A Nonparametric Non-classical Measurement Error Approach to Estimating Intergenerational Mobility Elasticities[*]

Yonghong An[†]   Le Wang[‡]   Ruli Xiao[§]

Abstract

This paper provides a framework for estimating intergenerational mobility elasticities (IGE) of children's income with respect to parental income. We allow the IGEs to be heterogeneous, by leaving the relationship of parental and child incomes unspecified, while acknowledging and addressing the *latent* nature of both child and parental *permanent* incomes and the resulted life-cycle bias. Our framework enables us to test the widely imposed assumption that the intergenerational mobility function is linear. Applying our method to the Panel Studies of Income Dynamics (PSID) data, we decisively reject the commonly imposed linearity assumption and find substantial heterogeneity in the IGEs across the population. We confirm an important finding that the IGEs with respect to parental income exhibit a U-shape pattern, which is occasionally highlighted in the analysis using transition matrices. Specifically, there is a considerable degree of mobility among the broadly defined middle class, but the children of both high- and low-income parents are more likely to be high- and low-income adults, respectively. This result also provides insights into the (intertemporal) Great Gatsby curve, suggesting that a higher level of inequality within one generation may lead to a higher level of social immobility in the next generation in the U.S..

Keywords: Intergenerational Mobility, Inequality, Measurement Error, Non-parametric Estimation

[†]Department of Economics, TAMU, College Station, TX 77843. E-mail: y.an@tamu.edu; Phone: (979) 845-8685; Fax: (979) 847-8757.

[‡]Department of Economics, University of Oklahoma, Norman OK, 73019. Email: Le.Wang.Econ@gmail.com; Phone: (405) 325-4348; Fax: (405) 325-5842.

[§]Department of Economics, Indiana University, 100 S. Woodlawn Ave. Bloomington, IN 47405. Email: rulixiao@indiana.edu; Phone: (812) 855-3213 ; Fax: (480) 855-3736. The authors would also like to thank Jim Heckman for his helpful comments.

1 Introduction

Equality of opportunity, the idea that individual success is determined by one's effort and motivation, but not one's family background, is a value central to the so-called "American Dream." To the extent that income is a good measure of success, the size of the correlation between (log) child income and (log) parental income (often called intergenerational mobility elasticity, IGE) is widely used to gauge the degree of equality of opportunity, and to provide important insights into the evolution of income distributions over generations. Economists and social scientists have long been interested in documenting this measure (Black and Devereux, 2011; Durlauf and Shaorshadze, forthcoming), and estimation of it presents much more than just an object for academic curiosity.

In this paper, we focus on two methodological issues in the estimation of IGE (among many): nonlinearity/heterogeneity and life-cycle bias *simultaneously*. Heterogeneity can matter. It is well documented that the family plays a vital role in shaping child outcomes through genetics, investment in human capital, and through choices of child environments (Cunha and Heckman, 2007). Economic theories have predicted that families with different socio-economic statuses differ in their resources and incentives to invest in their children, which in turn influence the future success of their children. Various mechanisms such as education and neighborhood quality imply that the intergenerational relationships between child income and parental income may be nonlinear and depend on the level of family income, thereby giving rise to heterogeneity in IGEs across socio-economic groups (Durlauf, 1996b,a).

The second issue is concerned with what the relevant income measure is in the estimation. As Solon (1992) notes, *the* relevant income measure for estimation is *permanent income*, a concept originally proposed by Friedman (1957). The original definitions of permanent and transitory incomes in Friedman (1957) are, however, only conceptual and not immediately operational in practice since both are considered as *latent*, and we observe only their imperfect measures – annual incomes. Use of annual incomes in estimations can lead to (non-classical) measurement errors, the so-called *life-cycle bias*, the magnitude of which varies systematically over the life-cycle, and hence bias the estimates. Note that this type of bias is distinct from *the self-reporting bias* inherent in the survey data, and cannot be addressed by use of data of better quality.

Modeling latent variables such as permanent income has long been a long-standing challenging issue in economics and econometrics (Chao, 2003), and even more so in our context. To see this, the above discussions suggest the following model

$$Y^* \;=\; g(X^*, Z) + \varepsilon,$$

where Y^*, X^* refer to child and parental *permanent* income, respectively. We do not know the functional form of $g(\cdot)$, nor do we observe either the dependent variable (Y^*) or the independent variable

(X^*). Addressing these two issues *simultaneously* is not trivial. It is therefore not surprising that the existing literature has often focused on only one issue while treating the other less systematically. One strand of literature focusing on nonlinearity has typically employed the "time averaging method", a two-step approach to address both issues: by first measuring permanent income with variants of averages of observed incomes and then plugging it in a particular type of nonlinear model in the second stage.[1] Another strand of the literature addresses the life-cycle bias but in the linear regression framework.[2] The former could fail because an average of observed annual incomes does not necessarily address the life-cycle bias (regardless of how many periods of data of high quality are available); the latter would fail when the intergenerational income-transmission model is nonlinear.

To address these issues altogether is challenging, and it calls for a different technique than conventional approaches (such as regression or nonparametric kernel estimation). We build on recent advances in the literature on nonparametric identification of latent variable models (Hu and Sasaki (2014)). Our approach is nonparametric allowing for nonlinearity by leaving the function $g(\cdot)$ unspecified, while directly acknowledging the *unobservable* nature of both child and parental *permanent* incomes. Our framework is also faithful to the original definitions of permanent and transitory incomes in Friedman (1957). We treat the annual incomes as a general linear combination of permanent income and transitory shocks, which accommodates non-classical measurement errors, and hence, addresses the life-cycle bias (and self-reporting bias) directly in estimation (as opposed to plugging in estimation a variant of averages of annual incomes).

We first establish the fact that normalization is *necessary* for any point identification of the mobility function and so the IGE. Examples of such normalization include imposing classical measurement-error assumptions. We then show that the intergenerational mobility function can be nonparametrically identified under a weaker normalization condition (i.e., one's annual income will reach her permanent income in the absence of transitory shocks at mid age). Our identification results lead to a fully non-parametric estimator of the mobility function with a closed form expression, which allows one to avoid numerical optimizations that often involve iterations. Our Monte Carlo simulation results illustrate and showcase the usefulness of our approach for modest-size samples.

An important feature of our approach is that only one measurement of child's permanent income and two measurements of parental permanent incomes are required for both identification and estimation. The three-measurement result is in line with the "major breakthrough in the measurement error literature", in which such models "can be non-parametrically identified under mild restrictions."

---

[1]These efforts include threshold regression (e.g., Durlauf et al., 2017), varying coefficient models (e.g., Kourtellos et al., 2019), regression tree methods (e.g., Cooper et al., 1994), spline regressions (e.g., Björklund et al., 2012), and standard nonparametric estimations (e.g., Corak and Heisz, 1999). An interesting paper by Prokhorov et al. (2015) considers a two-sample IV approach to estimate varying coefficient models to address nonlinearity while treating the life-cycle bias as a *classical* measurement error problem.

[2]Based on a linear model, one can derive the bias formula to correct for the bias (e.g., Haider and Solon, 2006). However, this actually assumes that we have data on permanent income, which is clearly unavailable! Additional distributional assumptions are also required for estimation of permanent income in Haider and Solon (2006).

(Hu, 2020). Our approach can be interpreted as nonparametric instrumental variable (IV) (in which one measurement is the outcome, one the endogenous variable, and the third the IV). Therefore, this feature is in stark contrast to the time averaging method that relies on classical measurement errors and a much longer panel of data. This feature is extremely useful when a large long-panel are not readily available to most researchers. For example, as noted in Mazumder (2018) and Chetty et al. (2014), while using a large administrative data, they have only five years of parental incomes and two years of child incomes; in this case, the time averaging method does not necessarily address the concern of classical measurement errors, let alone the life cycle bias (the non-classical ones). When a long panel dataset exists, our method is also flexible to use multiple measurements either through model averaging or using the time averages as one of the measurements.

We further the literature on the nonparametric models with non-classical measurement errors in three important ways to address issues specific to our empirical application. First, we allow *both* dependent *and* explanatory variables (child and parental permanent incomes) to be subject to measurement errors, instead of explanatory variables only. Such an extension is not trivial in the presence of non-classical measurement errors in dependent variables. A by-product of our paper is therefore a practical framework that is immediately applicable in many other similar contexts both within and outside labor economics (for example, nonparametric estimation of an educational production function where both student academic achievement and school quality are unobservable, and we observe measurements of each latent variable). Second, an important identification result of ours is that even when the normalization is misspecified, the general pattern of the heterogeneous IGEs using our closed-form solution remains invariant. This also leads us to focus our discussions on the uncovered pattern of the estimates, instead of the actual estimates themselves. Finally, together with this fact, our estimator enables us to propose a test of the linearity assumption underlying much of the existing empirical literature. To the best of our knowledge, our test is the first formal test of the linearity assumption in this context, allowing for not only the life-cycle bias, but also potentially misspecified normalization. Similar to Chen et al. (2016), the test of linearity is based on a flexible polynomial function fitted to the estimated slopes/derivatives of the *nonparametric* mobility function.

We illustrate our method using the PSID data and reach three main conclusions and implications. First, we decisively reject the linearity assumption often imposed in the estimation of the mobility function at the 1% significance level. This result highlights the importance of a nonparametric model to accommodate nonlinearity.

Second, we confirm a previous finding that is understood among practitioners in the field and is occasionally highlighted in published papers that use transition matrices (Corak et al., 2011; Isaacs et al., 2008). We also find there exists substantial heterogeneity and a U-shape pattern in IGEs across the population or within subgroups such as males and whites. Specifically, there is a considerable degree of mobility among the broadly defined middle earnings group, but both the children of high- and low-earning parents are more likely to grow up to be, respectively, high- and low-earning adults.

These results are remarkably robust to the choice of normalization and the ages of parents and of children.

Our results based on continuous IGEs also add to the U-shape finding. Our findings reassure that the results obtained using the transition-matrix in the previous literature are not an artifact of the floor-ceiling effect (Corak and Heisz, 1999); that is, that people in the very top and very bottom of the father's income distribution tend to have higher probabilities to stay in the same income category is not simply because those at the very top are restricted from further upward movement and those at the very bottom further downward movement. Instead, it is indeed a result of strong persistence in income associations among them.[3] Our non-parametric results characterize the magnitude of the persistence across the population in great details, which can potentially be masked by the approaches based on ranks or percentiles. Our estimates of IGEs facilitate the comparison to the literature based on regression approaches.[4]

Finally, in addition to being easy-to-interpret, our results can provide some insights into the Great Gatsby curve. Using our estimates, we perform simulation exercises to produce the relationship between inequality and mobility. The simulation results suggest an (intertemporal) Gatsby Curve, suggesting that a higher level of inequality within one generation may lead to a higher level of social immobility in the next. This result confirms that "an intertemporal Gatsby Curve is a salient feature of inequality in the United States." (Durlauf and Seshadri, 2017)

The outline of the paper is as follows. Section 2 lays out our theoretical results. Section 3 illustrates our approach using the PSID data and presents the results. Section 4 concludes. Proofs, tables, and figures are included at the end of the paper.

2 Empirical Strategy

2.1 The (Augmented) Model: Nonlinearity and Life-cycle Bias

Consider a model that describes a general intergenerational income transition:

$$Y^* = g(X^*, Z) + \varepsilon, \tag{1}$$

where Y^* and X^* are respectively the permanent (log) incomes of children and parents. Z is a vector of exogenous covariates of children: race, gender, and region; ε are other factors that might affect the

---

[3]Our results also avoid the issue of classification. When the classification of income groups is less refined, we may also exaggerate persistence when the movement is not necessarily drastic. Our findings also confirm that the persistence is not a result of such classification effects.

[4]With transition matrices, "there is no best way to summarize their content" (Hertz, 2005), or exists "a unique summary statistic of mobility" (Durlauf et al., 2017). More importantly, the estimated transition probabilities can capture only *rank mobility* and are not directly comparable to IGEs. For example, if the estimated probability from the second to the top quartiles is 25%, this figure cannot tell us by how much the child's income will increase should the father's income increase by a certain amount, as IGE does.

intergenerational income transition that we cannot control for. $g(\cdot)$ is an unspecified mobility function relating parental income X^* to child income Y^*. This non-parametric specification is consistent with the general nonlinear family investment income transmission model (Durlauf and Seshadri, 2017) and nests the linear specification commonly used in the literature as a special case. The IGE at a realized $Z = z$ is captured by the derivative of the mobility function, i.e., $\partial g(X^*, Z = z)/\partial X^* \mid_{X^*=x^*}$, which is allowed to vary with the parental income X^*.

To identify and estimate the IGE, we augment the above model by specifying the relationships between (latent) permanent income and observed incomes (see, e.g., Haider and Solon, 2006):

$$
\begin{aligned}
Y_t &= \delta_t Y^* + U_t, \\
X_t &= \alpha_t X^* + V_t, \quad t = 1, 2,
\end{aligned}
\tag{2}
$$

where t represents the stage of each individual in her life cycle such as ages; U_t and V_t are transitory shocks. δ_t and α_t, often called life-cycle coefficients, capture the weight of the income for age t in the permanent income. Note that $\delta_t > 0$ and $\alpha_t > 0$. $\delta_t \neq 1$ and $\alpha_t \neq 1$ implies non-classical measurement error or the life-cycle bias. This age-dependent property is due to heterogeneous age-earnings profiles. Y_t and X_t are the *observed incomes* at period t for the two generations, respectively. The structural error ε is assumed to be mean independent of one of the measurements of permanent parental income. Without loss of generality, we assume $E[\varepsilon|X_2] = 0$. This is actually a weaker assumption than what is typically assumed in the literature using an average of *all* observed incomes in the second stage nonlinear models. For example, Jenkins (1987) assumes that the random error is "independent of all other variables". Violation of this assumption simply means that we cannot necessarily interpret the result as indicating a causal relationship.

Equation (2) highlights the latent nature of permanent income and the life-cycle bias. We cannot observe the permanent income directly in the data. Instead, we can only have access to the actual annual incomes of each individual at different stages of their lives. As evident in Equation (2), these incomes are imperfect measures of the permanent income; how well annual income in a particular period approximating the permanent income varies with age.

The *ad hoc* time averaging method, which has been traditionally followed in the literature dating back to Friedman (1957) (Black and Devereux, 2011), cannot address the issue of perfectly approximating the permanent income and may even worsen it. Consider a simple case where one annual income equals the permanent income plus a classical measurement error, and the other is not (i.e., $X_1 = X^* + V_1, X_2 = \alpha X^* + V_2, \alpha \neq 1$). Averaging two years of observations results in $\overline{X} = \frac{1}{2}(X_1 + X_2) = \widetilde{\alpha} X^* + \widetilde{V}$, with $E[\overline{X} - X^*|X^*] \neq 0$ where $\widetilde{\alpha} = (\frac{1}{2} + \frac{1}{2}\alpha)$ and $\widetilde{V} = \frac{1}{2}(V_1 + V_2)$. This implies that the averaging method, although it may reduce the impact of transitory income shocks, leads to a non-classical measurement error structure, for which the direction of the bias is a *priori* unknown.

2.2 Identification of Mobility Function and IGE

We now turn to the identification of the IGE. Our identification method applies conditional on the vector of characteristics Z. For the sake of simplicity, we suppress Z below whenever there is no ambiguity. We first clarify a fundamental issue facing the literature on identifying IGE (or in a similar context) by making the following observation:

Claim 1. *[**Normalization is Required**] Not all the δ_t and α_t, $t = 1, 2$, are uniquely determined from the observed Y_t and X_t in model (1).*

The mobility function and the life-cycle biases are identified up to a scale. Additional restrictions or normalizations are necessary for point identification. Normalization is commonly required even in the linear case (see, e.g., Black and Smith, 2006; Cunha et al., 2010). For example, Black and Smith (2006) consider a linear generalized method of moments (GMM) to estimate the returns to the latent school quality. They normalize the variance of the underlying latent school quality to one. Such normalization is obviously not plausible in our context when the underlying variable is permanent income.

