T Y_{gt}$) rather than the log of summed incomes ($Y_g^{PL} = \log(\sum_t e^{Y_{gt}})$) for our analysis of the life-cycle estimator. This choice serves two purposes: first, it ensures comparability with the MI estimator, so both methods share the same lifetime income definition—a practical necessity when illustrating how the life-cycle estimator improves upon the MI estimator. Second, the practical differences between definitions appear limited in our application: the IGE estimates derived from Y_c^{PL} (log-sum) and Y_c^P (sum-of-logs) are nearly identical for the life-cycle estimator.

To illustrate their proposal, consider the following relationship between annual income and observables

$$Y_{ct} = m(\boldsymbol{X}_{ct}) + \epsilon_{ct}, \quad \mathbb{E}[\epsilon_{ct} \mid \boldsymbol{X}_{ct}] = 0, \tag{A28}$$

where, $\boldsymbol{X}_{ct}$ are family characteristics predictive of child income in year t, $m(\cdot)$ is the conditional mean of (log) annual income given $\boldsymbol{X}_{ct}$, and ϵ_{ct}[19] is the prediction error (or age shock), assumed to be mean independent of the observed family characteristics. For instance, if $\boldsymbol{X}_{ct}$ includes parental lifetime income, the prediction error in children's income would be mean independent of parental lifetime income by construction. As we will show, this implies that component (c) in equation (3) is equal to zero, and consequently, its product with (d) also vanishes, thereby eliminating both sources of bias. Accordingly, controlling for the interaction between parental income and child age reduces the correlation between the prediction error in child income with parental income.

The central distinction between the LC and MI estimands lies in how children's lifetime income is measured. While the MI approach uses average (log) income observed during mid-life, the LC estimator instead relies on the average of predicted (log) annual income over the life cycle. Formally, the LC estimand $\left(\beta^{LC}\right)$ corresponds to the slope coefficient of the projection of the child's estimated (log) lifetime income on the parental average (log) income during mid-life

$$\begin{aligned} \overline{Y}_c^P &= \alpha^{LC} + \beta^{LC}\tilde{Y}_f^P + u^{LC}, \quad \mathbb{E}\left[u^{LC}\tilde{Y}_f^P\right] = 0, \\ \overline{Y}_c^P &:= \sum_{t=1}^{T} \mathbb{E}[Y_{ct}|\boldsymbol{X}_{ct}], \quad \tilde{Y}_f^P := \frac{1}{T_f}\sum_{j\in\mathcal{M}_f} Y_{fj}D_{fj}, \end{aligned} \tag{A29}$$

where $D_{fj} = 1$ when Y_{fj} is observed and zero otherwise,[20] $\mathcal{M}_f$ is a set of pre-defined mid-life years for the father generation, and $T_f := \sum_{j\in\mathcal{M}_f} D_{fj}$ is the number of years used to average parental log annual income.

Applying the same logic as in equation (2), the LC estimand is expressed as follows:

$$\beta^{LC} = \frac{\mathbb{E}\left[\left(\overline{Y}_c^P - \mathbb{E}\left[\overline{Y}_c^P\right]\right)\left(\tilde{Y}_f^P - \mathbb{E}\left[\tilde{Y}_f^P\right]\right)\right]}{\mathbb{E}\left[\left(\tilde{Y}_f^P - \mathbb{E}\left[\tilde{Y}_f^P\right]\right)^2\right]}, \tag{A30}$$

[19]Although the prediction error ϵ_{ct} in equation (A28) can be interpreted as an age shock, it differs conceptually from the age shock v_{ct} in the GEIV model of equation (A1). Specifically, ϵ_{ct} captures the component of (log) annual income that is not explained by observed parental and own characteristics, that is, the residual from a predictive model based on observables. In contrast, v_{ct} reflects transitory deviations from an individual's lifetime income and arises within a latent factor structure that distinguishes between lifetime and transitory components of income.

[20]While $D_{fj} = 1$ is formally defined as indicating when Y_{fj} is observed, in practice, it also implicitly requires that Y_{fj} is used for estimation. This distinction arises because empirical studies often sample parental income selectively—for example, by focusing on log annual income during midlife (e.g., ages 30–50) to reduce lifecycle bias or measurement error. Thus, even if income is observed in other years, it may be excluded from estimation due to sampling design. This refinement clarifies that D_{fj} reflects both data availability and inclusion criteria, ensuring consistency with standard empirical approaches.

with the corresponding sample analogue (LC estimator) given by:

$$\text{(A31)} \qquad \hat{\beta}_n^{LC} = \frac{\mathbb{E}_n\left[\left(\hat{\tilde{Y}}_c^P - \mathbb{E}_n\left[\hat{\tilde{Y}}_c^P\right]\right)\left(\tilde{Y}_f^P - \mathbb{E}_n\left[\tilde{Y}_f^P\right]\right)\right]}{\mathbb{E}_n\left[\left(\tilde{Y}_f^P - \mathbb{E}_n\left[\tilde{Y}_f^P\right]\right)^2\right]}, \qquad \hat{\tilde{Y}}_c^P := \sum_{t=1}^{T} \hat{\mathbb{E}}\left[Y_{ct} | \boldsymbol{X}_{ct}, D_{ct} = 1\right]$$

where $\hat{\mathbb{E}}$ is an estimate of the conditional mean.

Similar to the MI estimand, we now state the underlying assumptions of the life-cycle approach. By leveraging family characteristics, the former relaxes the GEIV assumption for the children's generation, while maintaining it for the father generation.

Assumption 1-LC. *(Annual Income Process) For the fathers' generation, the relationship between annual and lifetime income remains as in Assumption 1-MI:*

$$\begin{aligned} &Y_{ft} = \lambda_t Y_f^P + v_{ft}, \quad \mathbb{E}\left[v_{ft} Y_f^P\right] = 0, \quad t = 1, \ldots, T, \\ &\lambda_t = 1, \quad \forall t \in \mathcal{M}_f. \end{aligned}$$

The model in equation (A29), together with Assumption 1-LC, involves a random, independent, and identically distributed sample of $W = \left(\boldsymbol{Y}_c \odot D_c, \boldsymbol{Y}_f \odot D_f, D_c, D_f, \boldsymbol{X}\right)$, unobserved components including: (i) the age shocks v_{ft} for the father's generation, indexed by $t = 1, \ldots, T$; (ii) the prediction errors ϵ_{ct} for the children's generation, indexed by $t = 1, \ldots, T$; (iii) the lifetime income Y_g^P for both generations $g \in c, f$; and (iv) the error term u^{LC} associated with the LC estimator. The model parameters consist of the parameter of interest β^{LC} and the nuisance parameter α^{LC}.

The following assumption formally sets out the conditional mean independence and orthogonality conditions required for the LC estimator to improve upon the MI estimator. As in Assumption 2-MI, parental age shocks during mid-life are assumed to be mean independent of children's lifetime income, conditional on the missingness status of parental income in that year. The second orthogonality condition in Assumption 2-MI is replaced by the assumption that the prediction error in children's annual income is orthogonal to parental income during mid-life. In contrast, the first and fourth orthogonality conditions from Assumption 2-MI are no longer imposed.

Assumption 2-LC. *(Conditional Mean Independence and Orthogonality)*

i. The following condition holds

$$\mathbb{E}\left[v_{fj} Y_c^P \middle| D_{fj}\right] = 0, \quad j \in \mathcal{M}_f.$$

ii. The prediction error of the children's annual income is uncorrelated to parental lifetime income

$$\mathbb{E}\left[\epsilon_{ct} Y_f^P\right] = 0, \quad t = 1, ..., T,$$

$$Y_{ct} = \mathbb{E}\left[Y_{ct} \mid \boldsymbol{X}_{ct}\right] + \epsilon_{ct}, \quad \mathbb{E}\left[\epsilon_{ct} \mid \boldsymbol{X}_{ct}\right] = 0, \quad t = 1, ..., T.$$

Assumption 2-LC.*ii* is the key to improving upon the MI estimator. By projecting children's annual income into the space of observables, the prediction error in child income is decorrelated with parental income, so that component (c), and its product with component (d), in equation (3) vanish.[21]

Because we do not observe complete income profiles, we require an additional unconfoundedness assumption: conditional on family characteristics, the observation of children's annual income must be independent of the income level itself. This assumption enables identification of the conditional expectation through the observed data, allowing us to estimate $\mathbb{E}\left[Y_{ct} \mid \boldsymbol{X}_{ct}\right]$ with $\mathbb{E}\left[Y_{ct} \mid \boldsymbol{X}_{ct}, D_{ct} = 1\right]$, which is identified since we observe family characteristics and snapshots of children's annual income. Furthermore, since this conditional mean has the equivalent representation

$$\mathbb{E}\left[Y_{ct} \mid \boldsymbol{X}_{ct}, D_{ct} = 1\right] = \frac{\mathbb{E}\left[Y_{ct} D_{ct} \mid \boldsymbol{X}_{ct}\right]}{p\left(D_{ct} = 1 \mid \boldsymbol{X}_{ct}\right)},$$

we also assume boundedness of the propensity score $p\left(D_{ct} = 1 \mid \boldsymbol{X}_{ct}\right)$ to ensure the denominator is well-defined and the estimator remains stable.

Assumption 3-LC. *(Missing at Random)*

i. The missingness of children's annual income Y_{ct} is as good as random once we control for $\boldsymbol{X}_{ct}$

$$Y_{ct} \perp D_{ct} \mid \boldsymbol{X}_{ct}, \quad t = 1, ..., T.$$

ii. Given family characteristics, there is both missing and non-missing children incomes for every age

$$0 < p\left(D_{ct} = 1 \mid \boldsymbol{X}_{ct}\right) < 1 \quad a.s \quad t = 1, ..., T.$$

The final assumption concerns the estimation of the age-income profiles. Since the LC estimator in equation (A31) relies on estimating $\mathbb{E}\left[Y_{ct} \mid \boldsymbol{X}_{ct}, D_{ct} = 1\right]$, we assume this conditional expectation is consistently estimated in order to characterize the probability limit of the LC estimator.

Assumption 4-LC. *(Consistent Estimation of Age-Income Profiles) The expected value of annual child's income conditional on family characteristics is consistently estimated for the observed income snapshots*

$$\hat{\mathbb{E}}\left[Y_{ct} | \boldsymbol{X}_{ct}, D_{ct} = 1\right] \xrightarrow{p} \mathbb{E}\left[Y_{ct} | \boldsymbol{X}_{ct}, D_{ct} = 1\right].$$

[21] Since the LC estimator measures parental lifetime income as average mid-life (log) income, Assumption 2-LC.*i* can be relaxed to $\mathbb{E}\left[\epsilon_{ct} Y_f^P\right] = 0$ for $t = 1, ..., T$, and $j \in \mathcal{M}_f$.

A8. Proof of Corollary 1.2

We aim to characterize the limit in probability of the MI estimator, defined as

$$\hat{\beta}_n^{LC} = \frac{\mathbb{E}_n\left[\left(\hat{\tilde{Y}}_c^P - \mathbb{E}_n\left[\hat{\tilde{Y}}_c^P\right]\right)\left(\tilde{Y}_f^P - \mathbb{E}_n\left[\tilde{Y}_f^P\right]\right)\right]}{\mathbb{E}_n\left[\left(\tilde{Y}_f^P - \mathbb{E}_n\left[\tilde{Y}_f^P\right]\right)^2\right]}, \quad \hat{\tilde{Y}}_c^P := \sum_{t=1}^{T} \hat{\mathbb{E}}\left[Y_{ct}|\boldsymbol{X}_{ct}, D_{ct} = 1\right].$$

Focusing on the numerator, we have

$$\begin{aligned}
\mathbb{E}\left[\bar{y}_c^P \tilde{y}_f^P\right] &= \mathbb{E}\left[\left(\sum_{t=1}^{T} \mathbb{E}\left[Y_{ct}|\boldsymbol{X}_{ct}, D_{ct} = 1\right]\right)\left(\frac{1}{T_f}\sum_{j\in\mathcal{M}_f} Y_{fj}D_{fj}\right)\right] \\
&\quad - \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{E}\left[Y_{ct}|\boldsymbol{X}_{ct}, D_{ct} = 1\right]\right]\mathbb{E}\left[\left(\frac{1}{T_f}\sum_{j\in\mathcal{M}_f} Y_{fj}D_{fj}\right)\right]. \qquad \text{(A32)}
\end{aligned}$$

We first analyze the first term of the last display

$$\begin{aligned}
&\mathbb{E}\left[\sum_{t=1}^{T} \mathbb{E}\left[Y_{ct}|\boldsymbol{X}_{ct}, D_{ct} = 1_t\right] \frac{1}{T_f}\sum_{j\in\mathcal{M}_f} Y_{fj}D_{fj}\right] \\
&= \frac{1}{T_f}\sum_{t=1}^{T}\sum_{j\in\mathcal{M}_f} \mathbb{E}\left[\mathbb{E}\left[Y_{ct}|\boldsymbol{X}_{ct}, D_{ct} = 1_t\right] Y_{fj}D_{fj}\right] \\
&= \frac{1}{T_f}\sum_{t=1}^{T}\sum_{j\in\mathcal{M}_f} \mathbb{E}\left[\mathbb{E}\left[Y_{ct}\boldsymbol{X}_{ct}\right] Y_{fj}D_{fj}\right] \\
&= \frac{1}{T_f}\sum_{t=1}^{T}\sum_{j\in\mathcal{M}_f} \mathbb{E}\left[\left(Y_{ct} - \epsilon_{ct}\right) Y_{fj}D_{fj}\right] \\
&= \frac{1}{T_f}\sum_{t=1}^{T}\sum_{j\in\mathcal{M}_f} \mathbb{E}\left[Y_{ct}Y_{fj}D_{fj}\right] \\
&= \frac{1}{T_f}\sum_{j\in\mathcal{M}_f} \mathbb{E}\left[Y_{fj}D_{fj}\left(\sum_{t=1}^{T} Y_{ct}\right)\right] \\
&= \frac{1}{T_f}\sum_{j\in\mathcal{M}_f} \mathbb{E}\left[Y_{fj}D_{fj}Y_c^P\right] \\
&= \mathbb{E}\left[\frac{1}{T_f}\sum_{j\in\mathcal{M}_f} \left(Y_f^P + v_{fj}\right) D_{fj}Y_c^P\right]
\end{aligned}$$

$$= \mathbb{E}\left[Y_f^P Y_c^P \frac{1}{T_f} \sum_{j \in \mathcal{M}_f} D_{fj} + \frac{1}{T_f} \sum_{j \in \mathcal{M}_f} v_{fj} D_{fj} Y_c^P\right]$$

$$= \mathbb{E}\left[Y_f^P Y_c^P\right] + \frac{1}{T_f} \sum_{j \in \mathcal{M}_f} \mathbb{E}\left[v_{fj} D_{fj} Y_c^P\right]$$

$$= \mathbb{E}\left[Y_f^P Y_c^P\right] \tag{A33}$$

Where the second equality follows by Assumption 3-LC, and the third by the definition of the prediction error in Assumption 2-LC. The fourth equality follows by Assumption 2-LC. The seventh equality follows by Assumption 1-MI. The ninth equality follows by the definition of T_f, and the last one by Assumption 2-LC and the LIE, since

$$\mathbb{E}\left[v_{fj} D_{fj} Y_c^P\right] = \mathbb{E}\left[\mathbb{E}\left[v_{fj} Y_c^P | D_{fj}\right] D_{fj}\right] = 0.$$

