Nonparametric Identification and Locally Robust Estimation of the Intergenerational Elasticity of Income*

Alejandro Puerta-Cuartas†

September 29, 2025

Click here for the latest version of the paper

The unobservability of lifetime income poses a fundamental challenge for estimating the intergenerational elasticity (IGE), forcing researchers to rely on annual income averages. We show that such proxy-based estimators converge to context-specific limits, compromising comparability across studies. To address this, we establish nonparametric identification of the IGE using partial income data and family characteristics. Building on this foundation, we construct a consistent two-step estimator that combines machine learning with cross-fitting and Neyman-orthogonal moments, ensuring that the resulting IGE estimates are locally robust to first-stage errors. This approach delivers valid inference and overcomes long-standing challenges in achieving comparability of IGE estimates across studies, time, and place. Applying it to the Panel Study of Income Dynamics, proxy-based estimators yield values of 0.38 and 0.51 for the United States, while our method produces an estimate of 0.69, in line with evidence from long-run parental income averages indicating that the true value exceeds 0.6.

JEL: J62; D63; C13; C14

Keywords: Intergenerational mobility, inequality, semiparametric inference, missing data.

* I am grateful to Juan Carlos Escanciano and Jan Stuhler for their guidance and support. I also thank participants of the International Association for Applied Econometrics 2025 Annual Conference, the Eleventh Italian Congress of Econometrics and Empirical Economics, the ENTER Jamboree 2025 Conference, and the III at 10: New Directions in Inequality Research for their valuable feedback.

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

I. Introduction

Many economic and causal parameters depend on lifetime outcomes, such as income, earnings, or consumption. However, survey and administrative data typically cover only a limited segment of individuals' working lives, posing significant challenges across fields such as household economics, education, and labor. Specifically, this data limitation hinders the separation of transitory and permanent shocks to consumption and savings (Jappelli and Pistaferri, 2010), and introduces bias into estimates of long-run educational returns (Heckman et al., 2006) and intergenerational income mobility (Solon, 1992). This paper addresses the challenge of unobserved lifetime income in estimating the intergenerational elasticity (IGE) by establishing its nonparametric identification and developing a consistent, locally robust estimator.

The intergenerational elasticity has traditionally served as the primary measure of income persistence across generations (Nybom and Stuhler, 2017). It is formally defined as the slope coefficient from a regression of a child's permanent income on that of the parent. However, due to data constraints, researchers commonly rely on proxies such as annual income averages, which introduces measurement error on both sides of the regression. While bias typically stems from errors in the independent variable, the IGE context presents two further complications. First, errors in the dependent variable may bias estimates if income growth varies by parental background, as documented for children from affluent families, who tend to experience steeper growth (Halvorsen et al., 2022). Second, life-cycle bias stems from observing fathers at older, more stable ages than sons, whose early-career income is a noisier proxy for permanent income (Haider and Solon, 2006).

Recent literature has made substantial progress in refining IGE estimates by addressing specific sources of bias, either by improving proxies for parental (Mazumder, 2016) or child income (Mello et al., 2024). Nevertheless, questions remain about the robustness of existing estimates and the reliability of comparisons across time and place (Mogstad and Torsvik, 2023). We formalize these concerns by showing that the magnitude of each source of bias depends on research design, income dynamics, and sampling rules. As a result, proxy-based estimators converge to context-specific limits, rather than a common population IGE, compromising the reliability and comparability of results across studies. This underscores the need to move beyond proxy refinement and focus on identification, which addresses all sources of bias and enables reliable and comparable estimation of the IGE across studies.

This paper shows nonparametric identification of the intergenerational elasticity using partially observed income data and family characteristics. We define permanent income as lifetime average log income, consistent with its interpretation as the permanent component of log income (Solon, 1992; Mazumder, 2005). Under this definition, we show that the IGE can be recovered through a moment function that depends on nuisance parameters, including the conditional expectation of income of both parent and child, as well as the conditional autocovariance of parental annual income, given family characteristics and income observability. Our identification result is nonparametric, relying on a theoretically grounded definition of permanent income rather than imposing functional form assumptions on the relationship between permanent and annual income, or between annual income and observable characteristics.

Identification relies on standard missing-at-random (MAR) and orthogonality assumptions. Empirical evidence from the Panel Study of Income Dynamics (PSID), the dataset used for the application, suggests that the MAR assumption is plausible when relevant characteristics are included (Fitzgerald, 2011; Schoeni and Wiemers, 2015). To assess the orthogonality conditions, we construct and implement two locally robust tests, both of which are not rejected in the PSID, supporting the plausibility of the assumptions. We also provide evidence that the IGE estimates remain robust when permanent income is defined alternatively as the log of average lifetime income. Taken together, these findings provide empirical support for the credibility of the identifying assumptions and the robustness to the permanent income definition.

We develop a consistent and locally robust two-step estimator of the IGE, enabling reliable and comparable measurement of income persistence across studies, time, and place. In the first step, we estimate nuisance parameters using machine learning (ML) to flexibly capture the heterogeneous and nonlinear dynamics of income while reducing the risk of misspecification. To avoid overfitting and own-observation bias, we implement cross-fitting. In the second step, the IGE is computed using the first-stage estimates via a Neyman-orthogonal moment, which we construct to guarantee local robustness, meaning that small errors in the first stage have a negligible effect on the final estimate. This property is especially relevant in settings where ML is applied in the first stage, to mitigate regularization and model selection bias introduced by nonparametric methods (Chernozhukov et al., 2022). We establish the asymptotic normality for the proposed estimator and provide locally robust hypothesis tests for two of the identifying assumptions. Our simulations suggest the estimator exhibits sound finite-sample performance, with negligible bias that vanishes as sample size increases and coverage rates close to nominal levels.

Applying our method to the United States using the core PSID sample yields an IGE estimate of 0.69. This aligns with previous evidence based on long-term parental income averages, which suggests that the U.S. IGE is likely above 0.6 (Mazumder, 2016). In contrast, two alternative estimators using income proxies produce considerably lower estimates of 0.38 and 0.51. A naive machine learning plug-in estimator, which relies on the standard identifying equation rather than the Neyman-orthogonal adjustment, delivers a lower estimate of 0.60. 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. To facilitate implementation and reproducibility, a user-friendly `R` package, `LRIGE`, is currently under development.

The remainder of the paper is as follows: Section II formalizes that proxy-based estimators converge to context-specific limits, compromising the reliability and comparability of results across studies. Section III shows nonparametric identification of the IGE, provides a locally robust estimator, establishes asymptotic normality for the proposed estimator, and provides hypothesis tests for two of the identifying assumptions. 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. Biases and Comparability in Intergenerational Elasticity Estimates

Our object of interest is the intergenerational elasticity of income, which captures the degree to which income differences between parents are associated with income differences among their children. Formally, the IGE is defined by the regression

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

where Y_f^P and Y_c^P denote the permanent component of log annual income for fathers and children (Solon, 1992), u is an idiosyncratic error term uncorrelated with parental income, and β_0 is the intergenerational elasticity. Henceforth, we will refer to Y^P as permanent income.

According to equation (1), the closed form solution for the IGE is given by

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

However, because permanent income data is rarely available in practice, researchers typically rely on short-term income snapshots (Mazumder, 2005), using either a single-year observation or a multi-year average as a proxy for permanent income. This naturally raises the question of what exactly the available estimators in the literature measure, and whether their estimates are comparable across settings.

To address this, we adopt the strategy that Mogstad and Torgovitsky (2024) refer to as "reverse engineering". In the context of instrumental variables (IV), this approach begins with a practical problem: when treatment effects exhibit unobserved heterogeneity (UHTE), the classical linear IV model is misspecified. Yet a linear IV estimate can still be computed. The reverse engineering framework then seeks to determine what, if anything, this estimator measures. Thus, it proceeds by starting with the tool, and it attempts to reverse engineer an interpretation for it under suitable assumptions.

The idea of reverse engineering has already been applied in the intergenerational mobility literature to formally characterize the behavior of estimators based on proxy measures of permanent income (Solon, 1992; Nybom and Stuhler, 2016). Specifically, it has been used to establish the probability limit of such estimators, thereby identifying the distinct sources of bias introduced by relying on imperfect proxies.

To set the grounds for our analysis, we follow the literature by first characterizing the probability limit of the estimator based on income proxies. We begin by defining the observed data typically available to researchers. 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 consist of an independent and identically distributed (i.i.d.) sample of $\boldsymbol{W} = (\boldsymbol{Y}_c \odot \boldsymbol{D}_c, \boldsymbol{Y}_f \odot \boldsymbol{D}_f, \boldsymbol{D}_c, \boldsymbol{D}_f, \boldsymbol{X})$, where $\boldsymbol{Y}_c$ and $\boldsymbol{Y}_f$ are T-dimensional random vectors containing information on (log) annual child and parental income, respectively, the vectors $\boldsymbol{D}_c$ and $\boldsymbol{D}_f$ are T-dimensional indicator vectors, with elements $D_{gt} = 1$ if Y_{gt} is observed and $D_{gt} = 0$ otherwise, for $g \in \{c, f\}$, $\odot$ denotes the element-wise product, so that $\boldsymbol{Y}_g \odot \boldsymbol{D}_g$ contains the observed entries of $\boldsymbol{Y}_g$ and zeros elsewhere, and the vector $\boldsymbol{X}$ contains observed characteristics for both generations.

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 permanent income as the average of T_f and T_c (log) annual income observations around mid-life for the fathers and the children, respectively. In the second step, it regresses the child's proxy measure on the parent's.

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

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

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, and $T_g := \sum_{j\in\mathcal{M}_g} D_{gj}$ is the number of years used for the average.

To establish the probability limit of the MI estimator using the reverse engineering approach (Mogstad and Torgovitsky, 2024), we now impose standard assumptions used in the literature. A comprehensive discussion of the MI estimator's definition, theoretical underpinnings, assumptions, and sources of bias can be found in Appendix A2.

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

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

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

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

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

The inconsistency of the MI estimator is well-documented in the literature. Proposition 1 restates this result to make explicit the four sources of bias that will be central to our discussion. While the main characterization of these biases is familiar, the proposition extends prior work by incorporating missing income data. Appendix A4 shows that Proposition 1 reduces to the results of Solon (1992) and coincides with Nybom and Stuhler (2016) under certain assumption variants. In addition, the proposition formalizes the empirical observation that IGE estimates are sensitive to sample inclusion criteria and missing income, by demonstrating how the observation probabilities $p_f\left(t, j \in \mathcal{M}_f\right)$ and $p_c\left(t \in \mathcal{M}_c\right)$ shape the asymptotic bias.

Proposition 1. *Under Assumptions 1-MI and 2-MI, the probability limit of the MI estimator 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 as defined in Assumption 1-MI, and $p_c\left(t \in \mathcal{M}_c\right)$ and $p_f\left(\{t, j\} \in \mathcal{M}_f\right)$ denote the probabilities of observing child income at mid-life year t, and parent income at mid-life years in years t and j, respectively.

Proposition 1 presents a formal statement of the biases already familiar from prior work, including (a) the measurement error and life-cycle bias of parental income (Solon, 1992; Mazumder, 2005), (b) the sensitivity of the IGE estimates to low, zero, and missing parental income observations (Couch and Lillard, 1998; Dahl and DeLeire, 2008; Chetty et al., 2014; Nybom and Stuhler, 2016), (c) the measurement error and life-cycle bias of children's 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). For ease of exposition, the proposition is stated under a classical errors-in-variables formulation, so that the age-shocks are treated as capturing both transitory fluctuations and the systematic life-cycle bias. In Appendix A4 (equation A28) we relax this restriction and allow for a generalized errors-in-variables structure that explicitly separates the life-cycle component. A brief discussion of each component that hinders the consistent estimation of the IGE by $\hat{\beta}_n^{MI}$ can be found in Appendix A2.

Identifying the distinct biases introduced by imperfect proxies has enabled the literature to refine estimation procedures by addressing specific sources of bias. For example, Mazumder (2005) addresses the measurement error and life-cycle bias of parental income (component (a) in equation (3)) by using long-term parental income averages centered at age 40. More recently, Mello et al. (2024) address life-cycle bias from using snapshots of children's income, captured by terms (c) and (d), by predicting children's income profiles from standard observables such as age and education, while allowing income growth to be steeper for children from more affluent families.

Despite these advances, questions remain about the robustness of existing estimates and the reliability of comparisons across time and place (Mogstad and Torsvik, 2023; Mello et al., 2024). While the literature has rightly emphasized the sensitivity of IGE estimates to the biases in Proposition 1, a more fundamental issue is that these biases systematically alter the target parameter itself. Crucially, the magnitude of each bias component varies with income dynamics, research design, and sampling rules. Due to heterogeneity in institutional contexts and study design choices, these factors differ across datasets, regions, countries, and time, rendering IGE estimates non-comparable. The following corollary makes this dependence explicit by formalizing that the target parameter is inherently dependent on the research design and the underlying income dynamics.

