STANFORD CENTER
on Poverty and Inequality Inequality.com

ESTIMATING THE INTERGENERATIONAL ELASTICITY OF EXPECTED INCOME WITH SHORT-RUN INCOME MEASURES: A GENERALIZED ERROR-IN-VARIABLES MODEL

Pablo A. Mitnik
(pmitnik@stanford.edu)

May, 2019

The Stanford Center on Poverty and Inequality is a program of the Institute for Research in the Social Sciences at Stanford University.

Estimating the Intergenerational Elasticity of Expected Income with Short-Run Income Measures: A Generalized Error-in-Variables Model

Pablo A. Mitnik[a]*

[a] *Center on Poverty and Inequality, Stanford University, USA*

*Contact: Pablo A. Mitnik. Email: pmitnik@stanford.edu. Mail: Center on Poverty and Inequality, Stanford University, 450 Serra Mall, Building 370, Room 212, Stanford, CA 94305, USA.

It has recently been argued that the intergenerational income elasticity (IGE) ubiquitously estimated in the economic mobility literature should be replaced by the IGE of expected income. This article derives an approximate closed-form expression for the probability limit of the Poisson Pseudo-Maximum Likelihood (PPML) estimator of constant elasticities and uses it to advance a generalized error-in-variables model for the estimation of the latter IGE with short-run income measures. Empirical analyses with data from the Panel Study of Income Dynamics offer clear support for the account of lifecycle and attenuation biases the model provides. Together, the model and the associated empirical evidence supply a methodological justification for the estimation of the IGE of the expectation with the PPML estimator and proxy income variables that satisfy some conditions. This eliminates the main obstacle for making this IGE the workhorse elasticity of the economic mobility field, which would (a) dissolve the selection-bias problem generated by the current expedient of dropping children with zero income from samples in order to make estimation of the conventional IGE possible, and (b) greatly stimulate the study of gender and marriage dynamics in intergenerational processes, something that the reliance on the conventionally estimated IGE essentially precludes.

Keywords: IGE of expected income; lifecycle bias; attenuation bias; Poisson Pseudo-Maximum Likelihood estimator; measurement error in nonlinear models

1. Introduction

The intergenerational elasticity (IGE) is the workhorse measure of intergenerational economic mobility. It has been widely estimated over the last four decades, very often with the goal of providing a summary assessment of the level of income or earnings mobility within a country (for reviews, see Solon 1999:1778-1788; Corak 2006; Mitnik et al. 2018:9-15). Beyond this elementary goal, the IGE has been used, among other purposes, to conduct comparative analyses of economic mobility and persistence across countries, regions, demographic groups, cohorts, and time periods (e.g., Chadwick and Solon 2002; Hertz 2005, 2007; Aaronson and Mazumder 2008; Björklund and Jäntti 2000; Mayer and Lopoo 2008; Bloome and Western 2011); to examine the relationship between cross-sectional economic inequality and mobility across generations (e.g., Corak 2013; Bloome 2016); and to theoretically model and empirically study the impact of social policies and political institutions on the persistence of economic status (e.g., Solon 2004; Bratsberg et al. 2007; Ichino et al. 2011; Landersø and Heckman 2017).

Despite the very central place that the IGE holds in the study of intergenerational mobility, Mitnik and Grusky (2019) have shown that this elasticity has been specified in a way that is inconsistent with the various interpretations imposed on its estimates. While the IGE has been construed as pertaining to the expectation of children's income conditional on their parents' income—as apparent, for instance, in its archetypical interpretation as a measure of regression to the (arithmetic) mean—in fact it pertains to the conditional geometric mean of the children's income. This not only invalidates the conventional interpretations of the IGE but, much more consequentially, it also generates serious methodological problems. At their root, these are the result of a simple fact, i.e., that the geometric mean is undefined for variables including zero in their support. This fact markedly hampers the study of gender and marriage dynamics in intergenerational processes, as it (a) makes it impossible to determine the extent to which parental economic advantage is transmitted through the labor market among women, and (b) greatly hinders research on the role that marriage plays in generating the observed levels of intergenerational persistence in family income. Equally important, estimation of the IGE with the annual or other short-run proxy income measures that are available is almost certainly affected by substantial selection biases, which result from mobility scholars' expedient of dropping children with zero income from samples (to address what is perceived as the problem of the logarithm of zero being undefined). This makes the (currently widespread) use of the IGE of sons' individual earnings as an index of economic persistence and mobility in a country a rather problematic practice.

Mitnik and Grusky (2019) have argued that these conceptual and methodological problems can be solved all at once by simply replacing the IGE of the geometric mean (the de facto estimated IGE) by the IGE of the expectation (the IGE consistent with the interpretations advanced by mobility scholars) as the workhorse intergenerational elasticity. They have also called for effectuating such replacement. Their call, however, confronts a key methodological obstacle.

IGEs are defined in terms of long-run income variables. However, such variables are very rarely available. Over the last 40 years or so, a large body of knowledge on the lifecycle and attenuation biases that may affect the estimation of the conventional IGE with short-run proxy income measures, and the methodological strategies that may be used to avoid them, has accumulated. A very central achievement in this regard is Haider and Solon's (2006) generalized error-in-variables (GEiV) model, which supplies the current methodological justification for the estimation of the conventional IGE with proxy variables that satisfy some conditions (see Nybom and Stuhler 2016 for an illuminating discussion). In stark contrast, the IGE of the expectation was estimated for the first time by Mitnik et al. (2015), and there is very little methodological knowledge regarding its estimation with short-run measures. In particular, there is nothing comparable to the GEiV model that would apply to the IGE of the expectation. As long as this is the case, broad estimation of this IGE is unfeasible. In this article, I remove this key methodological obstacle to estimating the IGE of the expectation with short-run proxy income variables.

Tackling this obstacle is made difficult by the lack of closed-form expressions for the probability limits of the estimators of the IGE of the expectation, and by the fact that in nonlinear models the bias induced by right-side measurement error may be in any direction. I deal with these difficulties by deriving an approximate closed-form expression for the probability limit of the Poisson Pseudo-Maximum Likelihood (PPML) estimator based on Taylor-series expansions, and by relying on substantive knowledge to sign some key quantities in conducting proofs. This allows me to advance a formal measurement-error model, the generalized error-in-variables model for the estimation of the IGE of the expectation (or GEiVE model), which plays for the IGE of the expectation the same role that the GEiV model plays for the conventional IGE. Further, I empirically evaluate the GEiVE model with U.S. data from the Panel Study of Income Dynamics (PSID). My results indicate that, when estimating the IGE of the expectation with proxy measures, the observed lifecycle and attenuation biases are in close agreement with the GEiVE model's predictions. Equally important, the results provide evidence on the patterns and magnitudes of lifecycle and attenuation biases that are relevant for the specification of estimation strategies and the interpretation of results obtained with the same and with other data.

Similarly to what Haider and Solon's (2006) GEiV model and the associated evidence did for the estimation of the conventional IGE, the GEiVE model and the evidence I present here supply a methodological justification for the estimation of the IGE of the expectation with proxy income variables that satisfy some conditions. These are that the parental income measure is based on several years of information—about 13 years in the case of survey data, almost certainly fewer years with administrative data—and that the income measures for both parents and children are obtained when they are close to 40 years old. This eliminates the main obstacle for the estimation of the IGE of the expectation with the data typically available and, therefore, for replacing the conventionally estimated IGE by the IGE of the expectation.

The structure of the rest of the article is as follows. I start by explaining Mitnik and Grusky's (2019) proposal to redefine the IGE and the motivation behind it. Next, I summarize the GEiV model, introduce the new GEiVE model, and present the empirical analyses. The last section distills the article's main conclusions. Mathematical proofs and supplementary materials are collected in an Online Appendix.

2. The IGE of what? Redefining the workhorse intergenerational elasticity

As already indicated, while mobility scholars have interpreted the conventionally estimated IGE as the elasticity of the expectation of children's income or earnings conditional on parental income (see textual evidence in Mitnik and Grusky 2019: Sec. II), that IGE pertains in fact to the conditional geometric mean. The standard population regression function (PRF) posited in the literature, which assumes the elasticity is constant across levels of parental income, is:

$$E(\ln Y\,|x) = \beta_0 + \beta_1 \ln x, \tag{1}$$

where Y is the child's long-run income or earnings, X is long-run parental income or father's earnings, β_1 is the IGE as specified in the literature, and I use expressions like "$Z|w$" as a shorthand for "$Z|W = w$" (where Z and W are any random variables). Due to Jensen's inequality, $E(\ln Y|x) \neq \ln E(Y|x)$ and the parameter β_1 is not, in the general case, the elasticity of the conditional expectation of the child's income. Instead, $E(\ln Y|x) = \ln \exp E(\ln Y|x) = \ln GM(Y|x)$, and Equation (1) is equivalent to

$$\ln GM(Y|x) = \beta_0 + \beta_1 \ln x, \tag{2}$$

where GM denotes the geometric mean operator. Therefore, β_1 is the elasticity of the conditional geometric mean—that is, the percentage differential in the geometric mean of children's long-run income with respect to a marginal percentage differential in parental long-run income.[1]

As the geometric mean is undefined when an income distribution includes zero in its support, the IGE is undefined as well when this is the case. Mitnik and Grusky (2019: Sec. IV) have shown that this has very negative consequences for the study of gender and marriage dynamics in intergenerational processes. First, because in many national-historical contexts a sizable share

---

[1] The parameter β_1 is (also) the IGE of the expectation only when the error term satisfies very special conditions (Santos Silva and Tenreyro 2006; Petersen 2017; Wooldridge 2002:17).

of women never join the (paid) labor force, the "zeros problem" makes it impossible to determine in those contexts the extent to which parental economic advantage—as measured by the IGE of long-run family-income—is transmitted through the labor market among women, and whether there are gender differences in this regard (as the relevant estimand, the IGE of daughters' long-run earnings, is undefined). Second, the zeros problem greatly hinders research on the role that marriage plays in generating the observed levels of intergenerational persistence in family income. This is so because a measure that is central for this research—the IGE of the income contributed by a spouse with respect to a child's own parental income—is also undefined as long as there are children that are single and spouses of married children that contribute no income.