As regards the second term in equation (A32), we have

$$\mathbb{E}\left[\sum_{t=1}^{T} \mathbb{E}\left[Y_{ct} | \boldsymbol{X}_{ct}, D_{ct} = 1\right]\right] \mathbb{E}\left[\left(\frac{1}{T_f} \sum_{j \in \mathcal{M}_f} Y_{fj} D_{fj}\right)\right] = \sum_{t=1}^{T} \mathbb{E}\left[\mathbb{E}\left[Y_{ct} \mid \boldsymbol{X}_{ct}\right]\right] \mathbb{E}\left[Y_f^P\right]$$

$$= \sum_{t=1}^{T} \mathbb{E}\left[Y_{ct}\right] \mathbb{E}\left[Y_f^P\right]$$

$$= \mathbb{E}\left[\sum_{t=1}^{T} Y_{ct}\right] \mathbb{E}\left[Y_f^P\right]$$

$$= \mathbb{E}\left[Y_c^P\right] \mathbb{E}\left[Y_f^P\right]. \tag{A34}$$

Thus, by plugging equations (A33) and (A34) into (A32) we have

$$\mathbb{E}\left[\bar{y}_c^P \tilde{y}_f^P\right] = \mathbb{E}\left[Y_f^P Y_c^P\right] - \mathbb{E}\left[Y_c^P\right] \mathbb{E}\left[Y_f^P\right]$$

$$= \mathbb{E}\left[\left(Y_c^P - \mathbb{E}\left[Y_c^P\right]\right)\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)\right]$$

$$= \beta_0 \mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right]. \tag{A35}$$

We now establish the limit in probability of the numerator in equation (A31)

$$\mathbb{E}_n\left[\hat{\bar{y}}_c^P \tilde{y}_f^P\right] = \mathbb{E}_n\left[\left(\sum_{t=1}^{T} \hat{\mathbb{E}}\left[Y_{ct} | \boldsymbol{X}_{ct}, D_{ct} = 1\right]\right)\left(\frac{1}{T_f} \sum_{j \in \mathcal{M}_f} Y_{fj} D_{fj}\right)\right]$$

$$- \mathbb{E}_n\left[\sum_{t=1}^{T} \hat{\mathbb{E}}\left[Y_{ct} | \boldsymbol{X}_{ct}, D_{ct} = 1\right]\right] \mathbb{E}_n\left[\left(\frac{1}{T_f} \sum_{j \in \mathcal{M}_f} Y_{fj} D_{fj}\right)\right]$$

$$\begin{aligned}
&\xrightarrow{p} \mathbb{E}_n\left[\left(\sum_{t=1}^{T}\mathbb{E}\left[Y_{ct}|\boldsymbol{X}_{ct}, D_{ct}=1\right]\right)\left(\frac{1}{T_f}\sum_{j\in\mathcal{M}_f}Y_{fj}D_{fj}\right)\right] \\
&- \mathbb{E}_n\left[\sum_{t=1}^{T}\mathbb{E}\left[Y_{ct}|\boldsymbol{X}_{ct}, D_{ct}=1\right]\right]\mathbb{E}_n\left[\left(\frac{1}{T_f}\sum_{j\in\mathcal{M}_f}Y_{fj}D_{fj}\right)\right] \\
&\xrightarrow{p} \mathbb{E}\left[\left(\sum_{t=1}^{T}\mathbb{E}\left[Y_{ct}|\boldsymbol{X}_{ct}, D_{ct}=1\right]\right)\left(\frac{1}{T_f}\sum_{j\in\mathcal{M}_f}Y_{fj}D_{fj}\right)\right] \\
&- \mathbb{E}\left[\sum_{t=1}^{T}\mathbb{E}\left[Y_{ct}|\boldsymbol{X}_{ct}, D_{ct}=1\right]\right]\mathbb{E}\left[\left(\frac{1}{T_f}\sum_{j\in\mathcal{M}_f}Y_{fj}D_{fj}\right)\right] \\
&= \mathbb{E}\left[\bar{y}_c^P\tilde{y}_f^P\right] \\
&= \beta_0\mathbb{E}\left[\left(Y_f^P-\mathbb{E}\left[Y_f^P\right]\right)^2\right], \qquad \text{(A36)}
\end{aligned}$$

where the first limit in probability follows by Assumption 4-LC and the CMT, the second from the LLN, and the last equality from equation (A35).

Finally, by equations (A21) and (A36), and the CMT, we have

(A37)

$$\hat{\beta}_n^{LC} \xrightarrow{p} \frac{\beta_0\mathbb{E}\left[\left(Y_f^P-\mathbb{E}\left[Y_f^P\right]\right)^2\right]}{\mathbb{E}\left[\left(Y_f^P-\mathbb{E}\left[Y_f^P\right]\right)^2\right]+\frac{1}{T_f^2}\sum_t\sum_j\mathbb{E}\left[v_{ft}v_{fj}\middle|D_{ft}=1, D_{fj}=1, \{t,j\}\in\mathcal{M}_f\right]\times p_f\left(\{t,j\}\in\mathcal{M}_f\right)}. Q.E.D.$$

Non-parametric Identification of the IGE and Local Robustness

B1. Proof of Theorem 2

To establish identification of the intergenerational elasticity, defined as

$$\beta_0 = \frac{\mathbb{E}\left[\left(Y_c^P-\mathbb{E}\left[Y_c^P\right]\right)\left(Y_f^P-\mathbb{E}\left[Y_f^P\right]\right)\right]}{\mathbb{E}\left[\left(Y_f^P-\mathbb{E}\left[Y_f^P\right]\right)^2\right]}, \qquad \text{(B1)}$$

we will express both the numerator and denominator into observable components. We start by showing identification of the conditional means of lifetime income for both generations $g\in\{c,p\}$:

$$\mathbb{E}\left[Y_g^P\right]=\mathbb{E}\left[\sum_{t=1}^{T}Y_{gt}\right]$$

$$
\begin{aligned}
&= \sum_{t=1}^{T} \mathbb{E}\left[Y_{gt}\right] \\
&= \sum_{t=1}^{T} \mathbb{E}\left[\mathbb{E}\left[Y_{gt} \mid \boldsymbol{X}_{gt}\right]\right] \\
&= \sum_{t=1}^{T} \mathbb{E}\left[\mathbb{E}\left[Y_{gt} \mid \boldsymbol{X}_{gt}, D_{gt} = 1\right]\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{E}\left[Y_{gt} \mid \boldsymbol{X}_{gt}, D_{gt} = 1\right]\right] \\
&:= \mu_g^P. \qquad \text{(B2)}
\end{aligned}
$$

Turning attention to the numerator in equation (B1), we exploit Assumptions 1-NP and 2-NP to express this term as:

$$
\begin{aligned}
\mathbb{E}\left[\left(Y_c^P - \mathbb{E}\left[Y_c^P\right]\right)\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)\right] &:= \mathbb{E}\left[\left(Y_c^P - \mu_c^P\right)\left(Y_f^P - \mu_f^P\right)\right] \\
&= \sum_{t=1}^{T}\sum_{j=1}^{T} \mathbb{E}\left[\left(Y_{ct} - \mu_c^P\right)\left(Y_{fj} - \mu_f^P\right)\right] \\
&= \sum_{t=1}^{T}\sum_{j=1}^{T} \mathbb{E}\left[\left(\mathbb{E}\left[Y_{ct} \mid \boldsymbol{X}_{ct}\right] + \epsilon_{ct} - \mu_c^P\right)\left(Y_{fj} - \mu_f^P\right)\right] \\
&= \sum_{t=1}^{T}\sum_{j=1}^{T} \mathbb{E}\left[\left(\mathbb{E}\left[Y_{ct} \mid \boldsymbol{X}_{ct}\right] - \mu_c^P\right)\left(Y_{fj} - \mu_f^P\right)\right] \\
&= \sum_{t=1}^{T}\sum_{j=1}^{T} \mathbb{E}\left[\left(\mathbb{E}\left[Y_{ct} \mid \boldsymbol{X}_{ct}, \boldsymbol{X}_{fj}\right] - \mu_c^P\right)\left(Y_{fj} - \mu_f^P\right)\right] \\
&= \sum_{t=1}^{T}\sum_{j=1}^{T} \mathbb{E}\left[\left(\mathbb{E}\left[Y_{ct} \mid \boldsymbol{X}_{ct}, \boldsymbol{X}_{fj}\right] - \mu_c^P\right)\left(\mathbb{E}\left[Y_{fj} \mid \boldsymbol{X}_{ct}, \boldsymbol{X}_{fj}\right] - \mu_f^P\right)\right] \\
&= \sum_{t=1}^{T}\sum_{j=1}^{T} \mathbb{E}\left[\left(\mathbb{E}\left[Y_{ct} \mid \boldsymbol{X}_{ct}\right] - \mu_c^P\right)\left(\mathbb{E}\left[Y_{fj} \mid \boldsymbol{X}_{fj}\right] - \mu_f^P\right)\right] \\
&= \sum_{t=1}^{T}\sum_{j=1}^{T} \mathbb{E}\left[\left(\mathbb{E}\left[Y_{ct} \mid \boldsymbol{X}_{ct}, D_{ct} = 1\right] - \mu_c^P\right)\left(\mathbb{E}\left[Y_{fj} \mid \boldsymbol{X}_{fj}, D_{fj} = 1\right] - \mu_f^P\right)\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T}\sum_{j=1}^{T} \left(\mathbb{E}\left[Y_{ct} \mid \boldsymbol{X}_{ct}, D_{ct} = 1\right] - \mu_c^P\right)\left(\mathbb{E}\left[Y_{fj} \mid \boldsymbol{X}_{fj}, D_{fj} = 1\right] - \mu_f^P\right)\right] \\
&:= \mathbb{E}\left[\sum_{t=1}^{T}\left(\mu_{ct}\left(\boldsymbol{X}_{ct}, 1\right) - \mu_c^P\right)\sum_{j=1}^{T}\left(\mu_{fj}\left(\boldsymbol{X}_{fj}, 1\right) - \mu_f^P\right)\right], \qquad \text{(B3)}
\end{aligned}
$$

where $\mu_{gt}(\boldsymbol{X}_{gt}, 1) := \mathbb{E}\left[Y_{gt} \mid \boldsymbol{X}_{gt}, D_{gt} = 1\right]$ for $g \in \{c, p\}$.

The first equality follows from the definition of the conditional means of lifetime income in equation (B2), while the second equality follows from the definition of lifetime income together with the linearity of expectation. The prediction error of the children's annual income in the third equality is defined according to Assumption 1-NP.*ii*. The fourth equality leverages the same assumption, which ensures that this error is uncorrelated with parental lifetime income, and has zero mean. The fifth equality exploits Assumption 1-NP.*i*, ensuring $\mathbb{E}\left[Y_{ct} \mid \boldsymbol{X}_{ct}, \boldsymbol{X}_{fj}\right] = \mathbb{E}\left[Y_{ct} \mid \boldsymbol{X}_{t}\right]$, while the sixth applies the same argument along with the law of iterated expectations. The seventh equality also uses Assumption 1-NP.*i*, while the eigth equality follows from Assumption 2-NP.*i* and Assumption 2-NP.*iii*, with the eight one resulting from another application of the linearity of expectations.

Turning to the denominator of equation (B1), we have

$$\begin{aligned}
\mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right] &:= \mathbb{E}\left[\left(Y_f^P - \mu_f^P\right)^2\right] \\
&= \sum_{t=1}^{T}\sum_{j=1}^{T} \mathbb{E}\left[\left(Y_{ft} - \mu_f^P\right)\left(Y_{fj} - \mu_f^P\right)\right] \\
&= \sum_{|t-j|\leq h} \mathbb{E}\left[\mathbb{E}\left[\left(Y_{ft} - \mu_f^P\right)\left(Y_{fj} - \mu_f^P\right) \mid \boldsymbol{X}_{ftj}\right]\right] \\
&= \sum_{|t-j|\leq h} \mathbb{E}\left[\mathbb{E}\left[\left(Y_{ft} - \mu_f^P\right)\left(Y_{fj} - \mu_f^P\right) \mid \boldsymbol{X}_{ftj}\right]\right] \\
&\quad + \sum_{|t-j|> h} \mathbb{E}\left[\mathbb{E}\left[\left(Y_{ft} - \mu_f^P\right)\left(Y_{fj} - \mu_f^P\right) \mid \boldsymbol{X}_{ftj}\right]\right], \qquad \text{(B4)}
\end{aligned}$$

where the third equality follows from LIE. Focusing on the second term of the last equality, we have

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{E}\left[\left(Y_{ft} - \mu_f^P\right)\left(Y_{fj} - \mu_f^P\right) \mid \boldsymbol{X}_{ftj}\right]\right] \\
=&\mathbb{E}\left[\mathbb{E}\left[\left(\mathbb{E}\left[Y_{ft}|\boldsymbol{X}_{ft}\right] + \epsilon_{ft} - \mu_f^P\right)\left(\mathbb{E}\left[Y_{fj}|\boldsymbol{X}_{fj}\right] + \epsilon_{fj} - \mu_f^P\right) \mid \boldsymbol{X}_{ftj}\right]\right] \\
=&\mathbb{E}\left[\left(\mathbb{E}\left[Y_{ft}|\boldsymbol{X}_{ft}\right] - \mu_f^P\right)\left(\mathbb{E}\left[Y_{fj}|\boldsymbol{X}_{fj}\right] - \mu_f^P\right)\right] + \mathbb{E}\left[\mathbb{E}\left[\epsilon_{ft}\epsilon_{fj} \mid \boldsymbol{X}_{ftj}\right]\right] \\
=&\mathbb{E}\left[\left(\mathbb{E}\left[Y_{ft}|\boldsymbol{X}_{ft}\right] - \mu_f^P\right)\left(\mathbb{E}\left[Y_{fj}|\boldsymbol{X}_{fj}\right] - \mu_f^P\right)\right] + \mathbb{E}\left[\epsilon_{ft}\epsilon_{fj}\right] \\
=&\left(\mathbb{E}\left[Y_{ft}|\boldsymbol{X}_{ft}\right] - \mu_f^P\right)\left(\mathbb{E}\left[Y_{fj}|\boldsymbol{X}_{fj}\right] - \mu_f^P\right). \qquad \text{(B5)}
\end{aligned}$$

In the second equality we have used Assumption 1-NP.*i*, and the fact that μ_f^P is constant w.r.t $\boldsymbol{X}_{ftj}$. The cross-terms involving ϵ_{ft} and ϵ_{fj} therefore vanish, since $\mathbb{E}\left[\epsilon_{ft} \mid \boldsymbol{X}_{ftj}\right] = \mathbb{E}\left[\epsilon_{ft} \mid \boldsymbol{X}_{ft}\right] = 0$. In the last equality, we have used Assumption 1-NP.*iii*.