Corollary 1.1 (Context-Dependence of the MI Estimand)**.** *Under the assumptions of Proposition 1, the probability limit of the Mid-Life Income estimator is a context-dependent parameter*

$$\hat{\beta}_n^{MI} \xrightarrow{p} \beta^{MI}(\eta) := \beta_0 \cdot \Delta(\eta), \tag{4}$$

where the distortion factor $\Delta(\eta) \neq 1$ *captures departure from consistency. Formally, the research-design/context vector*

$$\eta := (\mathcal{M}_f, \mathcal{M}_c, T_f, T_c, p, \Sigma),$$

collects all research-design choices and structural features of the income process: $\mathcal{M}_g$ *and* T_g *are the set of pre-defined mid-life years, and the number of years used for the average for generation* $g \in \{c, f\}$*, respectively;* $p = (p_f, p_c)$ *captures observation/availability and selection rules; and* Σ *summarizes how permanent income and transitory fluctuations in parents' and children's earnings* $\left(Y_f^P, v_{ft}, v_{ct}\right)$ *vary and relate to each other. Equation (3) provides the explicit form of* $\Delta(\eta)$*.*

Corollary 1.1 rationalizes concerns raised in the literature about the robustness of IGE estimates and the reliability of comparisons across studies, time, and place. Specifically, it shows that reliance on income proxies introduces systematic distortions, preventing identification of the true intergenerational elasticity. These distortions limit the reliability and comparability of resulting estimates across studies, as each converges to its own context-specific value $\beta^{MI}(\boldsymbol{\eta})$. A compelling example is the wide variation in recent U.S. estimates, which range from 0.35 to 0.65 (Mello et al., 2024). Corollary 1.1 makes explicit the potential drivers of this pattern. Even when components such as (a) and (c) in equation (3) are held constant, differences in the definition of mid-life income $\left(\mathcal{M}_g\right)$, the number of years averaged (T_g), or sample selection rules (p_g) alter the distortion factor $\Delta(\eta)$ in equation (4). As a result, each estimate converges to a different context-specific parameter $\beta^{MI}(\eta)$.

Distortions in $\Delta(\eta)$ also affect trend analyses, as both design choices and cohort-specific income dynamics can vary over time. In the U.S., the PSID's transition from annual (1968–1997) to biennial interviews illustrates how survey design changes can alter the estimand: reduced income observations for recent cohorts modify $\Delta(\eta)$ through the observation probability p_c, potentially distorting mobility trends. Empirical evidence from Sweden further supports the theoretical distortions highlighted in Corollary 1.1. Mello et al. (2024) show that MI-based estimates suggest a sharp decline in mobility for the 1950s–1970s cohorts, whereas their life-cycle estimator, which corrects for life-cycle bias in children's income, indicates stable mobility across these cohorts, and a modest increase for those born in the 1980s.

Finally, Corollary 1.1 formalizes how differences in study design and income dynamics can undermine cross-country comparisons. Even under the same definitions of mid-life income, differences in transitory shock persistence (a), children's income growth (c), and observation probabilities (b, d) alter the estimand $\beta^{MI}(\eta)$. This calls for caution in interpreting international patterns such as the Great Gatsby Curve: unlike scale-free measures like the Gini coefficient, MI-based IGE estimates reflect both underlying mobility and study-specific distortions captured by $\Delta(\eta)$.

Beyond its implications for common applications, Corollary 1.1 shows that eliminating individual biases alone does not guarantee comparability. Even after correcting for specific sources of bias, differences in study design, cohort composition, or the magnitude of residual distortions can still produce inconsistent estimates. This analysis highlights a fundamental shift in perspective: rather than addressing individual sources of bias, attention should be directed toward identifying the IGE, which simultaneously removes all biases and allows for reliable, comparable estimates.

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

A. *nonparametric Identification*

To establish identification of the intergenerational elasticity, we adopt the "forward-engineering" strategy proposed by Mogstad and Torgovitsky (2024). In their terminology, this approach begins with a model and then constructs estimators under the assumption that the model is correctly specified. In the context of the IGE, our proposal precisely follows this logic: begin by defining permanent income and then derive the conditions under which the IGE is identified, given that definition.

The literature has generally characterized permanent income rather than attempting to provide a precise definition. For instance, in line with Solon (1992), Mazumder (2005) interprets it as the permanent component of log earnings, capturing true long-term earning capacity. In contrast, Haider and Solon (2006) describes it as a long-run income variable, such as the log of the present discounted value of lifetime earnings. Other work (Black and Devereux, 2011; Corak, 2013) refer more generally to log permanent earnings without elaborating further. Reflecting this theoretical heterogeneity, empirical research has proxied permanent income differently; while some studies compute it as the log of average annual income (Dahl and DeLeire, 2008; Mazumder, 2016), others take the average of log annual income (Zimmerman, 1992; Bratberg et al., 2007).

It is crucial to clarify that our goal is not to define a unifying measure of permanent income or to identify a uniquely correct IGE. The intergenerational elasticity has never been directly observed; it has always been estimated under specific measurement choices and assumptions. Our objective is instead pragmatic: to adopt a definition of permanent income that is theoretically grounded, empirically tractable, and consistent with standard practice in the literature. By settling on a workable definition, researchers can generate estimates of the IGE that are more reliable and comparable across studies and datasets. In this sense, the forward-engineering approach provides a foundation for comparability and reliability, rather than a claim to recover the uniquely correct parameter.

Definition (Permanent income)**.** *For an individual of generation g (where $g \in \{c, f\}$ for child or father), permanent income $\left(Y_g^P\right)$ is defined as their average log annual income over a specific lifetime period from $t = 1$ to T:*

$$Y_g^P := \frac{1}{T} \sum_{t=1}^{T} Y_{gt}, \quad g \in \{c, f\}, \tag{5}$$

where Y_{gt} is log annual income in year t, with $t = 1$ indicating the start age and T the number of years covered.

Our definition of permanent income aligns with the literature that conceptualizes it as the permanent component of log earnings. More importantly, this definition provides an empirically tractable measure that facilitates the identification of the intergenerational elasticity. The linearity of the sum-of-logs specification is crucial, as it permits the use of standard missing-at-random (MAR) assumptions to recover permanent income from partially observed data. An alternative definition involving the log of the average introduces nonlinearities that preclude a similar identification strategy and require stronger assumptions about the joint distribution of income over lifetime. For a detailed discussion of these considerations, we refer the reader to Appendix A1. When applied to the life-cycle estimator of Mello et al. (2024), our definition yields results that are virtually identical to those obtained defining permanent income as the log of the average, as shown in Table A1.

The fundamental challenge in estimating (identifying) the IGE is its reliance on unobserved permanent income. Traditional approaches, such as the Generalized Error-in-Variables (GEIV) model (Haider and Solon, 2006), motivate the use of short-term averages of income—typically during mid-life—by assuming a parametric link between observed annual income and unobserved permanent income. In contrast, our definition underpins the nonparametric nature of our identification result, enabling us to recover the intergenerational elasticity without restrictive functional form assumptions.

While the IGE in equation (2) depends on unobserved permanent income for both generations, our definition allows us to reformulate the target parameter in terms of partially observed (log) annual incomes:

(6)
$$\beta_0 = \frac{\mathbb{E}\left[\left(Y_c^P - \mathbb{E}\left[Y_c^P\right]\right)\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)\right]}{\mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right]} = \frac{\sum_{t=1}^{T}\sum_{j=1}^{T}\mathbb{E}\left[\left(Y_{ct} - \mathbb{E}\left[Y_c^P\right]\right)\left(Y_{fj} - \mathbb{E}\left[Y_f^P\right]\right)\right]}{\sum_{t=1}^{T}\sum_{j=1}^{T}\mathbb{E}\left[\left(Y_{ft} - \mathbb{E}\left[Y_f^P\right]\right)\left(Y_{fj} - \mathbb{E}\left[Y_f^P\right]\right)\right]},$$

where the scaling component $1/T$ cancels out. 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

$$(7) \qquad Y_{gt} = \mathbb{E}\left[Y_{gt} \mid \boldsymbol{X}_{gt}\right] + \epsilon_{gt}, \quad \mathbb{E}\left[\epsilon_{gt} \mid \boldsymbol{X}_{gt}\right] = 0, \quad g \in \{c, f\}, \quad t = 1, \dots, T,$$

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, and ϵ_{ct} is the nonparametric prediction error. Although ϵ_{ct} can be interpreted as an age shock, it differs conceptually from the age shock v_{ct} in the GEIV model of Assumption 1-MI. 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 permanent income and arises within a latent factor structure that distinguishes between the permanent and transitory components of income.

To establish identification of the IGE, we proceed in two steps. First, we substitute the income decomposition into equation (6) 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 expectation $\mathbb{E}[Y_{gt} \mid \boldsymbol{X}_{gt}] = \mathbb{E}[Y_{gt} \mid \boldsymbol{X}_{gt}, D_{gt} = 1]$, where D_{gt} indicates income observability at time t. In a similar way, we are able to identify the conditional second moments arising in the denominator of equation (6) $\mathbb{E}[Y_{ft}Y_{fj} \mid \boldsymbol{X}_{ftj}] = \mathbb{E}[Y_{ft}Y_{fj} \mid \boldsymbol{X}_{ftj}, D_{ft} = 1, D_{fj} = 1]$, 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. $\mathbb{E}\left[Y_{ct} \mid \boldsymbol{X}_{ct}, \boldsymbol{X}_{cj}, \boldsymbol{X}_{fj}\right] = \mathbb{E}\left[Y_{ct} \mid \boldsymbol{X}_{ct}\right] \quad$ *for* $t, j = 1, ...T$,
2. $\mathbb{E}\left[Y_{ft} \mid \boldsymbol{X}_{ft}, \boldsymbol{X}_{ftj}, \boldsymbol{X}_{cj}\right] = \mathbb{E}\left[Y_{ft} \mid \boldsymbol{X}_{ft}\right] \quad$ *for* $\boldsymbol{X}_{fj} \subset \boldsymbol{X}_{ftj}, \quad t, j = 1, ...T$.

ii. The average covariance between children's prediction errors and parental permanent income across all observed years is zero

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

iii. The average covariance of parental income prediction errors for $|t-j| > h$ *is zero*

$$\frac{1}{T^2}\sum_{|t-j|>h} \mathbb{E}\left[\epsilon_{ft}\epsilon_{fj}\right] = 0, \quad \epsilon_{ft} \coloneqq Y_{ft} - \mathbb{E}\left[Y_{ft} \mid \boldsymbol{X}_{ft}\right].$$

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 permanent income constitutes a source of bias of the MI estimator (component (c) in equation (3)). One of the empirical patterns driving this dependence stems from 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 permanent income Y_f^P. Accordingly, Assumption 1-NP.*ii* imposes that thee average covariance between children's prediction errors and parental permanent income across all observed years is zero, once we have controlled for the relevant family characteristics. In Section III.D, we present a formal test for Assumption 1-NP*ii*, and in our U.S. application, the test does not reject the validity of this assumption.

The requirement of Assumption 1-NP.*iii* arises from the fundamental mismatch between the complete income profiles required by equation (6) 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. In Section III.D, we present a formal test for Assumption 1-NP*iii*, and in our U.S. application, the test does not reject the validity of this assumption.

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.

According to equation (6), the IGE depends on two distinct components: the covariance between parent and child income and the covariance within parental income, which implies that identification requirements differ across generations. For children, unconfoundedness needs only to hold for single income observations since the IGE exploits contemporaneous parent-child pairs, whereas for fathers, stronger conditions on income tuples are required to capture the temporal structure of their income process.

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*

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

Assumption 2-NP imposes a missing-at-random structure for child and parental incomes and a positivity condition for identification, similar to the conditional independence assumptions in Angrist and Imbens (1995). 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 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 1. *Under assumptions 1-NP and 2-NP and the definition of permanent income in equation (5), the IGE is nonparametrically identified as*

$$\beta_0 = \frac{\mathbb{E}\left[\sum_{t=1}^{T}\left(\mu_{ct}\left(\boldsymbol{X}_{ct},1\right)-\mu_c^P\right)\sum_{j=1}^{T}\left(\mu_{fj}\left(\boldsymbol{X}_{fj},1\right)-\mu_f^P\right)\right]}{\mathbb{E}\left[\sum_{|t-j|\leq h}\sigma_{tj}\left(\boldsymbol{X}_{ftj},1,1\right)+\sum_{|t-j|>h}\left(\mu_{ft}\left(\boldsymbol{X}_{ft},1\right)-\mu_f^P\right)\left(\mu_{fj}\left(\boldsymbol{X}_{fj},1\right)-\mu_f^P\right)\right]}, \tag{8}$$

where $\mu_{gt}(\boldsymbol{X}_{gt},1) := \mathbb{E}\left[Y_{gt} \mid \boldsymbol{X}_{gt}, D_{gt}=1\right]$, $\mu_g^P := \mathbb{E}\left[\sum_{t=1}^{T}\mu_{gt}(\boldsymbol{X}_{gt},1)\right]$, *and* $\sigma_{tj}\left(\boldsymbol{X}_{ftj},1,1\right) := \mathbb{E}\left[\left(Y_{ft}-\mu_f^P\right)\left(Y_{fj}-\mu_f^P\right) \mid \boldsymbol{X}_{ftj}, D_{ft}=1, D_{fj}=1\right]$.