The unwitting reliance of the mobility field on the IGE of the geometric mean also has very negative consequences for the widespread practice of using the IGE of sons' individual earnings and (less frequently) of sons' and daughters' family income as indices of economic persistence and mobility in a country. Indeed, the IGE is defined in terms of long-run earnings and family income, and it may be reasonable to assume, as an approximation, that these long-run measures are positive rather than nonnegative.[2] But estimation is nevertheless adversely affected by the ubiquitous need to use short-run measures of children's economic status (e.g., sons' earnings in a particular year) as proxies for their unavailable long-run counterparts (e.g., sons' average lifetime earnings). This is the case because the short-run measures almost always have a substantial probability mass at zero even when the corresponding long-run measures don't. Assuming that zero is not in the support of the annual or other short-run measures used for estimation is never a reasonable approximation in the case of sons' earnings (due to unemployment and other forms of nonemployment), and it's not a reasonable approximation in most cases—the potential

[2] As conditions like severe physical disability, severe mental illness, and imprisonment affect some children over their entire adult lives, strictly speaking zero is in the support of all measures of children's long-run income, including men's earnings and men's and women's family income.

exception being countries with a highly developed welfare state—in the case of family income. (For the prevalence of zero-income families in the United States, see Chetty et al. [2014: Online Appendix Table IV].) As argued in detail by Mitnik and Grusky (2019: Sec. III), dropping children without earnings or income from samples (the near-universal expedient employed to address this "short-run zeros problem") generates substantial selection biases, and all approaches that might be used to try to avoid those biases, including imputing "small values" to children with zero earnings or income, are very unattractive due to a combination of methodological and pragmatic reasons. [3]

To address these problems, and similarly to what Santos Silva and Tenreyro (2006) did in the field of international trade, Mitnik and Grusky (2019) have called for redefining the workhorse measure of economic mobility. This entails replacing the PRF of Equation (1) by a PRF whose estimation delivers estimates of the IGE of the expectation in the general case. Under the assumption of constant elasticity, that PRF can be written as:

$$ln\, E(Y|x) = \alpha_0 + \alpha_1\, ln\, x, \qquad (3)$$

where $Y \geq 0$, $X > 0$ and $\alpha_1 = d \ln E(Y|x)\, / d \ln x$ is the percentage differential in the expectation of children's long-run income with respect to a marginal percentage differential in parental long-run income. Crucially, (a) all interpretations incorrectly applied to the conventional IGE are correct or approximately correct under this formulation (Mitnik and Grusky 2019: Sec. V.A) while an attractive additional interpretation in terms of inequality of opportunity (Mitnik et al. 2019) is also available, and (b) the IGE of the expectation is fully immune to the methodological

[3] Zeros are much less of a problem for short-run parental measures, as (a) they typically are multiyear averages (more on this later), (b) parents with zero income for extended periods of time are unlikely to be able to raise their kids themselves, and (c) there are good reasons to use as parental income the income of the family in which the child was raised (see, e.g., Hertz 2007:35).

problems affecting the IGE of the geometric mean just discussed and, in particular, is very well suited for studying the role of marriage in the intergenerational transmission of advantage (see Mitnik and Grusky 2019: Sec. V.B for details; and Mitnik et al. 2015:64-68 for an application).

3. The GEiV model

When actual measures of long-run income or earnings for parents and children are available, Equation (1) may be consistently estimated by Ordinary Least Squares (OLS). However, such measures are almost never available, only short-run proxy variables are. Estimating Equation (1) by OLS after substituting the latter variables for the former opens the door to the two types of biases widely discussed in the literature. First, as income- and earnings-age profiles differ across economic origins, lifecycle biases result from estimating the conventional IGE with proxy measures taken when parents or children are too young or too old to represent lifetime differences well (e.g., Black and Devereux 2011). Second, in the case of the parental variables, and even in the absence of any lifecycle bias, the combination of transitory fluctuations and measurement error in the measure of short-run income (with respect to true short-run income) produce substantial attenuation bias (e.g., Solon 1999; Mazumder 2005). The joint analysis of these biases is provided by the GEiV model (Haider and Solon 2006; see also Nybom and Stuhler 2016).

For the sake of simplicity, I assume that all children are from the same cohort. I start by introducing the following population linear projections:

$$\ln Z_t = \lambda_{0t} + \lambda_{1t} \ln Y + V_t \tag{4}$$

$$\ln S_k = \eta_{0k} + \eta_{1k} \ln X + Q_k, \tag{5}$$

where $Z_t > 0$ is children's income at age *t*; $Y > 0$; $\lambda_{0t} + V_t$ is the (additive) measurement error in the logarithm of the short-run children's variable as a measure of the logarithm of the corresponding long-run variable when $\lambda_{1t} = 1$; λ_{1t} captures left-hand lifecycle bias and thus may be different from 1 and varies with *t*; $S_k > 0$ is parents' income at age *k*; $X > 0$; $\eta_{0k} + Q_k$ is the (additive) measurement error in the logarithm of the short-run parental variable as a measure of the logarithm of the corresponding long-run variable when $\eta_{1k} = 1$; and η_{1k} captures right-hand lifecycle bias and thus may be different from 1 and varies with parents' age.[4]

The empirical assumptions of the GEiV model are the following. For any *t* and *k*:

$$Cov(\ln X, V_t) = 0 \quad (6)$$

$$Cov(\ln Y, Q_k) = 0 \quad (7)$$

$$Cov(V_t, Q_k) = 0 \quad (8)$$

These assumptions are expected to hold imperfectly but still as good approximations, at least when $\lambda_{1t} \approx \eta_{1k} \approx 1$. (To simplify the notation, in what follows I drop the subscripts *t* and *k*.)

Following Haider and Solon (2006), I present separate analyses of measurement error in the left- and right-side variables of Equation (1). If instead of estimating the PRF of Equation (1), one estimates by OLS a PRF in which $\ln Z$ is substituted for $\ln Y$, or a PRF in which $\ln S$ is substituted for $\ln X$, the probability limits of the slope coefficients are respectively:

$$\tilde{\beta}_1 = \frac{Cov(\ln Z, \ln X)}{Var(\ln X)} = \lambda_1 \beta_1 + \frac{Cov(V, \ln X)}{Var(\ln X)} \quad (9)$$

$$\check{\beta}_1 = \frac{Cov(\ln Y, \ln S)}{Var(\ln S)} = \frac{\beta_1 \, \eta_1 \, Var(\ln X) + Cov(Q, \ln Y)}{(\eta_1)^2 \, Var(\ln X) + Var(Q)}. \quad (10)$$

Assuming Equation (6), it then follows from Equation (9) that $\tilde{\beta}_1 = \lambda_1 \beta_1$, which leads to the conclusion that left-side measurement error is unproblematic for estimation as long as the measurement age is among the "right points" of the child' lifecycle, i.e., as long as $\lambda_1 \cong 1$. In contrast, $\tilde{\beta}_1$ will be affected by an upward (downward) bias if $\lambda_1 > 1$ ($\lambda_1 < 1$). In turn, assuming Equation (7), it follows from Equation (10) that

---

[4] In the general case, the measurement errors are equal to $\lambda_{0t} + (\lambda_{1t} - 1) \ln Y + V_t$ (children) and $\eta_{0k} + (\eta_{1k} - 1) \ln X + Q_k$ (parents).

$$\breve{\beta}_1 = \beta_1 \frac{\eta_1}{(\eta_1)^2 + \frac{Var(Q)}{Var(\ln X)}}. \quad (11)$$

This equation shows that, as long as the parents' age is among the "right points" of their parents' lifecycle, i.e., as long as $\eta_1 \cong 1$, estimation with short-run proxies for long-run parental income is affected by a textbook form of attenuation bias, whose size depends on the relative magnitudes of $Var(Q)$ and $Var(\ln X)$ (i.e., on the size of what I later refer as the "variance ratio"). Values of η_1 larger than 1 will tend to exacerbate attenuation bias whereas values smaller than 1 will tend to reduce it—and, in fact, if η_1 is small enough, the result may even be amplification bias.

The GEiV model plays two main functions. First, it supplies the current methodological justification for the estimation of the conventional IGE by OLS with proxy variables that satisfy some conditions (Nybom and Stuhler 2016). Indeed, the GEiV model suggests that using income measures pertaining to specific ages should eliminate the bulk of lifecycle bias—while the available evidence (which pertains to men) indicates that estimating IGEs with parents' and children's information close to age 40 is the best approach (Haider and Solon 2006; Böhlmark and Lindquist 2006; Mazumder 2001; Nybom and Stuhler 2016). In the case of attenuation bias the GEiV model, and many analyses predating it (e.g. Solon 1992), suggests pushing $Var(Q)$ down by using a multiyear average of parents' income or earnings, rather than an annual measure, as the proxy measure *S*.[5] There is strong evidence that the bias can be substantially reduced if the average is computed over enough years but there is disagreement on how many years are necessary to eliminate the bulk of it (see Mazumder 2005; Chetty et al. 2014: 1582 and Online Appendix E; Mitnik et al. 2018: 9-15; Mazumder 2016).