Below we discuss how we can use repeated measurements of the latent variables to non-parametrically identify $g(\cdot)$ under a plausible normalization assumption. We then consider the identification when such normalization fails to hold. Identifying the shape of the mobility function, the main purpose of the paper, is still robust and can be tested even with misspecified normalization.

2.2.1 Assumptions

Here we discuss the assumptions for our identification. There are several types of assumptions: Assumption 1 is concerned with normalization of the life-cycle bias (only needed for point identification), Assumptions 2 and 3 are about zero-correlations between transitory shocks and other variables, and Assumption 4 imposes some regularity conditions for identification. As we discuss below, our assumptions are either weaker or similar to those typically imposed in the existing literature. These are only one set of possible assumptions, and alternative assumptions may be feasible for identification. But these assumptions provide a first step in the literature to treat the issues more systematically in the same framework. Below summarizes all the assumptions used in our identification.

Assumption 1. *[**Normalization**] There exist two periods, $t_0, t_1 \in \{1, 2\}$, such that (i) $X_{t_1} = X^* + V_{t_1}$, and (ii) $Y_{t_0} = Y^* + U_{t_0}$, where t_0 and t_1 are not necessarily the same. We call t_0 and t_1 baseline years for children and parents, respectively.*

Assumption 2. *[**Transitory Shocks to Income (I)**] (i) $Cov(U_t, X^*) = Cov(V_t, Y^*) = 0$, and (ii) $Cov(V_1, U_t) = Cov(U_1, V_t) = 0$, t=1, 2.*

Assumption 3. [*Transitory Shocks to Income (II)*] (i) X^* is independent of V_2, (ii) $E[V_1|X_2] = 0$, and (iii) $E[U_2|X_2] = 0$.

Assumption 4. [*Regularity Assumptions*] (i) ϕ_{X_2} does not vanish on $\mathbb{R}$, (ii) f_{X^*} and $\phi_{X^*}(\cdot)$ are continuous and absolutely integrable, where $\phi_{X^*}(\cdot)$ is the characteristic function of the density f_{X^*}, and (iii) The convolution of f_{X^*} and $g(\cdot)$, $f_{X^*} * g(\cdot)$ is continuous and absolutely integrable, and so is $\int_{-\infty}^{\infty}(f_{X^*} * g)dx$.

Assumption 1: The normalization assumes that, in the absence of any transitory shocks, one's income will eventually reach his permanent income at one point. Unlike the normalization on the variance, this assumption is more natural and economically plausible. There is also strong evidence supporting this normalization. For example, Haider and Solon (2006) estimate α_t and δ_t for the 1951-1991 Social Security Administration earnings histories of the members of the Health and Retirement Study sample. Assuming that the life-cycle coefficients of children and parents are the same, i.e., $\alpha_t = \delta_t$, they show that at the age of 32 the coefficient crosses one and remains insignificantly different from one until the middle or late forties. Böhlmark and Lindquist (2006) analyze the Swedish income tax data and find that the coefficient is closest to one at the age of 34. In fact, a widely used rule-of-thumb in the literature is that permanent income is better approximated at one's midlife than at early or later stages.

Our specification of the life-cycle bias is weaker than what is assumed in the conventional method such as averaging incomes. Unlike the averaging method, which implicitly imposes normalization on *all* time periods, we require only one period and allow for the scale of the annual incomes in any other years to differ from permanent income with a non-classical measurement error ($\alpha_2, \delta_2 \neq 1$). For ease of notation, we denote the normalization period as 1 and the other periods as 2 for both children and parents.

Assumptions 2 and 3 (transitory shocks, i.e., U_t, V_t): Following the literature (e.g., Haider and Solon, 2006; Jenkins, 1987), we impose independence assumptions on transitory shocks, which are plausible by the stochastic nature of these shocks. Mean independence assumptions are typically required for nonparametric identification and estimation like ours.

Most of these assumptions are rather standard but often implicit in the literature. Our discussions here clarify these assumptions and intend to evaluate them as a whole. These assumptions are plausible and faithful to the original definition of "transitory shocks" in Friedman (1957). While the dichotomy of observed income between permanent and transitory incomes in Equation (2) is not controversial, how to clearly define and separate them is unclear and open to debate. In fact, in his original work, Friedman admits that "The division [of permanent and transitory income] is, of course, in part arbitrary, and just where to draw the line may well depend on the particular application." (Friedman, 1957). Friedman (1957) defines the transitory income as a residual (measured income

minus permanent income). As stated in Friedman (1957, p.26-27), "[transitory incomes] have little substantive content and can almost be regarded as simply completing or translating the defnitions of transitory and permanent components [of income]; the qualitative notion that the transitory component is intended to embody is of an accidental and transient addition to or subtraction from income, which is almost equivalent to saying an addition or subtraction that is not correlated with the rest of income. The merging of errors of measurement with transitory components contributes further to the plausibility that these correlations are zero." Examples of such temporary income shocks include an inheritance, a lottery win, and a spell of unemployment.

Similar to how they are defined in many of the literature e.g., Haider and Solon (2006), Jenkins (1987) and Grawe (2006), these transitory shocks, U_t and V_t, are considered to be more of random shocks or "noises", and hence are unlikely to be correlated with structural errors or across generations as stated in Assumptions 2 and 3. These assumptions can often be readily satisfied by carefully choosing the measurements in practice.

Bearing these discussions in mind, we now look at each assumption more closely.

Assumption 2 Transitory Shocks to Income (I) (i) $Cov(U_t, X^*) = Cov(V_t, Y^*) = 0$, and (ii) $Cov(V_1, U_t) = Cov(U_1, V_t) = 0, t = 1, 2$.

Assumption 2 is standard in the literature. Parts (i) states that transitory shocks of one generation are uncorrelated with the permanent incomes of another. In other words, the parental (child) permanent income is uncorrelated with the transitory shocks to child (parental) income in both periods. Part (ii) states that transitory shocks are uncorrelated across generations. Unlike the literature (e.g., Jenkins, 1987 and Grawe, 2006), we need to impose such restrictions only on the transitory shocks in the baseline year (in which the observed income is equal to the permanent income plus transitory shocks). This assumption does not rule out the possibility of autocorrelation across generations, and it can be achieved by carefully choosing measurements that are far apart. To see this, suppose that the baseline year for both parents and children is age 34, and a second-year observation for children age 28. $Cov(U_2, V_1) = 0$ requires that the transitory shock to a child's income at age 28 is uncorrelated with any transitory shock to the parent's income at age 34. It is hard to imagine these two shocks will be correlated because they are about 20 years apart (assuming the parent is 25 at the children's birth). A similar argument applies to $Cov(U_1, V_1) = 0$. The discussion above can be readily extended to part (ii) for children.

Assumption 3: Transitory Shocks to Income (II) (i) X^* is independent of V_2, (ii) $E[V_1|X_2] = 0$, and (iii) $E[U_2|X_2] = 0$.

Part (i) of Assumption 3 is a textbook assumption on transitory shocks, requiring the permanent income X^* to be independent of the transitory shock V_2 at period 2. Note that in specification (2), we have $E[X_2 - X^*|X^*] = E[(\alpha_2 - 1)X^* + V_2|X^*] = E[(\alpha_2 - 1)X^*] \neq 0$ whenever $\alpha_2 \neq 1$. Thus,

the independence assumption in part (i) does not require the measurement error to be classical (i.e., $E[X_2 - X^*|X^*] = 0$).

Part (ii) of Assumption 3 imposes mean independence of the transitory shocks in the normalized period from the other observed income. It assumes essentially that the transitory shocks are mean independent across years. Again, it does not rule out the possibility of autocorrelation, but simply requires weaker autocorrelation that could disappear quickly when the number of time periods between the two measurements increases. In practice, we again have the flexibility to choose the period $t = 2$ to make this part of assumption more plausible. For example, given $t = 1$ being 34, we can choose $t = 2$ to be, say, 38, so that $E[V_1|X_2] = 0$ is plausibly met because it is likely that one's annual income at age 38 has no predictive power of the transitory shock to her income at age 34 (which, again, is random by nature).

Part (iii) of Assumption 3 is simply a strengthened version of no correlation of transitory shocks across generations (as in Assumption 2) required for nonparametric identification. It requires that the transitory shock to (second measurement of) children' income is mean-independent of the transitory shock to (second measurement of) parental income. This can be readily achieved by choosing the second measurement of child and parental income to be far apart from each other. Note that our Assumption 3 is less restrictive than the assumption of mutual independence of V_1, V_2, and X^* for parents, which is commonly imposed in the literature.

Assumption 4: Part (i) is a widely imposed restriction in the literature on measurement errors, where characteristic functions often appear as denominators. Many commonly used distribution families satisfy this requirement, such as exponential, gamma, chi-squared, and normal distributions. Parts (ii) and (iii) are commonly used regularity conditions that enable us to apply the Fourier transform and inversion to those functions.

2.2.2 Identification Results (I): Point Identification

Our identification proceeds with three steps. We first identify the life-cycle coefficients, δ_2 and α_2, then the *distribution* of parental permanent income, f_{X^*}, and finally the mobility function, $g(\cdot)$. Intuitively, the joint distribution of the observed incomes from multiple periods reveals information on the distribution of the permanent income, whereas the joint distribution of parental and children's incomes further helps us recover the unspecified mobility function $g(\cdot)$.

Lemma 1 (Step 1: Identification of Life-cycle Coefficients). *The coefficients δ_2 and α_2 are identified under Assumptions 1 (i) and 2 (i), and 1 (ii) and 2 (ii), respectively.*

$$\delta_2 = \frac{Cov(X_1, Y_2)}{Cov(X_1, Y_1)}, \quad \alpha_2 = \frac{Cov(X_2, Y_1)}{Cov(X_1, Y_1)}. \tag{3}$$

Once α_2 and δ_2 are identified, we obtain repeated measurements for the permanent income X^* and Y^*, which allow us to identify their distributions non-parametrically. Such an approach has been widely used in the literature of measurement errors. The basic idea is to use the one-to-one mapping between characteristic functions (hereinafter, ch.f.) and density functions.

Lemma 2 (**Step 2: Identification of Distribution of Parental Permanent Income**). *The density f_{X^*} is non-parametrically identified under Assumptions 1 (i), 2 (i), and 3(i)-(ii).*

$$f_{X^*}(x^*) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-isx^*} \phi_{X^*}(s)ds = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-isx^*} \exp\left(\int_0^s \frac{\frac{\partial}{\partial s_1}\phi_{X_1,X_2}(s_1, s_2/\alpha_2)\big|_{s_1=0}}{\phi_{X_1,X_2}(s_2/\alpha_2)\big|_{s_1=0}} ds_2 \right) ds. \quad (4)$$

Proof. See Appendix. ∎

The identification of the density f_{X^*} from two measurements of X^* follows the insights in Kotlarski (1966) and Rao (1992), in which the information on the latent variable X^* is explored in the characteristic functions of its measurements. Similarly, we may identify the density of the children's permanent income f_{Y^*} under a similar restriction. Nevertheless, as will be shown below, the density f_{Y^*} is not necessary for identifying the mobility function $g(\cdot)$.

Proposition 1 (**Step 3: Identification of Mobility Function**). [*Hu and Sasaki (2014)*] *Suppose Assumptions 1 (i), 2 (i), and 3-4 hold. The mobility function $g(X^*)$ is non-parametrically identified from two periods of parental incomes and one period of children's income.*

$$g(x^*) = \frac{1}{2\pi\delta_2 f_{X^*}} \int_{-\infty}^{+\infty} e^{-isx^*} \phi_{X^*}(s) \frac{\frac{\partial}{\partial v}\phi_{X_2,Y_2}(s/\alpha_2, v)\big|_{v=0}}{i\phi_{X_2}(s/\alpha_2)} ds. \quad (5)$$

Proof. See Appendix. ∎

Proposition 1 provides a closed-form estimator to the unknown function $g(X^*)$, conditional on the covariate Z. In an earlier paper, Schennach (2004) proposes a nonparametric estimator for $g(X^*)$ in a special case $\alpha_2 = 1$. Such a closed-form estimator does not rely on any optimization algorithms, so it has several advantages over a maximum likelihood estimation (MLE) or GMM. First, a closed-form estimator is global by construction. By contrast, an optimization algorithm, such as MLE, can only guarantee a local maximum or minimum even when a global solution exists. Second, it allows us to analyze how parameters affect the estimate constructively while this can only be done numerically for an estimator using optimization algorithms. Third, a closed-form estimator is computationally more convenient since most of the optimization algorithms involve iterations.

2.2.3 Identification Results (II): Misspecified Normalization

Following the literature, we have thus far considered a necessary normalization to *point* identify the mobility function. Below we show that a useful advantage of our closed-form solution is that it enables us to provide information on the mobility function even when such a normalization is mis-specified. Moreover, while misspecified normalization could certainly bias estimates of IGEs, it is inconsequential for identifying the pattern of interest, which is indeed our focus. The identification result for a mobility function without corrected normalization is innovative and important for testing linearity of the mobility function.

Recall that without normalization in Assumption 1, the observed income Y_1 and X_1 in the baseline years are given by $Y_1 = \delta_1 Y^* + U_1$, $X_1 = \alpha_1 X^* + V_1$, where δ_1 and α_1 are unknown life-cycle coefficients. The theorem below shows that the coefficients, δ_1 and α_1, affect the closed-form of function $g(\cdot)$ in a particular way such that $g(\cdot)$ could still be obtained without the correct information of α_1 and δ_1.

Theorem 1. *Suppose Assumptions 2(i), and 3-4 hold. Let $\widetilde{g}(\cdot)$ denote the mobility function identified based on Proposition 1 by falsely assuming that Assumption 1 holds. Then:*

$$\widetilde{g}(x^*) = \delta_1 g(x^*/\alpha_1) \ \text{ or equivalently } \ g(x^*) = \frac{1}{\delta_1}\widetilde{g}(\alpha_1 x^*). \tag{6}$$

Proof. See Appendix. ∎

This theorem expresses explicitly how an inaccurate relationship between permanent income and an observed annual income affects the mobility function. Interestingly, the coefficient for the children (δ_1) affects only the scale of the mobility function, and such effect is independent of the underlying mobility function $g(\cdot)$, whereas the effects of the coefficient for parents α_1 depend on $g(\cdot)$. Consider a numerical example: $Y^* = X^* + 2X^{*^2} + \epsilon$, $Y_1 = 0.5Y^* + U_1$, $X_1 = 0.3X^* + V_1$. Our theorem implies that the identified mobility function under misspecified normalization will become

$$Y^* = \delta_1 \cdot \frac{1}{\alpha_1}X^* + \delta_1 \cdot 2 \cdot (\frac{1}{\alpha_1}X^*)^2 + \epsilon = \frac{5}{3}X^* + \frac{100}{9} \cdot X^{*^2} + \epsilon.$$

The pattern of the derivatives (i.e., IGEs) remains the same under misspecified normalization.

The results in Theorem 1 have some important implications. First, it provides a convenient way to construct bounds on the mobility function should we have some imperfect knowledge of the range of the life-cycle coefficients (α_1 and δ_1), as opposed to having precise knowledge of in which period the coefficient equals to 1. For instance, the confidence interval of α_1 and δ_1, as in Haider and Solon (2006), can be used as a reasonable bound.