Plugging equation (B5) into (B4) identifies the covariance of parental lifetime income:

$$\mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right] = \sum_{|t-j|\leq h} \mathbb{E}\left[\mathbb{E}\left[\left(Y_{ft} - \mu_f^P\right)\left(Y_{fj} - \mu_f^P\right) \mid \boldsymbol{X}_{ftj}\right]\right]$$

$$
\begin{aligned}
&+ \sum_{|t-j|>h} \mathbb{E}\left[\left(\mathbb{E}\left[Y_{ft}|\boldsymbol{X}_{ft}\right] - \mu_f^P\right)\left(\mathbb{E}\left[Y_{fj}|\boldsymbol{X}_{fj}\right] - \mu_f^P\right)\right] \\
&= \sum_{|t-j|\leq h} \mathbb{E}\left[\mathbb{E}\left[\left(Y_{ft} - \mu_f^P\right)\left(Y_{fj} - \mu_f^P\right) \mid \boldsymbol{X}_{ftj}, D_{ft} = 1, D_{fj} = 1\right]\right] \\
&+ \sum_{|t-j|>h} \mathbb{E}\left[\left(\mathbb{E}\left[Y_{ft}|\boldsymbol{X}_{ft}, D_{ft} = 1\right] - \mu_f^P\right)\left(\mathbb{E}\left[Y_{fj}|\boldsymbol{X}_{fj}, D_{fj} = 1\right] - \mu_f^P\right)\right] \\
&:= \mathbb{E}\left[\sum_{|t-j|\leq h} \sigma_{tj}\left(\boldsymbol{X}_{ftj}, 1, 1\right) + \sum_{|t-j|>h} \left(\mu_{ft}\left(\boldsymbol{X}_{ft}, 1\right) - \mu_f^P\right)\left(\mu_{fj}\left(\boldsymbol{X}_{fj}, 1\right) - \mu_f^P\right)\right], \qquad \text{(B6)}
\end{aligned}
$$

where we have defined $\sigma_{tj}\left(\boldsymbol{X}_{ftj}, 1, 1\right) := \mathbb{E}\left[\left(Y_{ft} - \mu_f^P\right)\left(Y_{fj} - \mu_f^P\right) \mid \boldsymbol{X}_{ftj}, D_{ft} = 1, D_{fj} = 1\right]$, and used Assumptions 2-NP.*i* and 2-NP.*iii*.

Finally, plugging equations (B3) and (B6) into B1) yields

$$
\beta_0 = \frac{\mathbb{E}\left[\sum_{t=1}^{T}\left(\mu_{ct}(\boldsymbol{X}_{ct}, 1) - \mu_c^P\right)\sum_{j=1}^{T}\left(\mu_{fj}\left(\boldsymbol{X}_{fj}, 1\right) - \mu_f^P\right)\right]}{\mathbb{E}\left[\sum_{|t-j|\leq h}\sigma_{tj}\left(\boldsymbol{X}_{ftj}, 1, 1\right) + \sum_{|t-j|>h}\left(\mu_{ft}\left(\boldsymbol{X}_{ft}, 1\right) - \mu_f^P\right)\left(\mu_{fj}\left(\boldsymbol{X}_{fj}, 1\right) - \mu_f^P\right)\right]}. Q.E.D.
$$

B2. Locally Robust Moments

Before proposing a locally robust estimator for the IGE, we first illustrate the construction of locally robust moments, as proposed by Chernozhukov et al. (2022). The point of departure is GMM estimation of a parameter of interest θ, which depends on a nuisance parameter $\boldsymbol{\gamma}$, and W, a data observation with unknown cumulative distribution function (CDF) F_0. We assume that there is a known function $g(W, \gamma, \theta)$ of a possible realization w of W, $\boldsymbol{\gamma}$ and θ such that

$$
\mathbb{E}\left[g\left(W, \gamma_0, \theta_0\right)\right] = 0, \qquad \text{(B7)}
$$

where $\mathbb{E}[\cdot]$ is the expectation under F_0 and γ_0 is the probability limit (plim) under F_0 of a first step estimator $\hat{\boldsymbol{\gamma}}$. We also assume that θ_0 is identified by this moment, meaning that θ_0 is the unique solution to (B7) over θ in some set Θ.

Chernozhukov et al. (2022) provide a general procedure to construct orthogonal moment functions for GMM, where first steps have no effect, locally, on average moment functions. In particular, the authors show that an orthogonal (locally robust) moment function (ψ) can be constructed by adding the first step influence function (ϕ) to the identifying moment function (g)

$$
\psi\left(W, \gamma, \alpha, \theta\right) = g\left(W, \gamma, \theta\right) + \phi\left(W, \gamma, \alpha, \theta\right), \qquad \text{(B8)}
$$

where α is a function called the Riesz representer[22] of the functional γ, on which only the first step influence function[23] depends.

The vector of moment functions $\psi(W, \gamma, \alpha, \theta)$ is considered to be locally robust when (i) varying γ away from $\gamma_0 = \gamma(F_0)$ has no local effect on $\mathbb{E}[\psi(W, \gamma, \alpha_0, \theta)]$, and (ii) varying α away from α_0 has no local effect on $\mathbb{E}[\psi(W, \gamma_0, \alpha, \theta)]$, where $\gamma(F)$ is the limit in probability of $\hat{\gamma}$ for a possible CDF of the data W, denoted by F. The first condition is met when the set Γ of possible directions of departure of $\gamma(F)$ of γ_0 satisfy

$$\frac{d}{dt}\mathbb{E}[\psi(W, \gamma_0 + t\delta, \alpha_0, \theta)] = 0 \text{ for all } \delta \in \Gamma, \text{ and } \theta \in \Theta,$$

where t is a scalar, δ is a direction of deviation of $\gamma(b)$ of γ_0, and the derivative is evaluated at $t = 0$. The second condition is met when

$$\mathbb{E}[\phi(W, \gamma_0, \alpha, \theta)] = 0 \text{ for all } \theta \in \theta, \text{ and } \alpha \in \mathcal{A},$$

where the set $\mathcal{A}$ is given by the α_0's satisfying

$$\frac{d}{d\tau}\mathbb{E}[g(W, \gamma(F_\tau), \theta)] = \int \phi(\omega, \gamma_0, \alpha_0, \theta) H(d\omega),$$
$$\mathbb{E}[\phi(W, \gamma_0, \alpha_0, \theta)] = 0, \quad \mathbb{E}\left[\phi(W, \theta_0, \alpha_0, \theta)^2\right] < \infty,$$

for all H and all $\theta \in \Theta$, where H is an alternative distribution of $\boldsymbol{Z}$ different from its true distribution F_0, and $F_\tau = (1-\tau)F_0 + \tau H$ for $\tau \in [0, 1]$, where H is such that $\gamma(F_\tau)$ exists for τ small enough and regularity conditions are met.

B3. A Locally Robust Moment for the IGE in the Presence of Incomplete Income Data

Theorem 2 establishes an identification result for the intergenerational elasticity (IGE). However, estimating this parameter via the plug-in principle, e.g., using machine learning estimators for the conditional means, introduces model selection and regularization bias. To address this issue, we follow Chernozhukov et al. (2022) and construct a debiased machine learning estimator for equation (7). This estimator is based on an orthogonal moment function that corrects for the regularization bias in the estimation of β_0, which arises from the first-step estimation of the conditional expectations in our identification result.

As illustrated in Appendix B2, to find the orthogonal moment function corresponding

[22]We have assumed that θ_0 is identified by equation (B7). Thus, our object of interest can be expressed as $\theta_0 = \mathbb{E}[m(W, \gamma_0)]$. Under a continuity condition, we can express θ_0 as

$$\theta_0 = \mathbb{E}[\gamma_0\alpha_0], \quad \text{for all possible } \gamma_0,$$

where α_0 is called the Riesz representer of the functional γ_0.

[23]The first step influence function gives the effect of $\boldsymbol{\gamma}$ on average identifying moment functions under general misspecification. Therefore, adding the FSIF ($\phi(W, \gamma, \alpha, \theta)$) to the identifying moment $g(W, \gamma, \theta)$, provides an orthogonal moment, where first step estimation of $\boldsymbol{\gamma}$ has no effect, locally, on $\mathbb{E}[g(W, \gamma, \theta)]$.

to equation (8), it suffices to characterize the first step influence function of the identifying moment. To this end, we first define the following conditional expectations:

$$
\begin{aligned}
\mu_{gt}(F_\tau)(z) &:= \mathbb{E}_\tau\left[Y_{gt}|Z_t = z\right], \quad \boldsymbol{Z}_t := \left(\boldsymbol{X}_{gt}, D_{gt}\right), \quad g \in \{c, f\}, \quad t = 1, ..., T,\\
\sigma_{tj}(F_\tau)(z) &:= \mathbb{E}_\tau\left[\left(Y_{ft} - \mu_f^P\right)\left(Y_{fj} - \mu_f^P\right)|Z_{tj} = z\right], \quad \boldsymbol{Z}_{tj} := \left(\boldsymbol{X}_{ftj}, D_{ft}, D_{fj}\right), \quad t, j = 1, ..., T,\\
\mu_g^{1,T}(F_\tau) &:= \left(\mu_{1,t}(F_\tau)(z), ..., \mu_{gt}(F_\tau)(z)\right), \quad g \in \{c, f\},\\
\sigma^{t,1,T}(F_\tau) &:= \left(\sigma_{t,1}(F_\tau)(z), ..., \sigma_{t,T}(F_\tau)(z)\right), \quad t = 1, ..., T\\
\sigma^{1,T,1,T}(F_\tau) &:= \left(\sigma^{1,1,T}(F_\tau)(z), ..., \sigma^{T,1,T}(F_\tau)(z)\right),\\
\gamma(F_\tau) &:= \left(\mu_g^{1,T}(F_\tau), \gamma^{f,1,T}(F_\tau), \gamma^{f,1,T,1,T}(F_\tau)\right),
\end{aligned}
$$

where $\mathbb{E}_\tau$ denotes the expectation under $F_\tau = (1-\tau)F_0 + \tau H$. Thus, equation (7), which identifies our parameter of interest β_0, can be rewritten as

$$
\begin{aligned}
&\mathbb{E}\left[g_1\left(W, \gamma_0, \beta_0, \mu_c^P, \mu_f^P\right)\right] = 0,\\
&g_1\left(W, \gamma_0, \beta_0, \mu_c^P, \mu_f^P\right) = \beta_0 \sum_{|t-j|\le h} \sigma_{tj}(F_0)\left(\boldsymbol{X}_{ftj}, 1, 1\right)\\
&\qquad + \beta_0 \sum_{|t-j|>h} \left(\mu_{ft}(F_0)\left(\boldsymbol{X}_{ft}, 1\right) - \mu_f^P\right)\left(\mu_{fj}(F_0)\left(\boldsymbol{X}_{fj}, 1\right) - \mu_f^P\right)\\
&\qquad - \sum_{t=1}^{T} \left(\mu_{ct}(F_0)(\boldsymbol{X}_{ct}, 1) - \mu_c^P\right) \sum_{j=1}^{T} \left(\mu_{fj}(F_0)\left(\boldsymbol{X}_{fj}, 1\right) - \mu_f^P\right),
\end{aligned}
\tag{B9}
$$

where $\mathbb{E}[\cdot]$ is the expectation under the true distribution of W (F_0) and $\gamma_0 := \gamma(F_0)$ is the probability limit under F_0 of a first step estimator $\hat{\boldsymbol{\gamma}}$. Notice that β_0 also depends on the mean of children and parental income $\left(\mu_c^P, \mu_f^P\right)$. However, according to equation (B2) these two parameters are identified by

$$
\mu_g^P = \mathbb{E}\left[\sum_{t=1}^{T} \mu_{gt}(F_0)\left(\boldsymbol{X}_{gt}, 1\right)\right],
$$

so that the moment identifying μ_c^P can be expressed as

$$
\begin{aligned}
&\mathbb{E}\left[g_2(W, \gamma(F_0), \theta)\right] = 0,\\
&g_2(W, \gamma(F_0), \theta) = \sum_{t=1}^{T} \mu_{ct}(F_0)(\boldsymbol{X}_{ct}, 1) - \mu_c^P,
\end{aligned}
\tag{B10}
$$

and analogously for μ_f^P

$$
\mathbb{E}\left[g_3(W, \gamma(F_0), \theta)\right] = 0,
$$

$$g_3\left(W,\gamma\left(F_0\right),\theta\right)=\sum_{t=1}^{T}\mu_{ft}\left(F_0\right)\left(\boldsymbol{X}_{ft},1\right)-\mu_f^P. \tag{B11}$$

Accordingly, by defining the augmented parameter of interest $\theta_0 := \left(\beta_0,\mu_c^P,\mu_f^P\right)$, the identifying moment is given by

$$g\left(W,\gamma\left(F_\tau\right),\theta\right)=\begin{pmatrix} g_1\left(W,\gamma\left(F_\tau\right),\theta\right)\\ g_2\left(W,\gamma\left(F_\tau\right),\theta\right)\\ g_3\left(W,\gamma\left(F_\tau\right),\theta\right)\end{pmatrix}.$$

Thus, to characterize the FSIF it suffices to find ϕ and α_0 such that

$$\frac{d}{d\tau}\mathbb{E}\left[g\left(W,\gamma\left(F_\tau\right),\theta\right)\right]=\int\phi\left(\omega,\gamma_0,\alpha_0,\theta\right)H(d\omega) \tag{B12}$$

holds. In other words, to derive the locally robust moment for the intergenerational elasticity, along with the means of lifetime income, we must first characterize the influence function for each element in θ, and then augment their corresponding identifying moments with it.

We start by finding the FSIF for the nuisance parameter $\mu_{ct}\left(F_\tau\right)\left(\boldsymbol{X}_{ct},1\right)$ in the identifying equation $g_2\left(W,\gamma\left(F_\tau\right),\theta\right)$. To this end, we start by considering the left-hand side of equation (B12):

$$\begin{aligned}\frac{d}{d\tau}\mathbb{E}\left[g_2\left(W,\gamma\left(F_\tau\right),\theta\right)\right]&=\frac{d}{d\tau}\mathbb{E}\left[\sum_{t=1}^{T}\mu_{ct}\left(F_\tau\right)\left(\boldsymbol{X}_{ct},1\right)-\mu_c^P\right]\\ &=\sum_{t=1}^{T}\frac{d}{d\tau}\mathbb{E}\left[\mu_{ct}\left(F_\tau\right)\left(\boldsymbol{X}_{ct},1\right)\right],\end{aligned} \tag{B13}$$

where the interchange of differentiation and expectation is justified by the dominated convergence theorem under standard regularity conditions. We now express the expectation as

$$\begin{aligned}\mathbb{E}\left[\mu_{ct}\left(F_\tau\right)\left(\boldsymbol{X}_{ct},1\right)\right]&=\mathbb{E}\left[\mathbb{E}_\tau\left[Y_{ct}\mid\boldsymbol{X}_{ct},D_{ct}=1\right]\right]\\ &=\mathbb{E}\left[\frac{D_{ct}}{p\left(D_{ct}=1|\boldsymbol{X}_{ct}\right)}\mathbb{E}_\tau\left[Y_{ct}\mid\boldsymbol{X}_{ct},D_{ct}=1\right]\right]\\ &=\mathbb{E}\left[\frac{D_{ct}}{p\left(D_{ct}=1|\boldsymbol{X}_{ct}\right)}\mathbb{E}_\tau\left[Y_{ct}\mid\boldsymbol{X}_{ct},D_{ct}\right]\right]\\ &:=\mathbb{E}\left[\alpha_{0c,t}\left(\boldsymbol{X}_{ct},D_{ct}\right)\mu_{ct}\left(F_\tau\right)\left(\boldsymbol{X}_{ct},D_{ct}\right)\right],\end{aligned} \tag{B14}$$

where the first equality follows by definition, the second by the law of iterated expectations, and the third one by the MAR Assumption 2-NP.*i*. Furthermore, the term

$\alpha_{0c,t}(\boldsymbol{X}_{ct}, D_{ct})$ is the Riesz representer of the functional $\mu_{ct}(F_\tau)(\boldsymbol{X}_{ct}, 1)$.