The second function of the GEiV model is that it clearly identifies the various potential sources of bias in OLS estimates of the conventional IGE, and therefore provides an analytical framework to help investigate those biases with the very few datasets where long-run measures

[5] In some cases, a multiyear average of the logarithm of parental income has been used, which implies that the geometric mean of parental income over those years is the proxy measure *S*.

can be constructed. The model plays this function directly, as it points to values of λ_1 and η_1 not close enough to 1, and $Var(Q)$ not small enough, as sources of bias—which is what, explicitly or implicitly, the mobility literature has traditionally stressed. But the model also plays its second function indirectly, by virtue of being nested within a general expression—represented here by Equations (9) and (10)—for the probability limit of the OLS estimator of the conventional IGE with short-run variables. Thus, in contrast to the previous literature, Nybom and Stuhler (2016) have stressed that bias may result from the assumptions of the GEiV model not holding up.

For future reference, it's convenient to provide more details on the specific argument advanced by Nybom and Stuhler (2016:245-246). They argued that (a) the parameter λ_1 only captures how differences in annual and lifetime income relate on average across children, (b) there are good reasons to expect idiosyncratic deviations from this average relationship to correlate within families or with parental income, and (c) as a result, we should not expect that $Cov(V, \ln X) = 0$ when $\lambda_1 = 1$, which in turn entails that estimation of the IGE when $\lambda_1 = 1$ would not (fully) eliminate left-side lifecycle bias. Their empirical analyses with Swedish data are consistent with this argument, as when $\lambda_1 = 1$ (which happens when the children are about 35 years old) $Cov(V, \ln X) < 0$ and there is what we may call a "residual lifecycle bias" of about – 20 percent.[6] I later refer to this argument as the "correlated deviations argument."

4. The GEiVE model

After substituting short-run for long-run income measures in Equation (3), the IGE of the expectation can be estimated using several approaches. These include Nonlinear Least Squares, Generalized Method of Moments, and Pseudo-Maximum Likelihood (PML) based on a variety of distributions in the linear exponential family (e.g., Poisson, gamma).[7] Here I assume that estimation

---

[6] Nybom and Stuhler (2016) also found that lifecycle bias is very close to zero around age 40.

[7] PML estimators are consistent regardless of the actual distribution of the error term, provided that the mean function is correctly specified (Gourieroux, Monfort, and Trognon 1984).

is based on the PPML estimator employed by Minik et al. (2015; 2018; 2019) and Mitnik and Grusky (2019), as there are good reasons to prefer it over other estimators of constant-elasticity models (for these reasons, see Santos Silva and Tenreyro 2006, 2011).

What are the consequences of estimating the IGE of the expectation with the PPML estimator, when short-run proxy variables are substituted for the unavailable long-run variables? To answer, I advance here a measurement-error model aimed at playing for the IGE of the expectation the same role that the GEiV model has played for the estimation of the conventional IGE by OLS. Developing this model requires addressing difficulties absent in the case of the conventional IGE. First, a closed-form expression for the probability limit of the PPML estimator (or of any other estimator that could be used) is unavailable. Second, in nonlinear models the bias induced by right-side measurement error may be in any direction even when there is only one independent variable and the error is classical and additive in form (Carroll et al. 2006: Sec. 3.6; Schennach 2016). I address these difficulties by using the relevant population-moment condition and Taylor-series expansions to derive an approximate closed-form expression for the probability limit of the PPML estimator, and by relying on substantive knowledge (e.g., of the signs of various covariances) to conduct proofs and to derive the model's implications. This allows me to determine the direction of all potential biases and the factors driving them.

I start by introducing the following population linear projections:

$$Z_t = \theta_{0t} + \theta_{1t}\, Y + W_t \tag{12}$$

$$\ln\, S_k = \pi_{0k} + \pi_{1k} \ln X + P_k, \tag{13}$$

where $Z_t \geq 0; Y \geq 0;\; \theta_{0t} + W_t$ is the (additive) measurement error in the children's short-run income variable as a measure of the corresponding long-run variable when $\theta_{1t} = 1$; θ_{1t} captures left-hand lifecycle bias and thus may be different from 1 and varies with *t*; and Equation [13] is the same as Equation [5] but I have used a different notation for the parameters and the error term because the populations covered by the two equations will be different as long as there are children whose short-run income is zero.[8] Indeed, unlike in the case of the GEiV model in which Z_t and Y are positive, in the GEiVE model Z_t and Y are nonnegative. The reason is that, as the left-side measurement error concerns the short-run income variable rather than its logarithm, the model does not preclude it from being zero and children without any short-run income (e.g., those who are unemployed the year their incomes are measured) are included in all analyses.

The empirical assumptions of the GEiVE model are the following. For any t and k:

$$Cov(W_t, \ln X) = 0 \tag{14}$$

$$Cov(P_k, Y) = 0 \tag{15}$$

$$Cov(W_t, P_k) = 0. \tag{16}$$

Like in the case of the GEiV model, these assumptions are expected to hold imperfectly but still as good approximations, at least when $\theta_{1t} \approx \pi_{1k} \approx 1$. (As before, I omit in what follows the subscripts t and k.)

The analyses of the GEiV model compare (a) the probability limit of the OLS estimator of the conventional IGE with long-run income variables, with (b) the probability limit of the estimator when short-run proxy variables are used instead. Conducting this comparison is unproblematic, as it relies on the closed-form expression for the probability limit of the slope of an OLS regression. Although the corresponding probability limit is not available in closed form in the case of the IGE of the expectation, it's easy to show that α_1 (the probability limit of the PPML estimator $\hat{\alpha}_1$) solves the population-moment condition

$$\frac{E(X^{\alpha_1} \ln X)}{E(X^{\alpha_1})} = \frac{E(Y \ln X)}{E(Y)} \tag{17}$$

(Online Appendix, A). This suggests carrying out the analysis in terms of moment conditions.

---

[8] In the general case, the measurement error for children is $\theta_{0t} + (\theta_{1t} - 1)\, Y + W_t$. For parents, it's $\pi_{0k} + P_k$ when $\pi_{1k} = 1$ and $\pi_{0k} + (\pi_{1k} - 1) \ln X + P_k$ in the general case.

Although this is a feasible approach, it leads to long and convoluted proofs and makes the interpretation of results less straightforward. For this reason, I resort to approximate closed-form expressions based on moment conditions rather than working with the conditions themselves.

Without any loss of generality, I assume in what follows that $E(Z) = E(Y) = E(\ln S) = E(\ln X) = 1$.[9] An approximate expression for α_1 can be obtained by replacing the expectations on the left-hand side of Equation (17) by second-order Taylor-series approximations, and then solving the quadratic polynomial on α_1 that results. This yields (Online Appendix, A):

$$\alpha_1 \approx C_{\alpha_1} - \left[\left(C_{\alpha_1}\right)^2 - V_{\alpha_1}\right]^{\frac{1}{2}}, \tag{18}$$

where $C_{\alpha_1} = [Cov(Y, \ln X\,)]^{-1}$, $V_{\alpha_1} = 2\,[Var(\ln X)]^{-1}$, $\left(C_{\alpha_1}\right)^2 > V_{\alpha_1}$, $\partial\alpha_1/\partial Cov(Y, \ln X\,) > 0$, and $\partial\alpha_1/\partial\, Var(\ln X) < 0$.

For simplicity, and as Haider and Solon (2006) did, I next consider left- and right-side biases separately. Substituting Z for Y in Equation (3) yields the PRF $\ln E(Z|x) = \tilde{\alpha}_0 + \tilde{\alpha}_1 \ln X$. The probability limit of the PPML estimator, $\tilde{\alpha}_1$, is

$$\tilde{\alpha}_1 \approx C_{\tilde{\alpha}_1} - \left[\left(C_{\tilde{\alpha}_1}\right)^2 - V_{\alpha_1}\right]^{\frac{1}{2}}, \tag{19}$$

where

$$C_{\tilde{\alpha}_1} = \frac{1}{Cov(Z, \ln X\,)} = \frac{1}{\theta_1 Cov(\,Y, \ln X\,) + Cov(W, \ln X\,)},$$

$\left(C_{\tilde{\alpha}_1}\right)^2 > V_{\alpha_1}$, $\partial\tilde{\alpha}_1/Cov(Y, \ln X) > 0$, $\partial\tilde{\alpha}_1/\partial\, Cov(W,\ \ln X)\ > 0$, $\partial\tilde{\alpha}_1/\partial\theta_1 > 0$, and $\partial\tilde{\alpha}_1/\partial\, Var(\ln X) < 0$.

Equation (19) is the counterpart to Equation (9). The former equation identifies for the IGE of the expectation, as the latter does for the conventional IGE, the factors that may generate

---

[9] This entails no loss of generality because it can always be achieved by changing the monetary units used to measure income, i.e., by dividing the children's income variable by its mean and the parental income variable by the exponential of the mean of its logarithmic values minus 1.

left-side bias—and, in fact, together with Equation (18) it offers a full characterization of the relationship between α_1 and $\tilde{\alpha}_1$, while Equation (9) ignores selection bias (which is generated by the expedient of dropping children with zero short-run income) and thus fails to do the same for the relationship between β_1 and $\tilde{\beta}_1$. In particular, Equation (19) shows that lifecycle bias is fully eliminated if $\theta_1 = 1$ and $Cov(W, \ln X) = 0$. Indeed, comparing this equation with Equation (18) indicates that $\tilde{\alpha}_1$ and α_1 are approximated by the same Taylor-series-based expression when those conditions hold. Therefore, under the GEiVE model's empirical assumption specified by Equation (14)—and in strict analogy with what the GEiV model entails for the conventional IGE—it follows from Equation (19) that (the bulk of) left-side lifecycle bias will be eliminated as long as the measurement age is among the "right points" of the children's lifecycle, i.e., as long as $\theta_1 \approx 1$. Further, given that income-age profiles vary across people with different levels of human capital and that the latter are strongly associated to parental income, we expect θ_1 to be smaller than 1 when the children are younger and larger than 1 when they are older. Hence, as $\partial\tilde{\alpha}_1/\partial\theta_1 > 0$, we also expect that estimation with measures pertaining to when the children are too young (too old) will lead to underestimation (overestimation) of the IGE of the expectation, which is again similar to what the GEiV model entails for the conventional IGE (see the Online Appendix, B, for additional comments and for proofs for claims in the last two paragraphs).