Theorem 1 establishes the identification of the intergenerational elasticity in the presence of incomplete income data. Specifically, it shows that, under the conditional mean independence and orthogonality conditions in Assumption 1-NP and the standard missing-at-random assumptions in 2-NP, the IGE can be recovered from conditional expectations, including the conditional income profiles of parents and children, and the conditional covariance matrix of parental income.

To the best of our knowledge, the only existing identification result in this framework is that of An et al. (2022) who nonparametrically identify the mobility function relating children's to parents permanent 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 model (Assumption 1-MI with $\lambda_t = 1$) for two measurement periods, by exploiting the definition of permanent income. Second, we relax the assumption that transitory shocks to children's income are uncorrelated with parental permanent income and parental transitory shocks. In contrast, we assume that the prediction error of the children's annual income is uncorrelated to parental permanent income conditional on family characteristics (Assumption 1-NP). 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.

B. Locally Robust Estimation of the IGE

We estimate the intergenerational elasticity β_0 using a Generalized Method of Moments (GMM) approach based on Theorem 1. We begin by rearranging equation (8) to derive the moment condition that identifies β_0:

$$
\begin{aligned}
&\mathbb{E}\left[g_1\left(W,\gamma,\beta,\mu_c^P,\mu_f^P\right)\right]=0,\\
&\quad g_1\left(W,\gamma,\beta,\mu_c^P,\mu_f^P\right)=\beta\sum_{|t-j|\le h}\sigma_{tj}\left(\boldsymbol{X}_{ftj},1,1\right)+\beta\sum_{|t-j|>h}\left(\mu_{ft}\left(\boldsymbol{X}_{ft},1\right)-\mu_f^P\right)\left(\mu_{fj}\left(\boldsymbol{X}_{fj},1\right)-\mu_f^P\right)\\
&\qquad\qquad -\sum_{t=1}^{T}\left(\mu_{ct}\left(\boldsymbol{X}_{ct},1\right)-\mu_c^P\right)\sum_{j=1}^{T}\left(\mu_{fj}\left(\boldsymbol{X}_{fj},1\right)-\mu_f^P\right),
\end{aligned}
\tag{9}
$$

where $\gamma \coloneqq (\sigma_{tj}, \mu_{ft}, \mu_{fj})$. Thus, the moment identifying the IGE depends on the income profiles and parental income covariance structure, captured by the nuisance parameter γ, as well as the population mean permanent incomes $\left(\mu_c^P, \mu_f^P\right)$. These means are themselves identified by the moment conditions (see equation (B2)):

$$
\begin{aligned}
\mathbb{E}\left[g_2\left(W,\gamma,\mu_c^P\right)\right]=0,\quad g_2\left(W,\gamma,\mu_c^P\right)=\sum_{t=1}^{T}\mu_{ct}\left(\boldsymbol{X}_{ct},1\right)-\mu_c^P,\\
\mathbb{E}\left[g_3\left(W,\gamma,\mu_f^P\right)\right]=0,\quad g_3\left(W,\gamma,\mu_f^P\right)=\sum_{t=1}^{T}\mu_{ft}\left(\boldsymbol{X}_{ft},1\right)-\mu_f^P.
\end{aligned}
$$

Finally, we define the augmented parameter $\theta \coloneqq \left(\beta, \mu_c^P, \mu_f^P\right)$, and combine the moment conditions into a single system for GMM estimation

$$
g\left(W,\gamma,\theta\right)=\left(g_1\left(W,\gamma,\theta\right)\;\; g_2\left(W,\gamma,\theta\right)\;\; g_3\left(W,\gamma,\theta\right)\right)'.
$$

Because the nuisance parameter γ is unknown, a natural two-step estimation procedure is to first estimate the conditional expectations in γ, and then perform GMM estimation based on $g(W,\hat{\gamma},\theta)$. As in any two-step procedure, errors in the first step affect inference in the second. This issue is especially pronounced when machine learning (ML) is used, because regularization and model selection allow for bias to attain smaller variance. As a result, bias from the first step propagates to the second. This is formally captured by the sensitivity of the moment condition to small changes in the nuisance parameter:

$$
\frac{d}{d\tau}\mathbb{E}[g(W,\gamma_\tau,\theta)]\Big|_{\tau=0}\neq 0,
$$

indicating that the moment identifying θ is not locally robust to estimation error in γ_0.

Chernozhukov et al. (2022) provide a general procedure for constructing orthogonal moment functions for GMM, where the moment conditions are locally insensitive to first-step estimation. This property ensures that the resulting estimator is locally robust, meaning that estimation errors in the first step have no effect, locally, on the estimation of the parameter of interest. The authors show that an orthogonal (locally robust) moment function ψ can be constructed by augmenting the identifying moment function g with a correction term ϕ:

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

where α encompasses additional nuisance parameters introduced by ϕ.

Proposition 2. *There exists a function ϕ such that the augmented moment condition*

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

identifies the intergenerational elasticity, as well as the population mean of permanent incomes, and satisfies local robustness:

$$\frac{d}{d\tau}\mathbb{E}[\psi(W, \gamma_\tau, \alpha, \theta)]\Big|_{\tau=0} = 0,$$

where ψ is given by equation (B26). This latter property ensures that the estimator of the intergenerational elasticity is first-order insensitive to estimation error in the nuisance parameter γ.

The locally robust closed-form solution for the IGE follows directly from Proposition 2. In particular, solving for β in the orthogonal moment condition yields:

(10)
$$\beta = \frac{\mathbb{E}\left[\sum_{t=1}^{T}\left(\mu_{ct}(\boldsymbol{X}_{ct}, 1) - \mu_c^P\right)\sum_{j=1}^{T}\left(\mu_{fj}\left(\boldsymbol{X}_{fj}, 1\right) - \mu_f^P\right)\right] + \mathbb{E}[\phi_4 + \phi_5]}{\mathbb{E}\left[\sum_{|t-j|\le h}\sigma_{tj}\left(\boldsymbol{X}_{ftj}, 1, 1\right) + \sum_{|t-j|>h}\left(\mu_{ft}\left(\boldsymbol{X}_{ft}, 1\right) - \mu_f^P\right)\left(\mu_{fj}\left(\boldsymbol{X}_{fj}, 1\right) - \mu_f^P\right)\right] + \mathbb{E}[\phi_1 + \phi_2 + \phi_3]},$$

where

$$\phi_1 = \beta\sum_{|t-j|\le h}\frac{D_{ft}D_{fj}}{p\left(D_{ft} = 1, D_{fj} = 1|\boldsymbol{X}_{ftj}\right)}\left(\left(Y_{ft} - \mu_f^P\right)\left(Y_{fj} - \mu_f^P\right) - \sigma_{tj}\left(\boldsymbol{X}_{ftj}, 1, 1\right)\right),$$

$$\phi_2 = \sum_{|t-j|>h}\left(\mu_{fj}\left(\boldsymbol{X}_{fj}, 1\right) - \mu_f^P\right)\frac{D_{ft}}{p\left(D_{ft} = 1|\boldsymbol{X}_{ft}\right)}\left(Y_{ft} - \mu_{ft}\left(\boldsymbol{X}_{ft}, 1\right)\right),$$

$$\phi_3 = \sum_{|t-j|>h}\left(\mu_{ft}\left(\boldsymbol{X}_{ft}, 1\right) - \mu_f^P\right)\frac{D_{fj}}{p\left(D_{fj} = 1|\boldsymbol{X}_{fj}\right)}\left(Y_{fj} - \mu_{fj}\left(\boldsymbol{X}_{fj}, 1\right)\right),$$

$$\phi_4 = \sum_{t=1}^{T}\sum_{t=j}^{T}\left(\mu_{ct}\left(\boldsymbol{X}_{ct},1\right)-\mu_c^P\right)\frac{D_{fj}}{p\left(D_{fj}=1|\boldsymbol{X}_{fj}\right)}\left(Y_{fj}-\mu_{fj}\left(\boldsymbol{X}_{fj},1\right)\right),$$

$$\phi_5 = \sum_{t=1}^{T}\sum_{t=j}^{T}\left(\mu_{fj}\left(\boldsymbol{X}_{fj},1\right)-\mu_f^P\right)\frac{D_{ct}}{p\left(D_{ct}=1|\boldsymbol{X}_{ct}\right)}\left(Y_{ct}-\mu_{ct}\left(\boldsymbol{X}_{ct},1\right)\right),$$

$$\mu_c^P = \mathbb{E}\left[\sum_{t=1}^{T}\mu_{ct}\left(\boldsymbol{X}_{ct},1\right)\right]+\mathbb{E}\left[\sum_{t=1}^{T}\frac{D_{ct}}{p\left(D_{ct}=1|\boldsymbol{X}_{ct}\right)}\left(Y_{ct}-\mu_{ct}\left(\boldsymbol{X}_{ct},1\right)\right)\right],$$

$$\mu_f^P = \mathbb{E}\left[\sum_{t=1}^{T}\mu_{ft}\left(\boldsymbol{X}_{ft},1\right)\right]+\mathbb{E}\left[\sum_{t=1}^{T}\frac{D_{ct}}{p\left(D_{ft}=1|\boldsymbol{X}_{ft}\right)}\left(Y_{ft}-\mu_{ft}\left(\boldsymbol{X}_{ft},1\right)\right)\right].$$

The locally robust closed-form solution for β in equation (10) corresponds to the expression given in Theorem 1, augmented with correction terms that make it first-order insensitive to estimation errors in the nuisance parameters. The term ϕ_1 corrects for errors in estimating the conditional covariance of parental income for closely spaced periods ($|t-j| \leq h$), while ϕ_2 and ϕ_3 address errors in estimating parental income profiles for more distant periods ($|t-j| > h$). Similarly, ϕ_4 and ϕ_5 correct for errors in estimating the conditional income profiles of children and parents, respectively. A critical feature of all correction terms (ϕ_1 to ϕ_5) is their inherent adjustment for non-random missingness by weighting prediction errors by the inverse propensity score. Finally, the closed-form expressions for the population means of permanent incomes (μ_c^P and μ_f^P) also incorporate the corresponding prediction errors, ensuring that the estimator remains locally robust to first-step estimation mistakes.

Equation (10) illustrates why we define the augmented parameter $\theta := (\beta, \mu_c^P, \mu_f^P)$ rather than including μ_c^P and μ_f^P in the nuisance parameter γ. Each population mean μ_g^P depends not only on the conditional income profiles $\mu_{g,t}$ but also on the underlying population distribution. Consequently, small changes in the population distribution affect μ_g^P both through the conditional profiles and through the expectation itself. Thus, including μ_g^P in γ would therefore make the closed-form solution for β considerably more complex. By keeping μ_c^P and μ_f^P in θ, we separate the estimation of population permanent means from the first-step nuisance functions, which makes the locally robust solution more tractable.

To construct a debiased machine learning estimator for the IGE, we use the orthogonal moment condition in Proposition 2 (see equation (B26)) 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 pairs i, noting that a family may contribute multiple such pairs. 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 prediction errors in estimating the conditional expectations, 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.

C. Asymptotic Theory and Inference

Asymptotic Properties of the Locally Robust Estimator. — To provide rigorous justification for the empirical implementation of our proposed estimator, we examine its large-sample behavior. We begin by establishing consistency, which 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.

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, 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 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.

The asymptotic normality result in Lemma 2 is derived under the assumption of independently and identically distributed (i.i.d.) observations. In practice, however, datasets commonly include multiple children from the same family, introducing correlation within families. Accordingly, the asymptotic variance V should be estimated using a cluster-robust approach that accounts for this dependence structure. While the i.i.d. assumption is adopted here for ease of exposition and to align with the general theoretical framework of Chernozhukov et al. (2022), the core identification and estimation strategy remains sound. The cluster-robust extension is a straightforward implementation detail for the variance estimation, where the moment functions $\hat{\psi}_{i,tj}$ is aggregated at the family level before constructing the variance-covariance matrix $\hat{\Psi}$, accounting for correlation within families in the standard errors.

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 prediction errors and parental permanent income

$$H_0 : \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\epsilon_{ct}Y_f^P\right] = 0, \qquad \text{vs} \quad H_1 : \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\epsilon_{ct}Y_f^P\right] \neq 0, \tag{11}$$

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 permanent 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 $\theta_{cf} := \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\epsilon_{ct}Y_f^P\right]$. Second, we construct a locally robust estimator θ_{cf}. Finally, we provide a $t-$test based on $\hat{\theta}_{cf}$.

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

$$t_{cf,n} = \frac{\hat{\theta}_{cf,n}}{\sqrt{\hat{V}_{cf,n}/n}},$$

where $\hat{\theta}_{cf,n}$ is the argument solving the cross-fitted locally robust moment in equation (B30), and $\hat{V}_{cf,n}$ is a consistent estimator of the asymptotic variance of $\hat{\theta}_{cf,n}$, that accounts for dependence within families. Similar to the locally robust estimator for the IGE, this cluster-robust variance estimator is constructed by aggregating moment functions at the family level to allow for arbitrary correlation between observations from the same family, while maintaining independence across different families.