We now plug equation (B14) into (B13) to characterize the FSIF for

$$
\begin{aligned}
&\frac{d}{d\tau}\mathbb{E}\left[\mu_{ct}(F_\tau)(\boldsymbol{X}_{ct}, 1)\right] \\
=&\frac{d}{d\tau}\mathbb{E}\left[\alpha_{0c,t}(\boldsymbol{X}_{ct}, D_{ct})\mu_{ct}(F_\tau)(\boldsymbol{X}_{ct}, D_{ct})\right] \\
=&-\frac{d}{d\tau}\mathbb{E}\left[\alpha_{0c,t}(\boldsymbol{X}_{ct}, D_{ct})(Y_{ct} - \mu_{ct}(F_\tau)(\boldsymbol{X}_{ct}, D_{ct}))\right] \\
=&\frac{d}{d\tau}\mathbb{E}_\tau\left[\alpha_{0c,t}(\boldsymbol{X}_{ct}, D_{ct})(Y_{ct} - \mu_{ct}(F_\tau)(\boldsymbol{X}_{ct}, D_{ct}))\right] \\
=&\int \alpha_{0c,t}(\boldsymbol{X}_{ct}, D_{ct})(y_{ct} - \mu_{ct}(F_0)(\boldsymbol{x}_{ct}, d_{ct}))\,H(d\omega) \\
:=&\int \phi_{c,t}(\omega, \gamma, \alpha_{0c,t}, \theta)\,H(d\omega), \qquad \text{(B15)}
\end{aligned}
$$

where the second equality follows by the fact that Y_{ct} does not depend on τ. The third equality exploits the fact that the prediction errors of children's annual income are uncorrelated to any function of $(\boldsymbol{X}_{ct}, D_{ct})$, so that

$$
\mathbb{E}_\tau\left[\alpha_{0c,t}(\boldsymbol{X}_{ct}, D_{ct})(Y_{ct} - \mu_{ct}(F_\tau)(\boldsymbol{X}_{ct}, D_{ct}))\right] = 0,
$$

which in turn implies

$$
\begin{aligned}
&\frac{d}{d\tau}\mathbb{E}_\tau\left[\alpha_{0c,t}(\boldsymbol{X}_{ct}, D_{ct})(Y_{ct} - \mu_{ct}(F_\tau)(\boldsymbol{X}_{ct}, D_{ct}))\right] = 0 \iff \\
&\frac{d}{d\tau}\mathbb{E}_\tau\left[\alpha_{0c,t}(\boldsymbol{X}_{ct}, D_{ct})(Y_{ct} - \mu_{ct}(F_0)(\boldsymbol{X}_{ct}, D_{ct}))\right] \\
&= -\frac{d}{d\tau}\mathbb{E}\left[\alpha_{0c,t}(\boldsymbol{X}_{ct}, D_{ct})(Y_{ct} - \mu_{ct}(F_\tau)(\boldsymbol{X}_{ct}, D_{ct}))\right].
\end{aligned}
$$

Finally, the derivative of the expectation under the perturbed distribution F_τ corresponds to the integral with respect to the perturbation measure H

$$
\frac{d}{d\tau}\mathbb{E}_\tau\left[\alpha_{0c,t}(Y_{ct} - \mu_{ct})\right] = \int \alpha_{0c,t}(y_{ct} - \mu_{ct})H(d\omega),
$$

because we consider the linear perturbation $F_\tau = F_0 + \tau(H - F_0)$. Thus, the derivative w.r.t τ isolates the perturbation $H - F_0$, the expectation under H appears because we are evaluating the Gateaux derivative at $\tau = 0$, and similarly the terms involving F_0 vanish, as they are constant w.r.t τ.

According to equations (B13) and (B15), the FSIF of μ_c^P is thus given by

$$\sum_{t=1}^{T} \frac{D_{ct}}{p\left(D_{ct}=1 \mid \boldsymbol{X}_{ct}\right)}\left(Y_{ct}-\mu_{ct}\left(\boldsymbol{X}_{ct}, 1\right)\right),$$

which corrects for the prediction errors in children's annual income, weighted by the propensity scores.

Analogously, the the FSIF of μ_f^P is given by

$$\sum_{t=1}^{T} \frac{D_{ft}}{p\left(D_{ft}=1 \mid \boldsymbol{X}_{ft}\right)}\left(Y_{ft}-\mu_{ft}\left(\boldsymbol{X}_{ft}, 1\right)\right).$$

We now turn attention to the moment identifying the intergenerational elasticity (equation (B9)). In contrast to the two other identifying moments, this one involves the three nuisance parameters μ_{ct}, μ_{ft}, and $\sigma_{t,j}$. Thus, to find the FSIF, we start by decomposing the derivative on the left-hand side of (B12) as follows:

$$\begin{aligned}
\frac{d}{d\tau}\mathbb{E}\left[g\left(W, \gamma\left(F_\tau\right), \beta\right)\right] = {} & \beta \sum_{|t-j|\leq h} \underbrace{\frac{d}{d\tau}\mathbb{E}\left[\sigma_{tj}\left(F_\tau\right)\left(\boldsymbol{X}_{ftj}, 1, 1\right)\right]}_{(1)} \\
& + \beta \sum_{t=1}^{T}\sum_{j=t+1}^{\min(t+h,T)} \underbrace{\frac{d}{d\tau}\mathbb{E}\left[\left(\mu_{fj}\left(F_0\right)\left(\boldsymbol{X}_{fj}, 1\right)-\mu_f^P\right)\mu_{ft}\left(F_\tau\right)\left(\boldsymbol{X}_{ft}, 1\right)\right]}_{(2)} \\
& + \beta \sum_{t=1}^{T}\sum_{j=t+1}^{\min(t+h,T)} \underbrace{\frac{d}{d\tau}\mathbb{E}\left[\left(\mu_{ft}\left(F_0\right)\left(\boldsymbol{X}_{ft}, 1\right)-\mu_f^P\right)\mu_{fj}\left(F_\tau\right)\left(\boldsymbol{X}_{fj}, 1\right)\right]}_{(3)} \\
& - \sum_{t=1}^{T}\sum_{j=1}^{T} \underbrace{\frac{d}{d\tau}\mathbb{E}\left[\left(\mu_{ct}\left(F_0\right)\left(\boldsymbol{X}_{ct}, 1\right)-\mu_c^P\right)\mu_{fj}\left(F_\tau\right)\left(\boldsymbol{X}_{fj}, 1\right)\right]}_{(4)} \\
& - \sum_{t=1}^{T}\sum_{j=1}^{T} \underbrace{\frac{d}{d\tau}\mathbb{E}\left[\left(\mu_{fj}\left(F_0\right)\left(\boldsymbol{X}_{fj}, 1\right)-\mu_f^P\right)\mu_{ct}\left(F_\tau\right)\left(\boldsymbol{X}_{ct}, 1\right)\right]}_{(5)}.
\end{aligned} \tag{B16}$$

The FSIF is obtained by observing that each term (1)–(5) in Equation (B16) can be expressed in the form of the left-hand side of Equation (B12). Accordingly, we proceed as follows: for each term, we (i) identify its Riesz representer α_0, (ii) derive the corresponding FSIF ϕ by expressing the term as the right-hand side of Equation (B12), and (iii) substitute these results back into Equation (B16). This yields the required solution for ϕ and α_0 in Equation (B12). Below, we implement this procedure.

We start by analyzing the expectation in term (1) of equation (B16), which can be expressed as:

$$\mathbb{E}\left[\sigma_{tj}\left(F_\tau\right)\left(\boldsymbol{X}_{ftj}, 1, 1\right)\right]$$

$$
\begin{aligned}
&:=\mathbb{E}\left[\mathbb{E}_\tau\left[\left(Y_{ft}-\mu_f^P\right)\left(Y_{fj}-\mu_f^P\right)\mid \boldsymbol{X}_{ftj}, D_{ft}=1, D_{fj}=1\right]\right]\\
&:=\mathbb{E}\left[\mathbb{E}_\tau\left[U_{ft}U_{fj}\mid \boldsymbol{X}_{ftj}, D_{ft}=1, D_{fj}=1\right]\right]\\
&=\mathbb{E}\left[\mathbb{E}_\tau\left[U_{ft}U_{fj}\mid \boldsymbol{X}_{ftj}, D_{ft}, D_{fj}\right]\right]\\
&=\mathbb{E}\left[\frac{D_{ft}D_{fj}}{p\left(D_{ft}=1, D_{fj}=1\mid \boldsymbol{X}_{ftj}\right)}\mathbb{E}_\tau\left[U_{ft}U_{fj}\mid \boldsymbol{X}_{ftj}, D_{ft}, D_{fj}\right]\right]\\
&:=\mathbb{E}\left[\alpha_{01,tj}\left(\boldsymbol{X}_{ftj}, D_{ft}, D_{fj}\right)\sigma_{tj}\left(F_\tau\right)\left(\boldsymbol{X}_{ftj}, D_{ft}, D_{fj}\right)\right],
\end{aligned}
\tag{B17}
$$

where in the second equality we have defined $U_{ft} := Y_{fj}-\mu_f^P$, the third equality follows by Assumption 2-NP.*iii*, and the fourth by LIE and the fact that

$$
\begin{aligned}
&D_{ft}D_{fj}\mathbb{E}_\tau\left[U_{ft}U_{fj}\mid \boldsymbol{X}_{ftj}, D_{ft}=1, D_{fj}=1\right]\\
=&D_{ft}D_{fj}\mathbb{E}_\tau\left[U_{ft}U_{fj}\mid \boldsymbol{X}_{ftj}, D_{ft}, D_{fj}\right].
\end{aligned}
$$

Having characterized the Riesz representer for $\mathbb{E}\left[\sigma_{tj}\left(F_\tau\right)\left(\boldsymbol{X}_{ftj},1,1\right)\right]$, we now turn to derive its corresponding FSIF:

$$
\begin{aligned}
&\frac{d}{d\tau}\mathbb{E}\left[\sigma_{tj}\left(F_\tau\right)\left(\boldsymbol{X}_{ftj},1,1\right)\right]\\
=&\frac{d}{d\tau}\mathbb{E}\left[\alpha_{01,tj}\left(\boldsymbol{X}_{ftj}, D_{ft}, D_{fj}\right)\sigma_{tj}\left(F_\tau\right)\left(\boldsymbol{X}_{ftj}, D_{ft}, D_{fj}\right)\right]\\
=&-\frac{d}{d\tau}\mathbb{E}\left[\alpha_{01,tj}\left(\boldsymbol{X}_{ftj}, D_{ft}, D_{fj}\right)\left(U_{ft}U_{fj}-\sigma_{tj}\left(F_\tau\right)\left(\boldsymbol{X}_{ftj}, D_{ft}, D_{fj}\right)\right)\right]\\
=&\frac{d}{d\tau}\mathbb{E}_\tau\left[\alpha_{01,tj}\left(\boldsymbol{X}_{ftj}, D_{ft}, D_{fj}\right)\left(U_{ft}U_{fj}-\sigma_{tj}\left(F_\tau\right)\left(\boldsymbol{X}_{ftj}, D_{ft}, D_{fj}\right)\right)\right]\\
=&\int\alpha_{01,tj}\left(\boldsymbol{X}_{ftj}, d_{ft}, d_{fj}\right)\left(U_{ft}U_{fj}-\sigma_{tj}\left(F_0;\mu_f^P\right)\left(\boldsymbol{x}_{ftj}, d_{ft}, d_{fj}\right)\right)H(d\omega)\\
:=&\int\phi_{1,tj}\left(\omega,\gamma,\alpha_{01,tj}\right)H(d\omega),
\end{aligned}
\tag{B18}
$$

following the same arguments as equation (B15).

We now derive the FSIF for the second term in equation (B16)

$$
\begin{aligned}
&\mathbb{E}\left[\left(\mu_{fj}\left(F_0\right)\left(\boldsymbol{X}_{fj},1\right)-\mu_f^P\right)\mu_{ft}\left(F_\tau\right)\left(\boldsymbol{X}_{ft},1\right)\right]\\
=&\mathbb{E}\left[\left(\mathbb{E}\left[Y_{fj}\mid \boldsymbol{X}_{fj}, D_{fj}=1\right]-\mu_f^P\right)\mathbb{E}_\tau\left[Y_{ft}\mid \boldsymbol{X}_{ft}, D_{ft}=1\right]\right]\\
=&\mathbb{E}\left[\left(\mathbb{E}\left[Y_{fj}\mid \boldsymbol{X}_{fj}, D_{fj}=1\right]-\mu_f^P\right)\frac{D_{ft}}{p\left(D_{ft}=1|\boldsymbol{X}_{ft}\right)}\mathbb{E}_\tau\left[Y_{ft}\mid \boldsymbol{X}_{ft}, D_{ft}=1\right]\right]
\end{aligned}
$$

$$
\begin{aligned}
&= \mathbb{E}\left[\left(\mathbb{E}\left[Y_{fj} \mid \boldsymbol{X}_{fj}, D_{fj}=1\right]-\mu_f^P\right) \frac{D_{ft}}{p\left(D_{ft}=1 \mid \boldsymbol{X}_{ft}\right)} \mathbb{E}_\tau\left[Y_{ft} \mid \boldsymbol{X}_{ft}, D_{ft}\right]\right] \\
&= \mathbb{E}\left[\left(\mathbb{E}\left[Y_{fj} \mid \boldsymbol{X}_{fj}\right]-\mu_f^P\right) \frac{D_{ft}}{p\left(D_{ft}=1 \mid \boldsymbol{X}_{ft}\right)} \mathbb{E}_\tau\left[Y_{ft} \mid \boldsymbol{X}_{ft}, D_{ft}\right]\right] \\
&:= \mathbb{E}\left[\alpha_{02,tj}\left(\boldsymbol{X}_{fj}, \boldsymbol{X}_{ft}, D_{ft}\right) \mu_{ft}\left(F_\tau\right)\left(\boldsymbol{X}_{ft}, D_{ft}\right)\right],
\end{aligned}
$$

where we have followed the same arguments as those of equation (B17). Following an analogous procedure to equation (B18), yields

$$
\begin{aligned}
& \frac{d}{d\tau} \mathbb{E}\left[\left(\mu_{fj}\left(F_0\right)\left(\boldsymbol{X}_{fj}, 1\right)-\mu_f^P\right) \mu_{ft}\left(F_\tau\right)\left(\boldsymbol{X}_{ft}, 1\right)\right] \\
&= \frac{d}{d\tau} \mathbb{E}\left[\alpha_{02,tj}\left(\boldsymbol{X}_{fj}, \boldsymbol{X}_{ft}, D_{ft}\right) \mu_{ft}\left(F_\tau\right)\left(\boldsymbol{X}_{ft}, D_{ft}\right)\right] \\
&= -\frac{d}{d\tau} \mathbb{E}\left[\alpha_{02,tj}\left(\boldsymbol{X}_{fj}, \boldsymbol{X}_{ft}, D_{ft}\right)\left(Y_{ft}-\mu_{ft}\left(F_\tau\right)\left(\boldsymbol{X}_{ft}, D_{ft}\right)\right)\right] \\
&= \frac{d}{d\tau} \mathbb{E}_\tau\left[\alpha_{02,tj}\left(\boldsymbol{X}_{fj}, \boldsymbol{X}_{ft}, D_{ft}\right)\left(Y_{ft}-\mu_{ft}\left(F_\tau\right)\left(\boldsymbol{X}_{ft}, D_{ft}\right)\right)\right] \\
&= \int \alpha_{02,tj}\left(\boldsymbol{x}_{fj}, \boldsymbol{x}_{ft}, d_{ft}\right)\left(y_{ft}-\mu_{ft}\left(F_0\right)\left(\boldsymbol{x}_t, d_{ft}\right)\right) H(d\omega) \\
&:= \int \phi_{2,tj}\left(\omega, \gamma, \alpha_{02,tj}\right) H(d\omega). \qquad \text{(B19)}
\end{aligned}
$$