Moving to the right side, substituting S for X in Equation (3) yields the PRF $\ln E(Y|s) = \breve{\alpha}_0 + \breve{\alpha}_1 \ln S$. The probability limit of the PPML estimator, $\breve{\alpha}_1$, is (Online Appendix, C)

$$\breve{\alpha}_1 \approx C_{\breve{\alpha}_1} - \left[\left(C_{\breve{\alpha}_1}\right)^2 - V_{\breve{\alpha}_1}\right]^{\frac{1}{2}}, \qquad (20)$$

where

$$C_{\breve{\alpha}_1} = [Cov(Y, \ln S)]^{-1} = [\pi_1 \, Cov(Y, \ln X) + Cov(Y, P)]^{-1}$$

$$V_{\breve{\alpha}_1} = 2\,[Var(\ln S)]^{-1} = 2\,\{Var(\ln X)[(\pi_1)^2 + VR]\}^{-1},$$

$VR = Var(P)/Var(\ln X)$ is the "variance ratio," $\left(C_{\breve{\alpha}_1}\right)^2 > V_{\breve{\alpha}_1}$, $\partial\breve{\alpha}_1/Cov(Y, \ln X) > 0$,

$\partial\breve{\alpha}_1/Cov(Y,P\ ) > 0$, $\partial\breve{\alpha}_1/\partial\ Var(\ln X) < 0$, $\partial\breve{\alpha}_1/\partial\ VR < 0$, and $\partial\tilde{\alpha}_1/\partial\pi_1 < 0$ (the latter under some conditions specified below). Further, under the GEiVE model's empirical assumption specified by Equation (15), $C_{\breve{\alpha}_1}$ reduces to

$$C_{\breve{\alpha}_1} = [\pi_1\ Cov(Y, \ln X\ )]^{-1}.$$

Equation (20) is the counterpart to Equations (10) and (11). Similarly to what Equation (10) does for the conventional IGE, Equation (20) identifies the various right-side factors that may jointly affect estimation of the IGE of the expectation with short-run measures of parental income. Likewise, it indicates—as does Equation (11) in the case of the conventional IGE—that even when Equation (15) holds, estimation of the IGE of the expectation with short-run measures of parental income is still compatible with both attenuation and amplification bias, as $(\pi_1)^2 + VR$ may be larger or smaller than 1 when $\pi_1 < 1$.

There is evidence (e.g., Mazumder 2001; Mazumder 2005) that $Var(P)$ follows an asymmetric U pattern, reaching its minimum value close to age 40 and increasing significantly as parents get older; the variance ratio can be expected to exhibit the same pattern. In addition, for the same reason as in the left-side analysis, we expect π_1 to be smaller than 1 when the parents are younger and larger than 1 when they are older. As $\partial\breve{\alpha}_1/\partial VR < 0$ and a sufficient condition for $\partial\breve{\alpha}_1/\ \partial\pi_1 < 0$ is that π_1 is not too small and $VR \leq 1$, estimation with measures pertaining to when the parents are "too old" should lead to underestimating the IGE of the expectation whereas measures pertaining to when they are "too young" may lead to over or underestimating it, depending on the values of VR and π_1. All this is, again, exactly as in the case of the GEiV model.

Crucially, estimation with short-run proxies for long-run parental income will be affected by attenuation bias when $\pi_1 \approx 1$. Indeed, when this equality holds exactly, $C_{\breve{\alpha}_1} = C_{\alpha_1}$ and $V_{\breve{\alpha}_1}$ becomes:

$$V_{\breve{\alpha}_1} = 2\,\{Var(\ln X)[1 + VR]\}^{-1}.$$

Therefore $\breve{\alpha}_1 < \alpha_1$ and there is attenuation bias, with the magnitude of the bias determined by the size of the variance ratio (see Online Appendix, C, for additional comments and for proofs for claims in the last four paragraphs).

The GEiVE model plays for the IGE of the expectation the same functions that the GEiV model plays for the conventional IGE. First, together with the "nesting equations," i.e., Equations (19) and (20), it identifies the various potential sources of bias in estimates of the IGE of the expectation with the PPML estimator and supplies an analytical framework to help investigate those biases. Second, it provides a methodological justification for estimating the IGE of the expectation with proxy variables that satisfy some conditions. With regards to attenuation bias, the GEiVE model suggests using averages of parental income or earnings over several years, rather than annual measures, so as to reduce $Var(\mathrm{P})$ as much as possible (thus providing an IGE-of-the-expectation counterpart to the approach most often used with the conventional IGE). To minimize lifecycle biases, the GEiVE model suggests using measures of children's and parents' economic status pertaining to ages in which $\theta_1 \approx 1$ and $\pi_1 \approx 1$, respectively.

5. Empirical analyses

5.1 Data and estimation

The empirical analyses of this section are based on two different PSID samples. The first sample is geared towards (a) studying the patterns and magnitudes of left-side lifecycle biases that result when estimation relies on short-run variables, and (b) testing the GEiVE model's account of those biases. These goals require obtaining IGE estimates based on measures of children's long-run income. In line with this requirement, this sample includes information on children born between 1952 and 1960 and still present in the PSID at age 56.[10] The observations in this sample

[10] The PSID collects income information referring to the calendar year prior to the survey. Unless I indicate otherwise, I always refer to "income years" rather than "survey years." The PSID

are children-years pertaining to when the children were 24-56 years old.[11] I use information on children's family income (as adults) and on sons' individual earnings in all years they are available, and on parents' average family income and age when the children were 13-20 years old. I construct approximate long-run measures by averaging the children's income and earnings information across all available years. [12]

The second sample is geared towards (a) studying the patterns and magnitudes of right-side lifecycle and attenuation biases that result when estimation relies on short-run variables, and (b) testing the GEiVE model's account of those biases. These goals require estimating the IGE of the expectation with long-run measures of parental income. In accordance with this requirement, the sample includes information on children born between 1966 and 1975, for which 25 years of parental data—when the children were between 1 and 25 years old—are available. The observations in this sample are children-years covering the periods in which the children from each cohort were 36-41 years old. I use information on children's family income (as adults) and sons' individual earnings, and on parents' average family income and age when the children were 1-25 years old. I use the latter to construct approximate measures of long-run parental income.

Table 1 presents descriptive statistics for both samples. Section D of the Online Appendix provides more details on the samples and variables, and explains the approach used to address the problems generated by the PSID's switch to biannual data collection in survey year 1997.

I estimate the IGE of the expectation with the PPML estimator and employ sampling

---

switched to biannual data collection in survey year 1997. For cohorts that became 56 years old in years in which data are not available, I require that the children be present in the PSID at age 55.

[11] The analyses exclude children with fewer than 12 years of income or earnings (as relevant).

[12] For the 1953 and 1952 cohorts, parental information refers to when the children were 14-20 and 15-20 years old, respectively, rather than 13-20 years old.

Table 1: Descriptive Statistics (weighted values)

	Sample I: Birth cohorts 1952-1960		Sample II: Birth cohorts 1966-1975	
	All	Sons	All	Sons
Child's gender (% female)	49.1		48.5	
Child's age				
Mean	40.4	40.5	38.5	38.5
Standard deviation	9.5	9.4	1.7	1.7
Child's family income				
Mean	99,885		105,010	
Standard deviation	117,067		94,038	
Child's earnings				
Mean		66,082		71,044
Standard deviation		80,360		70,434
Average parental age				
Mean	45.9	46	40.2	40.2
Standard deviation	6.8	7.0	6	5.8
Average parental income				
Mean	91,115	89,229	87,362	84,465
Standard deviation	64,915	53,471	64,666	55,043
Number of children-years	32,204	14,992	5,554	2,514
Number of children	1,059	498	941	428

Monetary values are in 2012 dollars (adjusted by inflation using the Consumer Price Index for Urban Consumers - Research Series). In the first sample, average parental income and age pertain to when the children were 13-20 years old for all cohorts but the 1952 and 1953 cohorts, for which they pertain to when the children were 15-20 and 14-20 years old, respectively. Children's family income and sons' earnings pertain to when they were 26-55 years old. In the second sample, average parental income and age pertain to when the children were 1-25 years old, while children's family income and sons' earnings pertain to when they were 36-41 years old.

weights and compute cluster-corrected robust standard errors in all cases. As the relationship between long-run and short-run measures of income and earnings varies with the age at measurement (both for parents and children), it's customary to include polynomials on children's and parents' ages as controls when estimation is based on short-run measures. However, as all IGE estimates based on such measures pertain to samples in which the variation in children's ages is quite small, controlling for their age is unnecessary. Mitnik et al. (2015:34) have argued that the age at which parents have their children is not exogenous to their income, that parents' age is causally relevant for their children's life chances, and that insofar as we want persistence measures to reflect the gross association between parental and children's income we should not control for parental age. Here I present estimates from models without such controls but estimates from models that do include them are very similar (see the Online Appendix, E).