Second, if the *true* life-cycle coefficients for the baseline years are the same across generations (i.e., $\delta_1 = \alpha_1$), as in Haider and Solon (2006), mis-specified normalization does not bias the estimation of

the IGE as long as the true mobility function is linear. We summarize this result in the following corollary.

Corollary 1. *If the mobility function is linear, i.e., $g(x^*) = \beta_0 + \beta_1 x^*$, then (i) $\widetilde{g}(x^*) = \beta_0 \delta_1 + \beta_1 \frac{\delta_1}{\alpha_1} x^*$, and (ii) the IGE ($\beta_1$) is correctly identified if the two generations share the same relationship between permanent and observed income, i.e., $\delta_1 = \alpha_1$.*

Proof. See Appendix. ∎

The first part of Corollary 1 explicitly states that if the mobility function is linear, the linearity sustains even if one mis-specifies the baseline years. This result has an important implication for empirical analysis: our test results below do not depend on normalization. As long as we can reject the linearity assumption for the mobility function (i.e., constant IGE), this result holds true, whether or not the normalization is correct. This is one of the great advantages of our method, especially in the context of estimation of IGE in which the normalization is inevitable (see Claim 1).

In the second part of Corollary 1, the condition that the same income model for two generations is standard and widely imposed in the literature of intergenerational mobility of income. Under this condition, an interesting result arises: regardless of normalization, our method always identifies the true, linear mobility function.

2.3 Estimation and Testing of Linearity

2.3.1 Estimation of Life-cycle Coefficients, Mobility Function, and IGEs

The mobility function has closed-form identification, yielding a closed-form estimator. Given a sample of $\{X_{jt}, Y_{jt}\}_{j=1}^{N}, t = 1, 2$, the derivative of mobility function (i.e., IGE) is estimated in multiple steps. First, the coefficients α_2 and δ_2 can be directly estimated using the sample analog of Equation (3).

$$\widehat{\delta}_2 = \frac{\frac{1}{N}\sum_j X_{1j} Y_{2j} - \frac{1}{N}\sum_j X_{1j} \frac{1}{N}\sum_j Y_{2j}}{\frac{1}{N}\sum_j X_{1j} Y_{1j} - \frac{1}{N}\sum_j X_{1j} \frac{1}{N}\sum_j Y_{1j}}. \tag{7}$$

α_2 can be estimated analogously. Based on the estimate of α_2 and δ_2, together with the closed-form provided in Lemma 2, we estimate the densities of permanent income f_{X^*} in the second step.

$$\widehat{f}_{X^*}(x^*) = \frac{1}{2\pi} \int_{-S_N}^{S_N} e^{-isx^*} \exp\left(i \int_0^s \frac{\sum_{j=1}^N X_{1j} \exp(iX_{2j}v/\widehat{\alpha}_t)}{\sum_{j=1}^N \exp(iX_{2j}v/\widehat{\alpha}_2)} dv\right) \phi_K(s/S_N) ds, \tag{8}$$

where $i \equiv \sqrt{-1}$, $\exp\left(i \int_0^s \frac{\sum_{j=1}^N X_{1j}\exp(iX_{2j}v/\widehat{\alpha}_t)}{\sum_{j=1}^N \exp(iX_{2j}v/\widehat{\alpha}_t)} dv\right)$ is an estimate of $\phi_{X^*}(s)$, the characteristic function of f_{X^*}, $\phi_K(\cdot)$ is the characteristic function of the kernel function $K(\cdot)$, and S_N is a smoothing parameter depends on the sample size N. To ensure that the estimate of $\phi_{X^*}(s)$ uniformly converges

to its true function over $[-S_N, S_N]$ at a geometric rate with respect to the sample size N, Hu and Ridder (2008) suggest a form of $S_N = O(N/\log N)^\gamma$ for $\gamma \in (0, 1/2)$. Next, we estimate the mobility function using the closed-form expression in Proposition 1:

$$\widehat{g}(x^*) = \frac{1}{2\pi\widehat{\delta}_2\widehat{f}_{X^*}} \int_{-S_N}^{S_N} e^{-isx^*} \exp\left(i\int_0^s \frac{\sum_{j=1}^N X_{1j}\exp(iX_{2j}v/\widehat{\alpha}_2)}{\sum_{j=1}^N \exp(iX_{2j}v/\widehat{\alpha}_2)}dv\right) \frac{\sum_{j=1}^N Y_{2j}\exp(isX_{2j}/\widehat{\alpha}_2)}{\sum_{j=1}^N \exp(isX_{2j}/\widehat{\alpha}_2)}\phi_K(s/S_N)ds. \tag{9}$$

Once the mobility function is estimated, we estimate the IGEs, the derivative function of the mobility function, by employing the simple kernel method proposed in Rilstone and Ullah (1989).

$$\widehat{g}'(x^*) = \frac{\widehat{dg(x^*)}}{dx^*} = \frac{\widehat{g}(x^* + h_N) - \widehat{g}(x^* - h_N)}{2h_N}, \tag{10}$$

where h_N is a bandwidth such that $h_N \to 0$ and $Nh_N^3 \to \infty$ as $N \to \infty$. Our estimator has similar asymptotic properties to that of Hu and Sasaki (2014). Therefore, we only focus on its finite sample performance in Monte Carlo studies.

2.3.2 Testing for Linearity

The estimated derivative function $\widehat{g}'(\cdot)$ allows us to test the widely imposed linearity assumption of the mobility function. We adopt a simple approach of easy implementation involving a t- or F-test. Specifically, we fit the nonparametrically estimated derivative function $\widehat{g}'(\cdot)$ using a polynomial function, $\widehat{g}'(x^*; \rho) = \rho_0 + \rho_1 x^* + \cdots + \rho_k x^{*k} + u$, where $k \geq 1$ is specified by researchers. The null hypothesis that the mobility function is linear is equivalent to IGE $(g'(\cdot))$ is a constant, i.e., $\rho_1 = \cdots = \rho_k = 0$. Thus, we specify our null and alternative hypotheses as follows:

$$H_0 : \rho_1 = \cdots = \rho_k = 0 \quad v.s. \quad H_1 : H_0 \text{ is not true.}$$

Rejecting H_0 implies rejecting the linearity assumption. To compute the t- or F-test statistic, we estimate the vector of parameters $\rho \equiv (\rho_0, \rho_1, \cdots, \rho^k)$ using least squares and the variance using bootstrap. This method has been widely used for specification tests, e.g., Chen et al. (2016) for time-varying coefficient realized volatility models. While failing to reject the null does not necessarily imply that the mobility function is linear, rejecting the null reveals a strong signal suggesting a nonlinear mobility function. Thus, our test works particularly well when our goal is to detect nonlinearity and when the null hypothesis is rejected, as in our empirical application. When we fail to reject the null hypothesis, further test of linearity should be developed, as discussed in Li and Racine (2007).

More importantly, the conclusions based on our proposed test will not be affected even if the normalization is mis-specified. To see this, suppose that the true mobility function is linear, $g(x^*) =$

$\beta_0 + \beta_1 x^*$, but we mistakenly specify that $\delta_1 = \alpha_1 = 1$ while $\delta_1 \neq 1$ and $\alpha_1 \neq 1$. From Corollary 1, we have $\widetilde{g}(x^*) = \delta_1 g(x^*/\alpha_1) = \beta_0\delta_1 + \beta_1\frac{\delta_1}{\alpha_1}x^*$, the mobility function estimated from mis-normalization is still a linear function with intercept and slope being $\beta_0\delta_1$ and $\beta_1\delta_1/\alpha_1$, respectively. Thus testing the linearity of the mobility function estimated using the mis-normalization $\widetilde{g}(\cdot)$ is equivalent to testing that of the original mobility function $g(\cdot)$.

In the online appendix, we design several Monte Carlo studies to illustrate the finite-sample performance of our estimators and show the effectiveness of our test. These results showcase the satisfactory finite-sample performances of our proposed approach, even for modest size samples.

2.4 Further Discussions and Extensions

2.4.1 Continuous Control Variables

The results of identification, estimation, and testing in our paper are conditional on the children's exogenous covariates $Z = z$. Even though the identification results can be applied to a model with both discrete and continuous Z, the current estimation method is better suited to handle discrete covariates Z. This is primarily because, as with the literature, our primary goal here is to propose a method to characterize and document unconditional IGE and to document some interesting and easy-to-interpret variations in conditional IGEs. For example, the recent literature has focused on gender, geographic or racial differences in IGE (e.g., Chetty et al. (2014)).

While a *estimation* method for the full-blown nonparametric model with continuous variables is challenging and still underway, we discuss two approaches to incorporate continuous variables in estimations within our current framework. First, a straightforward approach is often to conduct estimation by meaningfully discretizing the variables. Whether discretization is sufficient or not depends on the empirical context and the nature of the exogenous variable. For example, while age is a continuous variable, Durlauf et al. (2017) discretize the age support to age cohorts to better characterize the dynamics and evolution of the IGE.

Second, when discretization is not completely satisfactory, one may need to impose more structure on the mobtliy function $g(\cdot)$ to incorporate Z. One possibility is to specify $g(\cdot)$ as a partially linear model as $g(X^*, Z) = Z'\beta + h(X^*)$, where $h(\cdot)$ is nonparametric. The model parameters $(\beta, h(\cdot))$ can be identified and estimated using our results in several steps. The estimation of such a partially linear model is also discussed in Section 4.2 of Härdle et al. (2012). Our Online Appendix discusses in more detail the identification and estimation for this particular model.

2.4.2 Sieve MLE as an Alternative Estimation Method

An alternative approach to estimate the mobility function $g(\cdot)$ is a semi-parametric sieve MLE, where $g(\cdot)$ is specified as a parametric function and the distributions of parental permanent income and

of measurement errors are nonparametric. The advantage of such a semi-parametric method is that $g(\cdot)$ can be estimated at a parametric rate. Despite its attractive theoretical properties, the sieve estimator requires a different set of assumptions for identification and estimation such as conditional independence, which are stronger than those required in our paper. In our empirical application, we adopt this approach to estimate $g(\cdot)$, and the main finding continues to hold. Our Online Appendix provides further clarifications and discussions of the sieve MLE.

2.4.3 Multiple Measurements

Our identification, specifically Lemma 2 and Proposition 1, requires only two measurements of parental income and one measurement of children's income. This feature is particularly useful for estimation in practice since long-panel data with a reasonably large sample size are also often not available. However, more measurements can be accommodated in our estimation. First, we can incorporate the time averaging method in our current framework by treating the time average as one of the measurements. Consistent with the existing literature, we can use the time average as the normalization in practice.

Second, we can further extend our method to use model averaging.[5] Let L be the number of models and $\widehat{g}'_i(\cdot)$ the estimated IGE from the ith model, where $i = 1, \ldots, L$. A model averaging IGE estimator can be obtained using $\widehat{g}'(\cdot) = \sum w_i \widehat{g}'_i(\cdot)$, where w_i is a weight based on the variance of all the L models. We illustrate its performance using Monte Carlo simulations. In the experiments, we consider six periods of incomes for parents and children, i.e., $(X_t, Y_t), t = 1, \cdots, 6$, where the life-cycle bias coefficients of X_1 and Y_1 are normalized to be 1. We first estimate the IGE for each of the triple $(X_1, X_t, Y_t), t = 2, \cdots, 6$, and then average the five IGEs. Figure (B.4) in the Online Appendix suggests that this weighted averages indeed reduce standard errors, especially when permanent income is at the tails (either small and very large). Nevertheless, the improvement is marginal, and the main message from the estimation remains unchanged. In our empirical analysis, our main findings of a (statistically significant) nonlinear mobility function can only be strengthened, as a result.

3 Empirical Application

3.1 Data

Our data are drawn from the Panel Study of Income Dynamics (PSID), which has been widely used for empirical studies of IGE (see Mazumder (2018) for an excellent account of the features and advantages of the PSID data for this purpose.). As in Solon (1992), we use only the Survey Research

---

[5]Another possibility is to employ the sieve MLE, but again, it requires an additional, stronger conditional independence assumption for it to work. Further, due to more periods of data, the number of parameters to be estimated substantially increases, and so does the computational burden.

Center component of the PSID, but exclude the Survey of Economic Opportunity (SEO) component to prevent over-representing the poverty sample. Previous literature has also pointed out some serious irregularities in the sampling of SEO respondents, which can "preclude easy generalization to any well-defined population" (Bloome, 2015).

Following the standard practice in the literature, we focus on (the logarithm of) family income for both parents and children, as it is "the more inclusive measure of economic status and preferable for most applications" (Bowles and Gintis, 2002); and it "reflects the objective of measuring mobility after accounting for secular income growth in a society" (Durlauf and Shaorshadze, forthcoming); see, also Chadwick and Solon (2002) and Lee and Solon (2009) (p.767) for further discussions of this point. In our analysis, we exclude zero or negative income and adjust for top-coded earnings by a factor of 1.4, as in Lemieux (2006). Because of the use of family income for both parents and children, we cannot separate the income of children from that of their parents' if the children live with their parents. Following the literature (e.g., Lee and Solon (2009)), we measure children's income at different ages by the family earnings in the household in which they become the head or head's spouse. Since family income is used, we need only map the household income of children to that of their parents.

We consider information on income for both children and parents at ages from 25 to 50. Note that the number of observations available is important when choosing baseline and alternative measurements of incomes in our estimations. As shown in Table 1, the number of observations on income available for children decreases with age monotonically, while the number of observations for parents exhibits an inverted-U shape. The number of observations would decrease further in estimations because we require that information on income be available for both child and parent. Thus, the combination of a child's and his parent's ages that we choose will determine the sample size. We discuss our choice further below.

3.1.1 The Issue of Zero Income

With the recent use of administrative data to study intergenerational mobility, the literature has also paid increasing attention to the issue of zero income, which merits further discussions prior to continuing. There is, to the best of our knowledge, not a particularly satisfactory approach to address this issue yet. However, there are three practices/approaches to minimize the possible bias in the estimation, two of which are specific only to linear or rank-rank regressions that provide summary measures of IGE and do not address the life-cycle bias and nonliearity.

The first approach is to assign a value to zeros (Chetty et al., 2014) or assign an absolute mobility rate of 100% to individuals with zero incomes in estimation (Chetty et al., 2016). The second framework is proposed in the recent work by Mitnik et al. (2015) in the *linear* framework, but the approach changes the estimand of the IGE that may answer a different question than the traditional

method (Chetty et al. 2014; Mazumder, 2016).

The final, and probably the most common, practice is to exclude zero incomes (for recent examples, see Chetty and Hendren (2017), Mazumder (2016) and Durlauf et al. (2017)). For example, Durlauf et al. (2017) restrict the sample to pairs of sons and fathers for whom there are at least three nonzero income observations. This is the practice that we follow here. However, excluding zero income does not mean that we simply ignore the issue. Instead, this practice can alleviate the potential impact on our estimates, while maintaining the proposed approach, in this particular context, given the our focus on heterogenous effects (as opposed to average effects), as well as our careful choice of data and income measure for empirical analysis. In fact, for our baseline specification, we have only one child-parent pair with son's income being zero, and one pair with father's income being negative.