The key distinction between equation (B18) and equation (B19) is that the latter requires

$$
\mathbb{E}\left[\alpha_{02,tj}\left(\boldsymbol{X}_{fj}, \boldsymbol{X}_{ft}, D_{ft}\right)\left(Y_{ft}-\mathbb{E}\left[Y_{ft} \mid \boldsymbol{X}_{ft}\right]\right)\right]=0.
$$

This orthogonality condition is satisfied by Assumption 1-NP.*i*, since the conditional expectation $\mathbb{E}[Y_{ft} \mid \boldsymbol{X}_{ft}, D_{ft}, \boldsymbol{X}_{fj}]$ reduces to $\mathbb{E}_\tau[Y_{ft} \mid \boldsymbol{X}_{ft}, D_{ft}]$, rendering the prediction error $Y_{ft} - \mathbb{E}_\tau[Y_{ft} \mid \boldsymbol{X}_t, D_{ft}]$ orthogonal to any function of $\boldsymbol{X}_{fj}$, $\boldsymbol{X}_{ft}$, and D_{ft}.

Observe that terms (2), (3), (4) and (5) in equation (B16) share an identical functional form, differing only in their superscripts (indicating generation $g \in \{c, f\}$) and time indices (j or t). This structural similarity implies that the derivations for terms (3), (4), and (5) have the same structure as term (2). Consequently, the FSIFs and Riesz representers for these terms are given by

(B20)

$$
\phi_{3,tj}\left(\omega, \gamma, \alpha_{03,tj}\right)=\alpha_{03,tj}\left(\boldsymbol{X}_{ft}, \boldsymbol{X}_{fj}, D_{fj}\right)\left(Y_{fj}-\mu_{fj}\left(\boldsymbol{X}_{fj}, D_{fj}\right)\right), \quad \alpha_{03,tj}=\left(\mu_{ft}\left(\boldsymbol{X}_{ft}, D_{ft}\right)-\mu_f^P\right) \frac{D_{fj}}{p\left(D_{fj}=1 \mid \boldsymbol{X}_{fj}\right)}
$$

(B21)

$$
\phi_{4,tj}\left(\omega, \gamma, \alpha_{04,tj}\right)=\alpha_{04,tj}\left(\boldsymbol{X}_{ct}, \boldsymbol{X}_{fj}, D_{fj}\right)\left(Y_{fj}-\mu_{fj}\left(\boldsymbol{X}_{fj}, D_{fj}\right)\right), \quad \alpha_{04,tj}=\left(\mu_{ct}\left(\boldsymbol{X}_{ct}, D_{ft}\right)-\mu_c^P\right) \frac{D_{fj}}{p\left(D_{fj}=1 \mid \boldsymbol{X}_{fj}\right)}
$$

(B22)

$$\phi_{5,tj}\left(\omega,\gamma,\alpha_{05,tj}\right)=\alpha_{05,tj}\left(X_{fj},X_{ct},D_{ct}\right)\left(Y_{ct}-\mu_{ct}\left(X_{ct},D_{ct}\right)\right),\qquad \alpha_{05,tj}=\left(\mu_{fj}\left(X_{fj},D_{fj}\right)-\mu_f^P\right)\frac{D_{ct}}{p\left(D_{ct}=1|\boldsymbol{X}_{ct}\right)}.$$

The orthogonality condition $\mathbb{E}[\phi_{k,tj}]=0$ also holds for each $k\in\{3,4,5\}$ by Assumption 1-NP.*i*.

Having characterized the first-step influence function for each term in equation (B16), we can plug equations (B18)–(B22) into equation (B16):

$$\begin{aligned}\frac{d}{d\tau}\mathbb{E}\left[g\left(W,\gamma\left(F_\tau\right),\theta\right)\right]=&\beta\sum_{|t-j|\le h}\int\phi_{1,tj}\left(\omega,\gamma,\alpha_{01,tj}\right)H(d\omega)+\beta\sum_{|t-j|>h}\int\phi_{2,tj}\left(\omega,\gamma,\alpha_{02,tj}\right)H(d\omega)\\&+\beta\sum_{|t-j|>h}\int\phi_{3,tj}\left(\omega,\gamma,\alpha_{03,tj}\right)H(d\omega)-\sum_{t=1}^{T}\sum_{j=1}^{T}\int\phi_{4,tj}\left(\omega,\gamma,\alpha_{04,tj}\right)H(d\omega)\\&-\sum_{t=1}^{T}\sum_{j=1}^{T}\int\phi_{5,tj}\left(\omega,\gamma,\alpha_{05,tj}\right)H(d\omega)\\:=&\int\phi_1\left(\omega,\gamma_0,\alpha_0,\theta\right)H(d\omega).\end{aligned}\tag{B23}$$

Equation (B23) defines the first step influence function of estimating γ on the moment identifying β, thereby allowing us to construct a locally robust moment to estimate the IGE. To illustrate this point, consider again equation (B8)

$$\psi\left(W,\gamma,\alpha,\theta\right)=g\left(W,\gamma,\theta\right)+\phi\left(W,\gamma,\alpha,\theta\right).$$

Thus, we construct the locally robust moment by adding $\phi_1\left(W,\gamma,\alpha,\theta\right)$ from equation (B23) to the identifying moment $g_1\left(W,\gamma,\theta\right)$ in equation (B9)

$$\begin{aligned}\psi_1\left(W,\gamma,\alpha,\theta\right)=&\beta\sum_{|t-j|\le h}\sigma_{tj}\left(\boldsymbol{X}_{ftj},1,1\right)+\beta\sum_{|t-j|>h}\left(\mu_{ft}\left(\boldsymbol{X}_{ft},1\right)-\mu_f^P\right)\left(\mu_{fj}\left(\boldsymbol{X}_{fj},1\right)-\mu_f^P\right)\\&-\sum_{t=1}^{T}\left(\mu_{ct}\left(\boldsymbol{X}_{ct},1\right)-\mu_c^P\right)\sum_{j=1}^{T}\left(\mu_{fj}\left(\boldsymbol{X}_{fj},1\right)-\mu_f^P\right)\\&+\beta\sum_{|t-j|\le h}\frac{D_{ft}D_{fj}}{p\left(D_{ft}=1,D_{fj}=1|\boldsymbol{X}_{ftj}\right)}\left(\left(Y_{ft}-\mu_f^P\right)\left(Y_{fj}-\mu_f^P\right)-\sigma_{tj}\left(\boldsymbol{X}_{ftj},1,1\right)\right)\\&+\beta\sum_{|t-j|>h}\left(\mu_{fj}\left(\boldsymbol{X}_{fj},1\right)-\mu_f^P\right)\frac{D_{ft}}{p\left(D_{ft}=1|\boldsymbol{X}_{ft}\right)}\left(Y_{ft}-\mu_{ft}\left(\boldsymbol{X}_{ft},1\right)\right)\\&+\beta\sum_{|t-j|>h}\left(\mu_{ft}\left(\boldsymbol{X}_{ft},1\right)-\mu_f^P\right)\frac{D_{fj}}{p\left(D_{fj}=1|\boldsymbol{X}_{fj}\right)}\left(Y_{fj}-\mu_{fj}\left(\boldsymbol{X}_{fj},1\right)\right)\\&-\sum_{t=1}^{T}\sum_{t=j}^{T}\left(\mu_{ct}\left(\boldsymbol{X}_{ct},1\right)-\mu_c^P\right)\frac{D_{fj}}{p\left(D_{fj}=1|\boldsymbol{X}_{fj}\right)}\left(Y_{fj}-\mu_{fj}\left(\boldsymbol{X}_{fj},1\right)\right)\end{aligned}$$

$$-\sum_{t=1}^{T}\sum_{t=j}^{T}\left(\mu_{fj}\left(\boldsymbol{X}_{fj},1\right)-\mu_f^P\right)\frac{D_{ct}}{p\left(D_{ct}=1|\boldsymbol{X}_{ct}\right)}\left(Y_{ct}-\mu_{ct}\left(\boldsymbol{X}_{ct},1\right)\right).$$

Similarly for μ_f^P and μ_c^P, we have

$$\psi_2\left(W,\gamma,\alpha,\theta\right)=\sum_{t=1}^{T}\mu_{ct}\left(\boldsymbol{X}_{ct},1\right)-\mu_c^P+\sum_{t=1}^{T}\frac{D_{ct}}{p\left(D_{ct}=1|\boldsymbol{X}_{ct}\right)}\left(Y_{ct}-\mu_{ct}\left(\boldsymbol{X}_{ct},1\right)\right)$$
$$\psi_3\left(W,\gamma,\alpha,\theta\right)=\sum_{t=1}^{T}\mu_{ft}\left(\boldsymbol{X}_{ft},1\right)-\mu_f^P+\sum_{t=1}^{T}\frac{D_{ct}}{p\left(D_{ft}=1|\boldsymbol{X}_{ft}\right)}\left(Y_{ft}-\mu_{ft}\left(\boldsymbol{X}_{ft},1\right)\right).$$

Thus, our locally robust moment for the parameter θ is given by

$$\psi\left(W,\gamma,\alpha,\theta\right)=\left(\psi_1\left(W,\gamma,\alpha,\theta\right),\psi_2\left(W,\gamma,\alpha,\theta\right),\psi_3\left(W,\gamma,\alpha,\theta\right)\right),$$

which yields a locally robust moment for the IGE, incorporating that it depends on μ_f^P and μ_c^P.

A debiased GMM estimator for θ is thus given by

$$\hat{\theta}=\arg\min_{\theta\in\Theta}\hat{\psi}(\theta)'\hat{\Upsilon}\,\hat{\psi}(\theta),$$

where $\hat{\Upsilon}$ is a positive semi-definite weighting matrix and Θ denotes the parameter space.

B4. Asymptotic Properties of the Locally Robust Estimator

To establish consistency for the locally robust estimator, we will impose the following assumption

Assumption C-NP. *(Boundedness and Regularity Conditions for Consistency)*

(i) ***Identification:*** $\mathbb{E}[\psi\left(W,\gamma_0,\alpha_0,\theta\right)]=0$ *if and only if* $\theta=\theta_0$*;*

(ii) ***Compactness:*** *The parameter space* $\Theta\subset\mathbb{R}^3$ *is compact;*

(iii) ***Regularity of the Identifying Moment:*** $\mathbb{E}[\|g\left(W,\gamma_0,\theta\right)\|]<\infty$ *and* $\int\left\|g\left(w,\hat{\gamma}^{(\ell)},\theta\right)-g\left(W,\gamma_0,\theta\right)\right\|dF_0(w)\xrightarrow{p}0$ *for all* $\theta\in\Theta$*, where* F_0 *denotes the unknown cumulative distribution function of the data* W*.*

(iv) ***Local Stability in*** θ***:*** *There exist a constant* $C>0$ *and a function* $d(W,\gamma)$ *such that for* $\|\gamma-\gamma_0\|$ *sufficiently small and* $\hat{\theta}_n^{LR},\theta\in\Theta$*,*

$$\left\|g\left(W,\gamma,\hat{\theta}_n^{LR}\right)-g\left(W,\gamma,\theta\right)\right\|\leq d(W,\gamma)\left\|\hat{\theta}_n^{LR}-\theta\right\|^{1/C},\quad \textit{with } \mathbb{E}[d(W,\gamma)]<C;$$

(v) ***Regularity of the Orthogonal Moment:***

(a) $\mathbb{E}[\|\psi(W, \gamma_0, \alpha_0, \theta_0)\|] < \infty$*;*

(b) The following hold:

$$\int \left\|\phi\left(w, \hat{\gamma}^{(\ell)}, \alpha_0, \theta_0\right) - \phi(w, \gamma_0, \alpha_0, \theta_0)\right\|^2 dF_0(w) \xrightarrow{p} 0,$$
$$\int \left\|\phi\left(w, \gamma_0, \hat{\alpha}^{(\ell)}, \theta_0\right) - \phi(w, \gamma_0, \alpha_0, \theta_0)\right\|^2 dF_0(w) \xrightarrow{p} 0,$$
$$\int \left\|\hat{\Delta}_\ell(w)\right\| dF_0(w) \xrightarrow{p} 0,$$

where $\hat{\Delta}_\ell(w) := \phi(w, \hat{\gamma}^{(\ell)}, \hat{\alpha}^{(\ell)}, \hat{\theta}_n^{LR}) - \phi(w, \gamma_0, \hat{\alpha}^{(\ell)}, \hat{\theta}_n^{LR}) - \phi(w, \hat{\gamma}^{(\ell)}, \alpha_0, \theta_0)$ $+\phi(w, \gamma_0, \alpha_0, \theta_0)$.

Identification ensures we are solving a well-posed moment problem. The compactness assumption is economically meaningful as the intergenerational income elasticity is theoretically bounded on the interval $(0, 1)$, reflecting imperfect but positive persistence of income across generations. The assumed bounds align with cross-country evidence, where estimates range from approximately 0.14 (Denmark) to 0.58 (Brazil), with most developed economies exhibiting elasticities between 0.2 and 0.5 (Stuhler et al., 2018). Moreover, permanent income cannot exceed the highest observed income in the data, nor can it be negative for individuals with any labor market participation.

The regularity of the identifying moment assumes integrability and L^1 continuity. The former guarantees the moment function remains well-defined in expectation across the entire parameter space, while the latter ensures that the difference between the moment function evaluated at the estimated nuisance parameter and its true value becomes negligible. Local Stability controls how the moment function varies with θ, preventing extreme sensitivity to parameter changes when nuisance estimates are near their true values. The integrability condition for the orthogonal moment matches that of the identifying moment, while the stronger L^2 convergence (compared to L^1) ensures the difference between the orthogonal moment function evaluated at the estimated nuisance parameters and their true value becomes negligible faster than for the identifying moment.