The following Theorem establishes the asymptotic properties of this locally robust $t-$test.

Theorem 2. *(Size, Consistency, and Local Power of the Locally Robust $t-$Test I) Under Assumptions 1-NP', 2-NP', 3-LR-7-LR and C-NP, the asymptotic properties of the locally robust t statistic*

$$t_{cf,n} = \frac{\hat{\theta}_{cf,n}}{\sqrt{\hat{V}_{cf,n}/n}}$$

are given by the following statements:

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

$$t_{cf,n} \xrightarrow{d} \mathcal{N}(0,1) \quad \textit{and} \quad \lim_{n\to\infty} \Pr\left(|t_{cf,n}| > z_{1-\alpha/2}\right) = \alpha,$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of the standard normal distribution.

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

$$\lim_{n\to\infty} \Pr\Big(|t_{cf,n}| > z_{1-\alpha/2}\Big) = 1.$$

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

$$t_{cf,n} \xrightarrow{d} \mathcal{N}\left(\frac{\delta}{\sqrt{V_{cf}}}, 1\right),$$

so the limiting power is

$$\lim_{n\to\infty} \Pr\Big(|t_{cf,n}| > z_{1-\alpha/2}\Big) = 2\left[1 - \Phi\left(z_{1-\alpha/2} - \frac{|\delta|}{\sqrt{V_{cf}}}\right)\right] > \alpha \quad \textit{whenever } \delta \neq 0,$$

where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function.

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

$$H_0 : \frac{1}{T^2} \sum_{t-j=h+1} \mathbb{E}\left[\epsilon_{ft}\epsilon_{fj}\right] = 0, \quad \text{vs} \quad H_1 : \frac{1}{T^2} \sum_{t-j=h+1} \mathbb{E}\left[\epsilon_{ft}\epsilon_{fj}\right] \neq 0 \tag{12}$$

In Appendix B5 we show that a locally robust $t-$test for H_0 in (12) is given by

$$t_{fh,n} = \frac{\hat{\theta}_{fh,n}}{\sqrt{\hat{V}_{fh,n}/n}}$$

where $\hat{\theta}_{fh,n}$ is the argument solving the cross-fitted orthogonal moment in equation (B33), and $\hat{V}_{fh,n}$ is a consistent estimator of the asymptotic variance of $\hat{\theta}_{cf,n}$, that accounts for dependence within families.

The following Corollary establishes the asymptotic properties of this locally robust $t-$test.

Corollary 2.1. *(Size, Consistency, and Local Power of the Locally Robust t−Test II) Under Assumptions 1-NP", 2-NP", 3-LR-7-LR, C-NP, and 8-NP", the asymptotic properties of the locally robust t statistic*

$$t_{fh,n} = \frac{\hat{\theta}_{fh,n}}{\sqrt{\hat{V}_{fh,n}/n}}$$

are given by the following statements:

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

$$t_{fh,n} \xrightarrow{d} \mathcal{N}(0,1) \quad \textit{and} \quad \lim_{n\to\infty} \Pr\left(|t_{fh,n}| > z_{1-\alpha/2}\right) = \alpha,$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of the standard normal distribution.

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

$$\lim_{n\to\infty} \Pr\left(|t_{fh,n}| > z_{1-\alpha/2}\right) = 1.$$

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

$$t_{fh,n} \xrightarrow{d} \mathcal{N}\left(\frac{\delta}{\sqrt{V_{fh}}}, 1\right),$$

so the limiting power is

$$\lim_{n\to\infty} \Pr\left(|t_{fh,n}| > z_{1-\alpha/2}\right) = 2\left[1 - \Phi\left(z_{1-\alpha/2} - \frac{|\delta|}{\sqrt{V_{fh}}}\right)\right] > \alpha \quad \textit{whenever } \delta \neq 0,$$

where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function.

IV. Simulations

We consider the following data generating process for generation g at age t:

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

Using the definition of permanent income:

$$Y_g^P = \gamma_{0,g} + \gamma_{1,g}X_{1,g} + \gamma_{2,g}X_{2,g} + \gamma_{3,g}\bar{t} + \gamma_{4,g}\bar{t^2} + \gamma_{5,g}X_{1,g}\bar{t} + \bar{\epsilon}_g, \qquad g \in \{c, f\},$$
$$\bar{t} = \frac{1}{T}\sum_{t=1}^{T} t, \quad \bar{t^2} = \frac{1}{T}\sum_{t=1}^{T} t^2, \quad \bar{\epsilon}_g = \frac{1}{T}\sum_{t=1}^{T} \epsilon_{gt}, \tag{13}$$

so the covariance between permanent incomes is given by

$$\begin{aligned}
\text{Cov}\left(Y_c^P, Y_f^P\right) &= \text{Cov}\left(\gamma_{1,c}X_{1,c}, \gamma_{1,f}X_{1,f}\right) + \text{Cov}\left(\gamma_{2,c}X_{2,c}, \gamma_{2,f}X_{2,f}\right) + \text{Cov}\left(\gamma_{5,c}X_{1,c}\bar{t}, \gamma_{5,f}X_{1,f}\bar{t}\right) \\
&\quad + \text{Cov}\left(\gamma_{1,c}X_{1,c}, \gamma_{5,f}X_{1,f}\bar{t}\right) + \text{Cov}\left(\gamma_{5,c}X_{1,c}\bar{t}, \gamma_{1,f}X_{1,f}\right) \\
&= \gamma_{1,c}\gamma_{1,f}\sigma_1 + \gamma_{2,c}\gamma_{2,f}\sigma_2 + \gamma_{5,c}\gamma_{5,f}\bar{t^2}\sigma_1 + \gamma_{1,c}\gamma_{5,f}\bar{t}\sigma_1 + \gamma_{5,c}\gamma_{1,f}\bar{t}\sigma_1,
\end{aligned} \tag{14}$$

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 (13), the variance of parental income corresponds to

$$\text{Var}\left(Y_f^P\right) = \gamma_{1,f}^2 + \gamma_{2,f}^2 + \gamma_{5,f}^2\bar{t^2} + 2\gamma_{1,f}\gamma_{5,f}\bar{t} + \sigma_\epsilon^2/T. \tag{15}$$

Finally, by plugging equations (14) and (15) into (2) yields

$$\begin{aligned}
\beta_0 &= \frac{\mathbb{E}\left[\left(Y_c^P - \mathbb{E}\left[Y_c^P\right]\right)\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)\right]}{\mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right]} \\
&= \frac{\sigma_1\left(\gamma_{1,c}\gamma_{1,f} + \gamma_{5,c}\gamma_{5,f}\bar{t^2} + \left(\gamma_{1,c}\gamma_{5,f} + \gamma_{5,c}\gamma_{1,f}\right)\bar{t}\right) + \gamma_{2,c}\gamma_{2,f}\sigma_2}{\gamma_{1,f}^2 + \gamma_{2,f}^2 + \gamma_{5,f}^2\bar{t^2} + 2\gamma_{1,f}\gamma_{5,f}\bar{t} + \sigma_\epsilon^2/T}.
\end{aligned} \tag{16}$$

Setting the parameter values to

$$\gamma_{0,c} = 8.5, \qquad \gamma_{0,f} = 5, \qquad \gamma_{1,c} = 0.275, \qquad \gamma_{1,f} = 0.4,$$

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

yields $T = 41$, and $\beta_0 = 0.50$ according to equation (16).

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:

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

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' permanent 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 permanent income as the average 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 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 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 1,138 child–father pairs, for children born between 1952 and 1960, using the same age span as in Mello et al. (2024). Following Lee and Solon (2009), we use the PSID core sample, corresponding to the Survey Research Center component, and define 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 nonpositive incomes, we bottom-code them at the sample 1st percentile, which affects 0.12% of the observations in the raw PSID data. 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), 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. If family income data during this period is insufficient, we use the closest available three-year window to ages 15–17. 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 covariates used to predict children's income satisfy:

$\mathbb{E}\left[Y_{ct} \mid \boldsymbol{X}_{ct}, \boldsymbol{X}_{cj}, \boldsymbol{X}_{fj}\right] = \mathbb{E}\left[Y_{ct} \mid \boldsymbol{X}_{ct}\right]$ for $t, j = 1, \ldots, T$, and those used to predict fathers' income satisfy $\mathbb{E}\left[Y_{ft} \mid \boldsymbol{X}_{ft}, \boldsymbol{X}_{ftj}, \boldsymbol{X}_{cj}\right] = \mathbb{E}\left[Y_{ft} \mid \boldsymbol{X}_{ft}\right]$ for $\boldsymbol{X}_{fj} \subset \boldsymbol{X}_{ftj}, t, j = 1, ...T$. 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 permanent 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 permanent 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%).

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 Permanent 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

The income measures exhibit tighter dispersion for fathers, with smaller standard deviations (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.

To assess the orthogonality conditions, we implement the two locally robust proposed tests, and for both tests, the null hypothesis is not rejected with our dataset, supporting the plausibility of the assumptions.

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' permanent income is proxied using a three-year average of log family income when the child was aged 15–17. 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 permanent 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.

Our 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, indicating 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 nonparametric 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. In contrast, the MI estimator produces a markedly lower IGE of 0.378, significantly underestimating the true intergenerational elasticity.

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 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 permanent income as the average 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 expectations, have 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. This results aligns with previous evidence derived from long-term parental income averages, which suggests that the U.S. IGE is likely above 0.6.

Our study highlights three 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 permanent income on parent permanent 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.

Finally, a central empirical challenge in economics is that many key parameters, from models of life-cycle income, savings, and consumption to measures of individual well-being, depend on latent lifetime outcomes, while only partial observations at certain ages are typically available. The methods proposed in this paper open the door for applications in other contexts with partially observed outcomes, providing a flexible framework for addressing similar empirical challenges.

References

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.

Angrist, J. and Imbens, G. (1995). Identification and estimation of local average treatment effects.

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.

Black, S. E. and Devereux, P. J. (2011). Recent developments in intergenerational mobility. *Handbook of labor economics*, 4:1487–1541.

Blanden, J., Haveman, R., Smeeding, T., and Wilson, K. (2014). Intergenerational mobility in the united states and great britain: 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.

Bratberg, E., Nilsen, Ø. A., and Vaage, K. (2007). Trends in intergenerational mobility across offspring's earnings distribution in norway. *Industrial Relations: A Journal of Economy and Society*, 46(1):112–129.

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.

Corak, M. (2013). Income inequality, equality of opportunity, and intergenerational mobility. *Journal of Economic Perspectives*, 27(3):79–102.

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.

Halvorsen, E., Ozkan, S., and Salgado, S. (2022). Earnings dynamics and its intergenerational transmission: Evidence from norway. *Quantitative Economics*, 13(4):1707–1746.

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.

Heckman, J. J., Lochner, L. J., and Todd, P. E. (2006). Earnings functions, rates of return and treatment effects: The mincer equation and beyond. *Handbook of the Economics of Education*, 1:307–458.

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

Jappelli, T. and Pistaferri, L. (2010). The consumption response to income changes. *Annu. Rev. Econ.*, 2(1):479–506.

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*.

Mogstad, M. and Torgovitsky, A. (2024). Instrumental variables with unobserved heterogeneity in treatment effects. In *Handbook of Labor Economics*, volume 5, pages 1–114. Elsevier.

Mogstad, M. and Torsvik, G. (2023). Family background, neighborhoods, and intergenerational mobility. *Handbook of the Economics of the Family*, 1(1):327–387.

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.

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.

Zimmerman, D. J. (1992). Regression toward mediocrity in economic stature. *The American Economic Review*, pages 409–429.

Appendix A

A1. Definition of Permanent Income

We define permanent income as the average log annual income over a specific lifetime period from $t = 1$ to T:

$$Y_g^P := \frac{1}{T}\sum_{t=1}^{T} Y_{gt}, \quad g \in \{c, f\}, \tag{A1}$$

where Y_{gt} is log annual income in year t, with $t = 1$ indicating the start age and T the number of years covered. A key advantage of this formulation is that it facilitates the identification of the Intergenerational Elasticity (IGE). Specifically, under this definition, permanent income depends only on the marginal distributions of log annual income, which can be partially observed and consistently estimated from the data. This property is particularly valuable in the presence of missing data, where income is not available for all individuals or years. The linearity of Y_g^P allows for the interchange of the summation and expectation operators, which is crucial for identification.

To illustrate this point, notice that 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:

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

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.

In contrast, consider an alternative definition of permanent income based on the log of average absolute income:

$$Y_g^{PL} := \log\left(\frac{1}{T}\sum_{t=1}^{T} e^{Y_{gt}}\right),$$

where $e^{Y_{gt}}$ denotes absolute annual income. This formulation complicates identification of the IGE because the nonlinearity introduced by the logarithm prevents expectations from decomposing into period-by-period components. By Jensen's inequality, we have

$$\mathbb{E}\left[Y_g^{PL}\right] = \mathbb{E}\left[\log\left(\frac{1}{T}\sum_{t=1}^{T} e^{Y_{gt}}\right)\right] \neq \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[Y_{gt}\right] = \mathbb{E}\left[Y_g^P\right].$$

Thus, 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. Accordingly, the expectation $\mathbb{E}\left[\log\left(\sum_t e^{Y_{gt}}\right)\right]$ cannot be reduced to the mean of conditional expectations of individual Y_{gt}. Instead, it requires 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.