First, we focus on family total income in our analysis. As noted in Mazumder (2016), this measure is not only a *preferred* measure since it measures total family resources available, but also can "help alleviate the problem of observing zero earnings or zero income as is common in administrative data." Survey data such as PSID include sources of income obtained from transfers and from other family members that are not available in tax data and can be non-zero even when earnings are not due to non-employment. Moreover, family income is still reported in the survey data even when it may be too low to be filed for tax purposes. As a result, "when using family income, instances of reports of zero income are relatively rare so the results are virtually immune to the inclusion of zeroes. Therefore the concerns about the sensitivity of results around how to handle years of zero income is effectively a non-issue when using family income." (Mazumder 2016)

Second, as noted Mazumder (2016), the issue of zero income arising from the administrative data is also largely a consequence of the coverage of their sample. For example, in Chetty et al. (2014), the information for child income is only available for children aged 29-32 during the years 2011 and 2012 when "unemployment was relatively high", and these children may be more likely to have zero incomes. On the other hand, PSID covers a wide range of stages over the life-cycle, and it is even rarer to have zeros in multiple years. Therefore, we can choose annual income during a child's prime earning years (mid-thirties) as our baseline measure, and also experiment with many different combinations of years (including time averages) to maximize the sample size and minimize the issue of zeroes. Our results are indeed rather robust to different combinations of years.

Finally, while there are not many instances of zero income in the PSID sample, it is also worth emphasizing that our focus here is on heterogenous effects with respect to permanent incomes, instead of summary measures that the conventional linear or rank-rank regressions do. We therefore avoid the need to assign an arbitrary mobility rate to summarize the heterogenous effects as in Chetty et al. (2016). We also assess the robustness of our approach to exclusion of the zeros in more general settings. While we cannot directly simulate log zeros in our model set-up, we note that the log of zero is negative infinity and therefore consider dropping the extreme values defined as bottom 5 and 10 percent of the distribution of parental observed incomes in our simulation experiments. Not

surprisingly, we find that exclusion of extreme values disproportionately impacts the estimates in the lower tail, which slightly biased downward. However, the estimates in the middle and the patterns of nonlinearity are rather robust. This may highlight the usefulness of our nonparametric approach even when using the administrative data with a large number of zeros. See Section C. in Online Appendix for details.

3.2 Results

Point identification and estimation requires the existence of one year when one's income reaches her permanent income (with some random transitory shocks). As emphasized above, while such normalization can impact the estimates, it has no effect on the pattern of the IGEs. **In what follows, we focus more on the pattern, instead of the actual estimates**.

3.2.1 Baseline Results

We follow the tradition in the literature to choose the year for normalization, which is plausible economically. The previous literature has typically found that the life-cycle bias coefficients become close to 1 in individuals' between their mid-thirties and forties (e.g., Haider and Solon, 2006; Böhlmark and Lindquist, 2006). As such, we choose $t_0 = t_1 = 34$ in our baseline estimation, roughly the first year where the empirical evidence from Haider and Solon (2006) and Böhlmark and Lindquist (2006) overlap, and show later that our results are robust to the choice of t_0 and t_1. For this choice of normalization, we estimate the density of parental and children's log income as an intermediate step, and present them in Figure 1. The log income of most of the observations ranges from 9 to 13, and the distribution is similar for the two generations.

We now turn to our baseline results of IGEs with the normalization set to be age 34 for both children and parents from the pooled sample. The second period is set to be age 28 for children and 38 for parents. This combination appears to be the best choice because it gives us the second largest sample for our empirical analysis, and in the meantime, ensures that the baseline and second years are not too close.[6] We uncover substantial variation in the IGEs across the population, a U-shape in particular. To visualize this finding, we plot the histogram of the derivatives of the estimated mobility function in the left panel of Figure 2, and report 10%, 25%, 50%, 75% and 90% quantiles (along with their corresponding bootstrapped 90% confidence intervals) in Table 2. IGEs differ across

---

[6]We want to maximize the number of child-parent pairs for the cohort of children born between 1952 and 1974, a cohort typically used in the previous literature (to facilitate comparison). Recall that we need information on incomes for *both* parent *and* child available not only for the baseline year, but for the second year as well. This requirement reduces the sample size. To visualize the patterns, we present the number of observations associated with each combination of the second years for children and parents in Table 1, and display these results in Figure **??** . There are two distinct patterns of how the sample sizes vary with age for both parents and children when the baseline year for parents is 34. Given the choice of age for parents, the sample size decreases with the child's age nearly monotonically. Given the choice for children, the sample size exhibits an inverted-U shape peaking at around 38.

all five quantiles. We in turn fit the IGE estimates to a quadratic function and presents the fitted coefficients in the first column of Table 3. We then test whether the estimated IGEs are constant (the mobility function is linear), and indeed reject such a null at the 1% significance level.

To further examine how the IGEs differ by parental income, we display the IGEs evaluated at each possible log income level of parents in the right panel of Figure 2. The black line represents our non-parametric estimates of IGEs, and the blue ones represent 90% bootstrapped confidence intervals.[7] For comparison purposes, we also estimate the IGE using the traditional linear approach with time averages of age 33 to 37, which yields an estimate of 0.48, the horizontal red line in Figure 2. The linear estimates from averaging different years range from 0.41 to 0.52, consistent with the previous literature using PSID. The summary measures, however, drastically conceals important features of the nonlinear relationship in incomes across the generations. Our non-parametric results in Figure 2 reveal a U-shape pattern of the IGEs. Specifically, we find that the income associations are high for the very poor and the very rich, but the degree of associations becomes relatively weak for the broadly defined middle-income class (whose log income is between 9.5 and 12.1). The finding of U-shape is also prevalent in the studies using the transition matrix approach (e.g., Hertz 2005), but masked in the studies using the linear regression approach focusing on summary measures.

3.2.2 Robustness Analysis

We re-estimate the IGEs using alternative second measurements and different sub-samples to assess the robustness of our results.[8] We find that the basic pattern is remarkably robust.

Alternative Second Measurements of Permanent Income. Given the number of annual incomes observed across ages available for both child and parent, there are many possible combinations of the second measurements of permanent incomes for children and parents. Here we experiment with two types of alternative combinations, while maintaining the base year being 34. The idea behind this experiment is to see whether or not our results are robust to departures from the original setting. We conduct four experiments to illustrate the robustness. First, we set the second period to be age 26 for children and 35 for parents; second, we set the second period to be age 32 for children and 40 for parents; third, we set the second period to be age 30 for children and 43 for parents; forth, we set the second period to be age 34 for children and 45 for parents. The estimates of IGEs are displayed in Figure 3, and the testing results are presented in the four columns under "Robustness check #1"

---

[7]The estimates for extremely low (log income less than 9.5) and high (log income greater than 12.5) parental income are excluded to facilitate presentation because they have been shown to be noisy and less reliable due to the small sample size as illustrated in the Monte Carlo simulations.

[8]It is worth emphasizing that although the simulation exercises have indicated the superior finite-sample performances in our approach, it is still a non-parametric approach, which is data demanding. Many of the decisions, when illustrating our approach, are made so as to obtain a reasonably large sample of the particular cohort of interest and hence more reliable estimates.

of Table 3. We find a U-shape IGEs with respect to parental income, and again decisively reject the linearity of the mobility function in all of the cases at the 1% significance level.

Alternative Normalization   Earlier we borrowed the insights from the literature on the life-cycle bias and consider annual income at age 34, the first year of the period that the literature has generally found to be good a representation of permanent income. Here we instead consider using different years as the base line year (33, 35, 36). The estimated IGEs are shown in Figure 4, and the testing results are presented in the three columns under "Robustness check #2" of Table 3. Again, we show that the U-shape pattern in our baseline results are robust to the alternative normalization.

Sub-sample Analysis I: Excluding Females.   One may be concerned that income is not necessarily a good indicator of daughters' economic success because of the relatively limited labor force participation of women (Mazumder, 2005). When lower-income women, who are more likely to come from lower-income families, need to work, we are more likely to observe higher persistence in the lower tail of the income distribution, but not in the upper tail. There are two implications of such negative selection for estimation. First, we are, on the average, more likely to observe an IGE of smaller magnitudes for women than for men (since the part of the population with large IGEs are not observed in the data)(Mazumder, 2005).[9] Second, inclusion of daughters in our non-parametric analysis will more likely bias the estimates of IGE in the lower tail of the income distribution. The results excluding daughters are displayed in the left panel of Figure 5. While finding much smaller persistence in the lower tail for sons, indicating more mobility once these low-income women who work are excluded, we continue to find the same nonlinear pattern in the IGE as above.

Sub-sample Analysis II: Excluding Non-Whites.   The transmission of income over generations may differ by races, and that there may be permanent differences in opportunities between families of different ethnic groups (Durlauf and Shaorshadze, forthcoming, Chetty et al. (2014), Bhattacharya and Mazumder (2011)). The results excluding non-whites are displayed in the right panel of Figure 5. The baseline result continues to hold when we focus only on whites. This should not be surprising, as PSID has a relatively small sample of non-whites, and thus exclusion of these observations should not drastically impact our estimates.[10] Furthermore, we also find that there is substantial heterogeneity even among whites. For both subsamples, our testing results are shown in the two columns under "Sub-sample analysis" of Table 3 convey the same message as the baseline result: the mobility function is highly nonlinear.

---

[9]Chadwick and Solon (2002) find that the differences between daughters and sons are not always statistically significant. This may be due to a much smaller estimation sample of daughters, and the inference is not as reliable. This is indeed what we find as well: the estimates are much smaller for daughters than for sons, but the estimates for daughters are very imprecise. The results are available from the authors upon request.

[10]Bhattacharya and Mazumder (2011) note that the intergenerational samples of blacks in PSID are indeed so small that it will likely inhibit research on them.

3.2.3 Multiple Measurements

Incorporating the time averaging method   We incorporate the time averaging method in our approach. Specifically, we use the five-year average of annual incomes with ages 32 to 36 as the normalization year, together with the second measurements in the baseline results. This exercise can thus be seen as an improvement of the averaging method for taking into account the life-cycle bias. The estimated derivative functions and the results of testing the linearity are shown in the left plot in Figure 6 and the last column in Table 3, respectively. Again, we show that the U-shape pattern in our baseline results are robust to the alternative normalization.

Model Averaging   To further take into multiple years of data, we perform model averaging of the baseline model and the four models in our robustness analysis using alternative second measurements of permanent income. Our main findings of a (statistically significant) nonlinear mobility function remains unchanged and even strengthened, as a result (see Figure 6).

3.3   Implications for Great Gatsby Curve

Our estimated mobility function and IGEs can also have implications for the so-called Great Gatsby Curve, observed by Corak (2013) and termed by Krueger (2012): a positive correlation between inequality and social immobility (captured by IGEs). The Great Gatsby Curve has received much attention since it suggests that the widely believed trade-off between inequality of outcomes and equality of opportunity is "illusory" (Durlauf and Seshadri, 2017).

The underlying reasons for the Great Gatsby Curve are unclear. The implied causality in the observed inequality/mobility relationship can run either way. However, as noted in Durlauf and Seshadri (2017), a linear mobility function of the form $Y^* = \alpha + \beta \cdot X^* + \epsilon$ may produce a Gatsby Curve, but "as a statistical subject", precludes the possibility that the causality can run from inequality to mobility. By contrast, such possibility can be accommodated by a nonlinear model. For example, Durlauf and Seshadri (2017) argue that "social influences on children create a nonlinear relationship between parental income and offspring income', and that "increases in inequality, by altering the ways in which family income determines and interacts with social influences, reduce mobility."

Regardless of the underlying mechanisms behind the nonlinear mobility function, it would be interesting to see if our econometric model can indeed produce the intertemporal Great Gatsby Curve. To this end, we first simulate data on parental income distributed from log normal with the mean value fixed (estimated using the PSID data) while varying values of variances (capturing inequality). Note that our simulation exercises do not depend on the value of the mean. Based on our model, each observation in a sample is associated with an estimate of IGE. We can then calculate the average IGE based on these estimates for each sample. Finally, the simulated pairs of average IGEs and inequality are fitted to both linear and quadratic curves. The results are plotted in Figure 7. Our

nonparametric model indeed generate a strong, positive relationship between inequality and IGEs, confirming the important finding in Durlauf and Seshadri (2017) that "an intertemporal Gatsby Curve is a salient feature of inequality in the United States."

4 Conclusions

Economists have long been interested in the estimation of IGE of children's income with respect to parental permanent income, a good indication of equality of opportunity in society. This paper provides a new framework for estimating the IGE. Our approach confronts two important issues surrounding the estimations that have yet to be addressed together in a systematic fashion, namely the heterogeneity/nonlinearity in IGE and the latent nature of the permanent income. Using the proposed method, we find that there exists substantial heterogeneity in IGEs. More importantly, we observe a U-shaped pattern for the IGE. This finding indicates substantial persistence in income across generations for both rich and poor families, but a lack of persistence for the majority of the population. It is our hope that our approach can be applied to explore within-group heterogeneity in IGEs in the case of datasets with a larger number of individuals for each type of characteristics, to ascertain the existence of the same nonlinearity in various other countries, which might be usefully compared in this respect, and other related topics.

References

Bhattacharya, D. and B. Mazumder, "A Nonparametric Analysis of Black-White Differences in Intergerational Income Mobility in the United States," *Quantitative Economics*, 2011, *2*, 335–379.

Björklund, A., J. Roine, and D. Waldenstrom, "Intergenerational Top Income Mobility in Sweden: Capitalist Dynasties in the Land of Equal Opportunity?," *Journal of Public Economics*, 2012, *96*, 474–484.

Black, Dan A and Jeffrey A Smith, "Estimating the returns to college quality with multiple proxies for quality," *Journal of Labor Economics*, 2006, *24* (3), 701–728.

Black, S.E. and P.J. Devereux, *Handbook of labor economics*, Vol. 4, Elsevier,

Bloome, Deidre, "Income Inequality and Intergenerational Income Mobility in the United States," *Social Forces*, 2015, *93* (3), 1047–1080.

Böhlmark, Anders and Matthew J Lindquist, "Life-Cycle Variations in the Association between Current and Lifetime Income: Replication and Extension for Sweden," *Journal of Labor Economics*, 2006, *24* (4), 879–896.

Bowles, Samuel and Herbert Gintis, "The Inheritance of Inequality," *Journal of Economic Perspectives*, 2002, *16* (3), 3–30.

Chadwick, Laura and Gary Solon, "Intergenerational Income Mobility Among Daughters," *American Economic Review*, 2002, *92* (1), 335–344.

Chao, Hsiang-Ke, "Milton Friedman and the Emergence of the Permanent Income Hypothesis," *History of Political Economy*, 2003, *1* (35), 77–104.

Chen, Xiangjin B, Jiti Gao, Degui Li, and Param Silvapulle, "Nonparametric Estimation and Forecasting for Time-Varying Coefficient Realized Volatility Models," *Journal of Business & Economic Statistics*, 2016, (just-accepted), 1–39.

Chetty, R. and N. Hendren, "The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects," *NBER Working Paper*, 2017.

_ , **D. Grusky, M. Hell, N. Hendren, R. Manduca, and J. Narang**, "The Fading American Dream: Trends in Absolute Income Mobility Since 1940," *NBER Working Paper*, 2016.

_ , **N. Hendren, P. Kline, and E. Saez**, "Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States," *Quarterly Journal of Economics*, 2014, *129* (4), 1553–1623.

Cooper, S., S. Durlauf, and P. Johnson, "On the Evolution of Economic Status Across Generations," in "American Statistical Association Papers and Proceedings" 1994, pp. 50–58.

Corak, M., "Income Inequality, Equality of Opportunity, and Intergenerational Mobility," *Journal of Economic Perspectives*, 2013, *27*, 79–102.

_ **and A. Heisz**, "The Intergenerational EArnings and Income Mobility of Canadian Men: Evidence from Longitudinal Income Tax Data," *Journal of Human Resources*, 1999, *34* (3), 504–533.

Corak, Miles, Lori Curtis, and Shelley Phipps, "Economic mobility, family background, and the well-being of children in the United States and Canada," *Persistence, privilege, and parenting: The comparative study of intergenerational mobility*, 2011, pp. 73–108.