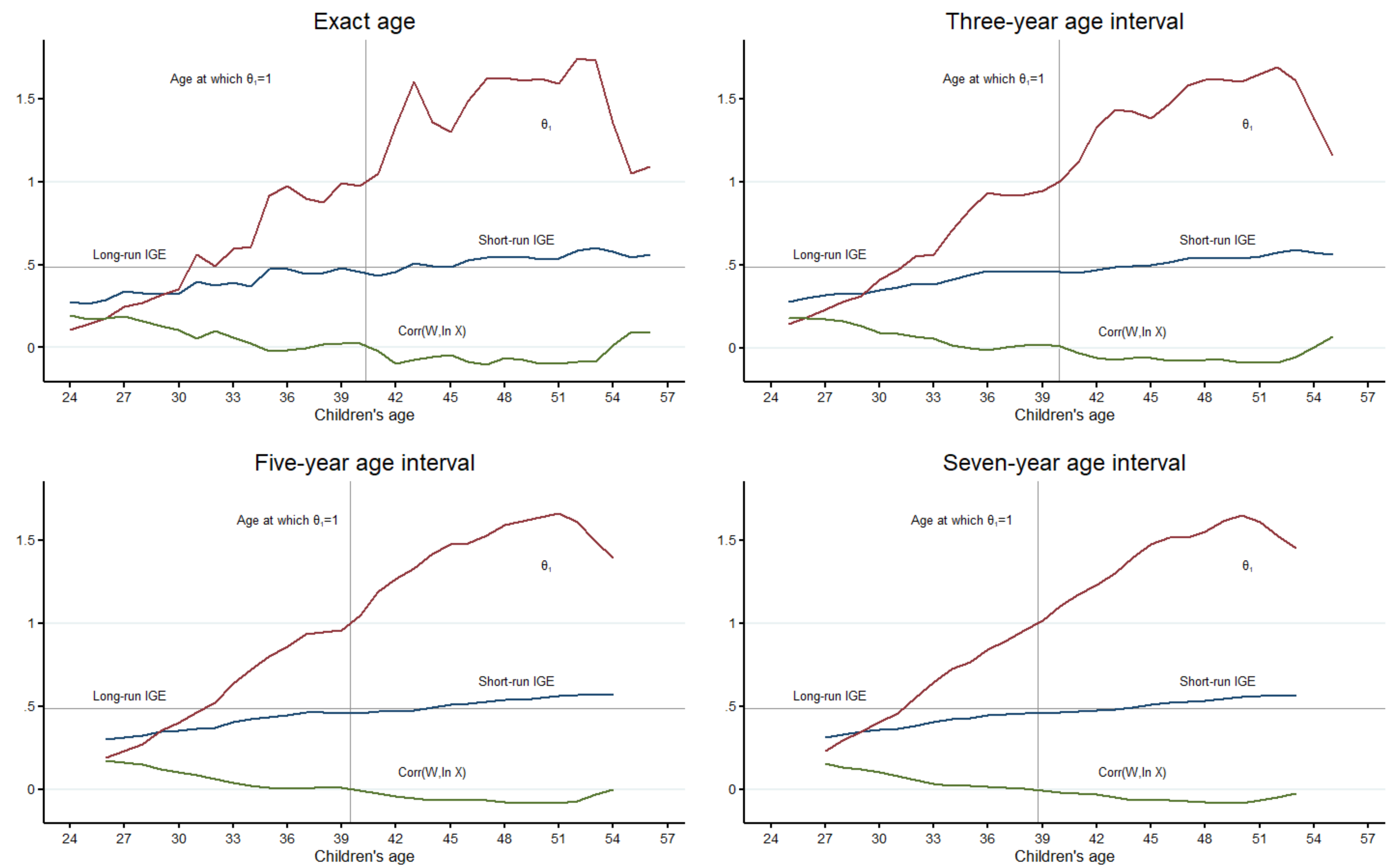
5.2. Left-side lifecycle bias

Figures 1 and 2 present evidence on left-side lifecycle biases, and its determinants, for the IGEs of expected family income and earnings (in the case of earnings I focus on sons, as is customary). In these figures, the long-run IGE is the IGE of the approximate long-run family income of children (Figure 1) or long-run earnings of sons (Figure 2) with respect to average parental income when the children were 13-20 years old. For the purposes of the analysis here, the latter income is assumed to be a measure of long-run parental income. Given that the parental measure is based on nine years of information, the estimated IGE values of 0.49 (family income) and 0.41 (earnings) are almost certainly affected by right-side attenuation bias (see evidence below). But given that I use the same measure of parental income to compute long-run and short-run IGEs (i.e., IGEs based on the long-run and the short-run measures of children's income or earnings, respectively), it's still possible to draw conclusions regarding left-side lifecycle bias.

The figures include four panels. In each of these panels, estimation is based on different sets of subsamples. In the upper-left panel of each figure, the subsamples only include children-income observations pertaining to the exact age indicated in the horizontal axis. In the other three panels estimates are based on progressively larger samples, which include observations pertaining to three-, five-, and seven-year intervals centered on the target age. As more observations are included in the subsamples, short-run point estimates are less affected by sampling variability and the age-profile of the left-side lifecycle bias becomes progressively clearer.

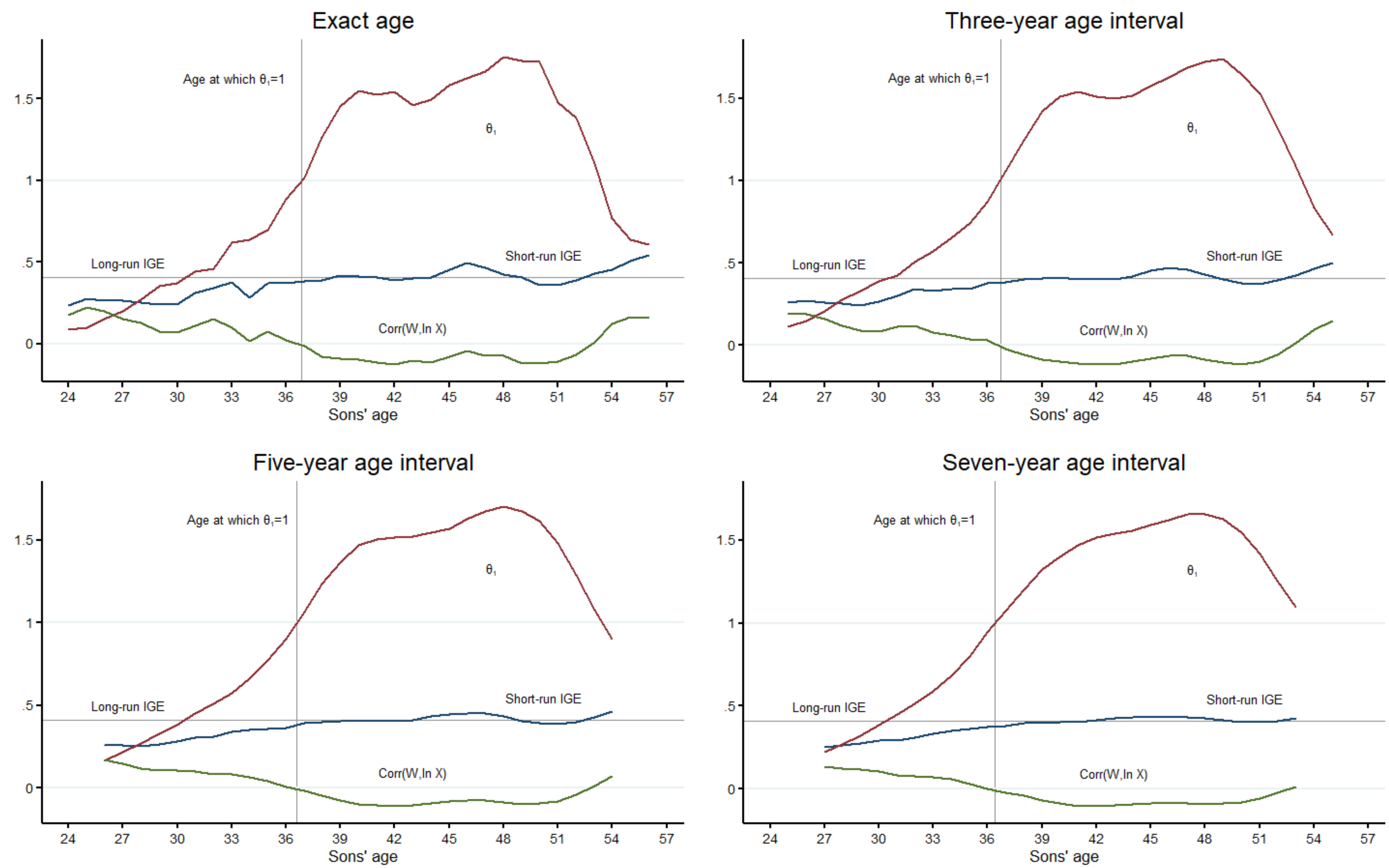
The middle curves in the four panels of Figures 1 and 2 show the relationship between

Figure 1: Left-side lifecycle bias in the estimation of the IGE of children's expected family income



The figure shows the relationship between the short- and long-run IGEs, θ_1, Corr(W, ln X), and children's ages. Parental income is parents' average income when their children were 13-21 years old. The darker-gray horizontal lines indicate the value of the long-run IGE (the IGE of the children's average family income at ages 24-56). The vertical lines indicate the ages at which $\theta_1 = 1$ (computed by interpolation). The four panels differ on whether exact ages or three different age intervals are used to select the observations in each estimation sample (the horizontal axis is either the exact age or the mid-point of the age interval).