When applied to the life-cycle estimator of Mello et al. (2024), our definition yields results that are virtually identical to those obtained using the log-sum specification. In their framework, predicted log annual incomes for children are first exponentiated to obtain absolute incomes, which are then averaged across years and logged to form permanent income. The IGE is subsequently estimated by regressing this measure on parental average log income. Table A1 depicts that the resulting IGE estimates under their log-average definition (Y_c^{PL}) are nearly indistinguishable from those obtained using our average-log definition (Y_c^P).

Table A1—: Estimation Results of Intergenerational Elasticity in the United States Using Alternative Permanent Income Definition for the Life-cycle Estimator

Life-cycle (log-avg)	**Life-cycle (avg-log)**
0.508	0.505
(0.475, 0.541)	(0.473, 0.538)

Sample size consists of 1,138 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 parentheses.

This close correspondence can be understood by rewriting the log-mean as

$$\log\left(\frac{1}{T}\sum_{t=1}^{T} e^{Y_{gt}}\right) = \log\left(\sum_{t=1}^{T} e^{Y_{gt}}\right) - \log(T)$$

$$
\begin{aligned}
&= \log\left(e^{Y_g^P}\sum_{t=1}^{T} e^{Y_{gt}-Y_g^P}\right) - \log(T) \\
&= Y_g^P + \log\left(\sum_{t=1}^{T} e^{Y_{gt}-Y_g^P}\right) - \log(T) \\
&= Y_g^P + \log\left(\frac{1}{T}\sum_{t=1}^{T} e^{Y_{gt}-Y_g^P}\right).
\end{aligned}
$$

When annual log income deviates minimally from its average ($Y_{gt} - Y_g^P \approx 0$), then $e^{Ygt-Y_g^P} \approx 1$ for all t. Consequently, the average inside the log term is approximately 1, so that

$$
\begin{aligned}
Y_g^{PL} &= \log\left(\frac{1}{T}\sum_{t=1}^{T} e^{Y_{gt}}\right) \approx Y_g^P + \log(1) \\
&= Y_g^P,
\end{aligned}
$$

explaining the similarity in IGE estimates of the life-cycle estimator.

A2. The Mid-life Income Estimator

The standard approach to estimating intergenerational elasticity, which we label the mid-life income (MI) estimator, proxies permanent 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 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' permanent 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)

$$
Y_{gt} = \lambda_t Y_g^P + v_{gt}, \quad \mathbb{E}\left[v_{gt} Y_g^P\right] = 0, \quad g \in c, f, \quad t = 1, \dots, T, \tag{A2}
$$

where Y_g^P is permanent income, λ_t captures varying persistence over the life cycle, and v_{gt} is an age-specific shock. As suggested by equation (A2), annual income at younger and older ages is a noisier measure of permanent income compared to mid-life income, as permanent 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 permanent income for fathers (e.g., Böhlmark and Lindquist (2006); Nybom and Stuhler (2017)).

For children's permanent 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 $\left(\tilde{Y}_f^P\right)$:

$$\text{(A3)} \qquad \begin{aligned} &\tilde{Y}_c^P = \alpha^{MI} + \beta^{MI} \tilde{Y}_f^P + u^{MI}, \quad \mathbb{E}\left[u^{MI}\left(1, \tilde{Y}_f^P\right)'\right] = 0, \\ &\tilde{Y}_g^P \coloneqq \frac{1}{T_g} \sum_{j \in \mathcal{M}_g} Y_{gj} D_{gj}, \quad g \in \{c, f\}, \end{aligned}$$

where $D_{gj} = 1$ when Y_{gj} is observed and zero otherwise,[1] $\mathcal{M}_g$ is a set of pre-defined mid-life years for generation g,[2] and $T_g \coloneqq \sum_{j \in \mathcal{M}_g} D_{gj}$ is the number of years used for the average. Accordingly, the closed-form expression for the MI estimand

$$\text{(A4)} \qquad \beta^{MI} = \frac{\mathbb{E}\left[\left(\tilde{Y}_c^P - \mathbb{E}\left[\tilde{Y}_c^P\right]\right)\left(\tilde{Y}_f^P - \mathbb{E}\left[\tilde{Y}_f^P\right]\right)\right]}{\mathbb{E}\left[\left(\tilde{Y}_f^P - \mathbb{E}\left[\tilde{Y}_f^P\right]\right)^2\right]} \coloneqq \frac{\mathbb{E}\left[\tilde{y}_c^P \tilde{y}_f^P\right]}{\mathbb{E}\left[\left(\tilde{y}_f^P\right)^2\right]},$$

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

$$\text{(A5)} \qquad \hat{\beta}_n^{MI} = \frac{\mathbb{E}_n\left[\left(\tilde{Y}_c^P - \mathbb{E}_n\left[\tilde{Y}_c^P\right]\right)\left(\tilde{Y}_f^P - \mathbb{E}_n\left[\tilde{Y}_f^P\right]\right)\right]}{\mathbb{E}_n\left[\left(\tilde{Y}_f^P - \mathbb{E}_n\left[\tilde{Y}_f^P\right]\right)^2\right]},$$

[1] 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.

[2] 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 $\mathcal{M}_c$.

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 (A2), which motivates the MI estimator, is correctly specified and assumes that persistency of permanent 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 permanent 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 permanent 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 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 (A3), together with Assumption 1-MI, involves a random i.i.d sample of $\tilde{W} = \left(\boldsymbol{Y}_c \odot D_c, \boldsymbol{Y}_f \odot D_f, \boldsymbol{D}_c, \boldsymbol{D}_f\right)$,[3] 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 permanent 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}.[4] 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*

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

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

[4]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.

Assumption 2-MI imposes a set of orthogonality conditions involving age shocks, missingness indicators, and unobserved permanent 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 permanent 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 permanent income.[5] 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, the probability limit of $\hat{\beta}_n^{MI}$ takes the form:

(A6)

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

The downward bias by measurement error highlighted by Solon (1992) and Mazumder (2005) corresponds to component (a) in equation (A6). Consider rewriting (a) as

(A7)

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

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

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

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 permanent income also affects the probability of observing a given tuple of years. If we were to observe permanent income, components (d) and (b) in equation (A6) 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 permanent status. Imposing such a restriction mechanically affects the probability of observing income in a given year, which corresponds to component (b) in equation (A6).

Even when using the same data, changes in the definition of mid-life alter the estimand in equation (A4), further limiting comparability. The transitory component of children's income depending on parental income is encompassed by (c) in equation (A6). Halvorsen et al. (2022) highlight that children from affluent families might experience faster income growth, which would cause the steepness of the income trajectory to depend on parental permanent income. Thus, the correlation between age shocks to children's (log) annual income and parental permanent income induces bias in estimating the IGE.

Component (d) in equation (A6) 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 permanent income. The second one, by the cardinality of $\mathcal{M}_c$ affecting the magnitude of component (d).

As shown in equation (A28), 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 (A6). However, relaxing this assumption allows our inconsistency result to capture another source of bias discussed in the literature.

A3. Proof of Proposition 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:

$$\begin{aligned}
\mathbb{E}\left[\tilde{y}_c^P \tilde{y}_f^P\right] &= \mathbb{E}\left[\tilde{Y}_c^P \tilde{Y}_f^P\right] - \mathbb{E}\left[\tilde{Y}_c^P\right]\mathbb{E}\left[\tilde{Y}_f^P\right] \\
&= \mathbb{E}\left[\left(\frac{1}{T_c}\sum_{t\in\mathcal{M}_c} Y_{ct}D_{ct}\right)\left(\frac{1}{T_f}\sum_{j\in\mathcal{M}_f} Y_{fj}D_{fj}\right)\right] - \mathbb{E}\left[\tilde{Y}_c^P\right]\mathbb{E}\left[\tilde{Y}_f^P\right] \\
&= \frac{1}{T_cT_f}\sum_{t\in\mathcal{M}_c}\sum_{j\in\mathcal{M}_f}\mathbb{E}\left[Y_{ct}Y_{fj}D_{ct}D_{fj}\right] - \mathbb{E}\left[\tilde{Y}_c^P\right]\mathbb{E}\left[\tilde{Y}_f^P\right] \\
&= \frac{1}{T_cT_f}\sum_{t\in\mathcal{M}_c}\sum_{j\in\mathcal{M}_f}\mathbb{E}\left[\left(\lambda_t Y_c^P + v_{ct}\right)\left(\lambda_j Y_f^P + v_{fj}\right)D_{ct}D_{fj}\right] - \mathbb{E}\left[\tilde{Y}_c^P\right]\mathbb{E}\left[\tilde{Y}_f^P\right] \\
&= \frac{1}{T_cT_f}\mathbb{E}\left[Y_c^P Y_f^P \sum_{t\in\mathcal{M}_c}\lambda_t D_{ct}\sum_{j\in\mathcal{M}_f}\lambda_j D_{fj}\right] + \frac{1}{T_cT_f}\mathbb{E}\left[Y_f^P\sum_{t\in\mathcal{M}_c} v_{ct}D_{ct}\sum_{j\in\mathcal{M}_f}\lambda_j D_{fj}\right] \\
&+ \frac{1}{T_cT_f}\sum_{t\in\mathcal{M}_c}\lambda_t\sum_{j\in\mathcal{M}_f}\mathbb{E}\left[Y_c^P v_{fj}D_{ct}D_{fj}\right] + \frac{1}{T_cT_f}\sum_{t\in\mathcal{M}_c}\sum_{j\in\mathcal{M}_f}\mathbb{E}\left[v_{ct}v_{fj}D_{ct}D_{fj}\right] \\
&- \mathbb{E}\left[\tilde{Y}_c^P\right]\mathbb{E}\left[\tilde{Y}_f^P\right],
\end{aligned} \tag{A8}$$

where we have used the definition of average (log) income during mid-life as given in equation (A3) in the second equality, and the fourth equality follows by Assumption 1-MI.

The first term in equation (A8) can be expressed as

$$\begin{aligned}
\frac{1}{T_cT_f}\mathbb{E}\left[Y_c^P Y_f^P \sum_{t\in\mathcal{M}_c}\lambda_t D_{ct}\sum_{j\in\mathcal{M}_f}\lambda_j D_{fj}\right] &= \frac{1}{T_cT_f}\mathbb{E}\left[Y_c^P Y_f^P \sum_{t\in\mathcal{M}_c} D_{ct}\sum_{j\in\mathcal{M}_f} D_{fj}\right] \\
&= \frac{1}{T_cT_f}\mathbb{E}\left[Y_c^P Y_f^P\right]\times T_cT_f \\
&= \mathbb{E}\left[Y_c^P Y_f^P\right],
\end{aligned} \tag{A9}$$

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 $T_g \coloneqq \sum_{j \in \mathcal{M}_g} D_{gj}$. As regards the second term in equation (A8), it can be simplified as follows:

$$\begin{aligned}
\frac{1}{T_c T_f}\mathbb{E}\left[Y_f^P \sum_{t\in\mathcal{M}_c} v_{ct} D_{ct} \sum_{j\in\mathcal{M}_f} \lambda_j D_{fj}\right] &= \frac{1}{T_c T_f}\mathbb{E}\left[Y_f^P \sum_{t\in\mathcal{M}_c} v_{ct} D_{ct} \sum_{j\in\mathcal{M}_f} D_{fj}\right] \\
&= \frac{1}{T_c T_f}\mathbb{E}\left[Y_f^P \sum_{t\in\mathcal{M}_c} v_{ct} D_{ct}\right] T_f \\
&= \frac{1}{T_c} \sum_{t\in\mathcal{M}_c} \mathbb{E}\left[Y_f^P v_{ct} D_{ct}\right] \\
&= \frac{1}{T_c} \sum_{t} \mathbb{E}\left[Y_f^P v_{ct} \middle| D_{ct} = 1, t \in \mathcal{M}_c\right] \times p\left(D_{ct} = 1 \middle| t \in \mathcal{M}_c\right) \\
&\coloneqq \frac{1}{T_c} \sum_{t} \mathbb{E}\left[Y_f^P v_{ct} \middle| D_{ct} = 1, t \in \mathcal{M}_c\right] \times p_c\left(t \in \mathcal{M}_c\right), \qquad \text{(A10)}
\end{aligned}$$

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

The third term in equation (A8) equals zero by the law of iterated expectations (LIE) and Assumption 2-MI

(A11)
$$\mathbb{E}\left[Y_c^P v_{fj} D_{ct} D_{fj}\right] = \mathbb{E}\left[\mathbb{E}\left[Y_c^P v_{fj} \middle| D_{ct}, D_{fj}\right] D_{ct} D_{fj}\right] = \mathbb{E}\left[\mathbb{E}\left[Y_c^P v_{fj} \middle| D_{fj}\right] D_{ct} D_{fj}\right] = 0.$$

Similarly, the fourth term in equation (A8) also equals zero by Assumption 2-MI and LIE

(A12)
$$\mathbb{E}\left[v_{ct} v_{fj} D_{ct} D_{fj}\right] = \mathbb{E}\left[\mathbb{E}\left[v_{ct} v_{fj} \middle| D_{ct}, D_{fj}\right] D_{ct} D_{fj}\right] = \mathbb{E}\left[\mathbb{E}\left[v_{ct} v_{fj} \middle| D_{ct}, D_{fj}\right] D_{ct} D_{fj}\right] = 0.$$

Finally, for the last term in equation (A8) we have

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

$$
\begin{aligned}
&= \left(\mathbb{E}\left[Y_c^P\right] + \frac{1}{T_c}\sum_{t\in\mathcal{M}_c}\mathbb{E}\left[v_{ct}D_{ct}\right]\right)\left(\mathbb{E}\left[Y_f^P\right] + \frac{1}{T_f}\sum_{j\in\mathcal{M}_f}\mathbb{E}\left[v_{fj}D_{fj}\right]\right)\\
&= \left(\mathbb{E}\left[Y_c^P\right] + \frac{1}{T_c}\sum_{t\in\mathcal{M}_c}\mathbb{E}\left[\mathbb{E}\left[v_{ct}|D_{ct}\right]D_{ct}\right]\right)\\
&\times\left(\mathbb{E}\left[Y_f^P\right] + \frac{1}{T_f}\sum_{j\in\mathcal{M}_f}\mathbb{E}\left[\mathbb{E}\left[v_{fj}|D_{fj}\right]D_{fj}\right]\right)\\
&= \mathbb{E}\left[Y_c^P\right]\mathbb{E}\left[Y_f^P\right], \qquad \text{(A13)}
\end{aligned}
$$

where the last equality follows by Assumption 1-MI.