Cunha, F., J.J. Heckman, and S. Schennach, "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 2010, *78* (3), 883–931.

Cunha, Flavio and James J. Heckman, "The Technology of Skill Formation," *American Economic Review*, 2007, *97* (2), 31–47.

Durlauf, S., *Dynamic Disequilibrium Modelling: Proceedings of the Ninth International Symposium on Economic Theory and Econometrics*, Cambridge: Cambridge University Press., 1996.

_ , "A Theory of Persistent Income Inequality," *Journal of Economic Growth*, 1996, *1*, 75–93.

_ **and A. Seshadri**, "Understanding the Great Gatsby Curve," *NBER Working Paper*, 2017.

_ **and I. Shaorshadze**, *Emerging Trends in the Social and Behavioral Sciences*, Sherman Oaks: Sage Publishing, forthcoming.

_ , **Andros Kourtellos, and Chih-Ming Tan**, "Status Trap," *Journal of Business and Economic Statistics*, 2017, *35* (2).

Friedman, M., *A Theory of Consumption Function*, Princeton University Press, 1957.

Grawe, Nathan D., "Lifecycle Bias in Estimates of Intergenerational Earnings Persistence," *Labour Economics*, 2006, *13*, 551–570.

Haider, Steven J and Gary Solon, "Life-Cycle Variation in the Association between Current and Lifetime Earnings," *American Economic Review*, 2006, *96* (4), 1308–1320.

Härdle, Wolfgang, Hua Liang, and Jiti Gao, *Partially linear models*, Springer Science & Business Media, 2012.

Hertz, T., *Unequal Chances: Family Background and Economic Success*, Princeton, NJ.: Princeton University Press,

Hu, Y. and G. Ridder, "On Deconvolution as a First Stage Nonparametric Estimator," *Econometric Reviews*, 2008, *29* (4), 1–32.

Hu, Yingyao, *The Econometrics of Unobservables: Latent Variable and Measurement Error Models and Their Applications in Empirical Industrial Organization and Labor Economics*, Unpublished Manuscript, 2020.

_ **and Yuya Sasaki**, "Closed-Form Estimation of Nonparametric Models with Non-Classical Measurement Errors," *Journal of Econometrics*, 2014, *185* (2), 392–408.

Isaacs, Julia B, Isabel V Sawhill, and Ron Haskins, "Getting ahead or losing ground: Economic mobility in America.," *Brookings Institution*, 2008.

Jenkins, Stephen, "Snapshots versus movies: "Lifecycle biases" and the estimation of intergenerational earnings inheritance," *European Economic Review*, 1987, *31* (5), 1149 – 1158.

Kotlarski, Ignacy, "On some characterization of probability distributions in Hilbert spaces," *Annali di Matematica Pura ed Applicata*, 1966, *74* (1), 129–134.

Kourtellos, Andros, C. Marr, and C. Tan, "Local Intergenerational Mobility," 2019. Unpublished Manuscript.

Krueger, A., "The Rise and Consequences of Inequality in the United States," *Unpublished Speech*, 2012.

Lee, C. and G. Solon, "Trends in Intergenerational Income Mobility," *Review of Economics and Statistics*, 2009, *91* (4), 766–772.

Lemieux, Thomas, "Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?," *American Economic Review*, 2006, *96* (3), 461–98.

Li, Qi and Jeffrey Scott Racine, *Nonparametric econometrics: theory and practice*, Princeton University Press, 2007.

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

_ , "Estimating the intergenerational elasticity and rank association in the United States: Overcoming the current limitations of tax data," *Research in Labor Economics*, 2016, *43*, 83–129.

_ , "Intergenerational mobility in the United States: What we have learned from the PSID," *The Annals of the American Academy of Political and Social Science*, 2018, *680* (1), 213–234.

Mitnik, Pablo, Victoria Bryant, Michael Weber, and David B Grusky, "New estimates of intergenerational mobility using administrative data," *Statistics of Income Division working paper, Internal Revenue Service*, 2015.

Prokhorov, Artem, I. Murtazashvili, and Di L., "Two-Sample Nonparametric Estimation of Intergenerational Income Mobility," *Canadian Journal of Economics*, 2015, *48* (5), 1733–1761.

Rao, B.L.S. Prakasa, *Identifiability in Stochastic Models: Characterization of Probability Distributions*, Cambridge, MA: Academic Press, 1992.

Rilstone, P and A Ullah, "Nonparametric estimation of response coefficients," *Communications in Statistics-Theory and Methods*, 1989, *18* (7), 2615–2627.

Schennach, Susanne M, "Nonparametric regression in the presence of measurement error," *Econometric Theory*, 2004, *20* (6), 1046–1093.

Solon, G., "Intergenerational Income Mobility in the United States," *American Economic Review,* 1992, *82* (3), 393–408.

Proofs

Proof of Lemma 2.   We only prove the identification of f_{X^*}. The proof for the density f_{Y^*} follows a symmetric argument. First consider the two measures of X^*

$$X_1 = X^* + V_1, \quad X_2 = \alpha_2 X^* + V_2,$$

with α_2 being known. First define the joint ch.f. of X_1 and X_2 as:

$$\phi_{X_1,X_2}(s_1, s_2) \equiv E[e^{is_1 X_1 + is_2 X_2}].$$

Take the derivate of the ch.f. above with respect to s_1 and evaluate it at $s_1 = 0$:

$$\frac{\partial \phi_{X_1,X_2}(s_1,s_2)}{\partial s_1}\bigg|_{s_1=0} = E[i(X^* + V_1)e^{is_2 X_2}] = E[iX^* e^{is_2 X_2}] + E[iV_1 e^{is_2 X_2}] = E[iX^* e^{is_2 \alpha_2 X^*}]E[e^{is_2 V_2}],$$

where $E[iV_1 e^{is_2 X_2}] = 0$ is due to the assumption $E[V_1|X_2] = 0$, and the third equality holds because of the independence of V_2 from X^*. Similarly, we have $\phi_{X_2}(s_2) = E[e^{is_2 \alpha_2 X^*}]E[e^{is_2 V_2}]$. Consider that

$$\frac{\frac{\partial \phi_{X_1,X_2}(s_1,s_2)}{\partial s_1}\big|_{s_1=0}}{\phi_{X_2}(s_2)} = \frac{E[iX^* e^{is_2 \alpha_2 X^*}]}{E[e^{is_2 \alpha_2 X^*}]} = \frac{\partial \ln E[e^{irX^*}]}{\partial r}\bigg|_{r=\alpha_2 s_2}.$$

Thus, the ch.f. of f_{X^*}, ϕ_{X^*}, can be recovered as

$$\phi_{X^*}(s) = \exp\left(\int_0^s \frac{\frac{\partial}{\partial s_1}\phi_{X_1,X_2}(s_1, s_2/\alpha_2)\big|_{s_1=0}}{\phi_{X_2}(s_2/\alpha_2)\big|_{s_1=0}} ds_2\right).$$

The density f_{X^*} can be recovered using inverse Fourier transform

$$f_{X^*}(x^*) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-isx^*}\phi_{X^*}(s)ds = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-isx^*}\exp\left(\int_0^s \frac{\frac{\partial}{\partial s_1}\phi_{X_1,X_2}(s_1, s_2/\alpha_2)\big|_{s_1=0}}{\phi_{X_1,X_2}(s_2/\alpha_2)\big|_{s_1=0}} ds_2\right)ds.$$

Proof of Proposition 1.   The proof follows Hu and Sasaki (2014) and we present here for completeness. Once f_{X^*} is identified, the identification of the mobility function $g(\cdot)$ can be obtained by using the joint distribution of only one measurement for both generations. Thus, one can identify the

mobility function using any combination of children and parental observed incomes, i.e., $\{X_1, Y_1\}$, $\{X_1, Y_2\}$, $\{X_2, Y_1\}$, or $\{X_2, Y_2\}$. We illustrate the identification using $\{X_2, Y_2\}$ where the life-cycle coefficients are not equal to 1.

We first define the joint ch.f. of X_2 and Y_2: $\phi_{X_2, Y_2}(s, v) \equiv E[e^{isX_2 + ivY_2}]$. Note that $Y_2 = \delta_2 Y^* + U_2 = \delta_2(g(X^*) + \epsilon) + U_2 = \delta_2 g(X^*) + \delta_2 \epsilon + U_2$. Taking derivative of the ch.f. $\phi_{X_2, Y_2}(s, v)$ with respect to v and evaluate the objective at $v = 0$.

$$
\begin{aligned}
\frac{\partial}{\partial v}\phi_{X_2, Y_2}(s, v)|_{v=0} &= iE\left[(\delta_2 g(X^*) + \delta_2 \varepsilon + U_2)e^{is(\alpha_2 X^* + V_2)}\right] \\
&= iE\left[\delta_2 g(X^*)e^{is(\alpha_2 X^* + V_2)}\right] + iE[(\delta_2 \varepsilon + U_2)e^{isX_2}] \\
&= iE\left[g(X^*)e^{i\alpha_2 sX^*}\right]\delta_2 E[e^{isV_2}] \\
&= iE\left[g(X^*)e^{i\alpha_2 sX^*}\right]\delta_2 \frac{\phi_{X_2}(s)}{\phi_{X^*}(\alpha_2 s)},
\end{aligned}
$$

where $\phi_{X_2}(s) \equiv E[e^{isX_2}]$ and $\phi_{X^*}(\alpha_2 s) \equiv E[e^{i\alpha_2 sX^*}]$ are ch.fs of X_2 and X^* evaluated at s and $\alpha_2 s$, respectively. The third equation holds because ϵ is assumed to be mean independent of one of the observed measurement of parental income. Without loss of generality, $E(\epsilon|X_2) = 0$, as well as Assumption 3 (iii) $(E(U_2|X_2) = 0)$. As a result, the equation above enables us to obtain:

$$
E\left[g(X^*)e^{i\alpha_2 sX^*}\right] = \phi_{X^*}(\alpha_2 s)\frac{\frac{\partial}{\partial v}\phi_{X_2, Y_2}(s, v)|_{v=0}}{i\delta_2 \phi_{X_2}(s)}.
$$

It can be equivalently expressed as,

$$
E\left[g(X^*)e^{isX^*}\right] = \int e^{isx^*}g(x^*)f_{X^*}(x^*)dx^* = \phi_{X^*}(s)\frac{\frac{\partial}{\partial v}\phi_{X_2, Y_2}(s/\alpha_2, v)|_{v=0}}{i\delta_2 \phi_{X_2}(s/\alpha_2)}.
$$

Using the inverse Fourier transform,

$$
g(x^*)f_{X^*}(x^*) = \frac{1}{2\pi\delta_2}\int_{-\infty}^{+\infty} e^{-isx^*}\phi_{X^*}(s)\frac{\frac{\partial}{\partial v}\phi_{X_2, Y_2}(s/\alpha_2, v)|_{v=0}}{i\phi_{X_2}(s/\alpha_2)}ds.
$$

Consequently, the mobility function $g(X^*)$ can be identified as

$$
g(x^*) = \frac{1}{2\pi\delta_2 f_{X^*}(x^*)}\int_{-\infty}^{+\infty} e^{-isx^*}\phi_{X^*}(s)\frac{\frac{\partial}{\partial v}\phi_{X_2, Y_2}(s/\alpha_2, v)|_{v=0}}{i\phi_{X_2}(s/\alpha_2)}ds,
$$

where $\phi_{X^*}(s)$ and $f_{X^*}(x^*)$ are identified in Lemma 2.

Proof of Theorem 1. Suppose the parental annual incomes are associated with the permanent

income as follows.

$$X_1 = \alpha_1 X^* + V_1, \ \ X_2 = \alpha_2 X^* + V_2, \ \ \alpha_1 \neq 1, \alpha_2 \neq 1.$$

Suppose that we falsely assume $\alpha_1 = 1$. Such a normalization is equivalent to treat $\widetilde{X}^* \equiv \alpha_1 X^*$ as the permanent income X^*. Under such a normalization, the association between the permanent and annual incomes can be reformulated as:

$$X_1 = \widetilde{X}^* + V_1, \ \ X_2 = \frac{\alpha_2}{\alpha_1}\widetilde{X}^* + V_2 \equiv \widetilde{\alpha}_2\widetilde{X}^* + V_2, \ \widetilde{\alpha}_2 \equiv \alpha_2/\alpha_1 \tag{11}$$

Following Lemma 2, we can identify the density of $\widetilde{X}^*$ from the joint distribution of X_1 and X_2. That is,

$$f_{\widetilde{X}^*}(u) = f_{X^*}(u/\alpha_1)/\alpha_1.$$

Similarly, we can reformulate children's incomes.

$$Y_1 = \delta_1 Y^* + U_1 \equiv \widetilde{Y}^* + U_1, \ \ Y_2 = \delta_2 Y^* + U_2 = \frac{\delta_2}{\delta_1}\widetilde{Y}^* + U_2 \equiv \widetilde{\delta}_2\widetilde{Y}^* + V_2. \tag{12}$$

We further consider that under the false normalization, the relationship between the observed child income and newly constructed parental permanent income can be reformulated as

$$\begin{aligned}
Y_2 &= \widetilde{\delta}_2\widetilde{Y}^* + U_2 = \widetilde{\delta}_2\delta_1 Y^* + U_2 = \widetilde{\delta}_2\delta_1 g(X^*) + \delta_1\varepsilon + U_2 \\
&= \widetilde{\delta}_2\left(\delta_1 g(\widetilde{X}^*/\alpha_1)\right) + \widetilde{\delta}_2\epsilon + U_2 \equiv \widetilde{\delta}_2\widetilde{g}(\widetilde{X}^*) + \widetilde{\delta}_2\varepsilon + U_2.
\end{aligned} \tag{13}$$

Equation (11) - (13) consist of a system with the mobility function $\widetilde{g}(\cdot)$, child permanent incomes $\widetilde{X}^*$, and parental permanent incomes $\widetilde{Y}^*$ under corrected normalization. Using Proposition 1, we can correctly identify the corresponding mobility function $\widetilde{g}(u)$, which has a relationship with the original mobility function $g(\cdot)$ as

$$\widetilde{g}(u) = \delta_1 g\left(u/\alpha_1\right). \tag{14}$$

Equivalently,

$$g(x^*) = \widetilde{g}(\alpha_1 x^*)/\delta_1. \tag{15}$$

Proof of Corollary 1. Theorem 1 states $\widetilde{g}(x^*) = \delta_1 g(x^*/\alpha_1)$, so the IGE satisfies $\widetilde{g}'(x^*) = \delta_1/\alpha_1 g'(x^*/\alpha_1)$. If the mobility function is linear, i.e., $g(x^*) = \beta_0 + \beta_1 x^*$ as widely used in the literature, then $\widetilde{g}'(x^*) = \delta_1\beta_1/\alpha_1$, which is also a constant for any x^*. If $\delta_1 = \alpha_1$, then the IGE

29

becomes $\widetilde{g}'(x^*) = \beta_1$. As a result, if we have a linear mobility function, mis-specifying the normalization does not affect the estimation of the IGE given the life-cycle coefficient is the same across generations.