As regards the conditions for assymptotic normality, we start by imposing regularity conditions that translate into concrete requirements within our intergenerational mobility framework. Specifically, we require the orthogonal moment function in equation (9) to be square-integrable at the true parameter values. This condition implies two key substantive requirements: first, the inverse propensity weights must be bounded, ensured by Assumption 3-LC, which restricts $p(D_{gt} = 1|\boldsymbol{X}_{gt})$ and $p(D_{ft} = 1, D_{fj}|\boldsymbol{X}_{fjt})$ away from zero and one; second, the income process must exhibit sufficient regularity, captured by weak temporal dependence (via mixing conditions) and finite higher-order moments, particularly for the cross-product terms $Y_{ft}Y_{fj}$ that enter the moment function.

In addition, we assume that the machine learning estimators for the nuisance parameters, such as the conditional expectations $\mathbb{E}[Y_{ct}|\boldsymbol{X}_{ct}, D_{ct} = 1]$ and the propensity scores, converge at suitable rates in mean-square error. These conditions are typically satisfied in longitudinal data settings where income dynamics are moderately dependent over time and income observation probability varies smoothly with covariates.

We now turn to establishing the assumptions required for asymptotic normality.

Assumption 3-LR. *(Boundedness and Regularity Conditions for Consistency) (i) The orthogonal moment function is square-integrable:*

$$\mathbb{E}[\|\psi(W, \beta_0, \gamma_0, \alpha_0)\|^2] < \infty.$$

(ii) The nuisance estimators are consistent in mean-square error:

$$\int \left\| g\left(w, \hat{\gamma}^{(\ell)}, \theta_0\right) - g(w, \gamma_0, \theta_0) \right\|^2 dF_0(w) \xrightarrow{p} 0,$$

$$\int \left\| \phi\left(w, \hat{\gamma}^{(\ell)}, \alpha_0, \theta_0\right) - \phi(w, \gamma_0, \alpha_0, \theta_0) \right\|^2 dF_0(w) \xrightarrow{p} 0,$$

$$\int \left\| \phi\left(w, \gamma_0, \hat{\alpha}^{(\ell)}, \hat{\theta}_n^{LR}\right) - \phi(w, \gamma_0, \alpha_0, \theta_0) \right\|^2 dF_0(w) \xrightarrow{p} 0,$$

where F_0 denotes the unknown cumulative distribution function of the data W.

We further impose a regularity condition that controls the remainder term arising from the interaction of first-step estimation errors. Specifically, Assumption 4-LR requires the correction term $\hat{\Delta}_\ell(w)$, which captures the deviation from exact orthogonality due to estimation of the nuisance parameters, to vanish sufficiently quickly in sample averages. This condition imposes a rate requirement on the interaction remainder $\hat{\Delta}_\ell(w)$, namely that its average must converge to zero faster than $1/\sqrt{n}$. This ensures that the remainder is asymptotically negligible and does not affect the limiting distribution of the estimator.

Assumption 4-LR. *(First-Step Remainder Control) For each fold $\ell = 1, \ldots, L$, define the correction term*

$$\hat{\Delta}_\ell(w) := \phi(w, \hat{\gamma}^{(\ell)}, \hat{\alpha}^{(\ell)}, \hat{\theta}_n^{LR}) - \phi(w, \gamma_0, \hat{\alpha}^{(\ell)}, \hat{\theta}_n^{LR}) - \phi(w, \hat{\gamma}^{(\ell)}, \alpha_0, \theta_0) + \phi(w, \gamma_0, \alpha_0, \theta_0).$$

We assume that this term satisfies at least one of the following conditions:

1) $\sqrt{n} \int \hat{\Delta}_\ell(w)\, dF_0(w) \xrightarrow{p} 0$ *and* $\int \left\|\hat{\Delta}_\ell(w)\right\|^2 dF_0(w) \xrightarrow{p} 0$

2) $\frac{1}{\sqrt{n}} \sum_{i \in I_\ell} \left\|\hat{\Delta}_\ell(W_i)\right\| \xrightarrow{p} 0$

3) $\frac{1}{\sqrt{n}} \sum_{i \in I_\ell} \hat{\Delta}_\ell(W_i) \xrightarrow{p} 0.$

To establish valid inference after machine learning estimation of nuisance parameters, we require a Neyman orthogonality condition that ensures our moment function remains insensitive to small estimation errors. This assumption requires the moment condition to hold at estimated nuisance parameters $\hat{\alpha}^{(\ell)}$, and (2) specifying alternative bias control conditions that adapt to different estimation scenarios. The first condition (affine linearity) covers classical doubly robust estimators, while the second (quadratic bound with $n^{-1/4}$ convergence) handles many semiparametric cases. The third condition provides a weaker alternative when the bias vanishes asymptotically. In our framework, these conditions will be satisfied through either the double robustness properties of our moment function or the convergence rates of our machine learning estimators, similar to standard results in the semiparametric literature.

Assumption 5-LR. *(Neyman Orthogonality and Bias Control) For each fold $\ell = 1, \ldots, L$, we require the orthogonality condition*

$$\int \phi(w, \gamma_0, \hat{\alpha}^{(\ell)}, \beta)\, dF_0(w) = 0$$

to hold with probability approaching one. In addition, one of the following conditions must be satisfied:

1) $\bar{\psi}(\gamma, \alpha, \beta) := \mathbb{E}\left[\psi\left(W, \gamma, \alpha, \beta\right)\right]$ *is affine in* γ
2) $|\bar{\psi}(\gamma, \alpha_0, \theta_0)| \leq C\|\gamma - \gamma_0\|^2$ *for all* γ *such that* $\|\gamma - \gamma_0\|$ *is sufficiently small, and* $\|\hat{\gamma}^{(\ell)} - \gamma_0\| = o_p(n^{-1/4})$
3) $\sqrt{n} \cdot \bar{\psi}(\hat{\gamma}^{(\ell)}, \alpha_0, \theta_0) \xrightarrow{p} 0.$

The following assumption ensures the consistency of the auxiliary components required for valid variance estimation. It serves two key purposes: first, it guarantees that the estimation error in $\hat{\theta}_n^{LR}$ becomes asymptotically negligible when substituted into the moment function; second, it requires the first-step remainder term $\hat{\Delta}_\ell(w)$ to vanish fast enough. Together, these conditions ensure that the variability introduced by cross-fitting and parameter estimation does not distort the asymptotic variance calculations.

Assumption 6-LR. *(Square-integrability) For each fold $\ell = 1, \ldots, L$*
$$\int \left\| g\left(w, \hat{\gamma}^{(\ell)}, \hat{\beta}_n^{LR}\right) - g\left(w, \hat{\gamma}^{(\ell)}, \beta_0\right) \right\|^2 dF_0(w) \xrightarrow{p} 0, \text{ and } \int \left\| \hat{\Delta}_\ell(w) \right\|^2 dF_0(w) \xrightarrow{p} 0.$$

Finally, we impose an assumption to guarantee the stability of the Jacobian matrix $G(\beta)$, ensuring the asymptotic normality of our estimator. Specifically, it requires: (1) convergence of the nuisance parameter estimates $\hat{\gamma}^{(\ell)}$ to their true values γ_0, (2) differentiability of the moment function $\psi(W, \gamma, \theta)$ in a neighborhood of θ_0, and (3) uniform convergence of the Jacobian over cross-fitting folds. These conditions ensure that the first-stage estimation of γ does not distort the asymptotic behavior of the estimator, even when machine learning methods are employed. Moreover, by controlling the sensitivity of the moment function to perturbations in both θ and γ, this assumption underpins the validity of inference in our cross-fitted setting.

Assumption 7-LR. *(Jacobian Stability) The Jacobian matrix* $G(\beta) := \mathbb{E}\left[\partial_\theta \psi(W, \gamma_0, \theta)\right]$ *exists, and there is a neighborhood* $\mathcal{N}$ *around* θ_0 *and a norm* $\|\cdot\|$ *such that the following conditions hold:*

1) *For each fold* ℓ*, the nuisance parameter estimate satisfies* $\|\hat{\gamma}^{(\ell)} - \gamma_0\| \xrightarrow{p} 0$*;*
2) *For all* $\|\gamma - \gamma_0\|$ *sufficiently small, the function* $\psi(W, \gamma, \theta)$ *is differentiable with respect to* θ *in* $\mathcal{N}$ *with probability approaching one. Moreover, there exists a constant* $C > 0$ *and a function* $d(W, \gamma)$ *such that for all* $\theta \in \mathcal{N}$ *and* $\|\gamma - \gamma_0\|$ *sufficiently small,*
$$\left\| \frac{\partial \psi(W, \gamma, \theta)}{\partial \beta} - \frac{\partial \psi(W, \gamma, \theta_0)}{\partial \beta} \right\| \leq d(W, \gamma) |\theta - \theta_0|^{1/C}, \quad \text{with } \mathbb{E}[d(W, \gamma)] < C;$$
3) *For each fold* $\ell = 1, \ldots, L$*,*
$$\int \left\| \frac{\partial \psi(w, \hat{\gamma}^{(\ell)}, \theta_0)}{\partial \beta} - \frac{\partial \psi(w, \gamma_0, \theta_0)}{\partial \beta} \right\| dF_0(w) \xrightarrow{p} 0.$$

Having established the assumptions for asymptotic normality, we now turn to derive a closed form solution for the asymptotic variance of $\sqrt{n}(\hat{\theta} - \theta_0)$, which according to Lemma 2 is given by

$$V = \left(G' \Upsilon G\right)^{-1},$$

where

$$G = \mathbb{E}\left[\frac{\partial g(W, \gamma, \theta)}{\partial \theta}\right] = \mathbb{E}\left[\begin{pmatrix} \frac{\partial g_1(W,\gamma,\theta)}{\partial \beta} & \frac{\partial g_1(W,\gamma,\theta)}{\partial \mu_c^P} & \frac{\partial g_1(W,\gamma,\theta)}{\partial \mu_f^P} \\ \frac{\partial g_2(W,\gamma,\theta)}{\partial \beta} & \frac{\partial g_2(W,\gamma,\theta)}{\partial \mu_c^P} & \frac{\partial g_2(W,\gamma,\theta)}{\partial \mu_f^P} \\ \frac{\partial g_3(W,\gamma,\theta)}{\partial \beta} & \frac{\partial g_3(W,\gamma,\theta)}{\partial \mu_c^P} & \frac{\partial g_3(W,\gamma,\theta)}{\partial \mu_f^P} \end{pmatrix}\right]$$

is the Jacobian matrix, and Υ is the efficient weighting matrix defined as $\Upsilon = \Psi^{-1}$,

$$\Psi = \frac{1}{n} \sum_{\ell=1}^{L} \sum_{i \in \mathcal{I}_\ell} \sum_{(tj) \in \mathcal{J}_i} \psi_{i,tj}^{(\ell)} \psi_{i,tj}^{(\ell)\prime}.$$

For the identifying moments in equations (B9), (B10), and (B11), we have

$$\frac{\partial g_1(W, \gamma, \theta)}{\partial \beta} = \sum_{|t-j| \leq h} \sigma_{tj}\left(\boldsymbol{X}_{ftj}, 1, 1\right) + \sum_{|t-j| > h} \left(\mu_{ft}(\boldsymbol{X}_{ft}, 1) - \mu_f^P\right)\left(\mu_{fj}(\boldsymbol{X}_{fj}, 1) - \mu_f^P\right),$$

$$\frac{\partial g_1(W, \gamma, \theta)}{\partial \mu_c^P} = -\sum_{j=1}^{T} \left(\mu_{fj}(\boldsymbol{X}_{fj}, 1) - \mu_f^P\right),$$

$$\frac{\partial g_1(W, \gamma, \theta)}{\partial \mu_f^P} = -\beta \sum_{t=1}^{T} \sum_{j=1}^{T} \left(\mu_{ft}(\boldsymbol{X}_{fj}, 1) + \mu_{fj}(\boldsymbol{X}_{fj}, 1) - 2\mu_f^P\right) + \sum_{t=1}^{T} \left(\mu_{ct}(\boldsymbol{X}_{ct}, 1) - \mu_c^P\right),$$

$$\frac{\partial g_2(W,\gamma,\theta)}{\partial \beta} = 0, \quad \frac{\partial g_2(W,\gamma,\theta)}{\partial \mu_c^P} = -1, \quad \frac{\partial g_2(W,\gamma,\theta)}{\partial \mu_f^P} = 0, \quad \frac{\partial g_3(W,\gamma,\theta)}{\partial \beta} = 0,$$
$$\frac{\partial g_3(W,\gamma,\theta)}{\partial \mu_c^P} = 0, \quad \frac{\partial g_3(W,\gamma,\theta)}{\partial \mu_f^P} = -1,$$
(B24)

where we have used

$$\begin{aligned}
\sigma_{tj}(X_{tj}, D_{ft}, D_{fj}) &= \mathbb{E}\left[Y_{ft}Y_{fj} - \mu_f^P(Y_{ft} + Y_{fj}) + \left(\mu_f^P\right)^2 \mid \boldsymbol{X}_{ftj}, D_{ft}, D_{fj}\right] \\
&= \mathbb{E}\left[Y_{ft}Y_{fj} \mid \boldsymbol{X}_{ftj}, D_{ft}, D_{fj}\right] + \left(\mu_f^P\right)^2 \\
&- \mu_f^P\left(\mathbb{E}\left[Y_{ft} \mid \boldsymbol{X}_{ftj}, D_{ft}, D_{fj}\right] + \mathbb{E}\left[Y_{fj} \mid \boldsymbol{X}_{ftj}, D_{ft}, D_{fj}\right]\right) \\
&= \mathbb{E}\left[Y_{ft}Y_{fj} \mid \boldsymbol{X}_{ftj}, D_{ft}, D_{fj}\right] + \left(\mu_f^P\right)^2 \\
&- \mu_f^P\left(\mathbb{E}\left[Y_{ft} \mid \boldsymbol{X}_{ftj}\right] + \mathbb{E}\left[Y_{fj} \mid \boldsymbol{X}_{ftj}\right]\right) \\
&= \mathbb{E}\left[Y_{ft}Y_{fj} \mid \boldsymbol{X}_{ftj}, D_{ft}, D_{fj}\right] + \left(\mu_f^P\right)^2 \\
&- \mu_f^P\left(\mathbb{E}\left[Y_{ft} \mid \boldsymbol{X}_{ft}\right] + \mathbb{E}\left[Y_{fj} \mid \boldsymbol{X}_{fj}\right]\right) \\
&= \mathbb{E}\left[Y_{ft}Y_{fj} \mid \boldsymbol{X}_{ftj}, D_{ft}, D_{fj}\right] + \left(\mu_f^P\right)^2 \\
&- \mu_f^P\left(\mathbb{E}\left[Y_{ft} \mid \boldsymbol{X}_{ft}, D_{ft} = 1\right] + \mathbb{E}\left[Y_{fj} \mid \boldsymbol{X}_{fj}, D_{fj} = 1\right]\right),
\end{aligned}$$

to find $\frac{\partial g_1(W,\gamma,\theta)}{\partial \mu_f^P}$. In the last expression, the third and fourth equality follow by the MAR Assumption 2-NP.*iii*, and Assumption 1-NP.*i*, which ensures $\mathbb{E}[Y_{ft} \mid \boldsymbol{X}_{ftj}] = \mathbb{E}[Y_{ft} \mid \boldsymbol{X}_t]$. Finally, the last equation also uses the MAR Assumption 2-NP.*iii*.