By plugging equations (A10), (A9), (A11), (A12), and (A13) into equation (A8), we have

$$
\begin{aligned}
\mathbb{E}\left[\tilde{y}_c^P\tilde{y}_f^P\right] &= \mathbb{E}\left[\left(Y_c^P - \mathbb{E}\left[Y_c^P\right]\right)\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)\right]\\
&\quad + \frac{1}{T_c}\sum_t \mathbb{E}\left[Y_f^P v_{ct}\middle|D_{ct} = 1, t\in\mathcal{M}_c\right]\times p_c\left(t\in\mathcal{M}_c\right). \qquad \text{(A14)}
\end{aligned}
$$

We now analyze the denominator in equation (A4)

$$
\begin{aligned}
\mathbb{E}\left[\left(\tilde{y}_f^P\right)^2\right] &= \mathbb{E}\left[\left(\frac{1}{T_f}\sum_{t\in\mathcal{M}_f}Y_{ft}D_{ft}\right)\left(\frac{1}{T_f}\sum_{j\in\mathcal{M}_f}Y_{fj}D_{fj}\right)\right] - \mathbb{E}\left[\left(\tilde{Y}_f^P\right)^2\right]\\
&= \frac{1}{T_f^2}\sum_{t\in\mathcal{M}_f}\sum_{j\in\mathcal{M}}\mathbb{E}\left[Y_{ft}Y_{fj}D_{ct}D_{fj}\right] - \mathbb{E}\left[\left(\tilde{Y}_f^P\right)^2\right]\\
&= \frac{1}{T_f^2}\sum_{t\in\mathcal{M}_f}\sum_{j\in\mathcal{M}_f}\mathbb{E}\left[\left(\lambda_t Y_f^P + v_{ft}\right)\left(\lambda_j Y_f^P + v_{fj}\right)D_{ft}D_{fj}\right] - \mathbb{E}\left[\left(\tilde{Y}_f^P\right)^2\right]\\
&= \frac{1}{T_f^2}\mathbb{E}\left[\left(Y_f^P\right)^2\sum_{t\in\mathcal{M}_f}\lambda_t D_{ft}\sum_{j\in\mathcal{M}_f}\lambda_j D_{fj}\right] + \frac{1}{T_f^2}\sum_{t\in\mathcal{M}_f}\sum_{j\in\mathcal{M}_f}\lambda_j\mathbb{E}\left[v_{ft}Y_f^P D_{ft}D_{fj}\right]\\
&\quad + \frac{1}{T_f^2}\sum_{t\in\mathcal{M}_f}\lambda_t\sum_{j\in\mathcal{M}_f}\mathbb{E}\left[Y_f^P v_{fj}D_{ft}D_{fj}\right] + \frac{1}{T_f^2}\sum_{t\in\mathcal{M}_f}\sum_{j\in\mathcal{M}_f}\mathbb{E}\left[v_{ft}v_{fj}D_{ft}D_{fj}\right]\\
&\quad - \mathbb{E}\left[\left(\tilde{Y}_f^P\right)^2\right]. \qquad \text{(A15)}
\end{aligned}
$$

The first term of equation (A15) can be expressed as

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

$$= \frac{1}{T_f^2}\mathbb{E}\left[\left(Y_f^P\right)^2\right]T_f^2$$

$$= \mathbb{E}\left[\left(Y_f^P\right)^2\right] \tag{A16}$$

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

$$\mathbb{E}\left[v_{ft}Y_f^P D_{ft}D_{fj}\right] = \mathbb{E}\left[\mathbb{E}\left[v_{ft}Y_f^P|D_{ft}, D_{fj}\right]D_{ft}D_{fj}\right] = 0, \tag{A17}$$

and the fourth term

$$\begin{aligned}\frac{1}{T_f^2}\sum_{t\in\mathcal{M}_f}\sum_{j\in\mathcal{M}_f}\mathbb{E}\left[v_{ft}v_{fj}D_{ft}D_{fj}\right] &= \frac{1}{T_f^2}\sum_t\sum_j\mathbb{E}\left[v_{ft}v_{fj}\middle|D_{ft}=1, D_{fj}=1, \{t,j\}\in\mathcal{M}_f\right]\\ &\times p\left(D_{ft}=1, D_{fj}=1\middle|\{t,j\}\in\mathcal{M}_f\right)\\ &:= \frac{1}{T_f^2}\sum_t\sum_j\mathbb{E}\left[v_{ft}v_{fj}\middle|D_{ft}=1, D_{fj}=1, \{t,j\}\in\mathcal{M}_f\right]\\ &\times p_f\left(\{t,j\}\in\mathcal{M}_f\right).\end{aligned} \tag{A18}$$

By the same arguments as those of equation (A13), the last term in equation (A15) boils down to

$$\mathbb{E}\left[\left(\tilde{Y}_f^P\right)^2\right] = \mathbb{E}\left[\left(Y_f^P\right)^2\right] \tag{A19}$$

By plugging equations (A16), (A17), (A18), and (A19) into equation (A15), we have

$$\begin{aligned}\mathbb{E}\left[\left(\tilde{y}_f^P\right)^2\right] &= \mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right]\\ &+ \frac{1}{T_f^2}\sum_t\sum_j\mathbb{E}\left[v_{ft}v_{fj}\middle|D_{ft}=1, D_{fj}=1, \{t,j\}\in\mathcal{M}_f\right]\times p_f\left(\{t,j\}\in\mathcal{M}_f\right).\end{aligned} \tag{A20}$$

Thus, by plugging equations (A14) and (A20) into (A4), we get

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

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

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

we study the convergence in probability of each of its components. The numerator converges in probability to

$$
\begin{aligned}
\mathbb{E}_n\left[\tilde{y}_c^P \tilde{y}_f^P\right] &\xrightarrow{p} \mathbb{E}\left[\left(\tilde{Y}_c^P - \mathbb{E}\left[\tilde{Y}_c^P\right]\right)\left(\tilde{Y}_f^P - \mathbb{E}\left[\tilde{Y}_f^P\right]\right)\right] \\
&= \mathbb{E}\left[\left(Y_c^P - \mathbb{E}\left[Y_c^P\right]\right)\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)\right] \\
&+ \frac{1}{T_c}\sum_t \mathbb{E}\left[Y_f^P v_{ct} \middle| D_{ct} = 1, t \in \mathcal{M}_c\right] \times p_c\left(t \in \mathcal{M}_c\right) \\
&= \mathbb{E}\left[\left(\beta_0\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right) + u\right)\left(\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)\right)\right] \\
&+ \frac{1}{T_c}\sum_t \mathbb{E}\left[Y_f^P v_{ct} \middle| D_{ct} = 1, t \in \mathcal{M}_c\right] \times p_c\left(t \in \mathcal{M}_c\right) \\
&= \beta_0 \mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right] + \frac{1}{T_c}\sum_t \mathbb{E}\left[Y_f^P v_{ct} \middle| D_{ct} = 1, t \in \mathcal{M}_c\right] \times p_c\left(t \in \mathcal{M}_c\right) \qquad \text{(A21)}
\end{aligned}
$$

where the convergence in probability follows by the Law of Large Numbers (LLN), the first equality from equation (A14), and the second and third by equation (1). As regards the denominator in equation (A5), it converges in probability to

$$
\begin{aligned}
\mathbb{E}_n\left[\left(\tilde{y}_f^P\right)^2\right] &\xrightarrow{p} \mathbb{E}\left[\left(\tilde{Y}_f^P - \mathbb{E}\left[\tilde{Y}_f^P\right]\right)^2\right] \\
&= \mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right] \\
&+ \frac{1}{T_f^2}\sum_t \sum_j \mathbb{E}\left[v_{ft} v_{fj} \middle| D_{ft} = 1, D_{fj} = 1, \{t, j\} \in \mathcal{M}_f\right] \times p_f\left(\{t, j\} \in \mathcal{M}_f\right) \qquad \text{(A22)}
\end{aligned}
$$

where the convergence in probability follows by the LLN, and the equality from equation (A20). Thus, by plugging equations (A21) and (A22) into (A4) and applying the Continuous Mapping Theorem (CMT) yields

(A23)

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

A4. Equivalence with Previous Inconsistency Results

Corollary 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 under an additional assumption. We then show that the two inconsistency results in Nybom and Stuhler (2016) are particular cases of Corollary 1 when variants of Assumption 1-MI are considered, and Assumption 2-MI is relaxed.

We now show that equation (3) in Corollary 1 simplifies to the inconsistency result in

(Solon, 1992, p. 400), if we further assume that parental permanent income is uncorrelated to child age shocks, for the observed years during mid-life. In particular, consider assuming

Assumption 3-S.

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

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 (A23) (which corresponds to equation (3) in Corollary 1) boils down to

$$\hat{\beta}_n^{MI} \xrightarrow{p} \frac{\beta_0 \mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right]}{\mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right] + \frac{1}{T_f^2} \sum_t \sum_j \mathbb{E}\left[v_{ft} v_{fj}\right]}, \tag{A24}$$

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 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 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 permanent income is governed by*

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

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 permanent income. Furthermore, in their result, the 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

$$\begin{aligned}\mathbb{E}\left[\tilde{y}_c^P \tilde{y}_f^P\right] &= \mathbb{E}\left[Y_{ct} Y_{fj}\right] - \mathbb{E}\left[Y_{ct}\right] \mathbb{E}\left[Y_{fj}\right]\\ &= \mathbb{E}\left[\left(Y_c^P + v_{ct}\right)\left(Y_f^P + v_{fj}\right)\right] - \mathbb{E}\left[\left(Y_c^P + v_{ct}\right)\right] \mathbb{E}\left[\left(Y_f^P + v_{fj}\right)\right]\\ &= \mathbb{E}\left[Y_c^P Y_f^P\right] + \mathbb{E}\left[Y_f^P v_{ct}\right] + \mathbb{E}\left[Y_c^P v_{fj}\right] + \mathbb{E}\left[v_{ct} v_{fj}\right] - \mathbb{E}\left[Y_c^P\right] \mathbb{E}\left[Y_f^P\right]\\ &= \mathbb{E}\left[\left(Y_c^P - \mathbb{E}\left[Y_c^P\right]\right)\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)\right] + \mathbb{E}\left[Y_f^P v_{ct}\right] + \mathbb{E}\left[Y_c^P v_{fj}\right] + \mathbb{E}\left[v_{ct} v_{fj}\right]\\ &= \beta_0 \mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right] + \mathbb{E}\left[Y_f^P v_{ct}\right] + \mathbb{E}\left[Y_c^P v_{fj}\right] + \mathbb{E}\left[v_{ct} v_{fj}\right], \end{aligned} \tag{A25}$$

and

$$\begin{aligned}\mathbb{E}\left[\left(\tilde{y}_f^P\right)^2\right] &= \mathbb{E}\left[\left(Y_{ft}\right)^2\right] - \mathbb{E}\left[Y_{ft}\right]^2\\ &= \mathbb{E}\left[\left(Y_f^P + v_{ft}\right)^2\right] - \mathbb{E}\left[\left(Y_f^P + v_{ft}\right)\right]^2\\ &= \mathbb{E}\left[\left(Y_f^P\right)^2\right] + \mathbb{E}\left[\left(v_{ft}\right)^2\right] + 2\mathbb{E}\left[Y_f^P v_{ft}\right] - \mathbb{E}\left[Y_f^P\right]^2\\ &= \mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right] + \mathbb{E}\left[v_{ft}^2\right] + 2\mathbb{E}\left[Y_f^P v_{ft}\right]\end{aligned} \tag{A26}$$