Table 1: Summary statistics

Var.		children					parent				
		Obs.	Mean	Std.	Min	Max	Obs.	Mean	Std.	Min	Max
gender		1839	1.51	0.50	1	2	1839	1	0	1	1
race		1839	1.15	0.49	1	3	1839	1.15	0.49	1	3
birthyear		1839	1973.02	9.52	1945	1985	1839	1945.31	10.16	1917	1966
	25	1077	10.52	0.84	0.30	12.57	955	10.62	0.78	0.61	11.71
	26	1053	10.58	0.86	0.24	12.59	989	10.66	0.66	0.30	11.80
	27	997	10.64	0.88	0.20	12.67	1015	10.72	0.51	6.70	11.94
	28	953	10.72	0.76	6.19	13.58	1040	10.73	0.72	1.13	12.23
	29	878	10.77	0.74	6.93	13.95	1087	10.75	0.78	0.83	12.30
	30	839	10.81	0.78	4.83	13.99	1123	10.78	0.75	0.70	12.56
	31	775	10.91	0.69	6.97	13.38	1171	10.83	0.62	0.61	12.45
	32	732	10.94	0.67	7.87	13.99	1232	10.83	0.84	0.38	12.64
	33	683	10.99	0.70	7.90	13.61	1219	10.89	0.55	8.07	13.01
	34	632	10.98	0.72	7.51	13.56	1216	10.90	0.63	0.47	12.63
	35	571	11.02	0.75	5.43	13.92	1250	10.95	0.65	0.44	12.68
	36	542	11.05	0.76	5.03	13.83	1306	10.94	0.83	0.42	13.07
Age	37	486	11.07	0.89	0.20	13.87	1368	10.98	0.92	0.17	13.88
	38	455	11.06	0.87	1.35	13.52	1399	10.98	0.86	0.42	13.53
	39	401	11.11	0.75	6.89	13.79	1408	11.03	0.83	0.34	13.99
	40	385	11.11	0.76	6.78	13.69	1440	11.07	0.69	0.44	13.37
	41	344	11.11	0.79	7.24	13.50	1429	11.07	0.93	0.24	13.42
	42	317	11.21	0.79	5.76	13.75	1440	11.10	0.98	0.06	13.63
	43	283	11.15	0.83	6.61	15.39	1411	11.13	0.92	0.30	13.69
	44	266	11.26	0.68	8.54	13.58	1395	11.19	0.64	7.63	13.73
	45	226	11.21	0.83	8.58	13.91	1350	11.13	0.97	0.38	13.94
	46	221	11.24	0.72	8.43	13.42	1376	11.23	0.66	6.78	13.53
	47	174	11.17	0.88	7.90	14.38	1317	11.22	0.70	6.69	13.71
	48	174	11.21	0.93	4.96	13.91	1257	11.20	1.00	0.06	14.09
	49	141	11.24	0.86	7.62	14.29	1247	11.24	0.80	0.38	13.43
	50	139	11.29	0.79	8.80	13.67	1207	11.24	0.70	7.48	13.95

Gender: male (1) and female (2); region: northeast+south(1), north central (2) and west (3); race: white (2) and nonwhite (2).
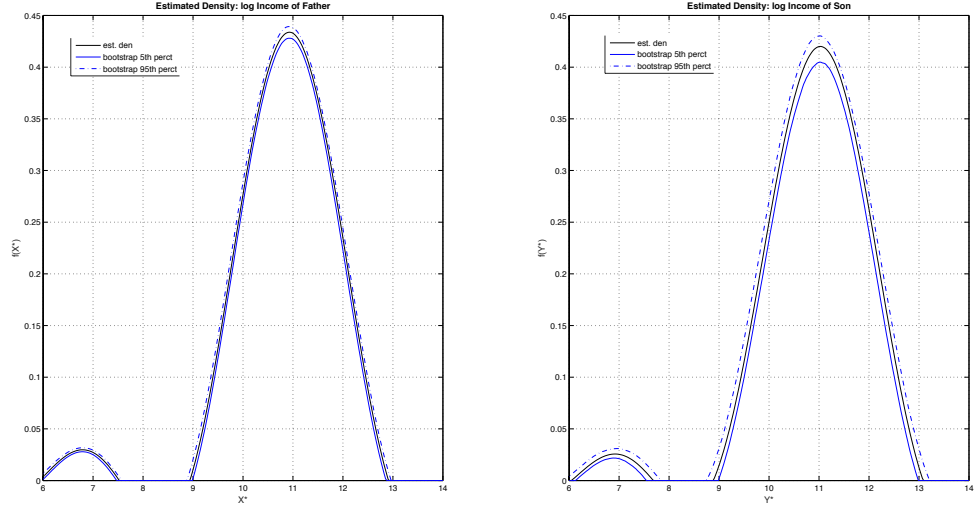
Figure 1: Estimated densities of permanent income

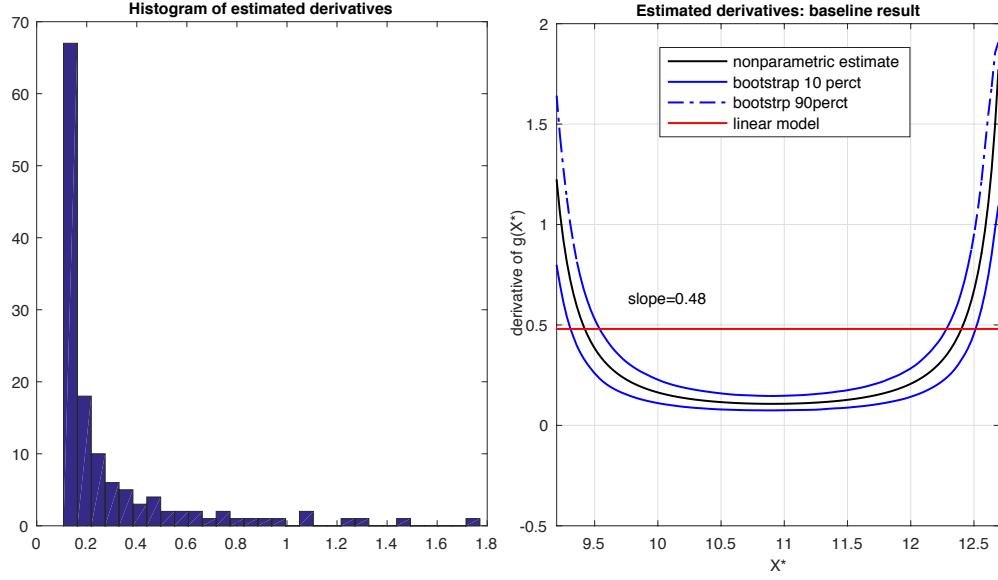Figure 2: Estimated derivatives of mobility function

Table 2: Quantiles of estimated derivatives

	10th		25th		50th		75th		90th	
	Est.	90% C.I.	Est.	90% C.I.	Est.	90% C.I.	Est.	90% C.I.	Est.	90% C.I.
Overall	0.108	[0.075, 0.147]	0.112	[0.077, 0.152]	0.129	[0.089, 0.173]	0.180	[0.127, 0.243]	0.318	[0.220,0.424]
Male	0.090	[0.054, 0.135]	0.093	[0.056, 0.140]	0.106	[0.064, 0.159]	0.147	[0.090, 0.217]	0.243	[0.148,0.354]
White	0.100	[0.065, 0.144]	0.103	[0.067, 0.149]	0.119	[0.078, 0.170]	0.169	[0.111, 0.238]	0.301	[0.191,0.412]

Figure 3: Robustness check #1
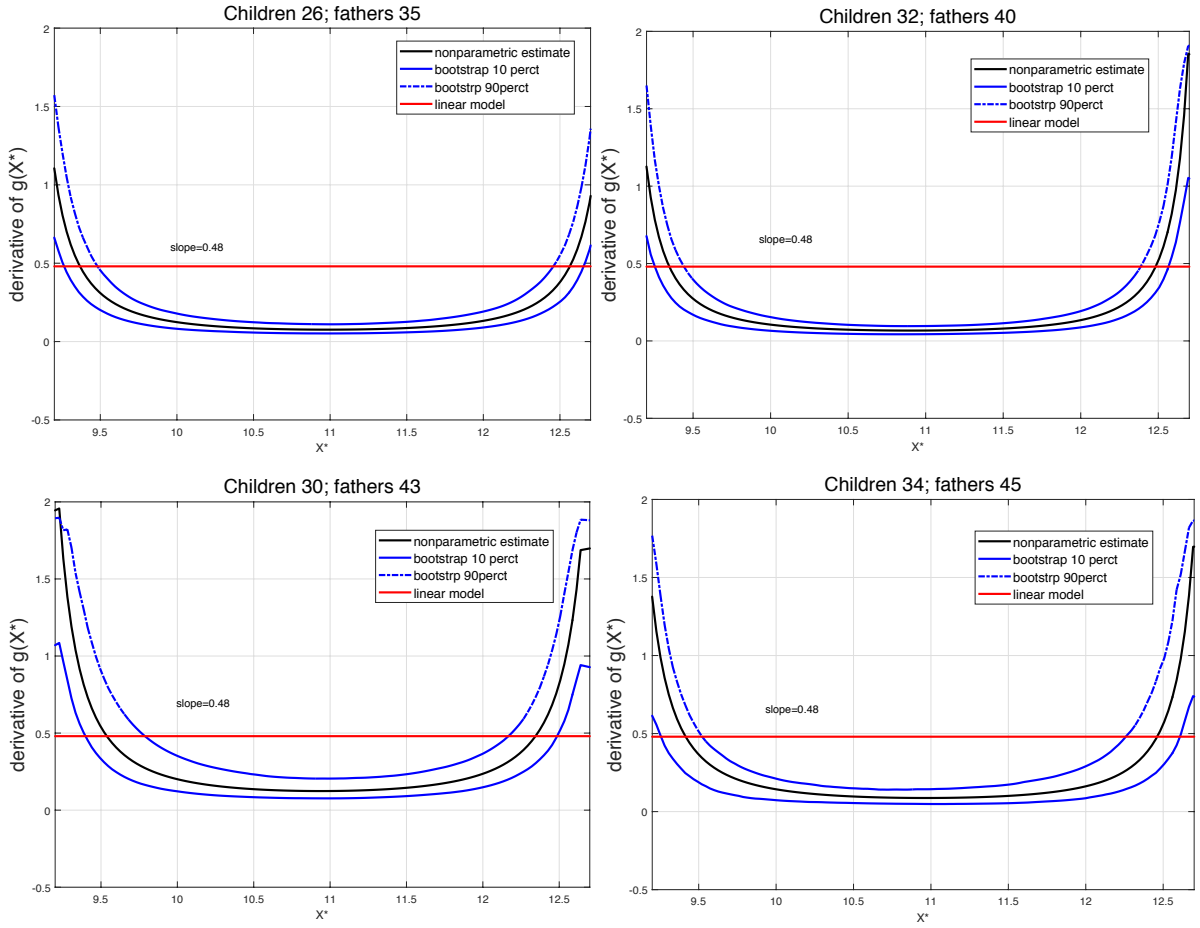
Table 3: Testing linearity of $g(\cdot)$

	Baseline	Robustness check #1				Robustness check #2			Sub-sample analysis		Averaging
		(1)	(2)	(3)	(4)	(1)	(2)	(3)	male	white	
constant	32.12***	22.08***	28.50***	49.03***	31.62***	18.15***	19.05***	23.32***	16.96***	29.12***	57.99***
	(6.29)	(5.24)	(6.08)	(12.31)	(7.11)	(5.16)	(4.04)	(4.67)	(4.28)	(6.45)	(8.64)
x^*	-5.91***	-4.01***	-5.24***	-8.95***	-5.78***	-3.32****	-3.44***	-4.19***	-3.12***	-5.33***	-10.63***
	(1.16)	(0.95)	(1.11)	(2.23)	(1.32)	(0.94)	(0.72)	(0.83)	(0.79)	(1.18)	(1.59)
x^{*2}	0.27***	0.18***	0.24***	0.41***	0.26***	0.15***	0.16***	0.19***	0.14***	0.24***	0.49***
	(0.05)	(0.04)	(0.05)	(0.10)	(0.06)	(0.043)	(0.032)	(0.037)	(0.037)	(0.054)	(0.074)
p-value	0.00	0.002	0.000	0.000	0.000	0.002	0.000	0.000	0.00	0.00	0.000

Note: Standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
The p-value is for testing the null that the mobility function is linear.
For robustness check #1, columns (1), (2), (3), and (4) are results for (children 26, parents 35), (children 32, parents 40), (children 30, parents 43), and (children 34, parents 45), respectively. For robustness check #2, columns (1), (2), and (3) are results for baseline years 33, 35, and 36, respectively. For averaging, we use the average income from 32 to 36 as the baseline (normalization).
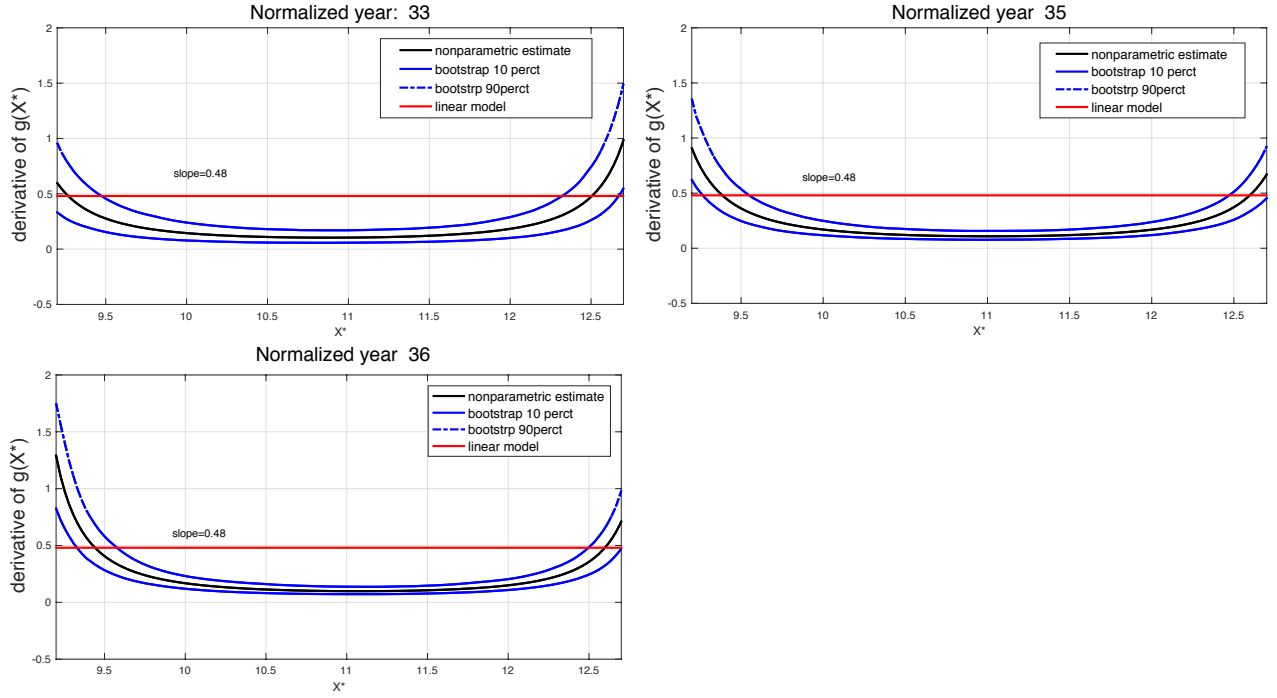
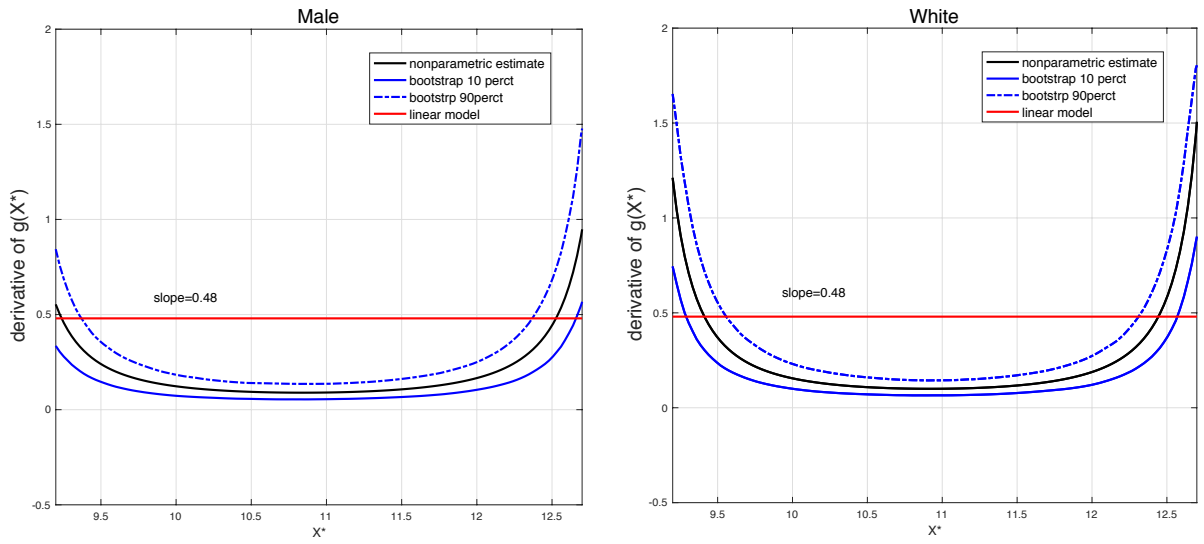Figure 4: Robustness check #2