Figure 2: Left-side lifecycle bias in the estimation of the IGE of sons' expected earnings



The figure shows the relationship between the short- and long-run IGEs, θ_1, Corr(W, ln X), and children's ages. Parental income is parents' average income when their children were 13-21 years old. The darker-gray horizontal lines indicate the value of the long-run IGE (the IGE of the sons' average earnings at ages 24-56). The vertical lines indicate the ages at which $\theta_1 = 1$ (computed by interpolation). The four panels differ on whether exact ages or three different age intervals are used to select the observations in each estimation sample (the horizontal axis is either the exact age or the mid-point of the age interval).

short-run IGEs and children's ages.[13] Estimates vary markedly with the age at which children's incomes and sons' earnings are observed, ranging from 0.27 to 0.60 in the case of family income and from 0.24 to 0.54 in the case of earnings; the values covered by each of these ranges entail very different levels of economic persistence. The shapes of the curves are consistent with expectations. For both income and earnings, short-run IGEs based on children's information from their mid-twenties greatly underestimate the long-run IGE, while the downward bias decreases steadily as the information is from closer to their late thirties. IGEs based on information pertaining to their late thirties and early forties appear nearly free of lifecycle bias, while those based on information past age 45 are affected by progressively larger upward biases (the latter is clear in the case of family income but less so with the much noisier earnings results, which are based on a sample that is only half as large); the biases at older ages, however, are comparatively smaller than those at younger ages, as evidenced by the fact that the curves are substantially steeper before age 36 than after age 45. Lastly, comparing the four panels within each figure indicates that, at least with small samples like those employed here, short-run estimates based on one year of children's information may be affected by relatively large upward or downward biases even when that information is obtained close to age 40, and that these biases can be reduced substantially, if not eliminated, by combining income information from a few years around that age.[14]

The top curve in each panel of Figures 1 and 2 shows how θ_1 changes with the children's age. The parameter increases with age up to their late forties or early fifties, at which point the curves turn downward. As expected, the values of θ_1 are much smaller than 1 when the children

[13] The estimates underlying the curves (and standard errors) as well as additional estimates from models with controls for parental age can be found in Tables E1 and E2 in the Online Appendix.

[14] A similar point was made by Nybom and Stuhler (2016:264) for the conventional IGE. They suggested averaging children's income information across years (within children), which can be expected to have the same effect as pooling years of information into one sample (as I do here).

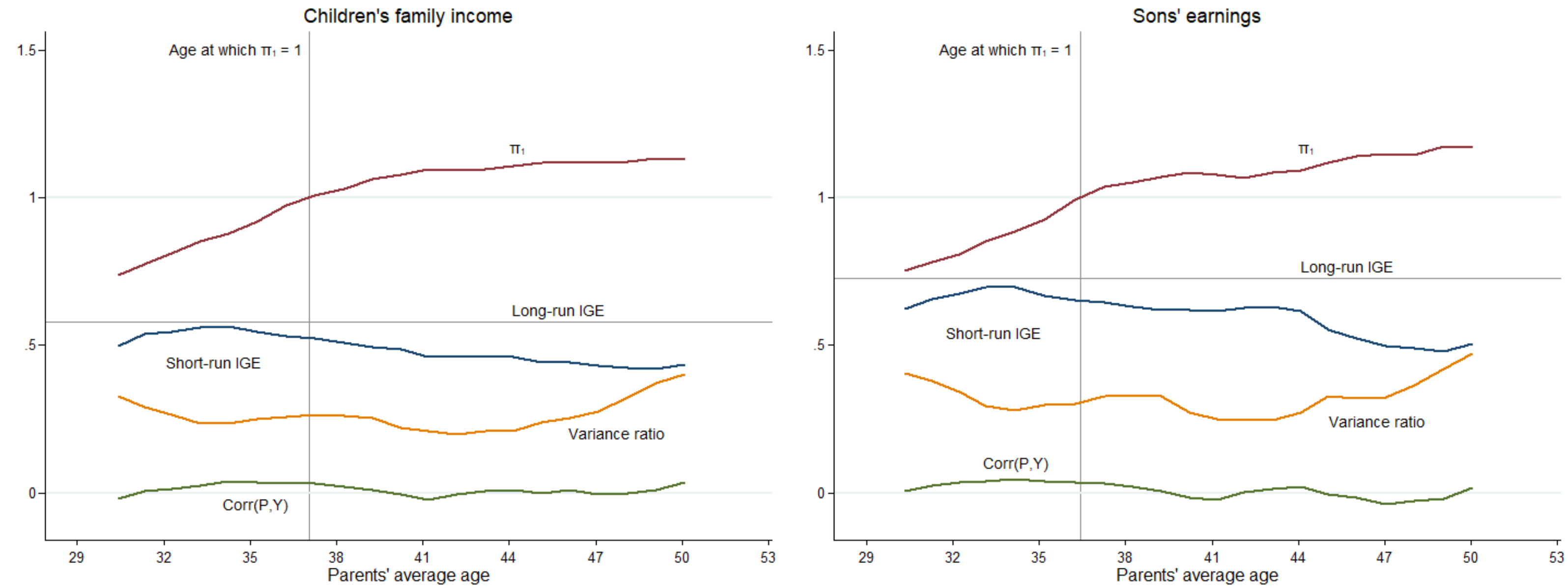
are younger and much larger than 1 when they are older (up to their early or mid-fifties). The parameter is equal to 1 close to ages 39-40 (family income) and 37 (earnings). The GEiVE model predicts that at $\theta_1 = 1$ the short-run IGE will be approximately equal to the long-run IGE, and that the former will tend to be smaller (larger) than the latter when $\theta_1 < 1$ ($\theta_1 > 1$). The prediction that the bulk of left-side lifecycle bias will disappear when $\theta_1 = 1$ is derived under the fallible assumption that $Cov(W, X) \approx 0$ at that value of θ_1. Moreover, Nybom and Stuhler's (2016) correlated-deviations argument also applies (mutatis mutandis) in the present context, thus providing a positive reason to doubt the appropriateness of that assumption. The assumption nevertheless holds with a high degree of approximation, as indicated by the bottom curves in all four panels of the figures; these curves show that when $\theta_1 = 1$, $Corr(\exp(W), X)$ is very close to zero in all cases.[15] As a result, the residual lifecycle biases when $\theta_1 = 1$ are in the -6 to -7 percent range for family income and in the -6 to -8 percent range for earnings (in both cases substantially less, in absolute value, than the residual bias of -20 percent reported by Nybom and Stuhler for the IGE of the geometric mean). The figures also show that when $\theta_1 < 1$ the short-run IGEs tend to underestimate the long-run IGE whereas the opposite happens when $\theta_1 > 1$, as expected. Importantly, short-run IGE estimates at ages in which θ_1 is somewhat larger or smaller than 1 also exhibit small lifecycle biases. This is important from a practical point of view, as it suggests that IGE estimates based on measures of children's income or earnings obtained close to age 40 will exhibit a small amount of bias even if θ_1 is not exactly 1 (of course, provided that the samples are large enough).

5.3. Right-side biases

Figure 3 presents evidence on right-side lifecycle bias and its determinants, for the IGE of children's family income (left panel) and sons' earnings (right panel). In this figure, the long-run IGEs are computed with children's income information for ages 36-41 and the approximate long-

[15] I substituted the correlation for the covariance for scaling purposes.

Figure 3: Right-side lifecycle bias in the estimation of the IGE of expected income



The figure shows the relationship between the short- and long-run IGEs, π_1, Corr(P,Y), the variance ratio, and the parents' age at the times their incomes were measured. The panel on the left pertains to the IGE of children's family income when they were 36-41 years old, while the panel on the right pertains to the IGE of sons' earnings at those ages. The horizontal axis is the average age of the parents in the sample. The vertical lines indicate the ages at which $\pi_1 = 1$ (computed by interpolation). The darker-gray horizontal lines indicate values of the long-run IGEs (as approximated by the IGEs computed with parents' information pertaining to when their children where 1-25 years old). Short-run IGEs are based on five years of parental information.

run parental income. The evidence discussed in the previous subsection provides a good reason for expecting these estimates to be unaffected by left-side lifecycle bias, even though they are based on short-run measures of children's income and earnings. In each panel, the second curve from the top shows the relationship between short-run IGEs, estimated with parental-income measures based on five years of information, and parental age.[16] The estimates vary significantly across ages, covering the ranges 0.42-0.56 (family income) and 0.48-0.70 (earnings). The location and shape of the short-run IGE curves are consistent with expectations. In the case of family income, the IGE curve first rises with age, up to age 34, and then decreases nearly monotonically up to age 50. In the case of earnings, the IGE estimates do not vary as much up to age 44 but fall very rapidly after that. Each curve is always below the long-run IGE line, reflecting the attenuation effect associated to using five years of parental information. Nevertheless, in the case of family income the bias is close to nil around age 34. This suggests that amplification bias—mentioned as a theoretical possibility when I presented the GEiVE model—may in fact occur at young ages if the parental income measure is based on more than five years of information. This is in fact the case (see Tables E3 and E4 in the Online Appendix), but the biases are quite small.

As expected, the π_1 curves increase with parental age while the variance-ratio curves exhibit an asymmetric-U shape. As $Corr(P, Y)$ is close to zero at all parental ages, the amount of lifecycle bias we observe at each age is mostly determined by π_1 and the variance ratio. Although the interaction between these two factors may lead to different shapes for the short-run IGE curve—as illustrated by the family-income and earnings IGE curves in Figure 3—the fact that π_1 and the variance ratio both increase when the parents are older suggests that, typically, IGE estimates will fall markedly when parental income pertains to older ages. The evidence in

[16] The estimates underlying the curves (and standard errors) as well as additional estimates generated with measures of parental income based on 1 to 25 years of information (at various parental ages) can be found in Tables A3 and A4 in the Online Appendix.

the figure suggests that this is indeed the case.