Then, by equation (A25)

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

and by equation (A26)

$$\mathbb{E}_n\left[\left(\tilde{y}_f^P\right)^2\right] \xrightarrow{p} \mathbb{E}\left[\left(\tilde{y}_f^P\right)^2\right]$$

$$= \mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right] + \mathbb{E}\left[v_{ft}^2\right] + 2\mathbb{E}\left[Y_f^P v_{ft}\right],$$

so that, under Assumption 1-NS, equation (A23) boils down to

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

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 permanent 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 permanent 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

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

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

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

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 (A23) becomes

$$\text{(A28)} \qquad \hat{\beta}_n^{MI} \xrightarrow{p} \frac{\beta_0\lambda_t\lambda_j\mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right] + \lambda_j\mathbb{E}\left[Y_f^P v_{ct}\right] + \lambda_t\mathbb{E}\left[Y_c^P v_{fj}\right] + \mathbb{E}\left[v_{ct}v_{fj}\right]}{\lambda_j^2\mathbb{E}\left[\left(Y_f^P - \mathbb{E}\left[Y_f^P\right]\right)^2\right] + \mathbb{E}\left[v_{fj}^2\right]},$$

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

Appendix B: nonparametric Identification of the IGE and Local Robustness

B1. Proof of Theorem 1

To establish identification of the intergenerational elasticity, defined as

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

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

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

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

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 permanent income in equation (B2), while the second equality follows from the definition of permanent 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 permanent 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

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

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

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

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 permanent income:

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

$$= \sum_{|t-j|\le h} \mathbb{E}\left[\mathbb{E}\left[\left(Y_{ft}-\mu_f^P\right)\left(Y_{fj}-\mu_f^P\right) \mid \boldsymbol{X}_{ftj}, D_{ft}=1, D_{fj}=1\right]\right]$$

$$+ \sum_{|t-j|>h} \mathbb{E}\left[\left(\mathbb{E}\left[Y_{ft}|\boldsymbol{X}_{ft}, D_{ft}=1\right]-\mu_f^P\right)\left(\mathbb{E}\left[Y_{fj}|\boldsymbol{X}_{fj}, D_{fj}=1\right]-\mu_f^P\right)\right]$$

(B6)
$$:= \mathbb{E}\left[\sum_{|t-j|\le h} \sigma_{tj}\left(\boldsymbol{X}_{ftj},1,1\right) + \sum_{|t-j|>h}\left(\mu_{ft}\left(\boldsymbol{X}_{ft},1\right)-\mu_f^P\right)\left(\mu_{fj}\left(\boldsymbol{X}_{fj},1\right)-\mu_f^P\right)\right],$$

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

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

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 $\boldsymbol{W}$, 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 $\boldsymbol{W}$ of $\boldsymbol{W}$, $\boldsymbol{\gamma}$ and θ such that

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

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)

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

where α is a function called the Riesz representer[6] of the functional γ, on which only the first step influence function[7] 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 $\boldsymbol{W}$, 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 1 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 (8). 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.

[6]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.

[7]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 $\mathbb{E}[g(W, \gamma, \theta)]$.

As illustrated in Appendix B2, to find the orthogonal moment function corresponding to equation (9), 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 (8), which identifies our parameter of interest β_0, can be rewritten as

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

where $\mathbb{E}[\cdot]$ is the expectation under the true distribution of $\boldsymbol{W}$ (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

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

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

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 permanent 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

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

$$:= \mathbb{E}\left[\alpha_{0c,t}\left(\boldsymbol{X}_{ct}, D_{ct}\right)\mu_{ct}\left(F_\tau\right)\left(\boldsymbol{X}_{ct}, D_{ct}\right)\right], \tag{B14}$$

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

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

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

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

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

which in turn implies

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

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:

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

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

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

following the same arguments as equation (B15).

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

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

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

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

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

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

$$
\phi_{3,tj}\left(\omega, \gamma, \alpha_{03,tj}\right)=\alpha_{03,tj}\left(\boldsymbol{X}_{ft}, \boldsymbol{X}_{fj}, D_{fj}\right)\left(Y_{fj}-\mu_{fj}\left(\boldsymbol{X}_{fj}, D_{fj}\right)\right),\quad \alpha_{03,tj}=\left(\mu_{ft}\left(\boldsymbol{X}_{ft}, D_{ft}\right)-\mu_f^P\right)\frac{D_{fj}}{p\left(D_{fj}=1|\boldsymbol{X}_{fj}\right)}
\tag{B20}
$$

(B21)
$$\phi_{4,tj}\left(\omega,\gamma,\alpha_{04,tj}\right)=\alpha_{04,tj}\left(\boldsymbol{X}_{ct},\boldsymbol{X}_{fj},D_{fj}\right)\left(Y_{fj}-\mu_{fj}\left(\boldsymbol{X}_{fj},D_{fj}\right)\right),\quad \alpha_{04,tj}=\left(\mu_{ct}\left(\boldsymbol{X}_{ct},D_{ft}\right)-\mu_c^P\right)\frac{D_{fj}}{p\left(D_{fj}=1|\boldsymbol{X}_{fj}\right)}$$

(B22)
$$\phi_{5,tj}\left(\omega,\gamma,\alpha_{05,tj}\right)=\alpha_{05,tj}\left(\boldsymbol{X}_{fj},\boldsymbol{X}_{ct},D_{ct}\right)\left(Y_{ct}-\mu_{ct}\left(\boldsymbol{X}_{ct},D_{ct}\right)\right),\qquad \alpha_{05,tj}=\left(\mu_{fj}\left(\boldsymbol{X}_{fj},D_{fj}\right)-\mu_f^P\right)\frac{D_{ct}}{p\left(D_{ct}=1|\boldsymbol{X}_{ct}\right)}.$$

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)

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

$$
\begin{aligned}
&-\sum_{t=1}^{T}\sum_{t=j}^{T}\left(\mu_{ct}\left(\boldsymbol{X}_{ct},1\right)-\mu_c^P\right)\frac{D_{fj}}{p\left(D_{fj}=1|\boldsymbol{X}_{fj}\right)}\left(Y_{fj}-\mu_{fj}\left(\boldsymbol{X}_{fj},1\right)\right)\\
\text{(B24)}\qquad &-\sum_{t=1}^{T}\sum_{t=j}^{T}\left(\mu_{fj}\left(\boldsymbol{X}_{fj},1\right)-\mu_f^P\right)\frac{D_{ct}}{p\left(D_{ct}=1|\boldsymbol{X}_{ct}\right)}\left(Y_{ct}-\mu_{ct}\left(\boldsymbol{X}_{ct},1\right)\right).
\end{aligned}
$$

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

$$
\begin{aligned}
\psi_2\left(W,\gamma,\alpha,\theta\right) &= \sum_{t=1}^{T}\mu_{ct}\left(\boldsymbol{X}_{ct},1\right)-\mu_c^P+\sum_{t=1}^{T}\frac{D_{ct}}{p\left(D_{ct}=1|\boldsymbol{X}_{ct}\right)}\left(Y_{ct}-\mu_{ct}\left(\boldsymbol{X}_{ct},1\right)\right)\\
\text{(B25)}\quad \psi_3\left(W,\gamma,\alpha,\theta\right) &= \sum_{t=1}^{T}\mu_{ft}\left(\boldsymbol{X}_{ft},1\right)-\mu_f^P+\sum_{t=1}^{T}\frac{D_{ct}}{p\left(D_{ft}=1|\boldsymbol{X}_{ft}\right)}\left(Y_{ft}-\mu_{ft}\left(\boldsymbol{X}_{ft},1\right)\right).
\end{aligned}
$$

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

$$
\text{(B26)}\qquad \psi\left(W,\gamma,\alpha,\theta\right)=\left(\psi_1\left(W,\gamma,\alpha,\theta\right),\psi_2\left(W,\gamma,\alpha,\theta\right),\psi_3\left(W,\gamma,\alpha,\theta\right)\right),
$$

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:

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

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 (B26) 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 2-NP, 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:

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

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

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

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* $\ell = 1, \ldots, L$
$\int \left\| g\left(w, \hat{\gamma}^{(\ell)}, \hat{\beta}_n^{LR}\right) - g\left(w, \hat{\gamma}^{(\ell)}, \beta_0\right) \right\|^2 dF_0(w) \xrightarrow{p} 0$, *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,*

$$\left\| \frac{\partial \psi(W, \gamma, \theta)}{\partial \beta} - \frac{\partial \psi(W, \gamma, \theta_0)}{\partial \beta} \right\| \leq d(W, \gamma) |\theta - \theta_0|^{1/C}, \quad \textit{with } \mathbb{E}[d(W, \gamma)] < C;$$

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

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

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

$$
\begin{aligned}
&\frac{\partial g_1(W,\gamma,\theta)}{\partial \mu_f^P} = -\beta \sum_{t=1}^{T}\sum_{j=1}^{T}\left(\mu_{ft}(\boldsymbol{X}_{fj},1) + \mu_{fj}(\boldsymbol{X}_{fj},1) - 2\mu_f^P\right) + \sum_{t=1}^{T}\left(\mu_{ct}(\boldsymbol{X}_{ct},1) - \mu_c^P\right),\\
&\frac{\partial g_2(W,\gamma,\theta)}{\partial \beta} = 0, \quad \frac{\partial g_2(W,\gamma,\theta)}{\partial \mu_c^P} = -1, \quad \frac{\partial g_2(W,\gamma,\theta)}{\partial \mu_f^P} = 0, \quad \frac{\partial g_3(W,\gamma,\theta)}{\partial \beta} = 0,\\
&\frac{\partial g_3(W,\gamma,\theta)}{\partial \mu_c^P} = 0, \quad \frac{\partial g_3(W,\gamma,\theta)}{\partial \mu_f^P} = -1,
\end{aligned}
$$

(B27)

where we have used

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

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 (B27), 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 $t-$ 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 permanent income. The formal hypothesis is specified as:

$$H_0 : \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\epsilon_{ct}Y_f^P\right] = 0, \qquad \text{vs} \quad H_1 : \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\epsilon_{ct}Y_f^P\right] \neq 0, \tag{B28}$$

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 permanent 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 $\theta_{cf} := \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\epsilon_{ct}Y_f^P\right]$. Second, we construct a locally robust estimator $\hat{\theta}_{cf}$. Finally, we provide a $t-$test based on $\hat{\theta}_{cf,n}$.

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:*

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

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 $\boldsymbol{X}_{gt}$*

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

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

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

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}:

$$
\begin{aligned}
\theta_{cf} &= \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\epsilon_{ct}Y_f^P\right] \\
&= \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\epsilon_{ct}\frac{1}{T}\sum_{j=1}^{T}Y_{fj}\right] \\
&= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{j=1}^{T}\mathbb{E}\left[\left(Y_{ct}-\mathbb{E}\left[Y_{ct}\mid \boldsymbol{X}_{ct}\right]\right)Y_{fj}\right] \\
&= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{j=1}^{T}\mathbb{E}\left[Y_{ct}Y_{fj}\right]-\mathbb{E}\left[\mathbb{E}\left[Y_{ct}\mid \boldsymbol{X}_{ct}\right]Y_{fj}\right] \\
&= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{j=1}^{T}\mathbb{E}\left[\mathbb{E}\left[Y_{ct}Y_{fj}\mid \boldsymbol{X}_{ct},\boldsymbol{X}_{fj}\right]\right]-\mathbb{E}\left[\mathbb{E}\left[Y_{ct}\mid \boldsymbol{X}_{ct}\right]\mathbb{E}\left[Y_{fj}\mid \boldsymbol{X}_{ct},\boldsymbol{X}_{fj}\right]\right] \\
&= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{j=1}^{T}\mathbb{E}\left[\mathbb{E}\left[Y_{ct}Y_{fj}\mid \boldsymbol{X}_{ct},\boldsymbol{X}_{fj},D_{ct}=1,D_{fj}=1\right]\right] \\
&- \frac{1}{T^2}\mathbb{E}\left[\mathbb{E}\left[Y_{ct}\mid \boldsymbol{X}_{ct},D_{ct}=1\right]\mathbb{E}\left[Y_{fj}\mid \boldsymbol{X}_{fj},D_{fj}=1\right]\right] \\
&:= \mathbb{E}\left[\frac{1}{T^2}\sum_{t=1}^{T}\sum_{j=1}^{T}\left(\mu_{cftj}\left(\boldsymbol{X}_{ct},\boldsymbol{X}_{fj},1,1\right)-\mu_{ct}\left(\boldsymbol{X}_{ct},1\right)\mu_{fj}\left(\boldsymbol{X}_{fj},1\right)\right)\right], \qquad \text{(B29)}
\end{aligned}
$$

where the second equality follows by the definition of permanent 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_{cf} that is locally robust to the nuisance parameters $\gamma_{cf} := \left(\mu_f^{1,T},\mu_c^{1,T},\mu_{cf}^{t,1,T}\right)$, where

$$
\begin{aligned}
\mu_g^{1,T} &:= \left(\mu_{g1},...,\mu_{gT}\right), \quad g\in\{c,f\}, \\
\mu_{cf}^{t,1,T} &:= \left(\mu_{cft},...,\mu_{cft}\right), \quad t=1,...,T \\
\sigma^{1,T,1,T} &:= \left(\mu_{cf}^{1,1,T},...,\mu_{cf}^{T,1,T}\right).
\end{aligned}
$$