Combining the results above, we obtain the closed-form solution for Jacobian, and thus for the asymptotic variance in Lemma 2. Accordingly, the asymptotic variance can be estimated as

$$\hat{V} = \left(\hat{G}'\hat{\Upsilon}\hat{G}\right)^{-1},$$

where $\hat{G}$ is a consistent estimator of the Jacobian characterized by equation (B24), and $\hat{\Upsilon} = \hat{\Psi}^{-1}$, where

$$\hat{\Psi} = \frac{1}{n}\sum_{\ell=1}^{L}\sum_{i\in\mathcal{I}_\ell}\sum_{(tj)\in\mathcal{J}_i}\hat{\psi}_{i,tj}^{(\ell)}\hat{\psi}_{i,tj}^{(\ell)\prime}.$$

B5. *Derivation of the Wald Tests for Assumptions 1-NP.ii and 1-NP.iii*

We begin by constructing a test for the orthogonality condition between children's income prediction errors and parental lifetime income. The formal hypothesis is specified as:

(B25)
$$H_0 : \mathbb{E}\left[\epsilon_{ct} Y_f^P\right] = 0, \quad \forall t = 1, \ldots, T \quad \text{vs} \quad H_1 : \mathbb{E}\left[\epsilon_{ct} Y_f^P\right] \neq 0 \text{ for some } t \in \{1, \ldots, T\},$$

where $\epsilon_{ct} \coloneqq Y_{ct} - \mathbb{E}\left[Y_{ct} \mid \boldsymbol{X}_{ct}\right]$ denotes the children's income prediction errors at time t and Y_f^P represents parental lifetime income. The main challenge in testing this hypothesis is that both random variables are unobserved, and their machine learning estimation introduces regularization and model selection bias when testing H_0. To address these issues, we propose a three stages procedure. First, we establish identification of the object of interest $\theta_{cft} \coloneqq \mathbb{E}\left[\epsilon_{ct} Y_f^P\right]$. Second, we construct a locally robust estimator $\hat{\theta}_{cft}$. Finally, we provide a Wald test based on $\hat{\theta}_{cft,n} = \left(\theta_{cft,n}, \ldots, \theta_{cfT,n}\right)$.

The assumptions required for identifying θ_{cft} differ from those needed for identifying the IGE. Accordingly, we now present variants of Assumptions 1-NP and 2-NP.

Assumption 1-NP'. *(Conditional Mean Independence) The observable characteristics satisfy:*
$$\mathbb{E}\left[Y_{ft} \mid \boldsymbol{X}_{ft}, \boldsymbol{X}_{cj}\right] = \mathbb{E}\left[Y_{ft} \mid \boldsymbol{X}_{ft}\right] \quad \textit{for}, \quad t, j = 1, \ldots T.$$

Assumption 2-NP'. *(Missing At Random)*

i. *The missingness of children's and parents annual income Y_{gt} is as good as random once we control for $\boldsymbol{X}_{gt}$*
$$Y_{gt} \perp D_{gt} \mid \boldsymbol{X}_{gt}, \quad t = 1, \ldots, T.$$

ii. *Given family characteristics, there is both missing and non-missing children and fathers' incomes for every age*
$$0 < p\left(D_{ct} = 1 \mid \boldsymbol{X}_{ct}\right) < 1 \quad a.s, \quad t = 1, \ldots, T.$$

iii. *The missingness of child–parent income pairs is as good as random once we control for covariates:*
$$\left(Y_{ct}, Y_{fj}\right) \perp \left(D_{ct}, D_{fj}\right) \mid \left(\boldsymbol{X}_{ct}, \boldsymbol{X}_{fj}\right), \quad t, j = 1, \ldots, T.$$

iv. *Given covariates, there is both missing and non-missing child–parent income pairs:*
$$0 < \Pr\left(D_{ct} = 1, D_{fj} = 1 \mid \boldsymbol{X}_{ct}, \boldsymbol{X}_{fj}\right) < 1 \quad a.s., \quad t, j = 1, \ldots, T.$$

With this variants of the assumptions in place, we now show identification of θ_{cft}:
$$\theta_{cft} = \mathbb{E}\left[\epsilon_{ct} Y_f^P\right]$$

$$
\begin{aligned}
&= \mathbb{E}\left[\epsilon_{ct}\sum_{j=1}^{T} Y_{fj}\right] \\
&= \sum_{j=1}^{T}\mathbb{E}\left[\left(Y_{ct}-\mathbb{E}\left[Y_{ct}\mid \boldsymbol{X}_{ct}\right]\right)Y_{fj}\right] \\
&= \sum_{j=1}^{T}\mathbb{E}\left[Y_{ct}Y_{fj}\right]-\mathbb{E}\left[\mathbb{E}\left[Y_{ct}\mid \boldsymbol{X}_{ct}\right]Y_{fj}\right] \\
&= \sum_{j=1}^{T}\mathbb{E}\left[\mathbb{E}\left[Y_{ct}Y_{fj}\mid \boldsymbol{X}_{ct},\boldsymbol{X}_{fj}\right]\right]-\mathbb{E}\left[\mathbb{E}\left[Y_{ct}\mid \boldsymbol{X}_{ct}\right]\mathbb{E}\left[Y_{fj}\mid \boldsymbol{X}_{ct},\boldsymbol{X}_{fj}\right]\right] \\
&= \sum_{j=1}^{T}\mathbb{E}\left[\mathbb{E}\left[Y_{ct}Y_{fj}\mid \boldsymbol{X}_{ct},\boldsymbol{X}_{fj},D_{ct}=1,D_{fj}=1\right]\right] \\
&- \mathbb{E}\left[\mathbb{E}\left[Y_{ct}\mid \boldsymbol{X}_{ct},D_{ct}=1\right]\mathbb{E}\left[Y_{fj}\mid \boldsymbol{X}_{fj},D_{fj}=1\right]\right] \\
&:= \mathbb{E}\left[\sum_{j=1}^{T}\left(\mu_{cftj}\left(\boldsymbol{X}_{ct},\boldsymbol{X}_{fj},1,1\right)-\mu_{ct}\left(\boldsymbol{X}_{ct},1\right)\mu_{fj}\left(\boldsymbol{X}_{fj},1\right)\right)\right], \qquad \text{(B26)}
\end{aligned}
$$

where the second equality follows by the definition of lifetime income. The fifth equality follows by LIE, while the sixth one follows by Assumption 2-NP'.*iv*.

Having established identification, we now construct a moment for θ_{cft} that is locally robust to the nuisance parameters $\gamma_{cft} := \left(\mu_f^{1,T}, \mu_{ct}, \mu_{cftj}\right)$, where $:= \left(\mu_{f1}, ..., \mu_{fT}\right)$. Building on the Riesz representer characterization for μ_{ct} in equation (B14) and following the arguments from Appendix B3, the first-step influence function for μ_{ct} in the the identifying moment in equation (B26) is

$$
-\sum_{j=1}^{T}\mu_{fj}\left(\boldsymbol{X}_{fj},1\right)\frac{D_{ct}}{p\left(D_{ct}=1|\boldsymbol{X}_{ct}\right)}\left(Y_{ct}-\mu_{ct}\left(\boldsymbol{X}_{ct},1\right)\right).
$$

Similarly, for μ_{ft}, we have

$$
-\mu_{ct}\left(\boldsymbol{X}_{ct},1\right)\sum_{j=1}^{T}\frac{D_{fj}}{p\left(D_{fj}=1|\boldsymbol{X}_{fj}\right)}\left(Y_{fj}-\mu_{fj}\left(\boldsymbol{X}_{fj},1\right)\right).
$$

Following the same argument in equation (B17), the FSIF for μ_{cftj} is given by

$$
\sum_{j=1}^{T}\frac{D_{ct}D_{fj}}{p\left(D_{ct}=1,D_{fj}=1|\boldsymbol{X}_{ct},\boldsymbol{X}_{fj}\right)}\left(Y_{ct}Y_{fj}-\mu_{cftj}\left(\boldsymbol{X}_{ct},\boldsymbol{X}_{fj},1,1\right)\right)
$$

Accordingly, the locally robust moment for θ_{cft} is given by

$$\begin{aligned}
\psi_{cft}\left(W,\gamma_{cft},\theta_{cft}\right) &= g_{cft}\left(W,\gamma_{cft},\theta_{cft}\right)+\phi_{cft}\left(W,\gamma_{cft},\alpha_{cft},\theta_{cft}\right)\\
&=\sum_{j=1}^{T}\left(\mu_{cftj}\left(\boldsymbol{X}_{ct},\boldsymbol{X}_{fj},1,1\right)-\mu_{ct}\left(\boldsymbol{X}_{ct},1\right)\mu_{fj}\left(\boldsymbol{X}_{fj},1\right)\right)-\theta_{cft}\\
&-\sum_{j=1}^{T}\mu_{fj}\left(\boldsymbol{X}_{fj},1\right)\frac{D_{ct}}{p\left(D_{ct}=1|\boldsymbol{X}_{ct}\right)}\left(Y_{ct}-\mu_{ct}\left(\boldsymbol{X}_{ct},1\right)\right)\\
&-\mu_{ct}\left(\boldsymbol{X}_{ct},1\right)\sum_{j=1}^{T}\frac{D_{fj}}{p\left(D_{fj}=1|\boldsymbol{X}_{fj}\right)}\left(Y_{fj}-\mu_{fj}\left(\boldsymbol{X}_{fj},1\right)\right)\\
&+\sum_{j=1}^{T}\frac{D_{ct}D_{fj}}{p\left(D_{ct}=1,D_{fj}=1|\boldsymbol{X}_{ct},\boldsymbol{X}_{fj}\right)}\left(Y_{ct}Y_{fj}-\mu_{cftj}\left(\boldsymbol{X}_{ct},\boldsymbol{X}_{fj},1,1\right)\right).
\end{aligned}$$

Thus, the locally robust moment for θ_{cf} is given by

$$\psi_{cf}\left(W,\gamma_{cf},\theta_{cf}\right)=\left(\psi_{cf1}\left(W,\gamma_{cf1},\theta_{cf1}\right),\cdots,\psi_{cfT}\left(W,\gamma_{cfT},\theta_{cfT}\right)\right),$$

and the debiased moment function is then computed as

$$\hat{\psi}_{cf}\left(\theta_{cf}\right)=\frac{1}{n}\sum_{\ell=1}^{L}\sum_{f\in\mathcal{F}_\ell}\sum_{i\in\mathcal{P}_f}\sum_{(t,j)\in\mathcal{J}_i}\hat{\psi}^{(\ell)}_{cf,i,t,j},\quad \hat{\psi}^{(\ell)}_{i,tj}:=g_{cf}(W_{i,tj},\hat{\gamma}^{(\ell)}_{cft},\theta_{cft})+\phi_{cf}(W_{i,tj},\hat{\gamma}^{(\ell)}_{cft},\hat{\alpha}^{(\ell)}_{cft},\theta_{cft}),$$

where $\mathcal{J}_i$ denotes the set of all tuples (t,j) observed for child–father pair i. Since the system is exactly identified, there is no need to compute fold-specific $\hat{\theta}^{(\ell)}_{cft}$. The locally robust estimator of θ_{cf} is thus obtained by solving

$$\hat{\theta}_{cf,n}=\arg\min_{\theta_{cf}\in\Theta_{cf}\subset\mathbb{R}^T}\hat{\psi}\left(\theta_{cf}\right)'\hat{\Upsilon}_{cf}\hat{\psi}\left(\theta_{cf}\right).$$

where $\hat{\Upsilon}_{cf}$ is a positive semi-definite weighting matrix, and Θ_{cf} denotes the set of parameter values.

To test the joint null hypothesis $H_0:\theta_{cf}=\mathbf{0}$ for $\theta_{cf0}=(\theta_{cf10},\ldots,\theta_{cfT0})'$, we implement a Wald test based on the estimator $\hat{\theta}_{cf,n}=\left(\hat{\theta}_{cf1,n},\ldots,\hat{\theta}_{cfT,n}\right)'$. Accordingly, the Wald statistic is given by

$$W_{cf,n}=n\hat{\theta}_{cf,n}\hat{V}^{-1}_{cf,n}\hat{\theta}_{cf,n},$$

where $\hat{V}_{cf,n}$ is a consistent estimator of the asymptotic variance of $\hat{\theta}_{cf,n}$.

Consistency of $\hat{\theta}_{cft,n}$ follows by Lemma 1. In particular, under Assumptions 1-NP', 2-NP' and C-NP, we have

$$\hat{\theta}_{cf,n} \xrightarrow{p} \theta_{cf0}.$$

Similarly, under Assumptions 1-NP', 2-NP', 3-LR-5-LR and 7-LR, $\hat{\theta}_n^{LR} \xrightarrow{p} \theta_0$, and non-singularity of Υ, the asymptotic normality of $\hat{\theta}_{cfn}$ directly follows from Theorem 9 of Chernozhukov et al. (2022). Specifically, we have:

$$\sqrt{n}\left(\hat{\theta}_{cf,n} - \theta_{cf0}\right) \xrightarrow{d} \mathcal{N}(0, V_{cf}),$$

where $V_{cf} = \Upsilon_{cf}^{-1}$, since $G_{cf} = \mathbb{E}[\partial_{\theta_{cf}} g_{cf}(W, \gamma_{cft}, \alpha, \theta_{cf})] = -1$, and $\hat{\Upsilon}_{cf}$ is the estimated efficient weighting matrix defined as $\hat{\Upsilon}_{cf} = \hat{\Psi}_{cf}^{-1}$ for

$$\hat{\Psi}_{cf} = \frac{1}{n} \sum_{\ell=1}^{L} \sum_{i \in \mathcal{I}_\ell} \sum_{(tj) \in \mathcal{J}_i} \hat{\psi}_{cf,i,t,j}^{(\ell)} \hat{\psi}_{cf,i,t,j}^{(\ell)\prime}.$$

In addition, if Assumption 6-LR holds, then $\hat{V}_{cf,n} \xrightarrow{p} V_{cf}$.

Under H_0 in equation (B25) and Assumptions 1-NP', 2-NP', 3-LR-5-LR, we have

$$\sqrt{n}\hat{\theta}_{cfn} = \sqrt{n}\left(\hat{\theta}_{cf,n} - \theta_{cf0}\right) \xrightarrow{d} \mathcal{N}(0, V_{cf}) \Rightarrow n\hat{\theta}_{cfn} V_{cf}^{-1} \hat{\theta}_{cf,n} \xrightarrow{d} \chi_T^2.$$

Moreover, if we further impose Assumption 6-LR, $\hat{V}_{cf,n} = V_{cf} + o_p(1)$, so that

$$\begin{aligned}
W_{cf,n} &= n\hat{\theta}_{cfn} \hat{V}_{cf,n}^{-1} \hat{\theta}_{cf,n} \\
&= n\hat{\theta}_{cfn} V_{cf}^{-1} \hat{\theta}_{cfn} + \sqrt{n}\hat{\theta}_{cfn} \left(\hat{V}_{cf,n}^{-1} - V_{cf}^{-1}\right) \sqrt{n}\hat{\theta}_{cfn} \\
&= n\hat{\theta}_{cfn} V_{cf}^{-1} n\hat{\theta}_{cfn} + O_p(1) o_p(1) O_p(1) \\
&= n\hat{\theta}_{cfn} V_{cf}^{-1} n\hat{\theta}_{cfn} + o_P(1) \\
&\xrightarrow{d} \chi_T^2.
\end{aligned}$$

Having established the asymptotic distribution of the test, we now turn to show that under Assumptions 1-NP', 2-NP', 3-LR-7-LR and C-NP, the test that rejects H_0 when $W_{cf,n} > \chi^2_{T,1-\alpha}$ (the $(1-\alpha)$–quantile of the χ_T^2 distribution) is consistent.

$$\begin{aligned}
W_{cf,n} &= n\hat{\theta}_{cfn} \hat{V}_{cf,n}^{-1} \hat{\theta}_{cf,n} \\
&= \left(\sqrt{n}\theta_{cf0} + O_p(1)\right)' \left(V_{cf}^{-1} + o_p(1)\right)\left(\sqrt{n}\theta_{cf0} + O_p(1)\right) \\
&= n\theta_{cf0}' V_{cf}^{-1} \theta_{cf0} + 2\sqrt{n}\theta_{cf0}' V_{cf}^{-1} O_p(1) + O_p(1)' V_{cf}^{-1} O_p(1) \\
&= n\theta_{cf0}' V_{cf}^{-1} \theta_{cf0} + O_p\left(\sqrt{n}\right)
\end{aligned}$$

Thus, under H_1 where $\theta_{cf0} \neq 0$:

$$\Pr\left(W_{cf,n} > \chi^2_{T,1-\alpha} \mid \theta_{cf0}\right) = \Pr\left(n\theta'_{cf0} V_{cf}^{-1} \theta_{cf0} + O_p\left(\sqrt{n}\right) > \chi^2_{T,1-\alpha}\right) \to 1,$$

since $n\theta'_{cf0} V_{cf}^{-1} \theta_{cf0}$ is strictly positive and diverges at rate n, it dominates the $O_p\left(\sqrt{n}\right)$ term as $n \to \infty$.

Having established the consistency of the test, we now analyze its behavior under local alternatives of the form

$$H_{1n} : \theta_{cf0} = \frac{\delta}{\sqrt{n}}, \quad \delta \in \mathbb{R}^T \text{ fixed.}$$

Under Assumptions 1-NP', 2-NP', 3-LR-7-LR and C-NP, we have

$$\sqrt{n}\left(\hat{\theta}_{cf,n} - \theta_{cf0}\right) \xrightarrow{d} \mathcal{N}\left(0, V_{cf}\right), \quad \hat{V}_{cf,n} \xrightarrow{p} V_{cf}.$$

Therefore,

$$\begin{aligned} W_{cf,n} &= n\hat{\theta}_{cfn} \hat{V}_{cf,n}^{-1} \hat{\theta}_{cf,n} \\ &= \left(\sqrt{n}\theta_{cf0} + \sqrt{n}\left(\hat{\theta}_{cf,n} - \theta_{cf0}\right)\right)' \left(V_{cf}^{-1} + o_p(1)\right)\left(\sqrt{n}\theta_{cf0} + \sqrt{n}\left(\hat{\theta}_{cf,n} - \theta_{cf0}\right)\right) \\ &= \left(\delta + O_p(1)\right)' V_{cf}^{-1} \left(\delta + O_p(1)\right) + o_p(1). \end{aligned}$$

By Slutsky's theorem,

$$W_{cf,n} \Rightarrow d\chi^2_T(\lambda), \quad \lambda = \delta' V_{cf}^{-1} \delta,$$

where $\chi^2_T(\lambda)$ denotes the non-central chi-square distribution with T degrees of freedom and non-centrality parameter λ.

Since $\lambda > 0$ whenever $\delta \neq 0$,

$$\Pr\left(W_{cf,n} > \chi^2_{T,1-\alpha} \mid \theta_{cf0} = \frac{\delta}{\sqrt{n}}\right) \to 1 - F_{\chi^2_T(\lambda)}\left(\chi^2_{T,1-\alpha}\right) > \alpha,$$

where $F_{\chi^2_T(\lambda)}$ is the CDF of the non-central $\chi^2_T(\lambda)$ distribution. Thus, the Wald test has asymptotic power strictly greater than its size against any local alternative with $\delta \neq 0$, and the power approaches one under fixed alternatives.

We now apply the same procedure to test whether parental income prediction errors are uncorrelated for observations separated by more than h years. In particular, we are interested in testing the assumption

$$H_0 : \mathbb{E}\left[\epsilon_{ft}\, \epsilon_{fj}\right] = 0 \quad \text{for all } |t - j| > h, \quad \text{vs} \quad H_1 : \mathbb{E}\left[\epsilon_{ft}\, \epsilon_{fj}\right] \neq 0 \quad \text{for some } |t - j| > h,$$

where the prediction errors of parental income at time t are defined as $\epsilon_{ft} := Y_{ft} -$

$\mathbb{E}\Big[Y_{ft}\,\Big|\,\boldsymbol{X}_{ft}\Big]$. However, joint observation of parental incomes (Y_{ft}, Y_{fj}) (i.e., $D_{ft} = 1, D_{fj} = 1$) occurs only for relatively close time periods, such as incomes observed between ages 25 and 35 for a given individual. Consequently, income pairs for distant periods ($|t - j| > h$) are systematically absent in available data. To address this limitation, we adopt the No Re-emergence of Dependence assumption, which posits that if the autocorrelation at lag $(h + 1)$ is negligible, then it remains negligible for all higher lags. In particular, we formalize it as

Assumption 8-NP". *(No Re-emergence of Dependence) Let $\{\epsilon_{ft}\}$ be covariance-stationary with autocorrelation function $\rho(k)$, $k \geq 1$. Assume:*

1) ***Bounded tail dependence:*** *There exists a nonincreasing sequence $u(k) \to 0$ such that*
$$|\rho(k)| \leq u(k) \quad \textit{for all } k \geq 1.$$
2) ***No re-emergence:*** *If $|\rho(h + 1)|$ is negligible (statistically indistinguishable from zero), then $|\rho(k)|$ is negligible for all $k > h + 1$.*

Under this assumption, we can test Assumption 1-NP.*iii* with the hypothesis

$$H_0 : \mathbb{E}[\epsilon_{ft}\,\epsilon_{fj}] = 0 \quad \text{for all } |t - j| = h + 1, \quad \text{vs} \quad H_1 : \mathbb{E}[\epsilon_{ft}\,\epsilon_{fj}] \neq 0 \quad \text{for some } |t - j| = h + 1. \tag{B27}$$

We now state the Assumptions required to identify $\theta_{ftj} := \mathbb{E}\left[\epsilon_{ft}\,\epsilon_{fj}\right]$

Assumption 1-NP". *(Conditional Mean Independence and Orthogonality) The observable characteristics satisfy*
$$\mathbb{E}\left[Y_{ft} \mid \boldsymbol{X}_{ft}, \boldsymbol{X}_{ftj}\right] = \mathbb{E}\left[Y_{ft} \mid \boldsymbol{X}_{ft}\right] \quad \textit{for } \boldsymbol{X}_{fj} \subset \boldsymbol{X}_{ftj}, \quad t, j = 1, ...T.$$

Assumption 2-NP". *(Missing At Random)*

i. *The missingness of parental annual income pairs $\left(Y_{ft}, Y_{fj}\right)$ is as good as random once we control for $\boldsymbol{X}_{ftj}$*
$$\left(Y_{ft}, Y_{fj}\right) \perp \left(D_{ft}, D_{fj}\right) \mid \boldsymbol{X}_{ftj}, \quad \textit{for all } t - j > h > 0,$$
where $\boldsymbol{X}_{ftj}$ are the family characteristics predictive of parental income covariance between years t and j, and $\boldsymbol{X}_{ftj} := \boldsymbol{X}_{ft}$ for $j = t$.

ii. *Given family characteristics, there is both missing and non-missing parental incomes for every age and its neighboring ages*
$$0 < p\left(D_{ft} = 1, D_{fj} = 1 \mid \boldsymbol{X}_{ftj}\right) < 1 \quad \textit{a.s.}, \quad \textit{for all } t - j > h > 0.$$

With the assumptions in place, we now show identification of θ_{ftj}:

$$
\begin{aligned}
\theta_{ftj} &= \mathbb{E}\left[\epsilon_{ft}\epsilon_{fj}\right] \\
&= \mathbb{E}\left[\left(Y_{ft} - \mathbb{E}\left[Y_{ft} \mid \boldsymbol{X}_{ft}\right]\right)\left(Y_{fj} - \mathbb{E}\left[Y_{fj} \mid \boldsymbol{X}_{fj}\right]\right)\right] \\
&= \mathbb{E}\left[Y_{ft}Y_{fj} - \mathbb{E}\left[Y_{ft} \mid \boldsymbol{X}_{ft}\right]Y_{fj} - Y_{ft}\mathbb{E}\left[Y_{fj} \mid \boldsymbol{X}_{fj}\right] + \mathbb{E}\left[Y_{ft} \mid \boldsymbol{X}_{ft}\right]\mathbb{E}\left[Y_{fj} \mid \boldsymbol{X}_{fj}\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[Y_{ft}Y_{fj} - \mathbb{E}\left[Y_{ft} \mid \boldsymbol{X}_{ft}\right]Y_{fj} - Y_{ft}\mathbb{E}\left[Y_{fj} \mid \boldsymbol{X}_{fj}\right] + \mathbb{E}\left[Y_{ft} \mid \boldsymbol{X}_{ft}\right]\mathbb{E}\left[Y_{fj} \mid \boldsymbol{X}_{fj}\right] \mid \boldsymbol{X}_{tfj}\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[Y_{ft}Y_{fj} \mid \boldsymbol{X}_{tfj}\right]\right] - \mathbb{E}\left[\mathbb{E}\left[Y_{ft} \mid \boldsymbol{X}_{ft}\right]\mathbb{E}\left[Y_{fj} \mid \boldsymbol{X}_{fj}\right]\right] \\
&- \mathbb{E}\left[\mathbb{E}\left[Y_{ft} \mid \boldsymbol{X}_{ft}\right]\mathbb{E}\left[Y_{fj} \mid \boldsymbol{X}_{fj}\right]\right] + \mathbb{E}\left[\mathbb{E}\left[Y_{ft} \mid \boldsymbol{X}_{ft}\right]\mathbb{E}\left[Y_{fj} \mid \boldsymbol{X}_{fj}\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[Y_{ft}Y_{fj} \mid \boldsymbol{X}_{tfj}\right]\right] - \mathbb{E}\left[\mathbb{E}\left[Y_{ft} \mid \boldsymbol{X}_{ft}\right]\mathbb{E}\left[Y_{fj} \mid \boldsymbol{X}_{fj}\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[Y_{ft}Y_{fj} \mid \boldsymbol{X}_{tfj}\right] - \mathbb{E}\left[Y_{ft} \mid \boldsymbol{X}_{ft}\right]\mathbb{E}\left[Y_{fj} \mid \boldsymbol{X}_{fj}\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[Y_{ft}Y_{fj} \mid \boldsymbol{X}_{ftj}, D_{ft} = 1, D_{fj} = 1\right] - \mathbb{E}\left[Y_{ft} \mid \boldsymbol{X}_{ft}, D_{ft} = 1\right]\mathbb{E}\left[Y_{fj} \mid \boldsymbol{X}_{fj}, D_{fj} = 1\right]\right] \\
&\coloneqq \mathbb{E}\left[\mu_{ftj}\left(\boldsymbol{X}_{ftj}, 1, 1\right) - \mu_{ft}\left(\boldsymbol{X}_{fj}, 1\right)\mu_{fj}\left(\boldsymbol{X}_{fj}, 1\right)\right].
\end{aligned} \tag{B28}
$$

The moment for θ_{ftj} that is locally robust to the nuisance parameter $\gamma_{ftj} \coloneqq \left(\mu_{ftj}, \mu_{ft}, \mu_{fj}\right)$ is given by

$$
\begin{aligned}
\psi_{ftj}\left(W, \gamma_{ftj}, \theta_{ftj}\right) = \mu_{ftj}\left(\boldsymbol{X}_{ftj}, 1, 1\right) - \mu_{ft}\left(\boldsymbol{X}_{fj}, 1\right)\mu_{fj}\left(\boldsymbol{X}_{fj}, 1\right) - \theta_{ftj} \\
+ \frac{D_{ft}D_{fj}}{p\left(D_{ft} = 1, D_{fj} = 1|\boldsymbol{X}_{ftj}\right)}\left(Y_{ft}Y_{fj} - \mu_{ftj}\left(\boldsymbol{X}_{ftj}, 1, 1\right)\right) \\
- \mu_{fj}\left(\boldsymbol{X}_{fj}, 1\right)\frac{D_{ft}}{p\left(D_{ft} = 1|\boldsymbol{X}_{ft}\right)}\left(Y_{ft} - \mu_{ft}\left(\boldsymbol{X}_{ft}, 1\right)\right) \\
- \mu_{ft}\left(\boldsymbol{X}_{ft}, 1\right)\frac{D_{fj}}{p\left(D_{fj} = 1|\boldsymbol{X}_{fj}\right)}\left(Y_{fj} - \mu_{fj}\left(\boldsymbol{X}_{fj}, 1\right)\right).
\end{aligned}
$$

Let θ_{fh} be the row vector stacking all θ_{ftj} for t and j such that $t - j = h + 1 < T$[24]:

$$\theta_{fh} = (\theta_{f,h+2,1},\ \theta_{f,h+3,2},\ \ldots,\ \theta_{f,T,T-h-1}).$$

Then, the locally robust moment for θ_{fh} is given by

$$\psi_{cf}\left(W, \gamma_{fh}, \theta_{fh}\right) = \left(\psi_{fh1}\left(W, \gamma_{fh1}, \theta_{fh1}\right), \cdots, \psi_{fhT}\left(W, \gamma_{fhT}, \theta_{fhT}\right)\right),$$

[24]since the covariance is symmetric, we do not need to test for all $|t - j| = h + 1 < T$, but rather for t and j such that $t - j = h + 1 < T$

and its corresponding estimator solves the cross-fitted orthogonal moment

$$\hat{\theta}_{fh,n} = \arg \min_{\theta_{fh} \in \Theta_{fh} \subset \mathbb{R}^{K_\theta}} \hat{\psi}\left(\theta_{fh}\right)' \hat{\Upsilon}_{fh} \hat{\psi}\left(\theta_{fh}\right),$$

where $\hat{\Upsilon}_{fh}$ is a positive semi-definite weighting matrix, Θ_{fh} denotes the set of parameter values, and K_θ is the dimension of θ_{fh}.

Consequently, the locally robust Wald test for H_0 in equation (B27) is given by

$$W_{fh,n} = n\hat{\theta}_{fh,n} \hat{V}_{fh,n}^{-1} \hat{\theta}_{fh,n},$$

where $\hat{V}_{fh,n}$ is a consistent estimator of the asymptotic variance of $\hat{\theta}_{fh,n}$. The asymptotic properties of $W_{fh,n}$ follow analogously from those established for $W_{cf,n}$ under Assumptions 1-NP", 2-NP", 3-LR-7-LR, C-NP, and 8-NP".