The GEiVE model entails that when $\pi_1 = 1$, the short-run IGEs will be necessarily affected by attenuation bias. This prediction is derived under the fallible assumption that $Cov(P,Y) \approx 0$ at that value of π_1. The prediction is confirmed in Figure 3, where $\pi_1 = 1$ close to age 37 for both income measures. However, this fact is not particularly informative, as all short-run estimates in the figure are affected by attenuation bias, regardless of parental age. As indicated above, this is not necessarily the case when measures based on more years of information are employed, which therefore provide a stronger test. Results included in Table 2, which are discussed below, show that IGE estimates are always affected by attenuation bias when $\pi_1 = 1$, regardless of the number of years

The standard approach utilized to assess the extent of attenuation bias that results from the use of short-run parental income, and to ascertain how many years of parental information are needed to eliminate the bulk of that bias, involves (a) estimating the IGE with parental-income measures based on progressively more years of information centered at age 40 (or some other fixed age), and (b) comparing these short-run estimates with the corresponding long-run estimate or, much more commonly, to an asymptotic IGE value extrapolated from the pattern of the short-run estimates.[17] This is a useful approach, but it has one limitation: As the age at which $\pi_1 = 1$ may be different from 40 and may vary with the number of years of information used to compute parental income, the measured attenuation bias most likely reflects not only the effect of the years of information but also a "lifecycle effect," i.e., the effect of π_1 being different from 1. An alternative approach, which isolates the effect of adding additional years of information and which can only be implemented when a long-run measure of parental income is available, is

[17] Reliance on an asymptotic value is implicit in arguments along the lines that the fact that the differences between IGE estimates based on *n*, *n* + 1 and *n* + 2 years of information are small and decreasing means that *n* + 2 years are enough to eliminate the bulk of attenuation bias.

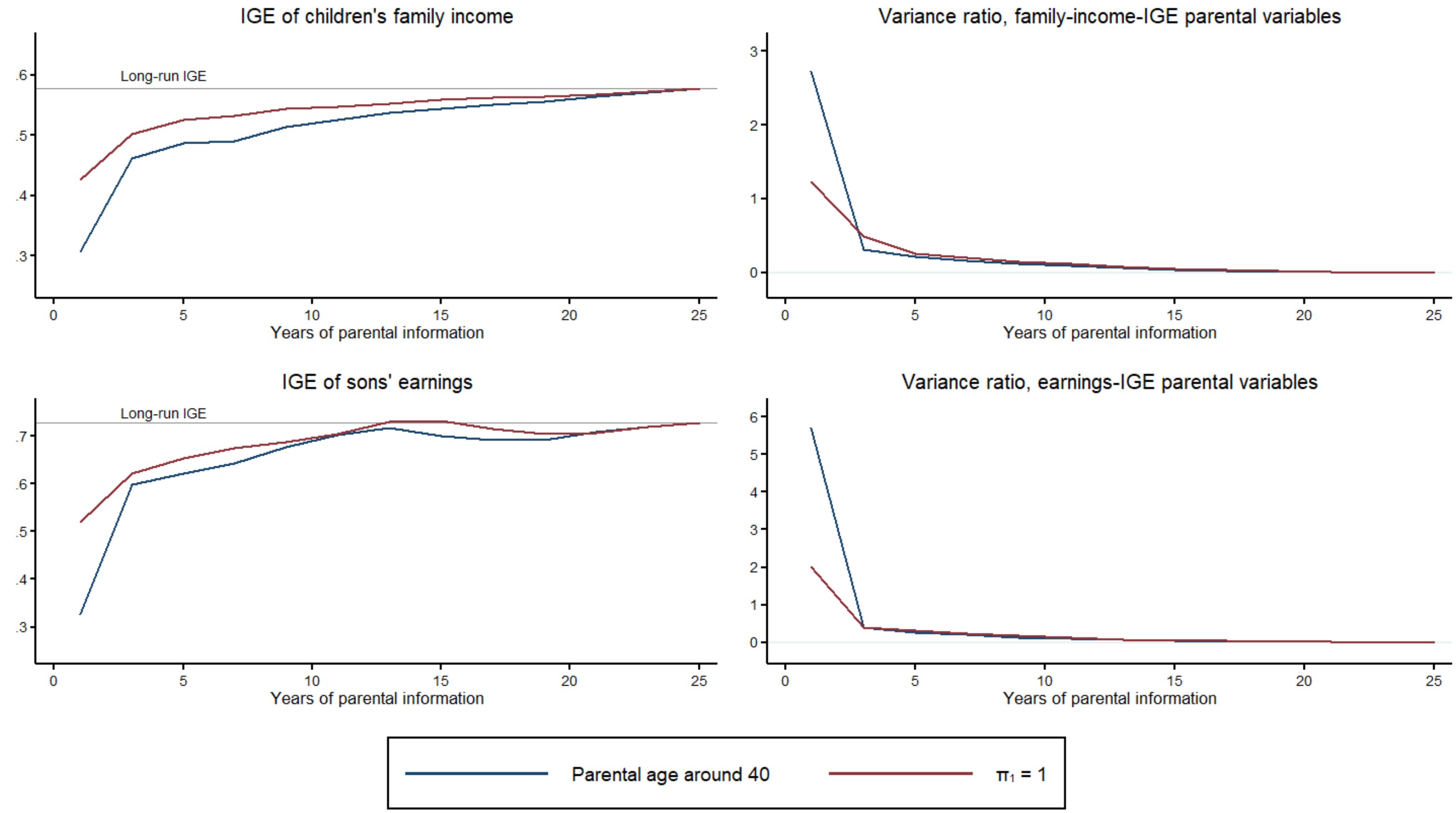
to keep π_1 fixed at 1 (rather than fixing the parents' age).

Figure 4 and Table 2, which present the results obtained under both approaches, provide clear evidence on the patterns and magnitudes of attenuation bias in the estimation of the IGE of the expectation. The panels on the left of the figure show how the estimates of the family-income IGE (top panel) and earnings IGE (bottom panel) change as more years of information are used, while the panels on the right show the corresponding values of the variance ratio. Table 2 displays the magnitudes of the attenuation biases, expressed as percentages, and the values of π_1 under the first approach and of parental age under the second.

With one year of parental information $Var(P)$ is larger than $Var(\ln X)$ under both approaches, and the variance ratios are very substantial in magnitude. Partly for this reason, the attenuation bias is nearly 47 percent (family income) and 55 percent (earnings) under the standard approach. However, these biases also reflect substantial lifecycle effects, as with one year of information the values of π_1 at age 40 are significantly larger than 1. The biases fall to about 26 percent (family income) and 28 percent (earnings) with the alternative approach (as π_1 is equal to 1 around age 35 rather than 40).

Assessing the magnitude of attenuation biases when parental income measures are based on three and on five years of information is important, as empirical research is quite often based on such measures. There is a very large fall in the variance ratios, and a substantial shrinkage of the values of π_1 towards 1 (of course, only under the standard approach), when three years of information are used instead of one year; moreover, the variance ratio is always substantially smaller than 1 if at least three years are employed, and decreases steadily as more years of information are added, under both approaches. The large fall in the variance ratio when moving from one to three years of information, together with the concomitant shift of π_1 towards 1 under the standard approach, translate into a large increase in IGE estimates (and is accompanied by a large reduction in the differences in results across approaches). Nevertheless, with three years of

Figure 4: Attenuation bias in the estimation of the IGE of expected income



The panels on the left of the figure show the relationship between IGEs estimated with short-run measures of parental income and the number of years of parental information used to compute those measures, for the IGEs of children's family income (top) and sons' earnings (bottom). The panels on the right of the figure show the relationship between the variance ratio and the years of parental information employed. Each of the four panels includes results from two approaches. The first fixes the age of parents across parental measures; here π_1 may be different from one, and its value may vary across parental measures. The second approach focuses on the IGE estimates that result when π_1 is exactly one; here parental age may vary across parental income measures. These estimates are obtained by interpolation. In the left panels the differences between the long-run IGE and the short-run IGEs estimated under the second approach reflect attenuation bias free of any lifecycle effects, as these are zero by construction. The short-run measures of parental income are based on odd numbers of years of information (1, 3, 5 . . .) and, in the first approach, are centered around the year the children were 13 years old (when mean parental age is about 40). Children's family income refers to when they were 36-41 years old. The (approximated) long-run IGEs are computed with parents' information pertaining to when their children were 1-25 years old.

Table 2: Attenuation biases and related quantities

Years of parental information	Parental age fixed at 40				π_1 fixed at 1			
	Children's family income		Sons' earnings		Children's family income		Sons' earnings	
	Attenuation bias (%)	π_1	Attenuation bias	π_1	Attenuation bias (%)	Parental age	Attenuation bias	Parental age
1	46.8	1.18	55.0	1.28	26.1	35.1	28.5	34.8
3	20.1	1.08	17.8	1.08	13.2	36.5	14.4	36.1
5	15.6	1.08	14.5	1.08	9.0	37.1	10.3	36.4
7	15.2	1.08	11.7	1.07	7.8	37.5	7.2	37.2
9	11.2	1.06	6.9	1.05	5.8	37.8	5.3	37.8
11	9.1	1.05	3.4	1.04	5.3	37.9	3.1	38.3
13	6.9	1.04	1.4	1.03	4.3	38.4	-0.3	39.1
15	5.9	1.03	3.7	1.02	3.2	38.7	-0.8	39.4
17	4.7	1.03	4.9	1.02	2.5	38.8	1.6	39.3
19	3.7	1.02	4.9	1.02	2.3	39.0	3.0	39.3
21	2.2	1.01	2.7	1.01	1.8	39.3	2.7	39.7
23	1.1	1.01	1.1	1.01	0.9	39.8	1.0	39.9

The table shows attenuation biases, and related quantities, as a function of the number of years used to compute short-run parental measures. Attenuation bias is computed as one minus the ratio between the short-run IGE, estimated with the number of years of parental information indicated in the first column, and the corresponding long-run IGE, and is expressed as a percentage. See the note to Figure 4 for the difference between the two approaches used (parental age fixed at 40 and π_1 fixed at 1), and for additional information on the parental income measures.

information the attenuation bias is still about 19 percent under the standard approach and about 14 percent under the alternative approach. With five years of information, attenuation bias is about 15 percent with the former approach and about 10 percent with the latter approach.

As indicated earlier, the issue of how many years of parental information are needed to eliminate the bulk of attenuation bias has been a central concern in the literature (e.g., Mazumder 2005; 2016). Stipulating that this happens when attenuation bias approaches five percent, the results of Table 2 suggests that, in the case of the IGE of the expectation, achieving this goal with survey data requires using about 13 years of parental information.

It is very likely, however, that fewer years of information are needed with tax and other administrative data, as Chetty et al. (2014) claimed to be the case for the conventional IGE. One reason is that administrative data may be less affected by measurement error (i.e., $Var(P)$ may be smaller with these data). Another reason, which does not seem to have been discussed previously in the literature, is that tax-based data cover much better than survey data the right tail of income distributions (Fixler and Johnson 2014). Therefore, we can safely expect $Var(\ln X)$ to be significantly larger with the former than with the latter data.

6. Conclusions

The IGE estimated in the mobility literature has been misinterpreted as pertaining to the expectation of children's income when it actually pertains to its geometric mean. Moreover, the (implicit) reliance on the geometric mean to index conditional income distributions makes studying gender and marriage dynamics in intergenerational processes a very difficult enterprise and leads to IGE estimates affected by substantial selection biases. For these reasons, Mitnik and Grusky (2019) have called for replacing the conventional IGE by the IGE of the expectation. To make this possible, mobility scholars need to have available a measurement-error model that plays for the estimation of the IGE of the expectation the role that Haider and Solon's (2006) GEIV model has played for the estimation of the conventional IGE.

After deriving an approximate closed-form expression for the probability limit of the PPML estimator—something that may be useful beyond the context of the current article—I have advanced here the needed formal model, that is, a generalized error-in-variables model for the estimation of the IGE of the expectation with short-run income variables. The GEIVE model provides a joint analysis of the biases that may affect estimation of the IGE of the expectation with short-run income measures and supplies a methodological justification for the estimation of that IGE with the PPML estimator and proxy variables that satisfy some conditions.

The results of the empirical analyses with PSID data offer strong support for the account of lifecycle and attenuation biases provided by the GEiVE model. They also indicate that the strategy most commonly employed to estimate the conventional IGE by OLS—using a multiyear measure of parental income centered around age 40 and a measure of children's income obtained around that age—also works well for estimating the IGE of the expectation with the PPML estimator. Further, the results suggests that—at least with survey data, and similarly to what has been reported for the conventional IGE (Mazumder 2016: Tables 1 and 2)—estimates of the IGE of the expectation with parental measures based on three to five years of information—that is, the measures often used by mobility scholars—would be affected by substantial attenuation biases. In fact, it appears that at least 13 years of data are needed to eliminate the bulk of that bias.

Our methodological knowledge regarding the estimation of the IGE of the expectation with short-run proxy measures needs to be developed further, and in several directions. Nevertheless, it seems appropriate to conclude that the measurement-error model and the associated evidence presented in this article eliminate the main obstacle for making the IGE of the expectation the workhorse elasticity of the economic mobility field, which would in turn dissolve the selection-bias problem and greatly stimulate the study of gender and marriage dynamics in intergenerational processes.

References

Aaronson, David and Bhashkar Mazumder. 2008. "Intergenerational Economic Mobility in the US: 1940 to 2000." *Journal of Human Resources* 43(1): 139-172.

Björklund, Anders and Markus Jäntti. 2000. "Intergenerational Mobility of Socio-Economic Status in Comparative Perspective." *Nordic Journal of Political Economy* 26: 3-32.

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

Black, Sandra and Paul Devereux. 2011. "Recent Developments in Intergenerational Mobility." *Handbook of Labor Economics*, Vol. 4b, edited by David Card and Orley Ashenfelter. Amsterdam: Elsevier.

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

Bloome, Deirdre and Western. 2011. "Cohort Change and Racial Differences in Educational and Income Mobility." *Social Forces* 90(2): 375-395

Bratsberg, Bernt, Knut Røed, Oddbjørn Raaum, Robin Naylor, Markus Jäntti, Tor Eriksson and Eva Österbacka. 2007. "Nonlinearities in Intergenerational Earnings Mobility. Consequences for Cross-Country Comparisons." *The Economic Journal* 117: C72-C92.

Carroll, Raymond, David Ruppert, Leonard Stefanski, and Ciprian Crainiceanu. 2006. *Measurement Error in Nonlinear Models: A Modern Perspective*. Boca Raton: Chapman.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014. "Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States." *The Quarterly Journal of Economics* 129(4): 1553-1623.

Chadwick, Laura and Gary Solon. 2002. "Intergenerational Income Mobility among Daughters." *The American Economic Review* 92(1): 335-344.

Corak, Miles. 2006. "Do Poor Children Become Poor Adults? Lessons from a Cross Country Comparison of Generational Earnings Mobility." IZA Discussion Paper No. 1993.

Corak, Miles. 2013. "Income Inequality, Equality of Opportunity, and Intergenerational Mobility." *Journal of Economic Perspectives* 27 (3):79-102.

Fixler, Dennis and David Johnson. 2014. "Accounting for the Distribution of Income in the U.S. National Accounts." In *Measuring Economic Sustainability and Progress*, edited by Dale

Jorgenson, J. Steven Landefeld, and Paul Schreyer. Chicago: Chicago University Press.

Gourieroux, C., A. Monfort and A. Trognon. 1984. “Pseudo Maximum Likelihood Methods: Theory.” *Econometrica* 52(3): 681-700.

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

Hertz, Tom. 2005. “Rags, Riches and Race: The Intergenerational Economic Mobility of Black and White Families in the United States.” In *Unequal Chances. Family Background and Economic Success*, edited by Samuel Bowles, Herbert Gintis, and Melissa Osborne Groves. New York, Princeton and Oxford: Russell Sage and Princeton University Press.

Hertz, Tom. 2007. “Trends in the Intergenerational Elasticity of Family Income in the United States.” *Industrial Relations* 46(1): 22-50.

Ichino, Andrea, Loukas Karabarbounis, and Enrico Moretti. 2011. “The Political Economy of Intergenerational Income Mobility." *Economic Inquiry* 49(1): 47-69.

Landersø, Rasmus and James Heckman. 2017. “The Scandinavian Fantasy: The Sources of Intergenerational Mobility in Denmark and the U.S.” *The Scandinavian Journal of Economics* 119(1): 178-230.

Mazumder, Bhashkar. 2001. “The Miss-measurement of Permanent Earnings: New Evidence from Social Security Earnings Data.” Federal Reserve Bank of Chicago Working Paper 2001-24.

Mazumder, Bhashkar. 2005. “Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data.” *The Review of Economics and Statistics* 87(2): 235-255.

Mazumder, Bhashkar. 2016. “Estimating the Intergenerational Elasticity and Rank Association in the United States: Overcoming the Current Limitation of Tax Data.” In *Inequality: Causes and Consequences*, edited by Lorenzo Cappellari, Solomon Polacheck, and Konstantinos Tatsiramos. Bingley: Emerald.

Mayer, Susan E., and Leonard Lopoo. 2008. “Government Spending and Intergenerational Mobility.” *Journal of Public Economics* 92: 139-58.

Mitnik, Pablo, Victoria Bryant, Michael Weber and David Grusky. 2015. “New Estimates of Intergenerational Mobility Using Administrative Data.” Statistics of Income Division, Internal Revenue Service.

Mitnik, Pablo and David Grusky. 2019. “The Intergenerational Elasticity of What? The Case for Redefining the Workhorse Measure of Economic Mobility.” *Sociological Methodology* 49(1). [This forthcoming article is available as a working paper released by The University of Chicago’s HCEO Global Working Group, www.hceconomics.org].

Mitnik, Pablo, Victoria Bryant, and Michael Weber. 2019 “The Intergenerational Transmission of Family-Income Advantages in the United States.” *Sociological Science* 6(15): 380-415.

Mitnik, Pablo, Victoria Bryant, and David Grusky. 2018. “A Very Uneven Playing Field: Economic Mobility in the United States.” Stanford Center on Poverty and Inequality Working Paper.

Nybom, Martin and Jan Stuhler. 2016. “Heterogeneous Income Profiles and Life-Cycle Bias in Intergenerational Mobility Estimation.” *The Journal of Human Resources* 15(1): 239-268.

Petersen, Trond. 2017. “Multiplicative Models for Continuous Dependent Variables: Estimation on Unlogged versus Logged Form.” *Sociological Methodology* 47:113-164.

Santos Silva, J. M. C. and Silvana Tenreyro. 2006. "The Log of Gravity." *The Review of Economics and Statistics* 88(4): 641-658.

Santos Silva, J. M. C. and Silvana Tenreyro. 2011. “Further Simulation Evidence on the Performance of the Poisson Pseudo-maximum Likelihood Estimator.” *Economics Letters* 112: 220-222.

Schennach, Susanne. 2016. “Recent Advances in the Measurement Error Literature.” *Annual Review of Economics* 8: 341-377.

Solon, Gary. 1992. “Intergenerational Income Mobility in the United States.” *American Economic Review* 82(3): 393-408.

Solon, Gary. 1999. “Intergenerational Mobility in the Labor Market.” *Handbook of Labor Economics*, Vol. 3A, edited by Orley C. Ashenfelter and David Card. Amsterdam: Elsevier.

Solon, Gary. 2004. “A Model of Intergenerational Mobility Variation over Time and Place.” In *Generational Income Mobility in North America in Europa*, edited by Miles Corak. Cambridge: Cambridge University Press.

Wooldridge, Jeffrey. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, Mass: The MIT Press.