Figure 5: Sub-sample analysis

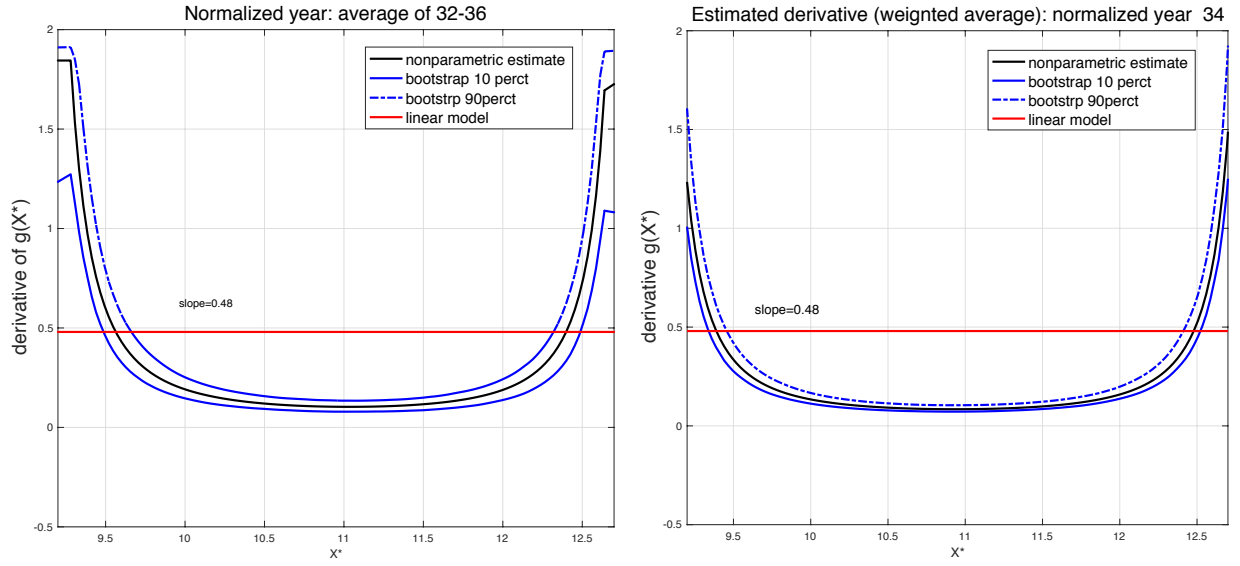Figure 6: Estimated derivatives of mobility function: multiple measurements

Figure 7: Simulation exercises for great gatsby curve

(a) Linear

(b) Quadratic
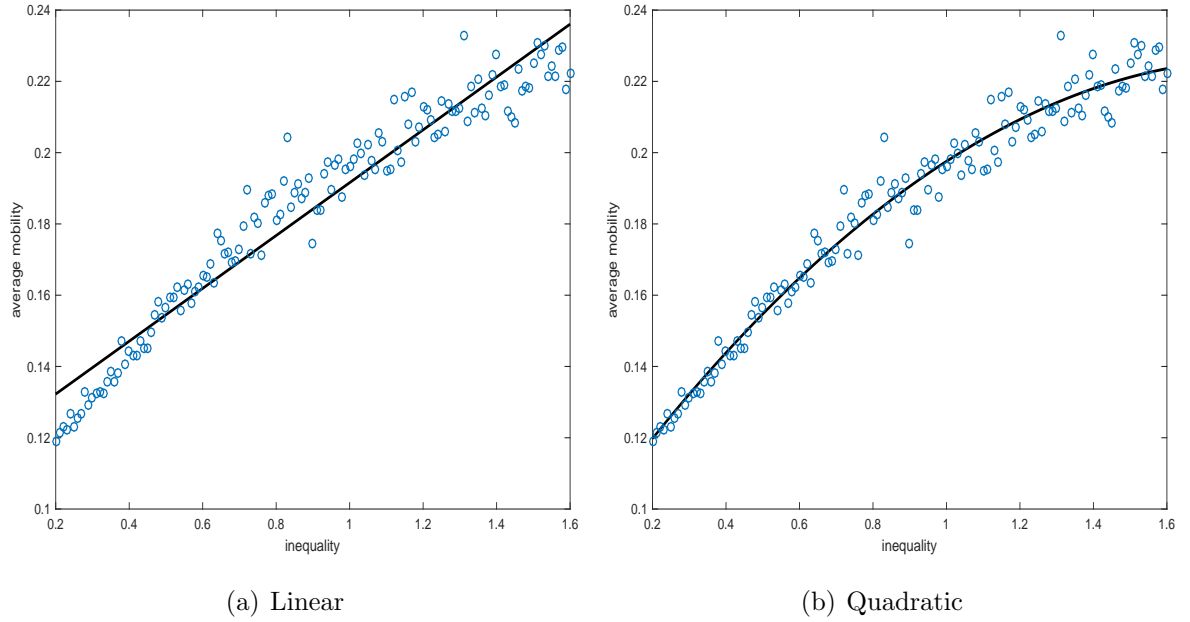
Part
Online Appendix to "A Nonparametric Non-classical Measurement Error Approach to Estimating Intergenerational Mobility Elasticities"

Table of Contents

Appendices

A   Monte Carlo Experiments

In the Appendix, we present Monte Carlo evidence to illustrate the finite-sample performance of the estimators in our paper and show our test's effectiveness. These results showcase the superior finite-sample performances of our proposed approach, even for modest size samples. Our designs accommodate measurement errors due to *both* self-reported *and* life-cycle bias, as well as nonlinearity.

We assume that the transitory shocks U_t and V_t are i.i.d. draws from the standard normal $\mathcal{N}(0,1)$. For illustrative purposes, we only simulate the data for two periods, and we fix $\alpha_2 = 1.10$ and $\delta_2 = 1.184$. Since our estimation procedure is conditional on Z, we need not worry about covariates in our Monte Carlo studies.

We consider three specifications of the function $g(X^*)$ and its derivative function (IGEs) $g'(X^*)$. (i) A linear mobility function and constant IGEs: $Y^* = g(X^*) + \varepsilon = 0.6X^* + \varepsilon, g'(X^*) = 0.6$. (ii) A quadratic mobility function and linear IGEs: $Y^* = g(X^*) + \varepsilon = X^* - 0.2X^{*2} + \varepsilon, g'(X^*) = 1 - 0.4X^*$. (iii) A cubic mobility function and quadratic IGEs: $Y^* = g(X^*) + \varepsilon = X^* - 0.2X^{*2} + 0.1X^{*3} + \varepsilon, g'(X^*) = 1 - 0.4X^* + 0.3X^{*2}$. In all the three specifications, X^* follows a normal distribution with mean zero and a standard deviation 2, and the structure error ε follows normal with mean zero and a standard deviation 0.1. The number of periods $T = 2$ and the sample size $N = 100, 300, 500$. We replicate each experiment for 1000 times.

In the first set of experiments, we maintain the normalization assumption $\alpha_1 = \delta_1 = 1$. We present the estimates of α_2 and δ_2 in Table A1. The results illustrate that our estimates perform very well even for a modest sample size of 300, which is smaller than the sample size used in most of the existing studies in this context; the superior finite-sample performances of our method highlight its potential use for empirical studies in this field. We present the nonparametrically estimated density of X^* in Figure A.1 for $N = 500$ under the three specifications. The results show that the estimates track the true density closely. The estimates of the derivative function $g'(\cdot)$ are presented in Figure A.2, where the linear, quadratic, and the cubic specifications are shown on the left, the middle, and the right, respectively. The figures showcase how the estimated functions closely capture the true derivative functions for both the linear and quadratic cases, and how the $[10\%, 90\%]$ point-wise confidence intervals improve significantly as the sample size increases.

Based on the estimated derivative functions, we formally test the functional form of the mobility function. The results are presented in Table A2. In all the tests, the null hypothesis is that $g'(\cdot)$ is a constant. When the true derivative is constant or linear, we use linear specification of the derivative as our alternative; when true model is quadratic, we use quadratic as our alternative. Thus the first row provides test size, and the next two rows are power.

Table A1: Estimate of α and δ

	Parameters			
	$\alpha_2 = 1.105$		$\delta_2 = 1.184$	
	Mean	Std. Dev.	Mean	Std. Dev.
Mobility function: $g(X^*) = 0.6X^*$				
$N = 100$	1.109	0.104	1.193	0.159
$N = 300$	1.108	0.057	1.187	0.087
$N = 500$	1.105	0.043	1.184	0.064
Mobility function: $g(X^*) = -0.2X^{*2} + X^*$				
$N = 100$	1.111	0.099	1.187	0.096
$N = 300$	1.109	0.055	1.185	0.052
$N = 500$	1.105	0.041	1.184	0.038
Mobility function: $g(X^*) = X^{*3} - 0.2X^{*2} + X^*$				
$N = 100$	1.109	0.092	1.184	0.008
$N = 300$	1.108	0.053	1.184	0.004
$N = 500$	1.107	0.042	1.184	0.003

The coefficients in period $t = 1$ are $\alpha_1 = \delta_1 = 1$ by construction. Standard errors are computed by using sample standard deviation of 1000 replications.

Table A2: Results of testing linearity

	$\alpha = \delta = 1$			$\alpha = \delta = 0.95$			$\alpha = \delta = 0.90$		
DGP	$N = 100$	$N = 300$	$N = 500$	$N = 100$	$N = 300$	$N = 500$	$N = 100$	$N = 300$	$N = 500$
constant (size)	0.0%	2%	4%	0.0%	1%	4%	2%	3%	5%
linear (power)	21%	63%	65%	26%	65%	69%	29%	58%	69%
quadratic (power)	10%	32%	65%	13%	40%	63%	18%	44%	66%

Note: "constant", "linear", and "quadratic" indicate that in the date generating process, the derivative of the mobility function is constant, linear, and quadratic, respectively. The null hypothesis is $g'(\cdot)$ that is a constant. When the true derivative is constant or linear, we use linear specification of the derivative as our alternative; when true model is quadratic, we use quadratic as our alternative. Thus the first row is test size, and the next two rows are test power.

The test results in Table A2 are satisfying. We first consider the cases where the normalization is correct, i.e., $\alpha = \delta = 1$. When the true derivative function is a constant, the size of the test for $N = 500$ is 4%, very close to the nominal size 5%. When the true derivative function is linear or quadratic, the power of the test increases significantly in sample size and it's about 70% for the moderate sample size $N = 500$. In the second set of experiments, we investigate the impact of mis-normalization on the test of linearity. Specifically, we assume that the normalization assumption $\alpha_1 = \delta_1 = 1$ does not hold in the data generating process, while in estimation we still impose this assumption. The testing results for $\alpha = \delta = 0.95$ and $\alpha = \delta = 0.90$ are similar to those for $\alpha = \delta = 1$. These results, again, showcase that the normalization does not affect our testing results, as discussed in Section 3, although the incorrect normalization affects the estimates and their confidence bands.

Figure A.1: Estimated density of X^*, $N = 500$
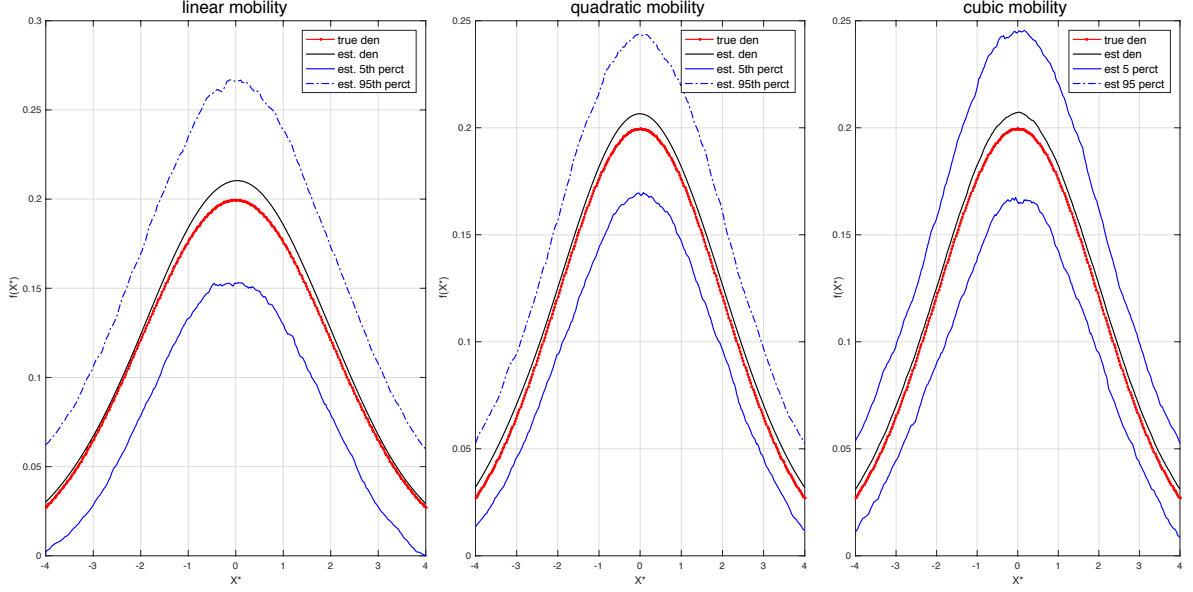
B   Further Discussions and Extensions

B.1   Continuous Variables

When the vector of covariates Z is continuous and discretizing it is not completely satisfactory in estimation, one may need to impose more structure on the mobility function $g(\cdot)$. Here, we consider a widely used partially linear model

$$Y^* = g(X^*, Z) + \epsilon = Z'\beta + h(X^*) + \epsilon, \tag{16}$$

with the associations of the annual income and permanent income being

$$
\begin{aligned}
Y_1 &= Y^* + U_1, \\
X_1 &= X^* + V_1, \\
X_2 &= \alpha_2 X^* + V_2.
\end{aligned}
\tag{17}
$$

Identification: For such a specification, we can identify the nonparametric function $h(\cdot)$ and the parameter β from the joint distribution of (Y_1, X_1, X_2, Z) in several steps.

First, the density of the permanent income X^*, $f_{X^*}(\cdot)$, can be identified from the joint distribution of X_1 and X_2 by applying Lemma 2. Next, the expectation of child income conditional on parental permanent income, $E(Y_1|X^*)$, can be identified using the identified density $f_{X^*}(x^*)$ together with Y_1, and X_1 based on Proposition 1. Following the similar argument, the mean of child characteristics Z

3

Figure A.2: Estimate of derivative functions
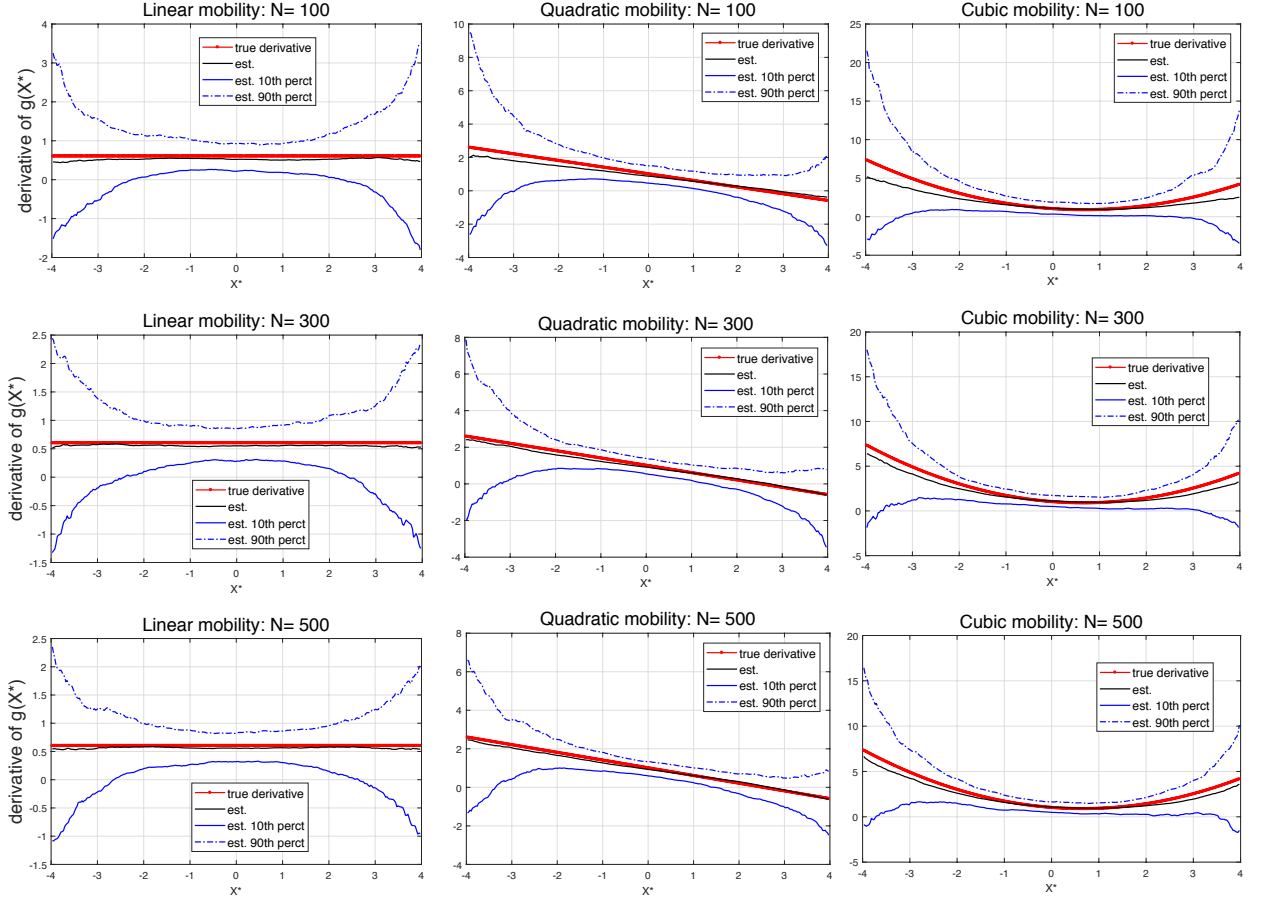
conditional on parental permanent income, $E(Z|X^*)$, can also be identified. It is easy to derive that

$$Y_1 - E(Y_1|X^*) = (Z' - E(Z'|X^*))\beta + (\epsilon + U_1).$$

If we assume that the matrix $(Z' - E(Z'|X^*))'(Z' - E(Z'|X^*))$ is full rank, then β is identified. Once β is identified, we can identify the non-parametric function $h(X^*)$ via the following closed-form expression

$$h(X^*) = E(Y_1|X^*) - E(Z'|X^*)\beta.$$

If Z and X^* are independent, the identification above fails because $(Z' - E(Z'|X^*))'(Z' - E(Z'|X^*))$ is zero. If this is the case, the model can be identified as follows. $E[Y_1|z_k] = h(X^*) + z'_k\beta, k = 1, 2,$ where z_1 and z_2 are two different realizations of Z. Then $E[Y|z_1] - E[Y|z_2] = (z'_1 - z'_2)\beta$. Under the assumption that the matrix $(z'_1 - z'_2)(z_1 - z_2)$ is full rank, β is identified. Given β is identified, we can restructure the original mobility model as $Y_1 - Z'\beta = h(X^*) + \epsilon + U_1$ and identify the mobility function $h(X^*)$ using Lemma 3 in our paper.

Estimation: Note that since the identification procedure above is constructive, one can follow the identification procedure to estimate the model directly, as we do in our paper. An alternative approach to estimate $(\beta, h(\cdot))$ is discussed in Section 4.2 of Härdle et al. (2012).

B.2   Sieve MLE as an Alternative Method

We have also estimated the mobility function in the empirical application by using a sieve MLE and present the result in Figure B.3 below, along with our nonparametric estimates. Both methods indicate a U-shape in the mobility elasticities with respect to parental income. The point-wise confidence intervals for sieve MLE are larger than our estimates, however.

Below, in addition to estimation details, we discuss why, despite its nice *theoretical* properties, the sieve MLE method may not necessarily be the best alternative *in practice* for this **particular** context. We also discuss the more restrictive assumptions required for the Sieve MLE to take into account multiple measurements, and then discuss an alternative approach within our framework.

1.   Sieve MLE Estimation Details: To be consistent with the data requirements of our nonparametric method, we use the joint distribution of Y, X_1, X_2 to conduct the sieve MLE. To be specific,

$$
\begin{aligned}
Y &= g(X^*) + U + \varepsilon \\
&= \beta_0 + \beta_1 x^* + \beta_2 x^{*2} + \beta_3 x^{*3} + \varepsilon' \\
X_1 &= X^* + V_1 \\
X_2 &= \alpha X^* + V_2
\end{aligned}
\tag{18}
$$

To proceed, we assume that Y, X_1 and X_2 are independent conditional on $X^* = x^*$. Under this assumption, the joint distribution of Y, X_1 and X_2 is

$$
\begin{aligned}
f_{Y,X_1,X_2}(y, x_1, x_2) &= \int f_{Y,X_1,X_2|X^*=x^*}(y, x_1, x_2|x^*) f_{X^*}(x^*) dx^* \\
&= \int f_{Y|X^*=x^*}(y|x^*) f_{X_1|X^*=x^*}(x_1|x^*) f_{X_2|X^*=x^*}(x_2|x^*) f_{X^*}(x^*) dx^* \\
&= \int f_{\varepsilon'}(y - g(x^*)) f_{V_1}(x_1 - x^*) f_{V_2}(x_2 - \alpha x^*) f_{X^*}(x^*) dx^*.
\end{aligned}
\tag{19}
$$

We first estimate the coefficient of life-cycle bias α as we do in the paper. Next, we use sieve MLE to estimate the parameters $\beta_j, j = 0, 1, 2, 3$ together with sieve parameters for the four unknown density functions $f_{\varepsilon'}, f_{V_1}, f_{V_2}$ and f_{X^*}. We use Hermite polynomials with order four for sieve approximations.

Figure B.3: Comparison between sieve MLE and nonparametric estimation
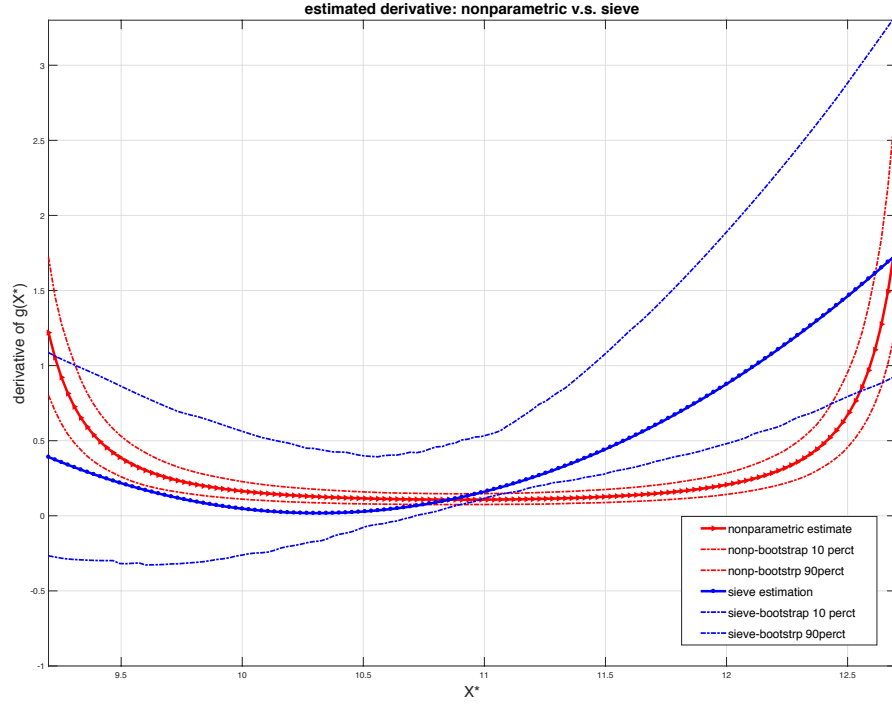
2. Further Discussions of Sieve MLE's practical issues: While sieve MLE is an excellent candidate for an alternative due to its nice theoretical properties, it has several practical disadvantages in this particular context of estimation of the mobility function. First, sieve MLE requires independence of Y, X_1, X_2 conditional on X^*, which is much more restrictive than our identification assumption. Second, our nonparametric estimation is a global one while sieve MLE is local. Possible local maximum may affect its performance in empirical applications. Finally, there is no theoretical guidance of choosing the order of the basis (Hermite polynomials in our case): the higher order it is, the smaller approximation error (and hence better performance of sieve MLE). Therefore, one faces the tradeoff between the approximation error and computational burden.

B.3 Multiple Measurement

In the last set of experiments, we investigate the possible improvement of our estimation by incorporating data of multiple periods by using model averaging. For this purpose, we first simulate data for six periods with $\alpha_1 = \delta_1 = 1, \alpha_2 = 1.10, \delta_2 = 1.184, \alpha_3 = \delta_3 = 0.8, \alpha_4 = \delta_4 = 0.9, \alpha_5 = \delta_5 = 1.2, \alpha_6 = \delta_6 = 1.05$. Then we estimate the five derivative functions using the joint distribution (Y_t, X_t, X_1) for $t = 2, \cdots, 6$. Finally, we take the weighted average of the five derivative functions through the inverse-variance weighting method. The results are presented in figure B.4, where the first plot is the weighted average and the next five ones are estimates for $t = 2, \cdots, 6$, respectively. As expected, the weighted average reduces standard errors, especially when the true income is extreme. Neverthe-

less, the improvement is relatively marginal, and the shape for each of the five derivative functions maintains.

Figure B.4: Estimate of derivative functions: Weighted average
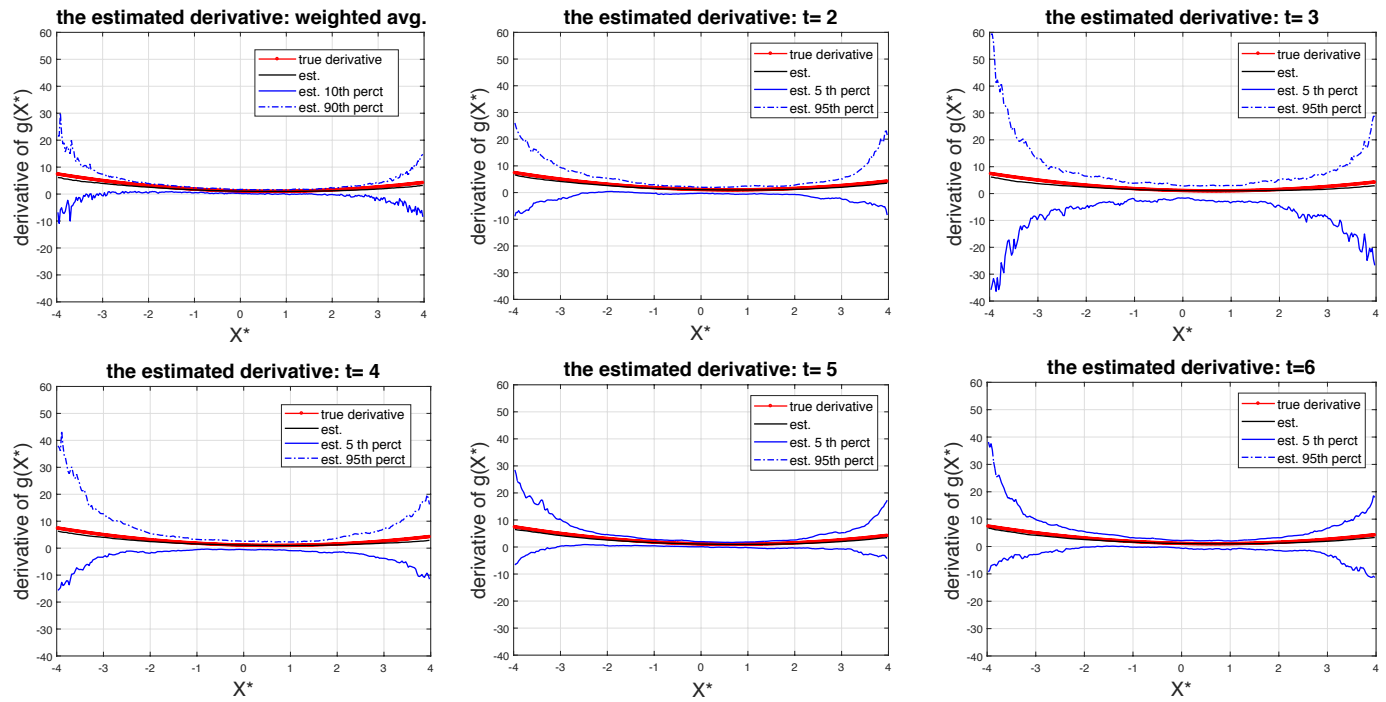
C Monte Carlo: Excluding Extremely Low Values

We also assess the robustness of our approach to exclusion of the zeros in more general settings via Monte Carlo experiments. Given our model set-up with the log income, we cannot directly simulate log zeros. Note that the log of zero is negative infinity, we therefore consider dropping the extreme values defined as bottom 5 and 10 percent of the distribution of observed (simulated) parental incomes in our simulation experiments (Figure C.5). Not surprisingly, we find that exclusion of extreme values disproportionately impacts the estimates in the lower tail, but the patterns of nonlinearity are rather robust. This may highlight the usefulness of our nonparametric approach in practice when using the administrative data.

Figure C.5: Monte Carlo simulation results dropping the bottom tail of the distribution

(a) LInear (b) Quadratic (c) Cubic

[**Five Percent**]

(d) Linear (e) Quadratic (f) Cubic

[**Ten Percent**]