Online Appendices

INTERGENERATIONAL INCOME ELASTICITIES, INSTRUMENTAL VARIABLE ESTIMATION, AND BRACKETING STRATEGIES

Pablo A. