INTERGENERATIONAL INCOME ELASTICITIES, INSTRUMENTAL VARIABLE ESTIMATION, AND BRACKETING STRATEGIES

Pablo A. Mitnik
(pmitnik@stanford.edu)

March, 2019

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

Abstract

The fact that the intergenerational income elasticity (IGE)—the workhorse measure of economic mobility—is defined in terms of the geometric mean of children's income generates serious methodological problems. This has led to a call to replace it by the IGE of the expectation, which requires developing the methodological knowledge necessary to estimate the latter with short-run measures of income. This article contributes to this aim. It advances a "bracketing strategy" for the set estimation of the IGE of the expectation that is equivalent to that used to set estimate (rather than point estimate) the conventional IGE with estimates obtained with the Ordinary Least Squares and Instrumental Variable (IV) estimators. The proposed bracketing strategy couples estimates generated with the Poisson Pseudo Maximum Likelihood estimator and a Generalized Method of Moments IV estimator of the Poisson or exponential regression model. To achieve its goal, the article develops a generalized error-in-variables model for the IV estimation of the IGE of the expectation, and compares it to the corresponding model underlying the IV estimation of the conventional IGE. By considering both bracketing strategies from the perspective of the partial-identification approach to inference, the article also specifies how to construct confidence intervals for the IGEs, in particular when the upper bound is estimated more than once with different sets of instruments. Lastly, using data from the Panel Study of Income Dynamics, the article shows that the bracketing strategies work as expected, and assesses the information they generate and how this information varies across instruments and short-run measures of parental income. Three computer programs made available as companions to the article make the set estimation of IGEs, and statistical inference, very simple endeavors.

Acknowledgments. Ivan Canay and Adam Rosen provided advice on inference in partial-identification contexts.

Introduction

The research on economic mobility across generations conducted in the last four decades has relied heavily on the intergenerational income elasticity (IGE). The IGE has been very extensively estimated, both to assess the levels of income and earnings mobility within countries (for reviews, see Solon [1999: 1778-1784]; Corak [2006]; Mitnik et al. [2018: 9-18]) and to conduct comparative analyses of income and earnings mobility across geographic areas, demographic groups and time periods (e.g., Björklund and Jäntti 2000; Chadwick and Solon 2002; Hertz 2005, 2007; Aaronson and Mazumder 2008; Mayer and Lopoo 2008; Bloome and Western 2011). More recently, the IGE has also been used to examine the relationship between cross-sectional economic inequality and mobility across generations (e.g., Bloome 2015) and to study the impact of social policies and political institutions on inequality of opportunity (e.g., Bratsberg et al. 2007; Landersø and Heckman 2016).

The IGE is defined in terms of the long-run incomes or earnings of parents and children but is almost always estimated with short-run proxy measures of income or earnings, that is, with variables affected by substantial measurement error. For this reason, the mobility field has been centrally concerned with the biases that may result and the methodological strategies that may be used to avoid them. Two central achievements in this regard are (a) the generalized error-in-variables (GEiV) model for the estimation of the IGE by Ordinary Least Squares (OLS) (Haider and Solon 2006; see also Nybom and Stuhler 2016), and (b) the analysis of the Instrumental Variable (IV) estimation of the IGE with invalid instruments (Solon 1992:Appendix; see also Nybom and Stuhler 2011:12-14). As I explain in detail later, these methodological analyses entail that if various empirical assumptions hold, then the probability limit of the OLS estimator of the IGE is affected by attenuation bias while the probability limit of the IV estimator is affected by amplification bias and, as Solon (1992:400) first argued, "the probability limits of the two estimators bracket the true value" of the IGE. This means that OLS and IV estimates of the IGE can be combined to "bracket" or, to use the terminology typically employed in more recent statistical and econometric literatures (e.g., Manski 2003; Manski 2008), to set estimate, rather than point estimate, the IGE. In other words, instead of producing a single IGE value as estimate, a "bracketing strategy" produces a set or range of values deemed to contain the IGE.

Although the IGE has long been the workhorse measure of intergenerational economic (i.e., earnings, income) mobility, Mitnik and Grusky (2017) have recently shown that this elasticity has been misinterpreted. Indeed, 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 oft-invoked interpretation as a measure of regression to the (arithmetic) mean. However, it pertains to the conditional *geometric mean* of the children's income.[1] This not only makes all conventional interpretations of the IGE invalid but also generates serious methodological problems.

At their root, these problems are the result of a very simple fact, i.e., that the geometric mean is undefined for variables including zero in their support. As Mitnik and Grusky (2017) have shown, this (a) makes it impossible to determine the extent to which parental income advantage is transmitted through the labor market among women (as many women have zero earnings), and (b) greatly hinders research on the role that marriage plays in generating the observed levels of intergenerational persistence in family income (as many people remain single or have nonworking spouses, and therefore cannot be included in analyses examining the relationship between people's parental income and the income contributed by their spouses).

As a result, the study of gender and marriage dynamics in intergenerational processes has been badly hampered.

Equally important, Mitnik and Grusky (2017) have shown that, as a consequence of mobility scholars' expedient of dropping children with zero earnings from samples (to address what is perceived as the problem of the logarithm of zero being undefined), estimation of earnings IGEs with short-run proxy earnings measures is almost certainly affected by substantial selection biases. This makes the widespread use of the IGE of men's individual earnings as an index of economic persistence and mobility in a country a rather problematic practice.

Mitnik and Grusky (2017) have argued that these conceptual and methodological problems can be solved in a straightforward manner by simply replacing the IGE of the geometric mean—the de facto estimated IGE—by the IGE of the expectation—the IGE that mobility scholars thought they were estimating—as the workhorse intergenerational elasticity. They have also called for effectuating such replacement. This requires, however, that the methodological knowledge necessary to estimate the IGE of the expectation with short-run income variables is made available.

Mitnik (2017) contributed to this goal by advancing a generalized error-in-variables model for the estimation of the IGE of the expectation with the Poisson Pseudo Maximum Likelihood (PPML) estimator (Santos Silva and Tenreyro 2006). Here I take on the complementary task of developing a generalized error-in-variables model for the IV estimation of that IGE. For reasons I discuss later, among various possible estimators I focus on the additive-error version of the Generalized Method of Moments (GMM) IV estimator of the Poisson or exponential regression model (Mullahy 1997; Windmeijer and Santos Silva 1997). I show that, under empirical assumptions fully equivalent to those made for the IV estimation of the conventional IGE, we can expect that GMM IV estimator to produce upward-biased estimates of the IGE of the expectation. By combining this result with Mitnik's (2017) result that the PPML estimation of the IGE of the expectation with short-run income measures is affected by attenuation bias, I further show that a bracketing strategy equivalent to that used with the conventional IGE can be employed to set estimate the IGE of the expectation. This strategy, and the generalized error-in-variables model for IV estimation on which it relies (which is of clear independent interest), are the first two contributions of this article.

In spite of the fact that the bracketing strategies generate set rather than point estimates of IGEs, it is nevertheless possible to construct confidence intervals for those IGEs. This can be achieved by considering the bracketing strategies from the perspective of the partial-identification approach to inference, a very active field of statistical and econometric research in recent times (see Tamer [2010] and Canay and Shaikh [2017] for reviews). I explain how to construct confidence intervals for the partially-identified IGEs—which are the confidence intervals of interest—and how they differ from confidence intervals for the "identified sets" (the ranges of values that lie between the probability limits of the bound or "bracketing" estimators). I address, in particular, the issue of how to construct confidence intervals for the IGEs when the upper bound is estimated repeatedly with different sets of instruments. The third contribution of the article is therefore to show how one central aspect of statistical inference may still be carried out when using the bracketing strategies to set estimate IGEs, and to connect the literature on IGEs to the bourgeoning literature on partial identification and set estimation (e.g., Manski 2008; Ho and Rosen 2017).

Although this does not seem to have been done before, when datasets with the necessary information are available it is possible to empirically evaluate whether a bracketing strategy works as expected. Here I rely on a U.S. sample from the Panel Study of Income Dynamics (PSID) to examine the performance of both bracketing strategies. Using various instruments, I show that the set estimates produced with the bracketing estimators do bound the corresponding long-run IGE estimates when they are expected do so. I also assess how the information supplied by the set estimates varies across instruments and short-run measures of parental income, and how informative the set estimates based on the minimum estimates of upper bounds across instruments are. These empirical analyses are the fourth contribution of the article.

The set estimation of IGEs—both the conventional IGE and the IGE of the expectation—can be carried out very easily with the statistical packages social scientists typically use. However, the computation of confidence intervals for the partially identified IGEs involves some difficulties, especially when it needs to account for the repeated estimation of an upper bound with different sets of instruments. For this reason, I have made available with this article three Stata programs that make the set estimation of both IGEs, and the computation of confidence intervals, a very simple endeavor. This is the fifth contribution of the article.

The structure of the rest of the article is as follows. I first explain why the conventional IGE pertains to the conditional geometric mean of children's income rather than to its conditional expectation, present the generalized error-in-variables models relevant for the set estimation of the conventional IGE, and fully specify the empirical conditions under which the bracketing strategy previously discussed in the literature works. Next, I introduce the IGE of the expectation (and discuss whether we need it in the toolbox of mobility scholars given the increasing popularity of the rank-rank slope, a different alternative to the conventional IGE), present the generalized error-in-variables model for its estimation with the PPML estimator advanced by Mitnik (2017), develop the new generalized error-in-variables model for its IV estimation contributed by this article, and fully specify the empirical conditions under which the new bracketing strategy I propose here works. After that I discuss statistical inference under the partial-identification approach and present the results of the empirical analyses. The last section draws the article's main conclusions.

The conventional IGE and its interpretation

The standard population regression function (PRF) posited in the mobility literature, which assumes the IGE is constant across levels of parental income, is:

$$E(\ln Y\,|x) = \beta_0 + \beta_1 \ln x, \qquad [1]$$

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 "$Z|W = w$." As already indicated, Mitnik and Grusky (2017) have shown that this conventionally estimated IGE has been misinterpreted. While mobility scholars have interpreted it as the elasticity of the expectation of children's income or earnings conditional on parental income, it pertains in fact to the conditional geometric mean.

Indeed, the parameter β_1 is not, in the general case, the elasticity of the conditional expectation of the child's income. This would hold as a general result only if $E(\ln Y|x) = \ln E(Y|x)$. But, due to Jensen's inequality, the latter is not the case. Instead, as $E(\ln Y\,|x) = \ln \exp E(\ln Y\,|x)$, and $GM(Y|x) = \exp E(\ln Y\,|x)$, Equation [1] is equivalent to

$$\ln GM(Y|x) = \beta_0 + \beta_1 \ln x, \qquad [2]$$

where GM denotes the geometric mean operator. Therefore, β_1 is the elasticity of the conditional geometric mean, i.e., 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.[2]

Set estimation of the IGE of the geometric mean

OLS estimation and the GEiV model

Estimation of Equation [1] by OLS, after substituting short-run income variables for the long-run variables, 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, using proxy measures taken when parents or children are too young or too old to represent lifetime differences well results in lifecycle biases (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 (see, 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).

In order to introduce the empirical assumptions of the GEiV model, it is necessary to first introduce the following population linear projections:

$$\ln Z_t = \lambda_{0t} + \lambda_{1t} \ln Y + V_t \qquad [3]$$

$$\ln S_k = \eta_{0k} + \eta_{1k} \ln X + Q_k, \qquad [4]$$

where $Z_t > 0$ is children's income at age *t*; $Y > 0$; $\lambda_{0t} + V_t$ is the 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 one and varies with *t*; $S_k > 0$ is parents' income at age *k*; $X > 0$; $\eta_{0k} + Q_k$ is the 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 one and varies with parents' age.[3]

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

$$Cov(\ln X, V_t) = 0 \qquad [5]$$

$$Cov(\ln Y, Q_k) = 0 \qquad [6]$$

$$Cov(V_t, Q_k) = 0. \qquad [7]$$

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*.)

In the general case, the probability limit of the "short-run OLS estimator" of the conventional IGE (i.e., the OLS estimator with the long-run income variables replaced by short-run variables), denoted by $\tilde{\beta}_1$, can be obtained by substituting Equations [3] and [4] in $\tilde{\beta}_1 \equiv \frac{Cov(\ln Z, \ln S)}{Var(\ln S)}$. Using that Equation [4] is a linear projection and therefore $Cov(\ln X, Q) = 0$, that probability limit is:

$$\tilde{\beta}_1 = \beta_1 \frac{\lambda_1\, \eta_1}{[\eta_1]^2 + VR} + \frac{\lambda_1 Cov(\ln Y, Q) + \eta_1 Cov(\ln X, V) + Cov(V, Q)}{Var(\ln X)[(\eta_1)^2 + VR]}, \qquad [8]$$

where $VR = \frac{Var(Q)}{Var(\ln X)}$ is what I will refer as the "variance ratio."

If the three empirical assumptions of the GEiV model hold, then Equation [8] reduces to:

$$\tilde{\beta}_1 = \beta_1 \frac{\lambda_1\, \eta_1}{[\eta_1]^2 + VR}.$$

If, in addition, both children's and parents' incomes are measured at the "right points" of their lifecycles, that is, if $\lambda_1 = \eta_1 = 1$, then $\tilde{\beta}_1 = \beta_1 \frac{1}{1 + VR}$ and the estimates obtained with the short-run OLS estimator can be expected to be affected by attenuation bias. The bias, however, falls with $Var(Q)$, which in turn depends on the exact short-run measure of parental income that is used.

The GEiV model supplies a methodological justification for the estimation of the conventional IGE by OLS with proxy variables that satisfy some conditions (Nybon and Stuhler 2016). Indeed, the GEiV model suggests that using measures of economic status (i.e., income, earnings) obtained at specific ages should eliminate the bulk of lifecycle bias—while the available evidence 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, rather than a single-year measure, as the proxy measure *S*. There is strong evidence that the bias can be substantially reduced this way, although 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. 2018:9-14; Mazumder 2016).

IV estimation and the GEiV-IV model

Most often mobility scholars have much fewer years of information available than what most believe are needed to nearly eliminate attenuation bias. Therefore, a natural alternative is to forgo estimation by OLS and, instead, address right-side measurement error by resorting to an IV estimator of β_1. Here, variables like parental education or occupational status are used as instruments for the error-ridden proxy measure of long-run parental income (e.g., Ng 2007; Mulligan 1997; Zimmerman 1992). The main concern with this strategy has been, however, that the instruments typically available are most likely endogenous. In this context, IV estimates may still be useful if the sign of their asymptotic bias can be established. An analysis by Solon (1992: Appendix)—carried out under the assumption that the measurement errors are classical—achieved that. I present next a more general version of this analysis that allows for lifecycle effects, to which I refer as the generalized error-in-variables model for the IV estimation of the conventional IGE, or GEiV-IV model. I also nest this model within a more general expression for the probability limit of the IV estimator of that IGE.[4]

Let L (e.g., parental years of education) be the instrument used for the IV estimation of β_1, and

$$\ln Y = \varrho_0 + \varrho_1 \ln X + \varrho_2 L + \kappa \qquad [9]$$

a population linear projection of $\ln Y$ on $\ln X$ and L. The GEiV-IV model makes the following empirical assumptions. For any *t* and *k*:

$$Cov(V_t, L) = 0 \qquad [10]$$

$$Cov(Q_k, L) = 0, \qquad [11]$$

at least when $\lambda_1 \approx \eta_1 \approx 1$.

I show in Online Appendix A that in the general case the probability limit of the IV estimator of the conventional IGE, $\ddot{\beta}_1$, may be written as:

$$\ddot{\beta}_1 = \frac{\lambda_1}{\eta_1}\left\{\beta_1\left[1 - \frac{Cov(Q,L)}{Cov(\ln S, L)}\right] + \varrho_2 \frac{SD(L)}{SD(\ln X)}\left[\frac{1 - [Corr(\ln X, L)]^2}{Corr(\ln X, L) + \frac{Cov(Q,L)}{\eta_1 SD(\ln X)\, SD(L)}}\right]\right\}$$
$$+ \frac{Cov(V,L)}{Cov(\ln S, L)}, \qquad [12]$$

where SD denotes the standard deviation operator. Therefore, taking for granted that $0 < |Corr(\ln X, L)| < 1$, Equation [12] shows that IV estimation is consistent when the following conditions are met: (a) L is uncorrelated with the measurement errors, that is, $Cov(V, L) = Cov(Q, L) = 0$, (b) both children's and parents' incomes are measured at the right points of their lifecycles, that is, $\lambda_1 = \eta_1 = 1$, and (c) L is a valid instrument, that is, $\varrho_2 = 0$. If, however, the instrument is invalid but Equations [10] and [11] do hold, the probability limit of the IV estimator is:

$$\ddot{\beta}_1 = \frac{\lambda_1}{\eta_1}\left\{\beta_1 + \varrho_2 \frac{SD(L)}{SD(\ln X)}\left[\frac{1-[Corr(\ln X\,, L)]^2}{Corr(\ln X\,, L)}\right]\right\}. \qquad [13]$$

If, in addition, $\lambda_1 = \eta_1 = 1$, then Equation [12] reduces to Solon's (1992) widely cited result:

$$\ddot{\beta}_1 = \beta_1 + \varrho_2 \frac{SD(L)}{SD(\ln X)}\left[\frac{1-[Corr(\ln X\,, L)]^2}{Corr(\ln X\,, L)}\right]. \qquad [14]$$

Parental education and other similar instruments are assumed to be positively correlated with log-parental income, i.e., it is assumed that $Corr(\ln X\,, L) > 0$. Then, Equation [14] indicates that we can expect $\ddot{\beta}_1$ to be larger or smaller than β_1, depending on the sign of ϱ_2. Under the additional substantive assumption that $\varrho_2 > 0$, IV estimates with years of parental education (and other similar instruments) have been typically interpreted as providing an upper bound for the conventional IGE.

Empirical assumptions of the bracketing strategy

Based on the last result, Solon (1992) suggested the bracketing strategy: To combine OLS and IV estimates to bracket—that is, set estimate—the true value of the conventional IGE. The analyses conducted here indicate that this strategy relies on nine empirical assumptions: (a) the assumptions of the GEiV model, i.e., Equation [5], [6] and [7]; (b) the assumptions of the GEiV-IV model, i.e., Equations [10] and [11]; (c) the lifecycle assumptions $\lambda_1 = \eta_1 = 1$; (d) the invalid-instrument assumption, i.e., $\varrho_2 > 0$; and (e) an auxiliary empirical assumption, i.e., $0 < Corr(\ln X\,, L) < 1$.

The IGE of the expectation

Due to the conceptual and methodological problems affecting the conventional IGE, Mitnik and Grusky (2017) have called for redefining the workhorse intergenerational elasticity. 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:

$$\ln E(Y|x) = \alpha_0 + \alpha_1 \ln x, \qquad [15]$$

where $Y \geq 0$, $X > 0$ and $\alpha_1 = \frac{d \ln E(Y|x)}{d \ln x}$ 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 (see Mitnik and Grusky 2017: Section V.A), (b) the IGE of the expectation is fully immune to the "zero problem" and the concomitant selection bias affecting the estimation of the IGE of the geometric mean, and (c) the IGE of the expectation is very well suited for studying the role of marriage in the intergenerational transmission of advantage (see Mitnik and Grusky 2017: Section V.B).

The IGE of the expectation is not the only mobility measure that is immune to the zero problem. Most notably, the increasingly popular "rank-rank slope" (RRS), a measure of relative-position mobility and of the transmission of relative-position advantages across generations (e.g., Dahl and DeLeire 2008; Chetty et al. 2014), is well defined when income variables include zero in their support. This may suggest that mobility scholars can rely on the RRS to circumvent the problems affecting the conventional IGE, and therefore that analyses based on the IGE of the expectation are unnecessary.

Nothing of the sort is the case. First, although the RRS is a clearly useful measure, it does not measure the same concept as does the IGE of the expectation—in other words, both of these measures are "mobility measures" in a very broad sense, not two alternative approaches for measuring the same thing. Second, while it is somewhat easier to deal with left-side and right-side lifecycle biases, as well as right-side attenuation bias, in the estimation of the RSS than in the estimation of intergenerational elasticities (Nybom and Stuhler 2017; Mazumder 2016), estimation of the RSS is affected by its own unique problem, left-side attenuation bias (Nybom and Stuhler 2017). This makes estimation of the RSS when only one annual measure of children's income is available to compute their ranks (a very common situation) an unadvisable course of action. Third, because ranks, unlike money flows, are not additive across income sources— and, more generally, because family-income rank cannot be derived in any way from the ranks of family members in the individual-earnings distribution—it is not possible to use the RSS to study the "channels" (e.g., labor market, marriage market) through which the intergenerational transmission of family-income advantage occurs, while this can be easily done with the IGE of the expectation (see Mitnik and Grusky 2017: Section V.B). Lastly, while intergenerational elasticities can be easily embedded, and have been frequently embedded, within policy-relevant theoretical models (see, e.g., Benabou [2000], Durlauf and Seshadri [2018] and Solon [2004] for the conventional IGE, and Mitnik [2018] for the IGE of the expectation), there aren't yet any comparable models embedding the RSS and, more crucially, developing them seems a quite daunting task. It follows that we need both the RRS and the IGE of the expectation in the toolbox of mobility scholars.

Set estimation of the IGE of the expectation

PPML estimation and the GEiVE model

After substituting short-run for long-run income measures in Equation [15], the IGE of the expectation can be estimated using several approaches. Here I assume that estimation is based on the PPML estimator.[5] In order to introduce the empirical assumptions of Mitnik's (2017) generalized error-in-variables model for the estimation of the IGE of the expectation with the PPML estimator, or GEiVE model, it is necessary to first introduce the following population linear projections:

$$Z_t = \theta_{0t} + \theta_{1t} Y + W_t \qquad [16]$$

$$\ln S_k = \pi_{0k} + \pi_{1k} \ln X + P_k , \qquad [17]$$

where $Z_t \geq 0$ is children's income at age *t*; *Y* is as defined earlier; $\theta_{0t} + W_t$ is the 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 one and varies with *t*; and Equation [17] is the same as Equation [4] but I have used a different notation for the parameters and the error term, both to avoid confusions and because the populations covered by the two equations will be different as long as there are children whose short-run income is zero.[6] Indeed, unlike in the case of the GEiV model in which $Z_t > 0$, in the GEiVE model $Z_t \geq 0$, i.e., as the left-side measurement error concerns the income variable rather than its logarithm, the model does not preclude that the observed income variable is zero, and therefore children without any short-run income (e.g., those who are unemployed the year in which their incomes are measured) are included in all analyses.

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

$$Cov(W_t, \ln X) = 0 \qquad [18]$$

$$Cov(P_k, Y) = 0 \qquad [19]$$

$$Cov(W_t, P_k) = 0. \qquad [20]$$

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* to simplify the notation.)

Mitnik (2017) has shown that, at the level of approximation provided by second-order Taylor-series expansions, the probability limit of the "long-run PPML estimator" of the IGE of the expectation (i.e., the PPML estimator with long-run income variables), is:

$$\alpha_1 \approx C_{\alpha_1} - \left[\left(C_{\alpha_1}\right)^2 - V_{\alpha_1}\right]^{\frac{1}{2}}, \qquad [21]$$

where $C_{\alpha_1} = [Cov(Y, \ln X\ )]^{-1}$ and $V_{\alpha_1} = 2\ [Var(\ln X)]^{-1}$.[7] His analyses also entail that, in the general case, the probability limit of the "short-run PPML estimator" of the IGE of the expectation (i.e., the PPML estimator with short-run income variables substituted for the long-run variables), denoted by $\tilde{\alpha}_1$, is:

$$\tilde{\alpha}_1 \approx C_{\tilde{\alpha}_1} - \left[\left(C_{\tilde{\alpha}_1}\right)^2 - V_{\tilde{\alpha}_1}\right]^{\frac{1}{2}}, \qquad [22]$$

where

$$C_{\tilde{\alpha}_1} = [\theta_1 \pi_1 Cov(Y, \ln X) + \theta_1 Cov(Y, P) + \pi_1 Cov(W, \ln X) + Cov(W, P)]^{-1}$$

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

and $VR = \frac{Var(P)}{Var(\ln X)}$ is again the variance ratio.[8]

If the three empirical assumptions of the GEiVE model (i.e., Equations [18]-[20]) hold, $C_{\tilde{\alpha}_1}$ reduces to

$$C_{\tilde{\alpha}_1} = [\theta_1 \pi_1 Cov(Y, \ln X)]^{-1}.$$

If, in addition, $\theta_1 = \pi_1 = 1$, i.e., if the children's and parents' incomes are measured at the right points of their lifecycles, $C_{\tilde{\alpha}_1} = C_{\alpha_1}$ and $V_{\tilde{\alpha}_1} = 2\ \{Var(\ln X)[1 + VR]\}^{-1}$. As $\frac{\partial \tilde{\alpha}_1}{\partial VR} < 0$, the estimates obtained with the short-run PPML estimator can be expected to be affected by attenuation bias.

The GEiVE model provides a methodological justification for estimating the IGE of the expectation with the PPML estimator and proxy variables that satisfy some conditions (exactly as the GEiV model does for the estimation of the conventional IGE by OLS). To minimize lifecycle biases, the GEiVE model suggests that researchers use measures of children's and parents' economic status (i.e., income, earnings) obtained at ages in which $\theta_1 \cong 1$ and $\pi_1 \cong 1$, respectively. Mitnik's (2017) empirical results suggest that, as in the case of the conventional IGE, this happens when both parents' and children's incomes are measured close to age 40. The GEiVE model also indicates that researchers should use measures of average parental income or earnings over several years, rather than annual measures, so as to reduce $Var(\mathrm{P})$ as much as possible. Mitnik's (2017) empirical results suggest that, with survey data, it is necessary to use at least 13 years of information to eliminate the bulk of that bias.

IV estimation and the GEiVE-IV model

As I pointed out earlier, in most cases mobility scholars have relatively few years of parental information available, so we can expect their estimates of the IGE of the expectation with the PPML estimator to be affected by (potentially substantial) attenuation biases. As in the case of the conventional IGE, an obvious alternative is to estimate the IGE of the expectation with an IV estimator. There are several such estimators that could be used to this effect. These include the multiplicative- and additive-error versions of the GMM IV estimator of the Poisson or exponential regression model (Mullahy 1997; Windmeijer and Santos Silva 1997), two two-step quasi-maximum-likelihood IV estimators of the same model (Wooldridge 1999:Sec. 6.1; Mullahy 1997:590-591), a two-step residual-inclusion estimator for nonlinear parametric models (Terza et al. 2008), and two different two-step regression-calibration IV estimators of generalized linear models with measurement error in covariates (Carroll et al. 2006:Ch. 6). The two GMM IV estimators have the upper hand, however, as they make weaker assumptions than any of the other estimators (as they neither require the functional-form assumptions that all other estimators need for their first estimation step, nor the additional assumptions that the "predictor-substitution" estimators of Mullahy 1997 and Carroll 2006 invoke). In addition, the GMM estimators involve standard asymptotic inferential procedures, while the other estimators require to account for the two-step nature of the estimation by using more complicated closed-form asymptotic variance estimators (e.g., Murphy and Topel 1985; Hardin 2002), or by resorting to resampling methods. In what follows I focus on the additive-error version of the GMM-IV estimator, to which I refer as the GMM-IVP estimator, and show that it is upward biased, asymptotically, when employed with the instruments typically available to mobility scholars.

I next advance a generalized error-in-variables model for the IV estimation of the IGE of the expectation, the GEiVE-IV model, in which I make empirical assumptions fully comparable to those of the GEiV-IV model.[9] The empirical assumptions are the following. For any *t* and *k*:

$$Cov(W_t, L) = 0 \qquad [23]$$

$$Cov(P_k, L) = 0. \qquad [24]$$

As usual, the assumptions are expected to hold imperfectly but still as good approximations, at least when $\theta_{1t} \approx \pi_{1k} \approx 1$. (In what follows, I drop the subscripts *t* and *k*.)

As with the PPML estimator, a closed-form expression for the probability limit of the GMM-IVP estimator is not available. In the case of the former estimator, Mitnik (2017) addressed this problem by deriving the approximate closed-form expression I introduced above from the population moment problem solved by the probability limit of the estimator. This approach has proved less convenient in the case of the GMM-IVP estimator, so here I work directly with the relevant population moment conditions. That is, I compare the population moment problems solved by (a) the probability limit of the PPML estimator with long-run income variables, and (b) the probability limit of the GMM-IVP estimator with short-run income variables (the "short-run GMM-IVP estimator") and the instruments typically available to mobility scholars. Like Mitnik (2017), however, I use second-order Taylor-series approximations to some expectations to derive my results.

The probability limit of the PPML estimator with long-run variables, α_1, solves the population-moment condition

$$\frac{E(X^{\alpha_1}\ln X)}{E(X^{\alpha_1})} = E(Y \ln X) \qquad [25]$$

(Mitnik 2017:15), while $\ddot{\alpha}_1$, the probability limit of the short-run GMM-IVP estimator, solves

$$\frac{E(S^{\ddot{\alpha}_1}L)}{E(S^{\ddot{\alpha}_1})} = E(Z\,L).^{10} \qquad [26]$$

Using equations [16] and [17] to substitute S and Z out in Equation [26] yields:

$$\frac{E\left(X^{\pi_1\ddot{\alpha}_1}\ln X\right)}{E(X^{\pi_1\ddot{\alpha}_1})}$$

$$\approx E(Y\ln X)\left\{\frac{1+F(\ddot{\alpha}_1)\,\left[\theta_1 Cov(Y,\ln X)+Var(\ln X)\dfrac{\theta_1 Cov(L,\Psi)+Cov(L,\mathrm{W})}{Cov(L,\ln X)}\right]}{1+Cov(Y,\ln X)}\right\}, \qquad [27]$$

where

$$F(\ddot{\alpha}_1) = \frac{E\left(X^{\pi_1\ddot{\alpha}_1}L\right)}{E(X^{\pi_1\ddot{\alpha}_1})}\left\{\frac{E\left(X^{\pi_1\ddot{\alpha}_1}[\exp(P)]^{\ddot{\alpha}_1}L\right)}{E(X^{\pi_1\ddot{\alpha}_1}[\exp(P)]^{\ddot{\alpha}_1})}\right\}^{-1},$$

and $\Psi = Y - \exp(\alpha_0)\,X^{\alpha_1}$ is an additive error based on Equation [15] (see Online Appendix A for the derivation of Equations [26] and [27]).

Equation [27] shows the (approximate) population moment problem solved by the probability limit of the short-run GMM-IVP estimator of the IGE of the expectation in the general case. Together, Equations [25] and [27] provide a counterpart to Equation [12]. As the latter does for the linear IV estimator of the IGE of the geometric mean, the former (a) identify the various factors determining the probability limit of the GMM-IVP estimator of the IGE of the expectation with short-run variables, and (b) indicate that, in the general case, this estimator may be upward or downward inconsistent.

Substituting second-order Taylor-series approximations for the four expectations in $F(\ddot{\alpha}_1)$ gives (see Online Appendix A):

$$F(\ddot{\alpha}_1) \approx \frac{1 + 0.5\,\{[\pi_1\ddot{\alpha}_1]^2 Var(\ln X) + [\ddot{\alpha}_1]^2\; Var(P)\}}{1 + 0.5\;[\pi_1\ddot{\alpha}_1]^2 Var(\ln X)} \frac{\pi_1\; Cov(\ln X\,, L)}{\pi_1 Cov(\ln X\,, L) + Cov(L, P)}, \qquad [28]$$

which may be larger, equal or smaller than one. Under the assumptions of the GEiVE-IV model, however, Equation [28] reduces to $F(\ddot{\alpha}_1) \approx 1 + \frac{0.5\;[\ddot{\alpha}_1]^2\; Var(P)}{1+0.5\;[\pi_1\ddot{\alpha}_1]^2\; Var(\ln X)} > 1$ and Equation [27] becomes:

$$\frac{E\left(X^{\pi_1\ddot{\alpha}_1}\ln X\right)}{E(X^{\pi_1\ddot{\alpha}_1})} \approx E(\ln X\;\; Y)\left\{\frac{1 + \left[\theta_1 Cov(Y,\ln X) + Var(\ln X)\dfrac{\theta_1 Cov(L,\Psi)}{Cov(L,\ln X)}\right]\left[1 + \dfrac{0.5\;[\ddot{\alpha}_1]^2\; Var(P)}{1 + 0.5\;[\pi_1\ddot{\alpha}_1]^2\; Var(\ln X)}\right]}{1 + Cov(Y,\ln X)}\right\}. [29]$$

If in addition $\theta_1 = \pi_1 = 1,$ the last equation reduces to:

$$\frac{E\left(X^{\ddot{\alpha}_1}\ln X\right)}{E(X^{\ddot{\alpha}_1})} \approx E(Y,\ln X) + Var(\ln X)\frac{Cov(L,\Psi)}{Cov(L,\ln X)}\left[1 + \frac{0.5\;[\ddot{\alpha}_1]^2\; Var(P)}{1 + 0.5\;[\ddot{\alpha}_1]^2\; Var(\ln X)}\right]$$

$$+ Cov(Y,\ln X)\frac{0.5\;[\ddot{\alpha}_1]^2\; Var(P)}{1 + 0.5\;[\ddot{\alpha}_1]^2\; Var(\ln X)}, \qquad [30]$$

where I have used $E(\ln X\;\; Y) = 1 + Cov(\ln X\,, Y)$.

Equations [25] and [30] provide a counterpart to Equation [14]. Indeed, as $\frac{\partial E\left(X^{\ddot{\alpha}_1}\ln X\right)\left[E\left(X^{\ddot{\alpha}_1}\right)\right]^{-1}}{\partial\ddot{\alpha}_1} > 0$ (see Online Appendix A), and safely assuming that $Cov(Y,\ln X) > 0$, those equations show that the probability limit of the short-run GMM-IVP estimator is an upper bound for the IGE of the expectation when the invalid instrument is positively correlated with the error term. The equations also show that the upper bound is tighter when (a) the covariance between the instrument and the error term, $Cov(L, \Psi)$, is smaller, (b) the slope of the linear projection of L on $\ln X$, $Cov(L, \ln X)/Var(\ln X)$, is larger, and (c) the short-run measure of parental income is less noisy, i.e., $Var(P)$ is smaller relative to $Var(\ln X)$.

Empirical assumptions of the bracketing strategy

The GEiVE-IV model just advanced provides a foundation for combining PPML and GMM-IVP estimates with the goal of set estimating the IGE of the expectation. The bracketing strategy relies in this case on the following empirical assumptions: (a) the assumptions of the GEiVE model, i.e., Equations [18], [19] and [20]; (b) the assumptions of the GEiVE-IV model, i.e., Equation [23] and [24]; (c) the lifecycle assumptions $\lambda_1 = \eta_1 = 1$; (d) the invalid-instrument assumption $Cov(L, \Psi) > 0$; and (e) the auxiliary empirical assumption $Cov(\ln X, L) > 0$. These assumptions are strictly isomorphic to those I identified in the case of the bracketing strategy for the IGE of the geometric mean.

Constructing confidence intervals for partially-identified IGEs

Under the empirical assumptions of the bracketing strategies discussed in the previous two sections, the probability limits of the OLS and IV estimators, and the probability limits of the PPML and GMM-IVP estimators, provide lower and upper bounds for the long-run IGEs of the geometric mean and expectation, respectively. This means that these IGEs are only "partially identified" (e.g., Manski 2003) by data on short-run incomes and the instruments: Even if we could obtain an unlimited number of observations, we would not be able to learn the true values of the long-run IGEs. For each IGE, we can only aim at learning what the range of values consistent with those data—the "identified set" defined by the relevant probability limits—is.

The second difficulty is, of course, that we do not have an unlimited number of observations available but, rather, need to estimate the bounds from a finite sample. This means we need to take into account not only the uncertainty due to partial identification but also the uncertainty regarding the estimated bounds. There are two different cases to consider. In the first, simpler, case, there is only one estimate of the upper bound. In the second case the upper bound is estimated more than once with different sets of instruments, which creates more serious complications.

A single estimate of the upper bound

Previous research using the bracketing strategy has reported separate confidence intervals for the bounds estimated by the OLS and IV estimators, i.e., for the bounds of the identified set. However, when we rely on a bracketing strategy we would like to provide just one confidence interval that (a) refers to the partially identified long-run IGE, and (b) reflects uncertainty due both to partial identification and to sampling variability. An approach that suggests itself is to construct such a confidence interval by using as lower bound the lower bound of the confidence interval associated to the OLS or PPML estimate, and as upper bound the upper bound of the confidence interval associated to the relevant IV estimate. This, however, would generate a confidence interval for the identified set (e.g., Stoye 2009:1300), not for the IGE itself. The probability that this interval covers the IGE is at least as large as the probability that it covers the identified set (Imbens and Manski 2004: Lemma 1). In other words, the suggested confidence interval is too conservative.

The intuition for why this is the case is that, asymptotically, the width of the identified set is large compared to the sampling error. Therefore, if the true IGE is not close to the lower bound, then the risk that the lower-bound estimate will be larger than the true value can be ignored. Likewise, if the true parameter is not close to the upper bound, then the risk that the upper-bound estimate will be lower than the true value can be ignored. As the true value cannot be close to both bounds, the noncoverage risk is effectively one-sided. Denoting the probability of type-I error by α, this entails that, asymptotically, a $100\ (1-\alpha)$ % confidence interval (e.g., a 90 percent confidence interval) for the identified set is a $100\ \left(1-\frac{\alpha}{2}\right)$ % confidence interval (e.g., a 95 percent confidence interval) for the IGE.

This suggests using $100\ (1-2\alpha)$ % confidence intervals for the identified sets as $100\ (1-\alpha)$ % confidence intervals for the IGEs.[11] This, however, has a shortcoming: For any finite sample size N, if the width of the identified set is short enough, the confidence interval will be shorter than if the IGE were point identified (i.e., shorter than the confidence interval that would result if the OLS or PPML estimator, and the relevant IV estimator, had identical probability limits). The reason for this is that the exact coverage probabilities do not converge to their nominal values uniformly across different values of the width of the identified set.

To address this problem, and following Imbens and Manski (2004), $100\ (1-\alpha)$ % confidence intervals for the long-run IGE of the geometric mean, denoted by $CI_{\beta_1}(\alpha)$, and for the long-run IGE of the expectation, denoted by $CI_{\alpha_1}(\alpha)$, can be constructed as follows:

$$CI_{\beta_1}(\alpha) \equiv \left[\hat{\beta}_1^{OLS} - c_{\beta_1}(\alpha)\ \widehat{SE}\left(\hat{\beta}_1^{OLS}\right), \hat{\beta}_1^{IV} + c_{\beta_1}(\alpha)\ \widehat{SE}\left(\hat{\beta}_1^{IV}\right)\right]$$

$$CI_{\alpha_1}(\alpha) \equiv \left[\hat{\alpha}_1^{PPML} - c_{\alpha_1}(\alpha)\ \widehat{SE}(\hat{\alpha}_1^{PPML}),\ \hat{\alpha}_1^{GMM-IVP} + c_{\alpha_1}(\alpha)\ \widehat{SE}(\hat{\alpha}_1^{GMM-IVP})\right],$$

where $c_{\beta_1}(\alpha)$ and $c_{\alpha_1}(\alpha)$ respectively solve

$$\Phi\left(c_{\beta_1}(\alpha) + RW_{\beta_1}\right) - \Phi\left(-c_{\beta_1}(\alpha)\right) = 1-\alpha$$

and

$$\Phi\left(c_{\alpha_1}(\alpha) + RW_{\alpha_1}\right) - \Phi\left(-c_{\alpha_1}(\alpha)\right) = 1-\alpha;$$

$RW_{\beta_1} = \frac{\hat{\beta}_1^{IV} - \hat{\beta}_1^{OLS}}{\max\left(\widehat{SE}(\hat{\beta}_1^{OLS}),\ \widehat{SE}(\hat{\beta}_1^{IV})\right)}$ and $RW_{\alpha_1} = \frac{\hat{\alpha}_1^{GMM-IVP} - \hat{\alpha}_1^{PPML}}{\max\left(\widehat{SE}(\hat{\alpha}_1^{PPML}),\ \widehat{SE}(\hat{\alpha}_1^{GMM-IVP})\right)}$ are (estimates of) the relative widths of the identified sets; SE is, as before, the standard error operator; $\Phi(.)$ denotes the CDF of the standard normal distribution; and the superscripts identify estimators. Here, $c_{\beta_1}(\alpha)$ and $c_{\alpha_1}(\alpha)$ are inversely related to RW_{β_1} and RW_{α_1}, respectively, and the coverage probabilities of the confidence intervals do converge to their nominal values uniformly across values of the width of the identified sets. With 95 percent confidence intervals, $c_{\beta_1}(\alpha)$ and $c_{\alpha_1}(\alpha)$ are close to $\Phi^{-1}(0.90) \cong 1.64$ when the width of the identified set is large compared to sampling error, and are equal to $\Phi^{-1}(0.95) \cong 1.96$ under point identification.

As companions to this article, I have made available two Stata programs that compute the confidence intervals just presented. One of them both set estimates the IGEs and computes the confidence intervals (which minimizes the amount of coding that is necessary), while the other computes the confidence intervals after the models have been estimated by the researcher (which provides more flexibility). With these programs, both set estimation and inference are very simple tasks.[12]

Multiple estimates of the upper bound

So far I have considered the case in which there is only one estimate of the upper bound. The upper bound, however, may be estimated with different sets of instruments. As the probability limit of the IV estimator is different in each case, the identified set is the intersection of all the sets that can be formed by combining one of these probability limits with the probability limit of the lower-bound estimator. In this “intersection-bounds context,” the two-step (2S) estimator that selects the minimum of the IV estimates across instruments is a consistent estimator of the upper bound (more on this below). However, as this estimator is based on multiple estimators, each with its own distribution, the construction of confidence intervals needs to take into account the sampling uncertainty about all upper-bound estimates, not just the estimate that happens to be the minimum in the current sample (e.g., Chernozhukov, Lee and Rosen 2013).

Closely related approaches for constructing valid confidence intervals in the intersection-bounds context have been developed by Nevo and Rosen (2012: Section IV) and Chernozhukov, Lee and Rosen (2013). Specialized to the problem at hand, Nevo and Rosen's approach is as follows:

(a) Identify and eliminate from further consideration the upper-bound estimates, if any, generated by estimators whose distributions are located "enough to the right" that they can be safely ignored in determining the asymptotic distribution of the two-step estimator.[13]

(b) Estimate the absolute width of the identified set, Δ, as the difference between the minimum upper-bound estimate and the estimate of the lower bound. Use this value to compute the probability $p = 1 - \Phi(\ln N\ \ \Delta)\ \alpha$ (instead of $p = 1 - 0.5\ \alpha$, as would be the case with a point estimate) that will be employed to determine the critical values for the construction of the confidence interval. This way of computing the probability takes into account that the IGE is set rather than point identified and is consistent with the goal of uniform convergence.

(c) Estimate the correlation matrix of the relevant upper-bound estimators (those not eliminated in the first step), and use it to simulate a large sample from a zero-mean multivariate normal distribution with as many dimensions as the number of relevant estimators of the upper bound.[14]

(d) Compute a new variable containing the maximum value obtained in each draw of the simulation, and determine the quantile of this variable corresponding to the probability p (i.e., determine the quantile by "inverting" the empirical CDF of the maximum-value variable). Let's denote the quantile in question as $q^u_{\beta_1}(\alpha)$ in the case of the long-run IGE of the geometric mean and as $q^u_{\alpha_1}(\alpha)$ in the case of the long-run IGE of the expectation.

(e) Compute the corresponding quantile for the lower bound by inverting the CDF of the univariate normal distributions, i.e., as $q = \Phi^{-1}(p)$. Let's denote this quantile as $q^l_{\beta_1}(\alpha)$ in the case of the long-run IGE of the geometric mean and as $q^l_{\alpha_1}(\alpha)$ in the case of the long-run IGE of the expectation.

(e) The $100\ (1-\alpha)$ % confidence intervals may then be constructed as follows:

$$CI_{\beta_1\cap}(\alpha) \equiv \left[\hat{\beta}_1^{OLS} - q^l_{\beta_1}(\alpha)\,\widehat{SE}\left(\hat{\beta}_1^{OLS}\right), \min_m\left(\hat{\beta}_1^{IV(m)} + q^u_{\beta_1}(\alpha)\,\widehat{SE}\left(\hat{\beta}_1^{IV(m)}\right)\right)\right]$$

$$CI_{\alpha_1\cap}(\alpha) \equiv \left[\hat{\alpha}_1^{PPML} - q^l_{\alpha_1}(\alpha)\,\widehat{SE}(\hat{\alpha}_1^{PPML}), \min_m\left(\hat{\alpha}_1^{GMM-IVP(m)} + q^u_{\alpha_1}(\alpha)\,\widehat{SE}\left(\hat{\alpha}_1^{GMM-IVP(m)}\right)\right)\right],$$

where $m = 1,2,\ldots,M$ indexes the sets of instruments used by the relevant upper bound estimators and $\cap$ denotes intersection (reflecting that these confidence intervals are valid in intersection-bounds contexts).

As a situation in which the upper bound is estimated with a single set of instruments qualifies as a (trivial) intersection-bounds context, the Nevo and Rosen's (2012) confidence intervals are also valid in that situation.[15] With large enough samples, they should be essentially identical to the Imbens and Manski's (2004) confidence intervals. With smaller samples, however, Nevo and Rosen's (2012) confidence intervals seem to be slightly more conservative. Therefore, Imbens and Manki's (2004) confidence intervals seem preferable.[16]

Although the 2S estimator of the upper bound is consistent (Nevo and Rosen 2012), it is not asymptotically unbiased—in fact, no estimator involving minimization or maximization can be unbiased (Hirano and Porter 2012). Chernozhukov, Lee and Rosen (2013) have proposed an alternative three-step (3S) estimator that is consistent and "half-median unbiased." In the context at hand, this means that the estimator has the property that at least half of its values across samples are above the true upper bound. The estimator selects the minimum of the estimates of the upper bound, but only after correcting them (this is the added step). The correction adds to each point estimate its standard error multiplied by a critical value.[17] Confidence intervals are computed exactly or almost exactly as I described above.[18]

While the risk with the 2S estimator is that in finite samples it may produce estimates that are significantly downward biased, the alternative 3S estimator has the shortcoming that it may tend to be overly conservative. The reason is that, unlike a median unbiased estimator (e.g., Birnbaum 1964), a half-median unbiased estimator of an upper bound may have the bulk of its distribution "way to the right." In the empirical analyses of the next section I report estimates based on both estimators.

Together with this article, I have made available a third Stata program that both set estimates the IGEs with multiple sets of instruments and computes the confidence intervals just discussed. Set estimation and inference in the intersection-bounds context is fully unproblematic with this program.[19]

Empirical analyses

The main goal of this section is to empirically assess whether the bracketing strategies work as the generalized error-in-variables models lead us to expect, the information supplied by the bounds the strategies generate, and how the bounds vary when different instruments and short-run measures of parental income are used. The analyses are preceded by a brief description of the data and the estimation approaches employed.

Data and estimation

The empirical analyses are based on a PSID sample that makes it possible to construct an approximate measure of long-run parental income but not of children's long-run income.[20] However, as I explain below, this sample still allows to shed light on questions of central interest for this paper. The sample includes information on children born between 1966 and 1974, for which 25 years of parental data centered on age 40 (obtained when the children were between 1 and 25 years old) are available. Children observed in the PSID when they were between 35 and 38 years old are included in the sample. I use information on the average family income of children when they were 35-38 years old, on parents' family income, age, and years of education when the children were 1-25 years old, and on fathers' occupation when the children were growing up (as reported by the latter). I do not use information on earnings because I only estimate family-income IGEs (in part, to minimize the selection bias that results when children with zero income or earnings are dropped from samples when estimating the IGE of the geometric mean). Table 1 presents descriptive statistics, while Online Appendix C provides additional details on the sample and variables and explains in more detail why I focus exclusively on family-income IGEs.

I use the OLS and Two-Stage Least Squares (TSLS) estimators to estimate the IGE of the geometric mean of children's family income, the PPML and GMM-IVP estimators to estimate the IGE of the expectation of children's family income, and employ sampling weights and compute robust standard errors in all cases. I construct the confidence intervals for the partially identified IGEs as explained in the previous section.[21]

I use the following instruments in my analyses: (a) parents' total years of education when the child was 15 years old, and in the time period covered by each short-run parental-income measure; (b) the household head's years of education when the child was 15 years old, and in the time period covered by each short-run parental-income measure; and (c) the father's occupation. The reason for employing both time-varying and at-age-15 parental-education variables as instruments is that, in the datasets used by mobility scholars, parental education is sometimes available all years in which parental income is measured and sometimes is only available for when the children were of some specific age, usually in the 12-16 range. It's then important to examine IV estimates generated with both types of parental-education variables. Online Appendix C explains why I use parents' education and the household head's education as instruments but not father's education, which is the typical approach in the literature.

The relationship between long-run and short-run measures of income varies with the age at measurement; for this reason, it is 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 I report are based on a sample in which the variation in children's ages is very small, controlling for children's age is unnecessary (see the next paragraph for a second reason for proceeding this way). Mitnik et al. (2015:34) have argued that the age at which parents have their children is not exogenous to their income, that parental 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 controls for parental age, but estimates from models with such controls are very similar.

As a measure of the long-run family income of children is not available, in all analyses, regardless of whether they pertain to short-run or long-run IGEs, I use the family income of children when they were 36-38 years old as their income measure. This is equivalent to making

$\lambda_1 = \theta_1 = 1, V = \mathrm{W} = 0$ (for all children), and $Cov(L, \mathrm{V}) = Cov(L, \mathrm{W}) = 0$ by construction. As a result, in the empirical analyses I am not able to assess any aspect of the generalized error-in-variables models pertaining to left-side measurement error. Nevertheless, I can still draw clear conclusions regarding right-side measurement error and the bracketing strategies.

Results

Figures 1 and 2 pertain to the IGE of the geometric mean. They present information that allows to assess the qualitative implications of the relevant generalized error-in-variables models, and how the bracketing strategy works when parental income is measured at different parental ages. Figures 3 and 4 do the same for the IGE of the expectation. Tables 2, 3a and 3b summarize the results obtained with the bracketing strategies, and present the estimates of the long-run IGEs. Table 2 reports the results obtained with what I will refer as the “ideal bracketing strategies,” as it displays IGE estimates pertaining to the average parental ages at which the measurement-error slopes η_1 and π_1 are equal to one. Those estimates are computed by interpolation, and therefore confidence intervals are not available. Tables 3a and 3b show estimates based on measures of parental income centered on the years the children were 13 years old, when the average age of the parents is close to 40, for the IGE of the geometric mean and the IGE of the expectation, respectively. Assessing the performance of the bracketing strategies in this context is of eminent practical interest, as mobility scholars normally do not know the ages at which the measurement-error slopes are equal to one with their data. So, when possible with those data, they simply use estimates based on income measures taken when parents and children are close to age 40 as their best guess (a guess informed by the results obtained with other, potentially quite different, data). I will refer to the bracketing strategies implemented this way as the “feasible bracketing strategies.” [22]

Each of Figures 1 to 4 includes four panels. The results presented in each of these panels are based on a different set of short-run measures of parental income. In the top-left panels, parental income was measured when the children were 1 or 2 or 3 . . . up to 25 years old. In the other three panels, the short-run measures of parental income are multiyear averages. In the top-right panels they are three-year averages, centered when the children were 2 or 3 or 4 . . . up to 24 years old. In the bottom panels the measures of parental income are five- and seven-year averages, centered similarly. The age in the horizontal axis is in all cases the average age of the parents in the sample. The bottom curve in each panel shows the relationship between OLS- or PPML-based IGE estimates and average parental ages, while the top curve shows the relationship between estimates of the relevant parental measurement-error slope and those ages. In Figures 1 and 3, the two middle curves in each panel show the relationship between IV IGE estimates obtained when parental income is instrumented by the at-age-15 parental education variables and average parental ages. In Figures 2 and 4 the middle curve in each panel is similar but pertains to IV IGE estimates generated with the father's occupation as instrument. As the short-run income measures rely on more years of information, from one to seven, estimates in all figures become less affected by transitory income fluctuations and the shapes of the curves become progressively clearer.

Income-age profiles vary in a well-known manner across people with different levels of human capital, while the latter are strongly associated to parental income. We therefore expect that η_1 and $\pi_1$will increase with parental age, and will be smaller than one when the parents are younger and larger than one when the parents are older (e.g., Harden and Solon 2006). Figures 1 to 4 fully confirm these expectations. In addition, consistent with what the two IV generalized error-in-variables models predict (see Equations [13] and [29]), the figures also show a clear inverse relationship between the measurement-error slopes and the IV IGE estimates: While the former increase, the latter fall with the parents' age.[23] Similarly, consistent with what the GEiV and GEiVE models predict (see Mitnik 2017:39-40), the OLS- and PPML-based IGE estimates first increase (at very low values of the measurement-error slopes) but then decrease with parents' age. Those slopes are equal to one when the parents are, on average, somewhat younger than 40, i.e., at average ages between 37.0 and 38.1 (see Table 2), while the slopes are in the 1.07–1.10 range at age 40 (see Tables 3a and 3b).

In all figures, the long-run IGE, represented by the darker-gray horizontal lines, is the IGE of the family income of children when they were 36-38 years old with respect to the (approximate) long-run family income of their parents. As explained earlier, for the purposes of the analyses here, the former income is assumed to be the true long-run income of children. Moreover, under the assumptions of the GEiV and the GEiVE models, and given what we know from previous research about the children's ages at which $\lambda_1 = 1$ and $\theta_1 = 1$ (e.g., Haider and Solon 2006; Mitnik 2017), it is likely that the estimates reported as long-run estimates in the figures and tables—0.7 in the case of the IGE of the geometric mean, 0.6 in the case of the IGE of the expectation—are quite close to the estimates that would be obtained if the long-run income of children were available.

Comparing the long-run-IGE line with the IGE curves based on short-run income measures in the four figures, makes apparent that the ideal bracketing strategies work as expected: In all 16 panels, without exception, the point estimate of the long-run IGE is covered by the corresponding set estimate (i.e., it is bracketed by the OLS and TSLS estimates, or by the PPML and GMM-IVP estimates, as applicable), when the relevant measurement-error slope is equal to one. This is confirmed by Table 2, which also includes results obtained with the time-varying instruments. This table also makes clear that, within IGE concepts, the location of the set estimates generated by the bracketing estimators (represented by the sets' midpoints), as well as their width, vary substantially across instruments and parental-income measures. This variation is driven, first, by the fact that, consistent with the implications of the GEiV and GEiVE models, the OLS and PPML estimates tend to increase with the number of years of information used to compute parental income.[24] Second, there is substantial variation across instruments—with some instruments generating much tighter upper bounds than others—and this variation is almost perfectly correlated across measures of parental income and IGE concepts. Father's occupation and parents' education when income was measured provide the tightest bounds, both household-head education variables provide the loosest bounds, while the upper bounds obtained with the at-age-15 parental-education variable are in between.

A very similar analysis applies when we focus on the results of estimating the IGEs with the feasible bracketing strategies. Although all short-run estimates tend to be somewhat lower close to age 40 than at the ages at which the measurement-error slopes are equal to one—the mid-points of the "feasible set estimates" are, on average, about 7.5 percent lower than the mid-points of the "ideal set estimates"—the former still cover, without any exception, the long-run estimates. Moreover, although the feasible set estimates are shifted downward, their widths are very similar to, and are highly correlated with, the widths of the ideal set estimates.[25] More generally, the short-run estimates bracket the long-run estimates at essentially all parental ages (between 29 and 53).

The foregoing shows that the upper bounds that different invalid instruments provide may be markedly different. It also suggests that, when implementing a feasible bracketing strategy, in addition to using a short-run measure of parental income centered around age 40 or so, and based on as many years of parental information as possible, mobility scholars should generate IV IGE estimates with multiple sets of instruments rather than rely in just one instrument or set of instruments.

The set estimates that result from the approach just suggested, using both the 2S and 3S estimators, are shown at the bottom of Tables 3a and 3b. They put the IGE of the geometric mean of children's family income in the 0.57-0.78 (2S estimator) and 0.57-0.83 (3S estimator) ranges, compared to the long-run IGE estimate of 0.7. Similarly, they put the IGE of children's expected family income in the 0.52-0.70 (2S estimator) and 0.52-0.76 (3S estimator) ranges, compared to the long-run IGE estimate of 0.6. Therefore, it is clear that the bracketing strategies provide highly informative set estimates for both long-run IGEs, even with the potentially conservative 3S estimator. At the same time, it is important to keep in mind that the confidence intervals for the long-run IGEs generated by the bracketing strategies are noticeably larger than those generated by the long-run estimators: 0.48-0.94 compared to 0.58-0.82, in the case of the IGE of the geometric mean; and 0.41-0.88 compared to 0.45-0.74, in the case of the IGE of the expectation. This is a consequence of the fact that the confidence interval of a partially identified parameter reflects not only sampling variability but also that the location of the parameter within the identified set cannot be determined by the data, regardless of the size of the sample.

In the previous paragraph I focused on the set estimates based on short-run income measures that average seven years of parental information. Often, fewer years of information are available; in some countries, the datasets used to estimate IGEs only include one year of parental information. Tables 3a and 3b show that although the set estimates obtained with the 2S and 3S estimators are wider, they are still highly informative when they are based on five- or even three-year parental-income measures: In the latter case, the set estimates are 0.52-0.79 (2S estimator) and 0.52-0.85 (3S estimator), for the IGE of the geometric mean, and 0.48-0.68 (2S estimator) and 0.48-0.74 (3S estimator), for the IGE of the expectation. The set estimates are much wider, however, when annual income measures are used, as in this case they put the first IGE in the 0.40-0.76 or 0.40-0.82 ranges, and the second IGE in the 0.40-0.68 or 0.40-0.74 ranges, depending on the estimator.

Discussion

The results of the empirical analyses have made clear that the generalized error-in-variables models underlying the bracketing strategies provide a very good account of the relationships between long-run IGE estimates, the short-run IGE estimates generated by the OLS, PPML and IV estimators, the parents' ages at which their income is measured, and the years of parental information used. Most crucially, the results have shown that both the ideal and the feasible versions of the bracketing strategies work exactly as those models lead us to expect. The empirical analyses also confirmed that, as predicted by the GEiV and GEiVE models, the lower bounds generated by the bracketing strategies become tighter as additional years of information are used to compute the short-run parental income measures.

Different invalid instruments can be expected to be differentially correlated to the logarithm of long-run parental income, and to the error terms of the long-run-IGE population regression functions (i.e., the error terms associated to Equations [1] and [15]). Therefore, we should also expect them to lead to different IV estimates, and to provide different upper bounds for the set estimates of the IGEs. Nevertheless, the magnitude of the differences revealed by the empirical analyses is quite striking. This suggests that mobility scholars should put a good amount of effort into searching for "best invalid instruments," as this effort may have a large payoff in terms of the tightness of the upper bounds supplied by the IV estimators. This may involve looking for additional instruments, beyond those typically employed by mobility researchers (i.e., parental education and occupation), and exploring the effects of alternative functional forms (e.g., entering an instrument in levels or in logarithms), as this has been shown to be very consequential in some contexts (Reiss 2016).

Although in the empirical analyses I simply focused on the instruments commonly employed in the previous literature, the set estimates generated by the bracketing strategies proved to be highly informative for both IGEs as long as they were based on short-run parental income measures relying on at least three years of information. At the same time, the fact that the confidence intervals for the long-run IGEs generated by those strategies are rather wide underscores the fact that obtaining satisfactory levels of precision with these strategies requires samples substantially larger than those that are required to obtain satisfactory levels of precision with the individual short-run estimators that the strategies combine.

Conclusion

The IGE conventionally estimated in the mobility literature pertains to the conditional geometric mean of children's income, which is at odds with the interpretations imposed on its estimates. In addition, the conventional IGE makes studying gender and marriage dynamics in intergenerational processes a very difficult enterprise, and leads to IGE estimates affected by selection biases. For these reasons, Mitnik and Grusky (2017) have called for replacing it by the IGE of the expectation. This requires that the methodological knowledge necessary to estimate this IGE with short-run income variables is made available.

In this paper I have contributed to this goal by advancing a generalized error-in-variables model for the IV estimation of the IGE of the expectation that makes empirical assumptions entirely comparable to those made for the IV estimation of the conventional IGE. Analogously to what its counterpart for the IV estimation of the IGE of the geometric mean does, this model (a) provides an account of the relationship between the ages at which children's and parents' income are measured and the GMM-IVP estimates of the IGE of the expectation, and (b) entails that when the measurement-error slopes are equal to one, estimation of the IGE of the expectation with the GMM-IVP estimator is upward inconsistent.

By combining the latter result with Mitnik's (2017) result that the PPML estimation of the IGE of the expectation with short-income measures is downward inconsistent in the same context, I have proposed a bracketing strategy fully equivalent to that used to set estimate the conventional IGE. The proposed bracketing strategy couples short-run estimates generated with the PPML and GMM-IVP estimators to generate a set estimate of the long-run IGE of children's expected income. As in the case of the IGE of the geometric mean, the feasible version of this strategy relies on estimates obtained with short-run income measures pertaining to when children and parents are close to 40 years old.

Previous research that estimated bounds for the conventional IGE with the OLS and IV estimators reported separate confidence intervals for those bounds. In contrast, by considering the bracketing strategies from the perspective of the partial-identification approach to inference, I have specified how to construct confidence intervals for the partially-identified long-run IGEs, in particular when the upper bound is estimated multiple times with different sets of instruments. These confidence intervals have the correct coverage and converge uniformly to their nominal values regardless of the width of the identified set.

The results of the empirical analyses with PSID data are fully consistent with the qualitative implications of the generalized error-in-variables models underlying the bracketing strategies (both the new strategy proposed here and the strategy previously used in the literature).

Most crucially, those analyses evaluated the performance of the feasible bracketing strategies by comparing their set estimates with point estimates of long-run IGEs. This indicated that those strategies work exactly as expected, and that the set estimates they generate may be highly informative.

Notes

[1] Mitnik and Grusky (2017:Sect. II) provide textual evidence of the field's misinterpretation of the IGE by reproducing and analyzing several quotations from prominent scholars of economic mobility. Two of those quotations are the following: (a) the IGE "measures the percentage differential in the son's expected income with respect to a marginal percentage differential in the income of the father" (Björklund and Jäntti 2011:497), and (b) the IGE provides "a parametric answer to questions like, if the parents' long-run earnings are 50% above the average in their generation, what percent above the average should we predict the child's long-run earnings to be in her or his generation?" (Solon 1999:1777).

[2] 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). For the sake of brevity, from now on I assume that Y and X pertain to children's and parents' income, rather than their income or earnings.

[3] Here, the measurement errors are additive and are equal to the difference between the logarithm of each short-run income variable and the logarithm of the corresponding long-run income variable. Therefore, in the general case, i.e., when λ_{1t} and η_{1k} are not equal to one, the measurement errors are $\lambda_{0t} + (\lambda_{1t} - 1)\ln Y + V_t$ (children) and $\eta_{0k} + (\eta_{1k} - 1)\ln X + Q_k$ (parents).

[4] These more general analyses do not seem to have appeared in the literature before. They are secondary contributions of this article, in addition to its main contributions discussed in the introduction.

[5] Like all Pseudo Maximum Likelihood estimators, when the variables are measured without error the PPML estimator is consistent, regardless of the actual distribution of the dependent

variable, provided that the mean function is correctly specified (Gourieroux, Monfort, and Trognon 1984). There are good reasons to prefer the PPML estimator to other possible estimators of constant-elasticity models (Santos Silvan and Tenreyro 2006; 2011).

[6] The measurement errors are additive and are equal to the difference between the short-run income variable (or its logarithm, as relevant) and the corresponding long-run variable (or its logarithm). When θ_{1t} is not equal to one, the measurement error in the children's short-run income variable as a measure of the corresponding long-run variable is $\theta_{0t} + (\theta_{1t} - 1)\, Y + W_t$. Similarly, when π_{1k} is not equal to one, the measurement error in the log of the short-run parental-income variable as a measure of the log of the long-run parental-income variable is $\pi_{0k} + (\pi_{1k} - 1) \ln X + P_k$.

[7] Equation [21] assumes $Cov(Y, \ln X\,) \neq 0$. From Equation A2 in Mitnik (2017: Online Appendix, A), it follows that when $Cov(Y, \ln X\,) = 0$ then $\alpha_1 \approx 0$. Here, however, we can safely assume that $Cov(Y, \ln X\,) > 0$.

[8] Equations [21] and [22] assume, without any loss of generality, that $E(Z) = E(Y) = E(\ln S) = E(\ln X) = 1$. This entails no loss of generality because it can always be achieved by simply changing the monetary units used to measure income, i.e., by dividing each children's income variable by its mean, and each parental income variable by the exponential of the mean of its logarithmic values minus one.

[9] In Online Appendix B, I advance a second generalized error-in-variables model, the GEiVE-IV-S model, which makes the stronger assumptions that are standard in the literature on measurement error in nonlinear models (e.g., Carroll et al. 2006). Both measurement-error models lead to the conclusion that a bracketing strategy is as feasible with the IGE of the expectation as with the IGE of the geometric mean. However, as I explain in Online Appendix B,

the models have different implications in other respects. I focus on the first measurement-error model here because it allows me to show that set estimation of the IGE of the expectation does not require stronger assumptions than those made for the set estimation of the conventional IGE, in spite of the fact that it involves nonlinear models.

[10] Here and in the rest of this section I assume, without any loss of generality, that the instrument L has been demeaned and that $E(Z) = E(Y) = E(\ln S) = E(\ln X) = 1$ (see note 8).

[11] For instance, constructing a 95 percent interval under this approach would involve computing the lower bound as the OLS or PPML estimate minus (approximately) 1.64 times its standard error, and the upper bound as the IV estimate plus (approximately) 1.64 times its standard error.

[12] The Stata programs, called "igeset" and "igesetci," are available as supplementary materials. They can also be installed directly from within Stata by using the following commands (when connected to the internet): "[redacted as URL identifies the author]" or "[idem]."

[13] This can be achieved by using a procedure proposed by Chernozhukov, Lee and Rosen (2013), which they dubbed "adaptive inequality selection" (AIS). Nevo and Rosen (2012: 666) used a different procedure, which has been superseded by AIS.

[14] Although Nevo and Rosen (2012) describe this step as involving the variance-covariance matrix rather than the correlation matrix, they meant the latter. Observe that although the components of the zero-mean random vector with a multivariate normal distribution are correlated, each of them follows a standard univariate normal distribution.

[15] In this trivial intersection-bounds context, the third and fourth steps described above should be replaced by the computation of the quantile by inversion of the CDF of the univariate normal distribution.

[16] These are the confidence intervals reported in the next section (in the empirical analyses in which the upper bound is estimated only once).

[17] This critical value is computed by following the same steps described above, when computing the critical value used to construct the upper bound of the Nevo and Rosen's confident interval, but using $p = 0.5$ instead of $p = 1 - \Phi(\ln N \ \ \Delta) \ \alpha$.

[18] Chernozhukov, Lee and Rosen (2013) compute the confidence intervals as described above. In my assessment, when the 3S estimator of the upper bound is used, it makes more sense to estimate the absolute width of the identified set by subtracting the estimate of the lower bound from the minimum of the *corrected* estimates of the upper bound. This is what I do in the empirical analyses of the next section (where it makes almost no difference).

[19] The Stata program is called "igeintb" and is available as supplementary material. It can also be installed directly from within Stata by using the following command (when connected to the internet): "[redacted as URL identifies the author]."

[20] It is not possible to select a PSID sample that (a) has the minimum size required by those analyses, and (b) can be used to construct long-run income measures for parents and children simultaneously.

[21] By "robust standard errors" I mean standard errors based on the sandwich estimator of variance (Huber 1967). I use the statistical package Stata to generate all estimates, and to produce the confidence intervals for the partially identified IGEs.

[22] The tables include results obtained with all instruments listed earlier. However, I have not included in the paper figures in which the IV estimates rely on instrumenting short-run parental income with time-varying measures of parents' education, as they do not add much information to that provided by the other figures.

[23] With the alternative, time-varying, parental-education variables, the IV IGE estimates tend to rise, in some cases markedly, when the parents are in their 50s. This is accounted by an increase in the covariance between the instrument and the error term at those ages, in the case of the IGE of the expectation (see Equation [29]); and by an increase in the parental-education coefficient at those ages, in the case of the IGE of the geometric mean (see Equation [13]).

[24] In contrast, the IV estimates do not show any clear trend in this respect. This is consistent with the GEIV-IV and the GEiVE-IV-S models (for the latter, see Online Appendix B), as neither leads us to expect that IV estimates will change in any particular way as more years of information are employed. The GEiVE-IV model, however, entails that the probability limit of the GMM-IVP estimator will fall with the number of years used (see Equation [30]). The fact that no trend is apparent should not be interpreted as evidence for the stronger assumptions of the GEiVE-IV-S model; it is possible that the effect is small and, given the size of the sample, the estimates may be too noisy for the effect to be visible.

[25] The average widths of the set estimates are 0.40 (ideal) and 0.35 (feasible), in the case of the IGE of the geometric mean; and 0.30 (ideal) and 0.29 (feasible), in the case of the IGE of the expectation. The correlations between widths are 0.95 and 0.94, respectively.

References

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

Benabou, Roland. 2000. "Unequal Societies: Income Distribution and the Social Contract." *American Economic Review* 90(1): 96-129.

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

Björklund, Anders and Markus Jäntti. 2011. "Intergenerational Income Mobility and the Role of Family Background". *The Oxford Handbook of Economic Inequality*, edited by B. Nolan, W. Salverda and T. Smeeding. Oxford: Oxford University Press.

Black, Sandra and Paul Devereux. 2011. "Recent Developments in Intergenerational Mobility." *Handbook of Labor Economics*, Volume 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.

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.

Birnbaum, Allan. 1964. "Median-Unbiased Estimators." *Bulletin of Mathematical Statistics* 11:25-34.

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 (March) C72-C92.

Canay, Ivan and Azeem Shaikh. 2017. "Practical and Theoretical Advances for Inference in Partially Identified Models." *Advances in Economics and Econometrics*, Vol. 2, edited by B. Honoré, A. Pakes, M. Piazzesi and L. Samuelson. Cambridge: Cambridge University Press.

Carroll, Raymond, David Ruppert, Leonard Stefanski, and Ciprian Crainiceanu. 2006. *Measurement Error in Nonlinear Models: A Modern Perspective* (second edition). Boca Raton: Chapman & Hall/CRC.

Chernozhukov, Victor, Sokbae Lee and Adam Rosen. 2013. "Intersection Bounds: Estimation and Inference." *Econometrica* 81(2): 667-737.

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, IZA.

Dahl, Molly and Thomas DeLeire. 2008. “The Association between Children’s Earnings and Fathers’ Lifetime Earnings: Estimates Using Administrative Data.” Institute for Research on Poverty Discussion Paper 1342-08, University of Wisconsin-Madison.

Durlauf, Steven and Ananth Seshadri. 2018. “Understanding the Great Gatsby Curve.” In *NBER Macroeconomics Annual 2017*, Vol. 32, edited by Martin Eichenbaum and Jonathan Parker. Chicago: University of Chicago Press.

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

Hardin, James. 2002. “The Robust Variance Estimator for Two-Stage Models.” *The Stata Journal* 2(3): 253-266.

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.

Hirano, Keisuke and Jack Porter. 2012. “Impossibility Results for Nondifferentiable Functionals.” *Econometrica* 80(4): 1769–1790.

Ho, Kate and Adam Rosen. 2017. “Partial Identification in Applied Research: Benefits and Challenges.” *Advances in Economics and Econometrics,* Vol. 2, edited by B. Honoré, A. Pakes, M. Piazzesi and L. Samuelson. Cambridge: Cambridge University Press.

Huber, P. J. 1967. “The Behavior of Maximum Likelihood Estimates under Nonstandard Conditions.” *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol 1, edited by Lucien M. Le Cam and Jerzy Neyman. Berkeley, CA: University of California Press.

Imbens, Guido and Charles Manski. 2004. “Confidence Intervals for Partially Identified Parameters.” *Econometrica* 72(6): 1845-1857.

Landersø, Rasmus and James Heckman. 2016. “The Scandinavian Fantasy: The Sources of Intergenerational Mobility in Denmark and the U.S.” NBER Working Paper 22465.

Manski, Charles. 2003. *Partial Identification of Probability Distributions*. New York: Springer-Verlag.

Manski, Charles. 2008. “Partial Identification in Econometrics.” *The New Palgrave Dictionary of Economics*, 2nd edition, edited by Steven Durlauf and Lawrence Blume. London: Palgrave Macmillan.

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.

Mullahy, John. 1997. “Instrumental-Variable Estimation of Count Data Models: Applications to Models of Cigarette Smoking Behavior.” *Review of Economics and Statistics* 79(4): 586-593.

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

Mitnik, Pablo. 2017. “Estimating the Intergenerational Elasticity of Expected Income with Short-Run Income Measures: A Generalized Error-in-Variables Model.” Stanford Center on Poverty and Inequality Working Paper.

Mitnik, Pablo. 2018. “Cross-National and Temporal Variation in the Intergenerational Elasticity of Expected Income: A Theoretical Model.” Stanford Center on Poverty and Inequality Working Paper.

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

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

Mitnik, Pablo and David Grusky. The Intergenerational Elasticity of What? The Case for Redefining the Workhorse Measure of Economic Mobility.” Stanford Center on Poverty and Inequality Working Paper.

Mulligan, Casey. 1997. *Parental Priorities and Economic Inequality*. Chicago: University of Chicago Press.

Murphy, Kevin and Robert Topel. 1985. “Estimation and Inference in Two-Step Econometric Models.” *Journal of Business & Economic Statistics* 3(4): 370-379.

Nevo, Aviv and Adam Rosen. 2012. “Identification with Imperfect Instruments.” *Review of Economics and Statistics* 94(3): 659-671.

Ng, Irene. 2007. “Intergenerational Income Mobility in Singapore,” *The B.E. Journal of Economic Analysis & Policy* 7(2): Article 3.

Nybom, Martin and Jan Stuhler. 2011. “Heterogeneous Income Profiles and Life-Cycle Bias in Intergenerational Mobility Estimation.” IZA Working Paper 5697.

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.

Nybom, Martin and Jan Stuhler. 2017. “Biases in Standard Measures of Intergenerational Income Dependence.” *The Journal of Human Resources* 52(3):.800-825.

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

Reiss, Peter. 2016, “Just How Sensitive are Instrumental Variable Estimates?”. *Foundations and Trends in Accounting* 10(2-4): 204–237.

Santos Silva, J. M. C. and Silvna 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.

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*, Volume 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.

Stoye, Jörg. 2009. “More on Confidence Regions for Partially Identified Parameters.” *Econometrica* 77(4):1299–1315.

Tamer, Elie. 2010. “Partial Identification in Econometrics.” *The Annual Review of Economics* 2: 167-195.

Terza, Joseph, Anirban Basu, Paul J. Rathouz. 2008. “Two-Stage Residual Inclusion Estimation: Addressing Endogeneity in Health Econometric Modeling.” *Journal of Health Economics* 27(3):531-543.

Windmeijer, F. A. G. and J. M. C. Santos Silva. 1997. “Endogeneity in Count Data Models: An Applications to Demand for Health Care.” *Journal of Applied Econometrics* 12: 281-294.

Wooldridge, Jeffrey. 1999. “Quasi-Likelihood Methods for Count Data.” In *Handbook of Applied Econometrics Volume II: Microeconomics*, edited by M. Hashem Pesaran and Peter Schmidt. Blackwell Publishing.

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

Zimmerman, David J. 1992. “Regression toward Mediocrity in Economic Stature.” *American Economic Review* 82 (3), pp. 409-429.

Table 1: Descriptive Statistics (unweighted values)

Child's gender (% female)	51.7	Father's occupation (%)			
Child's age		Management occs.	10.2	Construction trades	10.2
Mean	36.6	Business operations specialists	0.6	Extraction workers	0.4
Standard deviation	0.6	Financial specialists	1.8	Installation, maintenance, and repair workers	7.5
Child's family income		Computer and mathematical occs.	1.1	Production occupations	12.8
Mean	91,625	Architecture and engineering occs.	3.3	Transportation and material moving occs.	10.2
Standard deviation	88,398	Life, physical, and social science occs.	0.9	Military specific occupations	2.5
Average parental age		Community and social services occs.	0.9	Doesn't know, refused	11.6
Mean	40.1	Legal occupations	0.4	Not applicable	1.5
Standard deviation	6.7	Education, training, and library occs.	2.7		
Average parental income		Arts, design, entert., sports, and media occs.	1.6		
Mean	78,858	Healthcare practitioners and technical occs.	1.6		
Standard deviation	64,185	Protective service occupations	2.8		
Parents' years of education at child age 15		Food preparation and serving occupations	0.5		
Mean	21.8	Building and grounds clean. and maint. occs.	1.5		
Standard deviation	7.0	Personal care and service occs.	0.4		
Household head' years of education at child age 15		Sales occupations	8.6		
Mean	12.4	Office and administrative support occs.	3.0		
Standard deviation	3.0	Farming, fishing, and forestry occs.	1.9		
N	827				

Note: Monetary values in 2012 dollars (adjusted by inflation using the Consumer Price Index for Urban Consumers - Research Series). The average parental age and income pertain to when the children were 1-25 years old. The occupation is coded as "not applicable" when there was no father/surrogate, the father was deceased, or the father never worked

Table 2: Long-run estimates and set estimates when measurement-error slopes are equal to one

	Parental information				
	Long-run	One year	Three years	Five years	Seven years
IGE of geometric mean of children's income					
Long-run estimate	**0.70**				
	(0.58 - 0.82)				
Set estimates, with short-run income instrumented by:					
Parents' education when income was measured		**0.46 - 0.79**	**0.55 - 0.81**	**0.58 - 0.84**	**0.60 - 0.86**
Household head's education when income was measured		**0.46 - 1.11**	**0.55 - 1.11**	**0.58 - 1.09**	**0.60 - 1.09**
Parents' education when child was 15 yeard old		**0.46 - 0.90**	**0.55 - 0.91**	**0.58 - 0.91**	**0.60 - 0.91**
Household head's education when child was 15 years old		**0.46 - 1.05**	**0.55 - 1.02**	**0.58 - 1.01**	**0.60 - 1.01**
Father's occupation when child was growing up		**0.46 - 0.76**	**0.55 - 0.74**	**0.58 - 0.78**	**0.60 - 0.79**
Average parental age	**40.5**	**37.0**	**37.6**	**37.8**	**38.1**
IGE of children's expected income					
Long-run estimate	**0.60**				
	(0.45 - 0.74)				
Set estimates, with short-run income instrumented by:					
Parents' education when income was measured		**0.46 - 0.74**	**0.52 0.75**	**0.55 - 0.77**	**0.56 - 0.78**
Household head's education when income was measured		**0.46 - 0.95**	**0.52 0.93**	**0.55 - 0.93**	**0.56 - 0.92**
Parents' education when child was 15 yeard old		**0.46 - 0.80**	**0.52 0.80**	**0.55 - 0.81**	**0.56 - 0.81**
Household head's education when child was 15 years old		**0.46 - 0.92**	**0.52 0.89**	**0.55 - 0.90**	**0.56 - 0.90**
Father's occupation when child was growing up		**0.46 - 0.69**	**0.52 0.70**	**0.55 - 0.73**	**0.56 - 0.74**
Average parental age	**40.5**	**37.0**	**37.6**	**37.8**	**38.1**

Note: The set estimates were computed by interpolation. Point and set estimates are in bold, 95 percent confidence intervals (for the long-run estimates only) are in parentheses.

Table 3a: Long-run estimate and set estimates of the IGE of the geometric mean when average parental age is close to 40

	Parental information				
	Long-run	One year	Three years	Five years	Seven years
Long-run estimate	**0.7**				
	(0.58 - 0.82)				
Set estimates, with short-run income instrumented by:					
Parents' education when income was measured		**0.40 - 0.76**	**0.52 - 0.79**	**0.55 - 0.78**	**0.57 - 0.78**
		(0.29 - 0.90)	(0.44 - 0.94)	(0.47 - 0.90)	(0.48 - 0.90)
Household head's education when income was measured		**0.40 - 0.93**	**0.52 - 0.93**	**0.55 - 0.98**	**0.57 - 0.98**
		(0.29 - 1.12)	(0.44 - 1.11)	(0.47 - 1.17)	(0.48 - 1.17)
Parents' education when child was 15 yeard old		**0.40 - 0.82**	**0.52 - 0.82**	**0.55 - 0.82**	**0.57 - 0.82**
		(0.29 - 0.98)	(0.44 - 0.97)	(0.47 - 0.97)	(0.48 - 0.97)
Household head's education when child was 15 years old		**0.40 - 0.93**	**0.52 - 0.93**	**0.55 - 0.92**	**0.57 - 0.93**
		(0.29 - 1.11	(0.44 - 1.11)	(0.47 - 1.10)	(0.48 - 1.10)
Father's occupation when child was growing up		**0.40 - 0.80**	**0.52 - 0.82**	**0.55 - 0.83**	**0.57 - 0.84**
		(0.29 - 0.97)	(0.44 - 0.99)	(0.47 - 1.00)	(0.48 - 1.01)
η_1 estimates		**1.10**	**1.08**	**1.07**	**1.08**
		(1.02 - 1.18)	(1.01 - 1.15)	(1.02 - 1.13)	(1.03 - 1.12)
Set estimates based on all upper-bound estimates					
Two-step estimator		**0.40 - 0.76**	**0.52 - 0.79**	**0.55 - 0.78**	**0.57 - 0.78**
		(0.29 - 0.94)	(0.44 - 0.98)	(0.47 - 0.94)	(0.48 - 0.94)
Three-step estimator		**0.40 - 0.82**	**0.52 - 0.85**	**0.55 - 0.83**	**0.57 - 0.83**
		(0.29 - 0.94)	(0.44 - 0.98)	(0.47 - 0.94)	(0.48 - 0.94)

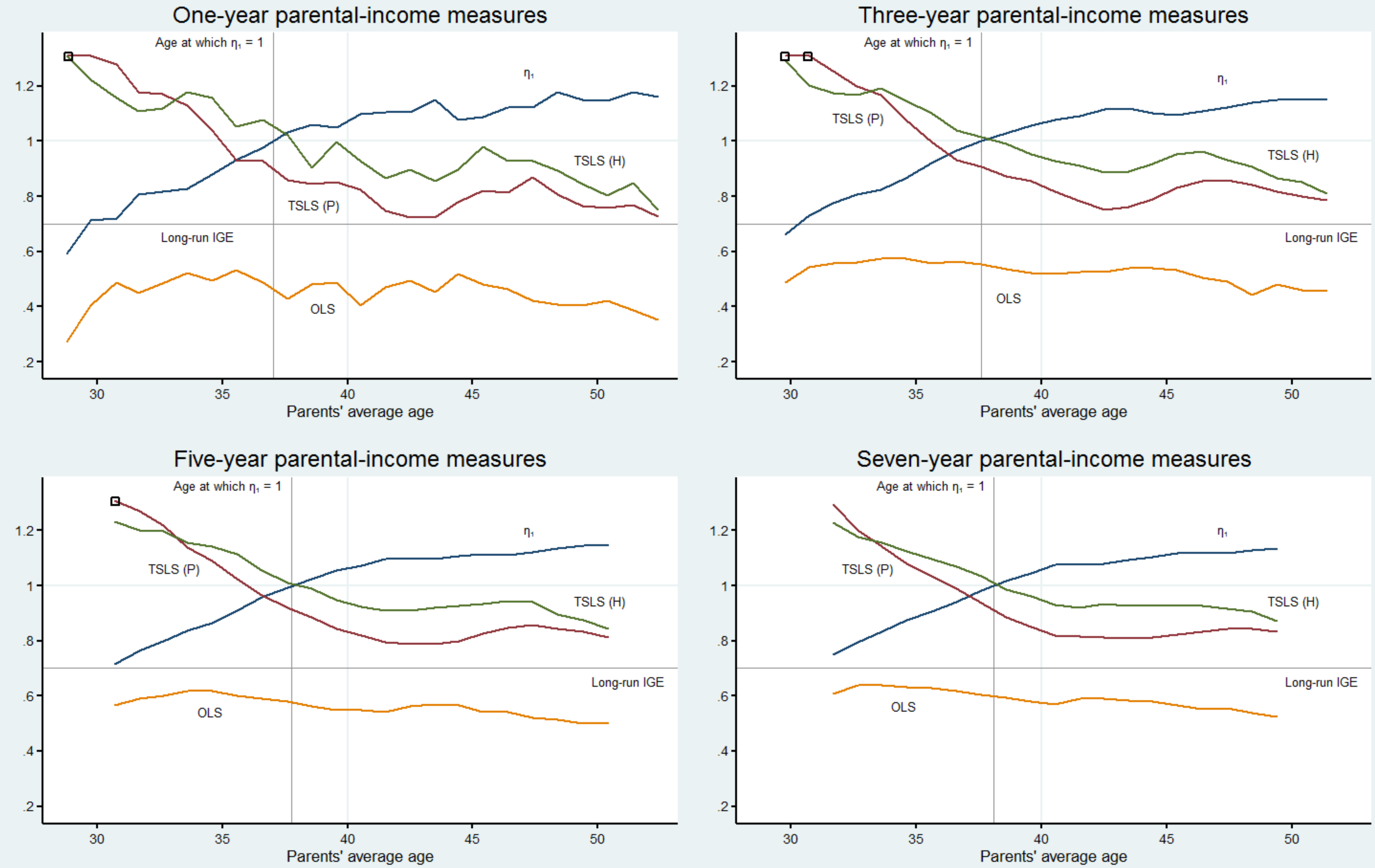
Note: The set estimates are based on measures of parental income centered on the years the children were 13 years old. Point and set estimates are in bold, 95 percent confidence intervals are in parentheses.

Table 3b: Long-run estimate and set estimates of the IGE of the expectation when average parental age is close to 40

	Parental information				
	Long-run	One year	Three years	Five years	Seven years
Long-run estimate	**0.6**				
	(0.45 - 0.74)				
Set estimates, with short-run income instrumented by:					
Parents' education when income was measured		**0.40 - 0.72**	**0.48 - 0.74**	**0.51 - 0.73**	**0.52 - 0.70**
		(0.31 - 0.86)	(0.39 - 0.87)	(0.41 - 0.87)	(0.41 - 0.84)
Household head's education when income was measured		**0.40 - 0.85**	**0.48 - 0.84**	**0.51 - 0.85**	**0.52 - 0.84**
		(0.31 - 1.04)	(0.39 - 1.00)	(0.41 - 1.01)	(0.41 - 0.99)
Parents' education when child was 15 yeard old		**0.40 - 0.77**	**0.48 - 0.75**	**0.51 - 0.74**	**0.52 - 0.72**
		(0.31 - 0.94)	(0.39 - 0.91)	(0.41 - 0.89)	(0.41 - 0.87)
Household head's education when child was 15 years old		**0.40 - 0.85**	**0.48 - 0.84**	**0.51 - 0.84**	**0.52 - 0.82**
		(0.31 - 1.04)	(0.39 - 1.01)	(0.41 - 1.00)	(0.41 - 0.98)
Father's occupation when child was growing up		**0.40 - 0.68**	**0.48 - 0.68**	**0.51 - 0.69**	**0.52 - 0.70**
		(0.31 - 0.82)	(0.39 - 0.83)	(0.41 - 0.82)	(0.41 - 0.85)
π_1 estimates		**1.10**	**1.08**	**1.07**	**1.08**
		(1.02 - 1.18)	(1.01 - 1.15)	(1.02 - 1.13)	(1.03 - 1.12)
Set estimates based on all upper-bound estimates					
Two-step estimator		**0.40 - 0.68**	**0.48 - 0.68**	**0.51 - 0.69**	**0.52 - 0.70**
		(0.31 - 0.87)	(0.38 - 0.88)	(0.41 - 0.87)	(0.41 - 0.88)
Three-step estimator		**0.40 - 0.74**	**0.48 - 0.74**	**0.51 - 0.75**	**0.52 - 0.76**
		(0.31 - 0.87)	(0.39 - 0.88)	(0.41 - 0.87)	(0.41 - 0.88)

Note: The set estimates are based on measures of parental income centered on the years the children were 13 years old. Point and set estimates are in bold, 95 percent confidence intervals are in parentheses.

Figure 1: Estimates of the IGE of the geometric mean of children's income, OLS and TSLS estimators
Instruments: Parents' education (P) and household head's education (H) when children were 15 years old



Note: The true values of the estimates identified by squares are larger than shown. They were replaced by the value 1.31 to improve the graphical representation of the results.

Figure 2: Estimates of the IGE of the geometric mean of children's income, OLS and TSLS estimators
Instrument: Father's occupation (as reported by children)

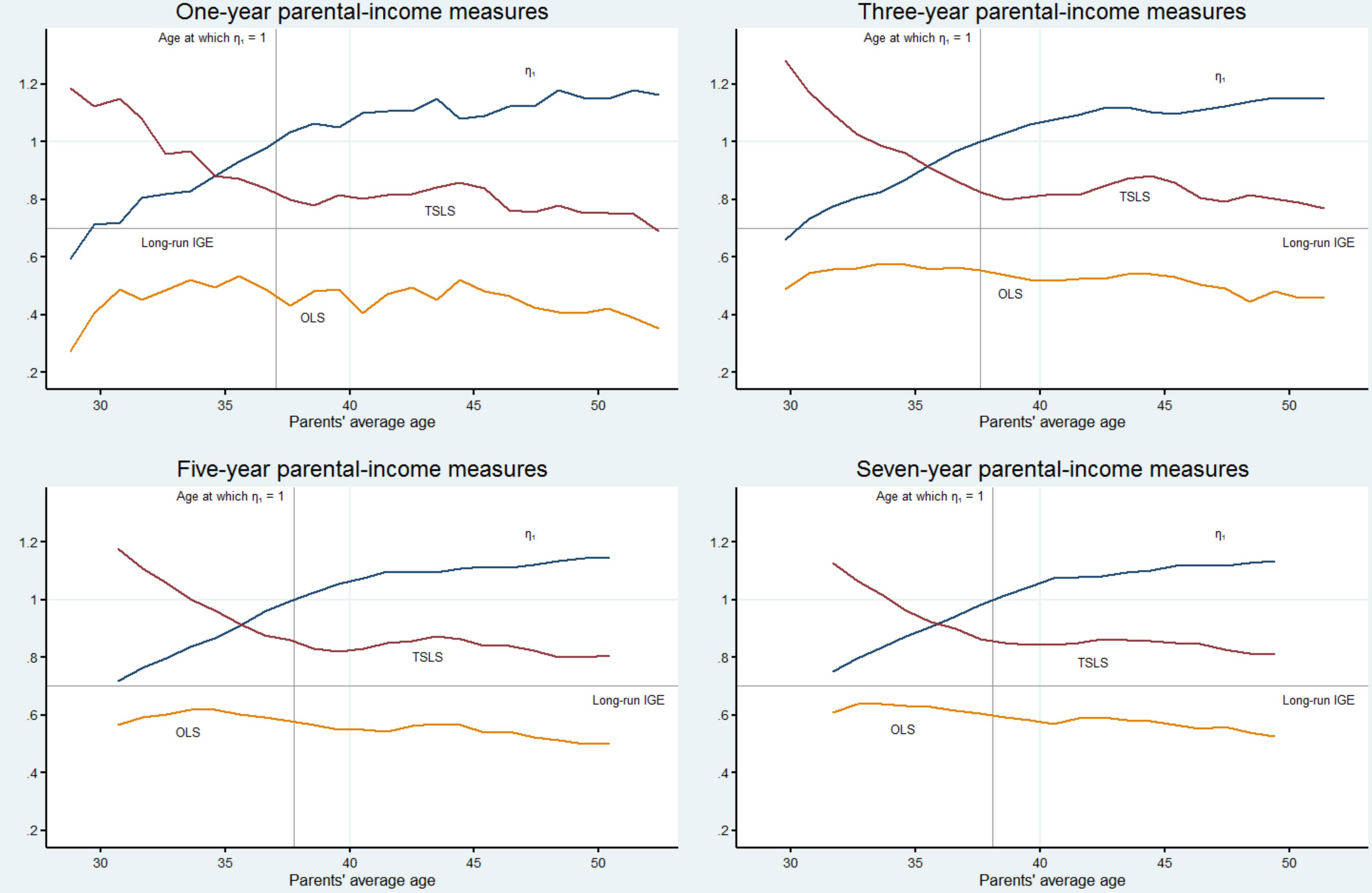
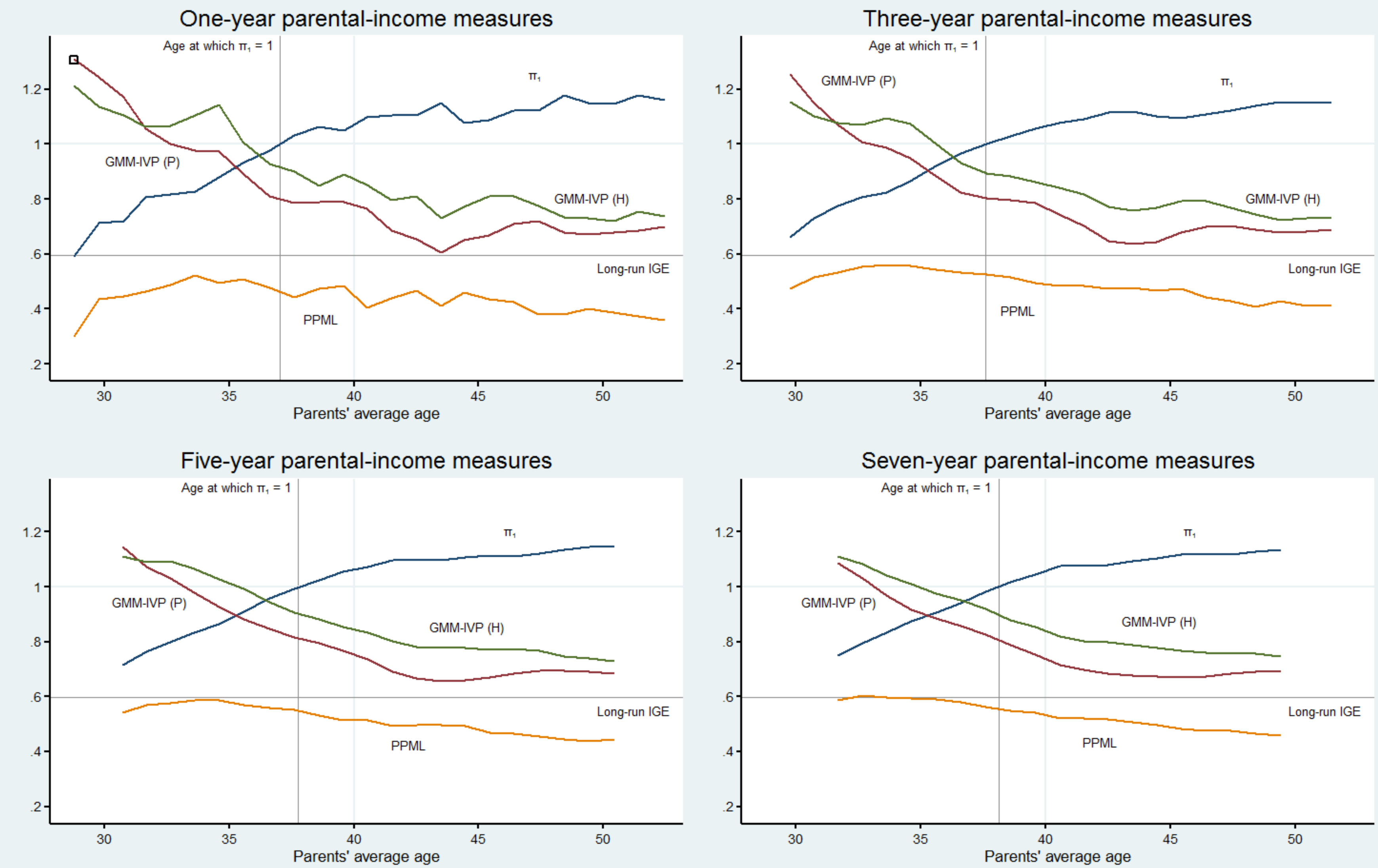
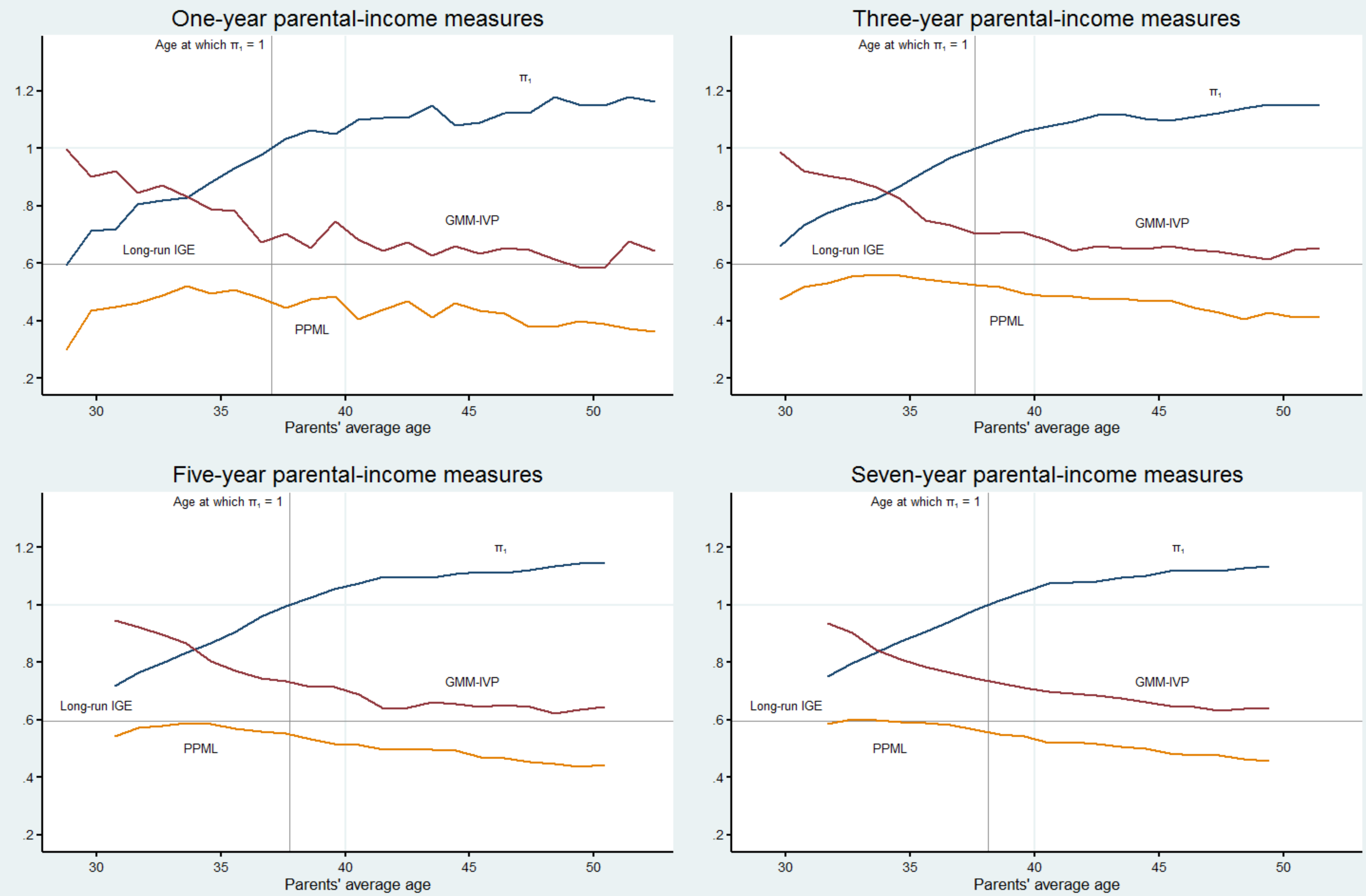


Figure 3: Estimates of the IGE of children's expected income, PPML and GMM-IVP estimators
Instruments: Parents' education (P) and household head's education (H) when children were 15 years old



Note: The true value of the estimate identified by a square is larger than shown. It was replaced by the value 1.31 to improve the graphical representation of the results.

Figure 4: Estimates of the IGE of children's expected income, PPML and GMM-IVP estimators
Instrument: Father's occupation (as reported by children)



Appendices

INTERGENERATIONAL INCOME ELASTICITIES, INSTRUMENTAL VARIABLE ESTIMATION, AND BRACKETING STRATEGIES

Pablo A. Mitnik
(pmitnik@stanford.edu)

March, 2019

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

A. Mathematical proofs for results presented in the main text

I refer to equations presented in the main text using the equation numbers employed there. When equations from the main text are reproduced in this appendix, I rely on their original numbers in the main text.

IV estimation of the IGE of the geometric mean

I derive here Equation [12]. In doing so I use Equations [3] and [4] and the fact that, because these are linear projections, it follows that $E(V) = 0$, $Cov(V, \ln Y) = 0$, $E(Q) = 0$, and $Cov(Q, \ln X) = 0$. I also use that the relationship between β_1 and the parameters in Equation [9] is provided by the omitted-variable formula:

$$\beta_1 = \varrho_1 + \varrho_2 \frac{Cov(\ln X, L)}{Var(\ln X)}.$$

The probability limit of the IV estimator in the general case can be derived as follows:

$$\ddot{\beta}_1 = \frac{Cov(\ln Z, L)}{Cov(\ln S, L)}$$

$$= \frac{Cov(\lambda_1 \ln Y + V, L)}{Cov(\ln S, L)}$$

$$= \frac{Cov(\lambda_1 \varrho_0 + \lambda_1 \varrho_1 \ln X + \lambda_1 \varrho_2 L + \lambda_1 \kappa, L)}{Cov(\ln S, L)} + \frac{Cov(V, L)}{Cov(\ln S, L)}$$

$$= \lambda_1 \varrho_1 \frac{Cov(\ln X, L)}{Cov(\ln S, L)} + \lambda_1 \varrho_2 \frac{Var(L)}{Cov(\ln S, L)} + \frac{Cov(V, L)}{Cov(\ln S, L)}$$

$$= \frac{\lambda_1}{\eta_1} \varrho_1 \frac{Cov(\ln S - Q, L)}{Cov(\ln S, L)} + \lambda_1 \varrho_2 \frac{Var(L)}{Cov(\ln S, L)} + \frac{Cov(V, L)}{Cov(\ln S, L)}$$

$$= \frac{\lambda_1}{\eta_1} \left[\beta_1 - \varrho_2 \frac{Cov(\ln X, L)}{Var(\ln X)}\right] + \lambda_1 \varrho_2 \frac{Var(L)}{Cov(\ln S, L)} + \frac{Cov(V, L)}{Cov(\ln S, L)} - \frac{\lambda_1}{\eta_1} \varrho_1 \frac{Cov(Q, L)}{Cov(\ln S, L)}$$

$$= \frac{\lambda_1}{\eta_1} \beta_1 + \frac{\lambda_1}{\eta_1} \varrho_2 \left[\frac{\eta_1 Var(L)}{Cov(\ln S, L)} - \frac{Cov(\ln X, L)}{Var(\ln X)}\right] + \frac{Cov(V, L)}{Cov(\ln S, L)} - \frac{\lambda_1}{\eta_1} \varrho_1 \frac{Cov(Q, L)}{Cov(\ln S, L)}. \quad [A1]$$

Focusing for now on the term in square brackets:

$$\frac{\eta_1 Var(L)}{Cov(\ln S\,, L)} - \frac{Cov(\ln X\,, L)}{Var(\ln X)} = \frac{\eta_1 Var(L)}{\eta_1 Cov(\ln X\,, L) + Cov(Q, L)} - \frac{Cov(\ln X\,, L)}{Var(\ln X)}$$

$$= \frac{1}{\frac{SD(\ln X)}{SD(L)} Corr(\ln X\,, L) + \frac{Cov(Q, L)}{\eta_1 Var(L)}} - \frac{SD(L)}{SD(\ln X)} Corr(\ln X\,, L)$$

$$= \frac{SD(L)}{SD(\ln X)} \left[\frac{1}{Corr(\ln X\,, L) + \frac{SD(L)}{SD(\ln X)} \frac{Cov(Q, L)}{\eta_1 Var(L)}} - Corr(\ln X\,, L) \right]$$

$$= \frac{SD(L)}{SD(\ln X)} \left[\frac{1 - [Corr(\ln X\,, L)]^2 - \frac{Cov(Q, L)}{\eta_1 SD(\ln X)\, SD(L)} Corr(\ln X\,, L)}{Corr(\ln X\,, L) + \frac{Cov(Q, L)}{\eta_1 SD(\ln X)\, SD(L)}} \right]$$

$$= \frac{SD(L)}{SD(\ln X)} \left[\frac{1 - [Corr(\ln X\,, L)]^2}{Corr(\ln X\,, L) + \frac{Cov(Q, L)}{\eta_1 SD(\ln X)\, SD(L)}} \right]$$

$$- \frac{SD(L)}{SD(\ln X)} \left[\frac{\frac{Cov(Q, L)}{\eta_1 SD(\ln X)\, SD(L)} Corr(\ln X\,, L)}{Corr(\ln X\,, L) + \frac{Cov(Q, L)}{\eta_1 SD(\ln X)\, SD(L)}} \right]$$

$$= \frac{SD(L)}{SD(\ln X)} \left[\frac{1 - [Corr(\ln X\,, L)]^2}{Corr(\ln X\,, L) + \frac{Cov(Q, L)}{\eta_1 SD(\ln X)\, SD(L)}} \right]$$

$$- \frac{SD(L)}{SD(\ln X)} \left[\frac{\frac{Cov(Q, L)}{\eta_1 SD(\ln X)\, SD(L)} Corr(\ln X\,, L)}{\frac{Cov(\ln S\,, L)}{\eta_1 SD(\ln X)\, SD(L)}} \right]$$

$$= \frac{SD(L)}{SD(\ln X)} \left[\frac{1 - [Corr(\ln X\,, L)]^2}{Corr(\ln X\,, L) + \frac{Cov(Q, L)}{\eta_1 SD(\ln X)\, SD(L)}} \right] - \frac{SD(L)}{SD(\ln X)} Corr(\ln X\,, L) \frac{Cov(Q, L)}{Cov(\ln S\,, L)}$$

$$= \frac{SD(L)}{SD(\ln X)}\left[\frac{1-[Corr(\ln X, L)]^2}{Corr(\ln X, L)+\frac{Cov(Q,L)}{\eta_1 SD(\ln X)\, SD(L)}}\right]-\frac{Cov(\ln X, L)}{Var(\ln X)}\frac{Cov(Q,L)}{Cov(\ln S, L)}. \quad [A2]$$

Finally, substituting Equation [A2] into Equation [A1]:

$$\ddot{\beta}_1 = \frac{\lambda_1}{\eta_1}\beta_1 + \frac{\lambda_1}{\eta_1}\varrho_2\frac{SD(L)}{SD(\ln X)}\left[\frac{1-[Corr(\ln X, L)]^2}{Corr(\ln X, L)+\frac{Cov(Q,L)}{\eta_1 SD(\ln X)\, SD(L)}}\right]$$

$$-\frac{\lambda_1}{\eta_1}\varrho_2\frac{Cov(\ln X, L)}{Var(\ln X)}\frac{Cov(Q,L)}{Cov(\ln S, L)}+\frac{Cov(V,L)}{Cov(\ln S, L)}-\frac{\lambda_1}{\eta_1}\varrho_1\frac{Cov(Q,L)}{Cov(\ln S, L)}$$

$$= \frac{\lambda_1}{\eta_1}\beta_1 + \frac{\lambda_1}{\eta_1}\varrho_2\frac{SD(L)}{SD(\ln X)}\left[\frac{1-[Corr(\ln X, L)]^2}{Corr(\ln X, L)+\frac{Cov(Q,L)}{\eta_1 SD(\ln X)\, SD(L)}}\right]+\frac{Cov(V,L)}{Cov(\ln S, L)}$$

$$-\frac{\lambda_1}{\eta_1}\frac{Cov(Q,L)}{Cov(\ln S, L)}\left[\varrho_1+\varrho_2\frac{Cov(\ln X, L)}{Var(\ln X)}\right]$$

$$= \frac{\lambda_1}{\eta_1}\beta_1 + \frac{\lambda_1}{\eta_1}\varrho_2\frac{SD(L)}{SD(\ln X)}\left[\frac{1-[Corr(\ln X, L)]^2}{Corr(\ln X, L)+\frac{Cov(Q,L)}{\eta_1 SD(\ln X)\, SD(L)}}\right]+\frac{Cov(V,L)}{Cov(\ln S, L)}$$

$$-\frac{\lambda_1}{\eta_1}\beta_1\frac{Cov(Q,L)}{Cov(\ln S, L)}$$

$$\ddot{\beta}_1 = \frac{\lambda_1}{\eta_1}\left\{\beta_1\left[1-\frac{Cov(Q,L)}{Cov(\ln S, L)}\right]+\varrho_2\frac{SD(L)}{SD(\ln X)}\left[\frac{1-[Corr(\ln X, L)]^2}{Corr(\ln X, L)+\frac{Cov(Q,L)}{\eta_1 SD(\ln X)\, SD(L)}}\right]\right\}$$

$$+\frac{Cov(V,L)}{Cov(\ln S, L)}. \quad [12]$$

IV estimation of the IGE of the expectation

I derive here Equations [26], [27] and [28] and the sign of $\frac{\partial E(X^{\ddot{\alpha}_1}\ln X)\,[E(X^{\ddot{\alpha}_1})]^{-1}}{\partial \ddot{\alpha}_1}$. I assume, without any loss of generality, that $E(Y)=E(Z)=E(\ln X)=E(\ln S)=1$ (see note 8 in the main text) and that the instrument L has been demeaned. I use Equations [16] and [17] and the fact that, because these are linear projections, it follows that $E(W) = 0$, $Cov(W, Y) = 0$, $E(P) = 0$, and $Cov(P, \ln X) = 0$.

When resorting to the GMM-IVP estimator to estimate the IGE of the expectation with short-run measures, estimation is based on the sample analog of the following population moment conditions (where I assume only one instrument, L, is employed):

$$E\left(\left[Z - \exp(\ddot{\alpha}_0)\, S^{\ddot{\alpha}_1}\right]\right) = 0$$

$$E\left(\left[Z - \exp(\ddot{\alpha}_0)\, S^{\ddot{\alpha}_1}\right] L\right) = 0.$$

This means that $\ddot{\alpha}_1$ solves:

$$\frac{E(S^{\ddot{\alpha}_1} L)}{E(S^{\ddot{\alpha}_1})} = \frac{E(Z\ L)}{E(Z)}$$

$$\frac{E(S^{\ddot{\alpha}_1} L)}{E(S^{\ddot{\alpha}_1})} = E(Z\ L). \qquad [26]$$

Using Equations [16] and [17] to substitute S and Z out in Equation [26] yields:

$$\frac{E([\exp(\pi_0 + \pi_1 \ln X + P)]^{\ddot{\alpha}_1} L)}{E([\exp(\pi_0 + \pi_1 \ln X + P)]^{\ddot{\alpha}_1})} = E([\theta_0 + \theta_1 Y + W]\ L)$$

$$\frac{[\exp(\pi_0)]^{\ddot{\alpha}_1} E([\exp(\pi_1 \ln X)]^{\ddot{\alpha}_1} [\exp(P)]^{\ddot{\alpha}_1} L)}{[\exp(\pi_0)]^{\ddot{\alpha}_1} E([\exp(\pi_1 \ln X)]^{\ddot{\alpha}_1} [\exp(P)]^{\ddot{\alpha}_1})} = \theta_0 E(L\ ) + \theta_1 E(YL) + E(WL)$$

$$\frac{E(X^{\pi_1 \ddot{\alpha}_1} [\exp(P)]^{\ddot{\alpha}_1} L)}{E(X^{\pi_1 \ddot{\alpha}_1} [\exp(P)]^{\ddot{\alpha}_1})} = \theta_1 E(YL) + Cov(L, W)\ .$$

Let’s now define:

$$F(\ddot{\alpha}_1) = \frac{E\left(X^{\pi_1 \ddot{\alpha}_1} L\right)}{E(X^{\pi_1 \ddot{\alpha}_1})} \left\{\frac{E\left(X^{\pi_1 \ddot{\alpha}_1} [\exp(P)]^{\ddot{\alpha}_1} L\right)}{E(X^{\pi_1 \ddot{\alpha}_1} [\exp(P)]^{\ddot{\alpha}_1})}\right\}^{-1}.$$

We may now write:

$$\frac{E\left(X^{\pi_1\ddot{\alpha}_1}L\right)}{E(X^{\pi_1\ddot{\alpha}_1})} = [\theta_1 E(Y\,L) + Cov(L, W)]F(\ddot{\alpha}_1). \qquad [A3]$$

Let $L = \gamma_0 + \gamma_1 \ln X + \ddot{L}$ be the population linear projection of L on $\ln X$. Replacing L by this expression in Equation [A3] gives:

$$\frac{E\left(X^{\pi_1\ddot{\alpha}_1}[\gamma_0 + \gamma_1 \ln X + \ddot{L}]\right)}{E(X^{\pi_1\ddot{\alpha}_1})} = \left\{\theta_1 E\left(Y\,[\gamma_0 + \gamma_1 \ln X + \ddot{L}]\right) + Cov(L, W)\right\}F(\ddot{\alpha}_1)$$

$$\gamma_0 + \gamma_1 \frac{E\left(X^{\pi_1\ddot{\alpha}_1}\ln X\right)}{E(X^{\pi_1\ddot{\alpha}_1})} + \frac{E\left(X^{\pi_1\ddot{\alpha}_1}\ddot{L}\right)}{E(X^{\pi_1\ddot{\alpha}_1})}$$

$$= \left\{\theta_1\gamma_0 + \theta_1\gamma_1 E(\ln X\,Y) + \theta_1 E\left(\ddot{L}Y\right) + Cov(L, W)\right\}F(\ddot{\alpha}_1).$$

Using now $Y = \exp(\alpha_0)\,X^{\alpha_1} + \Psi$ to substitute Y out, we obtain:

$$\gamma_0 + \gamma_1 \frac{E\left(X^{\pi_1\ddot{\alpha}_1}\ln X\right)}{E(X^{\pi_1\ddot{\alpha}_1})} + \frac{E\left(X^{\pi_1\ddot{\alpha}_1}\ddot{L}\right)}{E(X^{\pi_1\ddot{\alpha}_1})}$$

$$= \left\{\theta_1\gamma_0 + \theta_1\,\gamma_1 E(\ln X\,Y) + \theta_1 E\left(\ddot{L}\,[\exp(\alpha_0)\,X^{\alpha_1} + \Psi]\right) + Cov(L, W)\right\}F(\ddot{\alpha}_1)$$

$$-\gamma_1 + \gamma_1 \frac{E\left(X^{\pi_1\ddot{\alpha}_1}\ln X\right)}{E(X^{\pi_1\ddot{\alpha}_1})} + \frac{E\left(X^{\pi_1\ddot{\alpha}_1}\ddot{L}\right)}{E(X^{\pi_{\ddot{1}}\alpha_1})}$$

$$= \left[-\gamma_1\theta_1 + \theta_1\gamma_1 E(\ln X\,Y) + \theta_1 \frac{E\left(X^{\alpha_1}\ddot{L}\right)}{E(X^{\alpha_1})} + \theta_1 E\left(\ddot{L}\Psi\right) + Cov(L, W)\right]F(\ddot{\alpha}_1),$$

where I have used $\exp(\alpha_0) = \frac{E(Y)}{E(X^{\alpha_1})}$ and $\gamma_0 = -\gamma_1$; the latter follows from the demeaning of L and $E(\ln X) = 1$.

As $E\left(\ddot{L}\right) = 0$ and $Cov\left(\ddot{L}, \ln X\right) = 0$ by construction, second-order Taylor-series approximations to $E\left(X^{\pi_1\ddot{\alpha}_1}\ddot{L}\right)$ and $E\left(X^{\alpha_1}\ddot{L}\right)$ around the expectations of $\ddot{L}$ and $\ln X$ are zero as well. We may then write:

$$-\gamma_1 + \gamma_1 \frac{E\left(X^{\pi_1\ddot{\alpha}_1}\ln X\right)}{E(X^{\pi_1\ddot{\alpha}_1})} \approx \left[-\gamma_1\theta_1 + \theta_1\,\gamma_1 E(\ln X\,Y) + \theta_1 E\left(\ddot{L}\Psi\right) + Cov(L, W)\right]F(\ddot{\alpha}_1)$$

$$\gamma_1\left[\frac{E\left(X^{\pi_1\ddot{\alpha}_1}\ln X\right)}{E(X^{\pi_1\ddot{\alpha}_1})}-1\right]$$

$$\approx\{\gamma_1\theta_1[E(\ln X\, Y)-1]+\theta_1 Cov(L-\gamma_0-\gamma_1\ln X\,,\Psi)+Cov(L,W)\}F(\ddot{\alpha}_1)$$

$$\frac{E\left(X^{\pi_1\ddot{\alpha}_1}\ln X\right)}{E(X^{\pi_1\ddot{\alpha}_1})}-1$$

$$\approx\left[\theta_1[E(\ln X\, Y)-1]+\frac{\theta_1 Cov(L,\Psi)}{\gamma_1}-\theta_1 Cov(\ln X\,,\Psi)+\frac{Cov(L,W)}{\gamma_1}\right]F(\ddot{\alpha}_1)$$

$$\frac{E\left(X^{\pi_1\ddot{\alpha}_1}\ln X\right)}{E(X^{\pi_1\ddot{\alpha}_1})}\approx 1+F(\ddot{\alpha}_1)\left[\theta_1 Cov(\ln X,Y)+\frac{\theta_1 Cov(L,\Psi)}{\gamma_1}+\frac{Cov(L,W)}{\gamma_1}\right]$$

$$\frac{E\left(X^{\pi_1\ddot{\alpha}_1}\ln X\right)}{E(X^{\pi_1\ddot{\alpha}_1})}$$

$$\approx E(\ln X\,,Y)\left\{\frac{1+F(\ddot{\alpha}_1)\left[\theta_1 Cov(\ln X,Y)+\frac{\theta_1 Cov(L,\Psi)}{\gamma_1}+\frac{Cov(L,W)}{\gamma_1}\right]}{1+Cov(\ln X\,,Y)}\right\},\qquad [A4]$$

where I have used $E(\ln X\,,Y)=1+Cov(\ln X\,,Y)$ and the fact that $E(\Psi|\mathrm{x})=0$ entails that $Cov(\ln X\,,\Psi)=0$.

Substituting γ_1 out in Equation [A4] yields:

$$\frac{E\left(X^{\pi_1\ddot{\alpha}_1}\ln X\right)}{E(X^{\pi_1\ddot{\alpha}_1})}$$

$$\approx E(\ln X\ \ Y)\left\{\frac{1+F(\ddot{\alpha}_1)\ \left[\theta_1 Cov(\ln X,Y)+Var(\ln X)\,\frac{\theta_1 Cov(L,\Psi)+Cov(L,\mathrm{W})}{Cov(L,\ln X)}\right]}{1+Cov(\ln X\,,Y)}\right\}.\qquad [27]$$

Computing now second-order Taylor-series approximations around $E(\ln X)=1$, $E(L)=0$, and $E(P)=0$ for the four expectations in $F(\ddot{\alpha}_1)$ gives:

$$E\left(X^{\pi_1\ddot{\alpha}_1}\right)=E\left([\exp(\ln X)]^{\pi_1\ddot{\alpha}_1}\right)$$

$$\approx\exp(\pi_1\ddot{\alpha}_1)+0.5\,[\pi_1\ddot{\alpha}_1]^2\exp(\pi_1\ddot{\alpha}_1)\,Var(\ln X)$$

$$E(X^{\pi_1\ddot{\alpha}_1}L) = E\big([\exp(\ln X)]^{\pi_1\ddot{\alpha}_1}L\big)$$

$$\approx 0 + 0.5\ 0\ \ Var(\ln X) + 0.5\ 0\ Var(L) + \exp(\pi_1\ddot{\alpha}_1)\,\pi_1\ddot{\alpha}_1\ Cov(\ln X\,, L)$$

$$\approx \exp(\pi_1\ddot{\alpha}_1)\ \ \pi_1\ddot{\alpha}_1\ Cov(\ln X\,, L)$$

$$E\big(X^{\pi_1\ddot{\alpha}_1}[\exp(P)]^{\ddot{\alpha}_1}\big) = \ E\big([\exp(\ln X)]^{\pi_1\ddot{\alpha}_1}[\exp(P)]^{\ddot{\alpha}_1}\big)$$

$$\approx \exp(\pi_1\ddot{\alpha}_1) + \ 0.5\ [\pi_1\ddot{\alpha}_1]^2\exp(\pi_1\ddot{\alpha}_1)\,Var(\ln X)$$

$$+\ 0.5\ [\ddot{\alpha}_1]^2\exp(\pi_1\ddot{\alpha}_1)\ \ Var(P) + [\ddot{\alpha}_1]^2\pi_1\exp(\pi_1\ddot{\alpha}_1)\ \ Cov(\ln X\,, P)$$

$$\approx \exp(\pi_1\ddot{\alpha}_1) + \ 0.5\ [\pi_1\ddot{\alpha}_1]^2\exp(\pi_1\ddot{\alpha}_1)\,Var(\ln X)$$

$$+\ 0.5\ [\ddot{\alpha}_1]^2\exp(\pi_1\ddot{\alpha}_1)\ \ Var(P) + [\ddot{\alpha}_1]^2\pi_1\exp(\ddot{\alpha}_1)\ \ 0$$

$$\approx\ \exp(\pi_1\ddot{\alpha}_1) + 0.5\ [\ddot{\alpha}_1]^2\exp(\pi_1\ddot{\alpha}_1)\,\{[\pi_1]^2Var(\ln X) + \ Var(P)\}$$

$$E\big(X^{\pi_1\ddot{\alpha}_1}[\exp(P)]^{\ddot{\alpha}_1}L\big) = \ E\big([\exp(\ln X)]^{\pi_1\ddot{\alpha}_1}[\exp(P)]^{\ddot{\alpha}_1}L\big)$$

$$\approx 0 + \ 0.5\ 0\ Var(\ln X) + 0.5\ 0\ Var(P) + 0.5\ 0\ Var(L) + 0\ Cov(\ln X\,, P)$$

$$+ \exp(\pi_1\ddot{\alpha}_1)\,\pi_1\ddot{\alpha}_1\ Cov(\ln X\,, L) + \exp(\pi_1\ddot{\alpha}_1)\ \ \ddot{\alpha}_1\ Cov(L, P)$$

$$\approx \exp(\pi_1\ddot{\alpha}_1)\ \ \ddot{\alpha}_1[\pi_1 Cov(\ln X\,, L) + Cov(L, P)]\,.$$

Substituting these approximations in the expression for $F(\ddot{\alpha}_1)$:

$$F(\ddot{\alpha}_1) \approx \frac{\dfrac{\exp(\pi_1\ddot{\alpha}_1)\ \ \pi_1\ddot{\alpha}_1\ Cov(\ln X\,, L)}{\exp(\pi_1\ddot{\alpha}_1) + 0.5\ [\pi_1\ddot{\alpha}_1]^2\exp(\pi_1\ddot{\alpha}_1)Var(\ln X)}}{\dfrac{\exp(\pi_1\ddot{\alpha}_1)\ \ \ddot{\alpha}_1[\pi_1 Cov(\ln X\,, L) + Cov(L, P)]}{\exp(\pi_1\ddot{\alpha}_1) + 0.5\ [\ddot{\alpha}_1]^2\exp(\pi_1\ddot{\alpha}_1)\,\{[\pi_1]^2Var(\ln X) + \ Var(P)\}}}$$

$$\approx \frac{\dfrac{\pi_1\ Cov(\ln X\,, L)}{1 + 0.5\ [\pi_1\ddot{\alpha}_1]^2Var(\ln X)}}{\dfrac{\pi_1 Cov(\ln X\,, L) + Cov(L, P)}{1 + 0.5\ [\ddot{\alpha}_1]^2\{[\pi_1]^2Var(\ln X) + \ Var(P)\}}}$$

$$\approx \frac{1 + 0.5\ \{[\pi_1\ddot{\alpha}_1]^2Var(\ln X) + \ [\ddot{\alpha}_1]^2\ Var(P)\}}{1 + 0.5\ [\pi_1\ddot{\alpha}_1]^2Var(\ln X)}\frac{\pi_1\ Cov(\ln X\,, L)}{\pi_1 Cov(\ln X\,, L) + Cov(L, P)}.$$

This is the result shown in Equation [28].

In the main text I used that $\frac{\partial E(X^{\ddot{\alpha}_1}\ln X)\,[E(X^{\ddot{\alpha}_1})]^{-1}}{\partial \ddot{\alpha}_1} > 0$, which I prove next. Employing integral expressions for expectations, Leibniz's rule for differentiation under the integral sign, and usual derivative rules, we have:

$$\frac{\partial E\left(X^{\ddot{\alpha}_1}\ln X\right)\left[E\left(X^{\ddot{\alpha}_1}\right)\right]^{-1}}{\partial \ddot{\alpha}_1}$$

$$= \frac{E\left(X^{\ddot{\alpha}_1}\right)\frac{dE\left(X^{\ddot{\alpha}_1}\ln X\right)}{d\ddot{\alpha}_1} - E\left(X^{\ddot{\alpha}_1}\ln X\right)\frac{dE\left(X^{\ddot{\alpha}_1}\right)}{d\ddot{\alpha}_1}}{[E(X^{\ddot{\alpha}_1})]^2}$$

$$= \frac{E\left(X^{\ddot{\alpha}_1}\right)\int_{x>0}\frac{dX^{\ddot{\alpha}_1}\ln X}{d\ddot{\alpha}_1}f_X(x)dx - E\left(X^{\ddot{\alpha}_1}\ln X\right)\int_{x>0}\frac{dX^{\ddot{\alpha}_{1_1}}}{d\ddot{\alpha}_1}f_X(x)dx}{[E(X^{\ddot{\alpha}_1})]^2}$$

$$= \frac{E\left(X^{\ddot{\alpha}_1}\right)E\left(X^{\ddot{\alpha}_{1_1}}[\ln X]^2\right) - E\left(X^{\ddot{\alpha}_1}\ln X\right)E\left(X^{\ddot{\alpha}_1}\ln X\right)}{[E(X^{\ddot{\alpha}_1})]^2}.$$

I need to show that the numerator of the last expression is positive, and to this end I use a "symmetrization trick." Let's define the following expectation:

$$C_X \equiv E\left(\left[\ln \dot{X} - \ln X\right]^2 \dot{X}^{\ddot{\alpha}_1}X^{\ddot{\alpha}_1}\right),$$

were $\dot{X}$ is an independent copy of X. It is clearly the case that $C_X > 0$. I show next that C_X is twice the numerator in question, which entails that the latter is positive:

$$\begin{aligned}
C_X &= E\left(\left[\ln \dot{X} - \ln X\right]\left[\ln \dot{X} - \ln X\right]\dot{X}^{\ddot{\alpha}_1}X^{\ddot{\alpha}_1}\right)\\
&= E\left(\left[\dot{X}^{\ddot{\alpha}_1}\ln \dot{X} - \dot{X}^{\ddot{\alpha}_1}\ln X\right]\left[X^{\ddot{\alpha}_1}\ln \dot{X} - X^{\ddot{\alpha}_1}\ln X\right]\right)\\
&= E\left(X^{\ddot{\alpha}_{1_1}}\dot{X}^{\ddot{\alpha}_1}\ln \dot{X}\,\ln \dot{X} - X^{\ddot{\alpha}_1}\ln X\,\dot{X}^{\ddot{\alpha}_1}\ln \dot{X} - \dot{X}^{\ddot{\alpha}_1}\ln \dot{X}\,X^{\ddot{\alpha}_{1_1}}\ln X + \dot{X}^{\ddot{\alpha}_1}X^{\ddot{\alpha}_1}\ln X\ln X\right)\\
&= E\left(X^{\ddot{\alpha}_1}\right)E\left(\dot{X}^{\ddot{\alpha}_1}\ln \dot{X}\ln \dot{X}\right) - E\left(X^{\ddot{\alpha}_1}\ln X\right)E\left(\dot{X}^{\ddot{\alpha}_1}\ln \dot{X}\right) + E\left(\dot{X}^{\ddot{\alpha}_1}\right)E\left(X^{\ddot{\alpha}_1}\ln X\,\ln X\right) -\\
&\quad E\left(\dot{X}^{\ddot{\alpha}_1}\ln \dot{X}\right)E\left(X^{\ddot{\alpha}_1}\ln X\right)\\
&= 2\left[E\left(X^{\ddot{\alpha}_1}\right)E(X^{\ddot{\alpha}_1}\ln X\ln X) - E\left(X^{\ddot{\alpha}_1}\ln X\right)E(X^{\ddot{\alpha}_1}\ln X)\right],
\end{aligned}$$

which completes the proof.

B. A second generalized error-in-variables model for the IV estimation of the IGE of the expectation

I advance here a second generalized error-in-variables models for the IV estimation of the IGE of the expectation, in which I make the stronger assumptions typically employed in the literature on measurement error in nonlinear models (e.g., Carroll et al. 2006). As I indicated in the main text, both measurement-error models lead to the conclusion that a bracketing strategy is as feasible with the IGE of the expectation as with the IGE of the geometric mean. However, the models have different implications in other respects, which I discuss here. I refer to the second measurement-error model as the GEiVE-IV-S model (the "S" stands for "stronger assumptions").

I refer to equations presented in the main text using the equation numbers employed there. When equations from the main text are reproduced in this appendix, I rely on their original numbers in the main text.

Like the GEiVE-IV model, the GEiVE-IV-S model makes the following empirical assumption:

$$Cov(W_t, L) = 0. \qquad [23]$$

Instead of assumption [24], however, the GEiVE-IV-S model assumes

$$P_k \perp X \qquad [B1]$$

$$P_k \perp L, \qquad [B2]$$

that is, that the measurement error in the log of the short-run parental-income variable is independent of both the long-run parental income and the instrument. (From here on, I drop the subscripts *t* and *k*.)

Let's recall what $F(\ddot{\alpha}_1)$ denotes:

$$F(\ddot{\alpha}_1) = \frac{E\left(X^{\pi_1\ddot{\alpha}_1}L\right)}{E(X^{\pi_1\ddot{\alpha}_1})}\left\{\frac{E\left(X^{\pi_1\ddot{\alpha}_1}[\exp(P)]^{\ddot{\alpha}_1}L\right)}{E(X^{\pi_1\ddot{\alpha}_1}[\exp(P)]^{\ddot{\alpha}_1})}\right\}^{-1}.$$

Given that P is assumed independent of both L and X, and using $T = \exp(P)$ to simplify the notation, we have that the term inside curly brackets is:

$$\frac{E\left(X^{\pi_1\ddot{\alpha}_1}T^{\ddot{\alpha}_1}L\right)}{E(X^{\pi_1\ddot{\alpha}_1}T^{\ddot{\alpha}_1})} = \frac{E\left(X^{\pi_1\ddot{\alpha}_1}\, L\right)E\left(T^{\ddot{\alpha}_1}\right) + Cov\left(X^{\pi_1\ddot{\alpha}_1}L, T^{\ddot{\alpha}_1}\right)}{E(X^{\pi_1\ddot{\alpha}_1})E(T^{\ddot{\alpha}_1})}$$

$$\approx \frac{E\left(X^{\pi_1\ddot{\alpha}_1}\, L\right)E\left(T^{\ddot{\alpha}_1}\right) + E(L)\, Cov\left(X^{\pi_1\ddot{\alpha}_1}, T^{\ddot{\alpha}_1}\right) + E\left(X^{\pi_1\ddot{\alpha}_1}\right)Cov\left(L, T^{\ddot{\alpha}_1}\right)}{E(X^{\pi_1\ddot{\alpha}_1})E(T^{\ddot{\alpha}_1})}$$

$$\approx \frac{E\left(X^{\pi_1\ddot{\alpha}_1}L\right)}{E(X^{\pi_1\ddot{\alpha}_1})},$$

where I have used the following approximation: $Cov(M, N\, R\,) \approx E(N)\, Cov(M, R) + E(R)\, Cov(M, N)$, where M, N and R are any random variables.[1] It follows that $F(\ddot{\alpha}_1) \approx 1$, and that in the GEiVE-IV-S model Equation [27] is replaced by:

$$\frac{E\left(X^{\pi_1\ddot{\alpha}_1}\ln X\right)}{E(X^{\pi_1\ddot{\alpha}_1})} \approx E(Y \ln X)\left\{\frac{1 + \theta_1\left[Cov(Y, \ln X) + Var(\ln X)\frac{Cov(L, \Psi)}{Cov(L, \ln X)}\right]}{1 + Cov(Y, \ln X)}\right\}. \qquad [B3]$$

Hence, assuming—as in the analysis with the IGE of the geometric mean—that $Cov(L, \ln X) > 0$, a comparison of Equations [25], [27] and [B3] indicates that the short-run GMM-IVP estimator is consistent when the following conditions are met: (a) the assumptions of the GEiVE-IV-S model hold, that is, $Cov(W, L) = 0$, $P \perp X$, and $P \perp L$, (b) both children's and parents' incomes are measured at the right points of their lifecycles, that is, $\theta_1 = \pi_1 = 1$, and (c) L is a valid instrument, that is, $Cov(L, \Psi) = 0$.

Equation [B3] shows a first important difference between the two measurement-error models. If the instrument is valid, the GEIVE-IV's Equation [30] reduces to:

$$\frac{E\left(X^{\ddot{\alpha}_1}\ln X\right)}{E(X^{\ddot{\alpha}_1})} \approx E(Y \ln X) + Cov(Y, \ln X)\frac{0.5\,[\ddot{\alpha}_1]^2\, Var(P)}{1 + 0.5\,[\ddot{\alpha}_1]^2\, Var(\ln X)}. \qquad [B5]$$

Therefore, unlike with the GEiVE-IV-S model, under the weaker assumptions of the GEiVE-IV model the short-run GMM-IVP estimator is upward inconsistent even when the instrument is valid, and the magnitude of the asymptotic bias increases with $Var(P)$.

If the instrument is invalid but both the empirical assumptions of the GEiVE-IV-S model and the lifecycle assumptions hold, Equation [B3] becomes:

$$\frac{E\left(X^{\ddot{\alpha}_1}\ln X\right)}{E(X^{\ddot{\alpha}_1})} \approx E(Y\ln X) + Var(\ln X)\frac{Cov(L,\Psi)}{Cov(L,\ln X)}, \qquad [B4]$$

where I have used $E(\ln X\ \ Y) = 1 + Cov(\ln X\,, Y)$. Equations [25] and [B4] provide a counterpart to Equation [14] under the stronger assumptions of the GEiVE-IV-S model. Indeed, as $\frac{\partial E\left(X^{\ddot{\alpha}_1}\ln X\right)\left[E\left(X^{\ddot{\alpha}_1}\right)\right]^{-1}}{\partial\ddot{\alpha}_1} > 0$ (see above), Equations [25] and [B4] indicate that, given an invalid instrument positively correlated with log-parental income, we can expect $\ddot{\alpha}_1$ to be larger or smaller than α_1, depending on the sign of $Cov(L,\Psi)$. Under the additional substantive assumption that $Cov(L,\Psi) > 0$ when L is parental education (and other similar instruments), we can expect estimates obtained with the GMM-IVP estimator to provide an upper bound for the IGE of the expectation.

Equations [25] and [B4] further show that the bias increases with $Var(\ln X)\,\frac{Cov(L,\Psi)}{Cov(L,\ln X)}$. Therefore, as with the GEiVE-IV model, the smaller the covariance between the instrument and the error term, and the larger the slope of the linear projection of L on $\ln X$, the tighter that upper bound will be. Unlike with the GEiVE-IV model, however, with the GEiVE-IV-S model the level of noise in the short-run measure of parental income has no influence on the tightness of the upper bound. This is a second important difference between the two measurement-error models.

C. Sample, variables and related issues

Like Hertz (2007), I define "child" broadly to include anyone of the right age reported in the PSID to be either the son, daughter, stepson, stepdaughter, nephew, niece, grandson or granddaughter of the household head or his wife (or long-term partner).[2] As Hertz (2007:35) put it, "the idea is to look at the relation between children's income and the income of the households in which they were raised, even if that household was not, or not always, headed by their mother or father." Similarly, when the children are 1-17 years old, the "father" is the household head (if the head is male), while the "mother" is either the household head (if the head is female) or the head's wife or long-term partner. When the children are older than 17, the father and mother are those determined to be the father and mother at age 17.

The annual measures of family income are based on the PSID notion of "total family income." But as the income components the PSID used to compute total family income are effectively affected by top coding in the period 1970-1978 (i.e., top codes were not only in place but were "binding" in that period for some people), and the PSID-computed total-family income for those years is based on these top-coded values, I proceeded as follows: (a) I addressed the top-coding of all income components in 1970-1978 by using Pareto imputation (Fichtenbaum and Shahidi 1988), and (b) I recomputed total family income for those years with the Pareto-imputed component variables.

I only estimate IGEs of family income, not of earnings. There are two reasons for proceeding this way. First, the IGEs of children's individual earnings need to be estimated separately by gender, but the available PSID sample is rather small for IV estimation even when men and women are pooled (as I do in all analyses). Second, in the case of the IGE of the geometric mean, short-run estimates are affected by (potentially severe) selection biases because children with zero income or earnings need to be dropped (Mitnik and Grusky 2017). This problem, however, is much more serious with earnings than with family income, as the share of children with zero earnings is much larger than the share of children with zero family income. To further address this issue, instead of using as dependent variable the logarithm of an annual measure of children's family income when estimating the IGE of the geometric mean, I use the logarithm of the average family income of children when they were 35-38 years old (thus further reducing the number of children with zero income). For the sake of consistency (as children with zero income pose no problem in this case), I also use the average family income as dependent variable when estimating the IGE of the expectation.

I use as instruments the household head's years of education when the child was 15 years old and at the times parental income was measured, but not the father's years of education, which is the instrument most often used in the mobility literature. I do not use this instrument because that would require dropping from the sample those children who grew up without a father, which is likely to generate selection bias. Nevertheless, if the father is present in the household, in the vast majority of cases he is coded as household head by the PSID. Therefore, the household head's years of education is similar, but not identical, to the father's years of education. I do use father's occupation as instrument. This is not a problem because this variable is categorical; children that did not provide information on their fathers' occupation (regardless of the reason) can be coded in a separate category, and this is what I do (see Table 1).

Notes

[1] Bohrnstedt and Goldberger's (1969) attribute the approximation to Kendall and Stuart (1963); the former's analysis entails that the approximation involves assuming that $E(\Delta M\ \Delta N\ \Delta R) \approx 0$, where $\Delta i = i - E(i)$ for $i = M, N, R$. It is easy to see that a second-order Taylor-series approximation for $E(\Delta M\ \Delta N\ \Delta R)$ around the expectations of M, N and R is equal to zero.

[2] By convention, the PSID always codes a man as household head and his spouse as wife, i.e., it does not code a woman as household head and his spouse as husband.

References

Bohrnstedt and Goldberger. 1969. “On the Exact Covariance of Products of Random Variables.” *Journal of the American Statistical Association* 64 (328): 1439-1442.

Fichtenbaum, Rudy and Hushang Shahidi. 1998. “Truncation Bias and the Measurement of Income Inequality.” *Journal of Business & Economic Statistics* 6(3):335-337.

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

Kendall, Maurice and Alan Stuart. 1963. *The Advanced Theory of Statistics*, Vol. I (second edition). London: Charles Griffin.

Mitnik, Pablo and David Grusky. The Intergenerational Elasticity of What? The Case for Redefining the Workhorse Measure of Economic Mobility.” Stanford Center on Poverty and Inequality Working Paper.