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 (B29) is

$$
-\frac{1}{T^2}\sum_{t=1}^{T}\sum_{j=1}^{T}\mu_{fj}\left(\boldsymbol{X}_{fj},1\right)\frac{D_{ct}}{p\left(D_{ct}=1|\boldsymbol{X}_{ct}\right)}\left(Y_{ct}-\mu_{ct}\left(\boldsymbol{X}_{ct},1\right)\right).
$$

Similarly, for μ_{ft}, we have

$$-\frac{1}{T^2}\sum_{t=1}^{T}\mu_{ct}\left(\boldsymbol{X}_{ct},1\right)\sum_{j=1}^{T}\frac{D_{fj}}{p\left(D_{fj}=1|\boldsymbol{X}_{fj}\right)}\left(Y_{fj}-\mu_{fj}\left(\boldsymbol{X}_{fj},1\right)\right).$$

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

$$\frac{1}{T^2}\sum_{t=1}^{T}\sum_{j=1}^{T}\frac{D_{ct}D_{fj}}{p\left(D_{ct}=1,D_{fj}=1|\boldsymbol{X}_{ct},\boldsymbol{X}_{fj}\right)}\left(Y_{ct}Y_{fj}-\mu_{cftj}\left(\boldsymbol{X}_{ct},\boldsymbol{X}_{fj},1,1\right)\right)$$

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

$$\begin{aligned}
\psi_{cf}\left(W,\gamma_{cf},\theta_{cf}\right) &= g_{cf}\left(W,\gamma_{cf},\theta_{cf}\right)+\phi_{cf}\left(W,\gamma_{cf},\alpha_{cf},\theta_{cf}\right)\\
&= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{j=1}^{T}\left(\mu_{cftj}\left(\boldsymbol{X}_{ct},\boldsymbol{X}_{fj},1,1\right)-\mu_{ct}\left(\boldsymbol{X}_{ct},1\right)\mu_{fj}\left(\boldsymbol{X}_{fj},1\right)\right)-\theta_{cf}\\
&+\frac{1}{T^2}\sum_{t=1}^{T}\sum_{j=1}^{T}\frac{D_{ct}D_{fj}}{p\left(D_{ct}=1,D_{fj}=1|\boldsymbol{X}_{ct},\boldsymbol{X}_{fj}\right)}\left(Y_{ct}Y_{fj}-\mu_{cftj}\left(\boldsymbol{X}_{ct},\boldsymbol{X}_{fj},1,1\right)\right)\\
&-\frac{1}{T^2}\sum_{t=1}^{T}\sum_{j=1}^{T}\mu_{ct}\left(\boldsymbol{X}_{ct},1\right)\frac{D_{fj}}{p\left(D_{fj}=1|\boldsymbol{X}_{fj}\right)}\left(Y_{fj}-\mu_{fj}\left(\boldsymbol{X}_{fj},1\right)\right)\\
&-\frac{1}{T^2}\sum_{t=1}^{T}\sum_{j=1}^{T}\mu_{fj}\left(\boldsymbol{X}_{fj},1\right)\frac{D_{ct}}{p\left(D_{ct}=1|\boldsymbol{X}_{ct}\right)}\left(Y_{ct}-\mu_{ct}\left(\boldsymbol{X}_{ct},1\right)\right).
\end{aligned}\tag{B30}$$

and the debiased moment function is then computed as

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

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 $\hat{\theta}^{(\ell)}_{cf}$. Accordingly, the locally robust estimator $\hat{\theta}_{cf,n}$ is the solution to the sample moment condition $\hat{\psi}_{cf}\left(\theta_{cf}\right)=0$.

To test the null hypothesis $H_0:\theta_{cf}=0$, we implement a $t-$test based on the estimator $\hat{\theta}_{cf,n}$. Accordingly, the t statistic is given by

$$t_{cf,n}=\frac{\hat{\theta}_{cf,n}}{\sqrt{\hat{V}_{cf,n}/n}},$$

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

Consistency of $\hat{\theta}_{cf,n}$ 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, $\hat{\theta}_n^{LR} \xrightarrow{p} \theta_0$, the asymptotic normality of $\hat{\theta}_{cf,n}$ 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 $V_{cf} = \mathbb{E}\left[\psi_{cf}^2\left(W, \gamma_{cf}, \theta_{cf}\right)\right]$. In addition, if Assumption 6-LR holds, then $\hat{V}_{cf,n} \xrightarrow{p} V_{cf}$.

Under H_0 in equation (B28) and Assumptions 1-NP', 2-NP', 3-LR-6-LR, we have

$$\sqrt{n}\left(\hat{\theta}_{cf,n} - \theta_{cf0}\right) \xrightarrow{d} \mathcal{N}(0, V_{cf}) \Rightarrow \frac{\hat{\theta}_{cf,n}}{\sqrt{\hat{V}_{cf,n}/n}} \xrightarrow{d} \mathcal{N}(0, 1),$$

where $\hat{V}_{cf,n}$ is a consistent estimator of V_{cf} that accounts for dependence within families.

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 $|t_{cf,n}| > z_{1-\alpha/2}$ is consistent.

$$\begin{aligned} t_{cf,n} &= \frac{\hat{\theta}_{cf,n}}{\sqrt{\hat{V}_{cf,n}/n}} \\ &= \frac{\theta_{cf0} + O_p(n^{-1/2})}{\sqrt{(V_{cf} + o_p(1))/n}} \\ &= \frac{\sqrt{n}\theta_{cf0} + O_p(1)}{V_{cf}^{1/2} + o_p(1)} \\ &= \sqrt{n}\,\theta_{cf0}\,V_{cf}^{-1/2} + O_p(1). \end{aligned}$$

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

$$\begin{aligned} \Pr\left(|t_{cf,n}| > z_{1-\alpha/2} \mid \theta_{cf0}\right) &= \Pr\left(\sqrt{n}\,\theta_{cf0}\,V_{cf}^{-1/2} + O_p(1) > z_{1-\alpha/2}\right) \\ &\to 1, \end{aligned}$$

since $\sqrt{n}\,\theta_{cf0}\,V_{cf}^{-1/2}$ diverges as $n \to \infty$ and dominates the $O_p(1)$ term.

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

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

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, the t-statistic satisfies

$$\begin{aligned}
t_{cf,n} &= \frac{\hat{\theta}_{cf,n}}{\sqrt{\hat{V}_{cf,n}/n}} \\
&= \frac{\theta_{cf0} + (\hat{\theta}_{cf,n} - \theta_{cf0})}{\sqrt{\hat{V}_{cf,n}/n}} \\
&= \frac{\delta/\sqrt{n} + (\hat{\theta}_{cf,n} - \theta_{cf0})}{\sqrt{V_{cf}/n} + o_p(n^{-1/2})} \\
&= \frac{\delta + \sqrt{n}(\hat{\theta}_{cf,n} - \theta_{cf0})}{V_{cf}^{1/2} + o_p(1)} \\
&= \frac{\delta}{V_{cf}^{1/2}} + \frac{\sqrt{n}(\hat{\theta}_{cf,n} - \theta_{cf0})}{V_{cf}^{1/2}} + o_p(1) \\
&= \frac{\delta}{V_{cf}^{1/2}} + O_p(1).
\end{aligned}$$

By Slutsky's theorem, under H_{1n} we then have

$$t_{cf,n} \xrightarrow{d} \mathcal{N}\left(\frac{\delta}{V_{cf}^{1/2}}, 1\right),$$

so that the t-test has asymptotic power

$$\lim_{n\to\infty} \Pr\left(|t_{cf,n}| > z_{1-\alpha/2} \mid \theta_{cf0} = \frac{\delta}{\sqrt{n}}\right) = 2\left[1 - \Phi\left(z_{1-\alpha/2} - \frac{|\delta|}{V_{cf}^{1/2}}\right)\right] > \alpha \quad \text{whenever } \delta \neq 0,$$

where Φ denotes the CDF of the standard normal distribution.

Thus, the t-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

$$H_0 : \frac{1}{T^2} \sum_{|t-j|>h} \mathbb{E}\left[\epsilon_{ft}\epsilon_{fj}\right] = 0, \quad \text{vs} \quad H_1 : \frac{1}{T^2} \sum_{|t-j|>h} \mathbb{E}\left[\epsilon_{ft}\epsilon_{fj}\right] \neq 0,$$

where the prediction errors of parental income at time t are defined as $\epsilon_{ft} := Y_{ft} - \mathbb{E}\left[Y_{ft} \,\middle|\, \boldsymbol{X}_{ft}\right]$. 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[8]

$$H_0 : \frac{1}{T^2} \sum_{t-j=h+1} \mathbb{E}\left[\epsilon_{ft}\epsilon_{fj}\right] = 0, \quad \text{vs} \quad H_1 : \frac{1}{T^2} \sum_{t-j=h+1} \mathbb{E}\left[\epsilon_{ft}\epsilon_{fj}\right] \neq 0 \tag{B31}$$

We now state the Assumptions required to identify $\theta_{fh} := \frac{1}{T^2} \sum_{t-j=h+1} \mathbb{E}\left[\epsilon_{ft}\,\epsilon_{fj}\right]$

Assumption 1-NP". *(Conditional Mean Independence and Orthogonality) The observable characteristics satisfy*

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

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 $\boldsymbol{X}_{ftj}$

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

[8]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 $t - j = h + 1 < T$

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_{fh}:

$$\begin{aligned}
\theta_{fh} &= \frac{1}{T^2} \sum_{t-j=h+1} \mathbb{E}\left[\epsilon_{ft}\epsilon_{fj}\right] \\
&= \frac{1}{T^2} \sum_{t-j=h+1} \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] \\
&= \frac{1}{T^2} \sum_{t-j=h+1} \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]\right]\right) \\
&= \frac{1}{T^2} \sum_{t-j=h+1} \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] \\
&= \frac{1}{T^2} \sum_{t-j=h+1} \left(\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]\right) \\
&- \frac{1}{T^2} \sum_{t-j=h+1} \left(\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]\right) \\
&= \frac{1}{T^2} \sum_{t-j=h+1} \left(\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]\right) \\
&= \frac{1}{T^2} \sum_{t-j=h+1} \left(\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]\right) \\
&= \mathbb{E}\left[\frac{1}{T^2} \sum_{t-j=h+1} \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]\right) \\
&:= \mathbb{E}\left[\frac{1}{T^2} \sum_{t-j=h+1} \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)\right]. \qquad \text{(B32)}
\end{aligned}$$

The moment for θ_{fh} that is locally robust to the nuisance parameter $\gamma_{fh} := \left(\mu_f^{1,T}, \mu_{fh}\right)$ is given by

$$\begin{aligned}
\psi_{fh}\left(W, \gamma_{fh}, \theta_{fh}\right) = &\frac{1}{T^2} \sum_{t-j=h+1} \left(\mu_{ftj}\left(\boldsymbol{X}_{ftj}, 1, 1\right) - \mu_{ft}\left(\boldsymbol{X}_{fj}, 1\right)\mu_{fj}\left(\boldsymbol{X}_{fj}, 1\right)\right) - \theta_{fh} \\
&+ \frac{1}{T^2} \sum_{t-j=h+1} \frac{D_{ft}D_{fj}}{p\left(D_{ft} = 1, D_{fj} = 1 | \boldsymbol{X}_{ftj}\right)} \left(Y_{ft}Y_{fj} - \mu_{ftj}\left(\boldsymbol{X}_{ftj}, 1, 1\right)\right) \\
&- \frac{1}{T^2} \sum_{t-j=h+1} \mu_{fj}\left(\boldsymbol{X}_{fj}, 1\right) \frac{D_{ft}}{p\left(D_{ft} = 1 | \boldsymbol{X}_{ft}\right)} \left(Y_{ft} - \mu_{ft}\left(\boldsymbol{X}_{ft}, 1\right)\right)
\end{aligned}$$

$$- \frac{1}{T^2} \sum_{t-j=h+1} \mu_{ft}\left(\boldsymbol{X}_{ft}, 1\right) \frac{D_{fj}}{p\left(D_{fj} = 1 | \boldsymbol{X}_{fj}\right)} \left(Y_{fj} - \mu_{fj}\left(\boldsymbol{X}_{fj}, 1\right)\right), \tag{B33}$$

and its corresponding estimator solves the cross-fitted orthogonal moment. Consequently, the locally robust $t-$test for H_0 in equation (B31) is given by

$$t_{fh,n} = \frac{\hat{\theta}_{fh,n}}{\sqrt{\hat{V}_{fh,n}/n}},$$

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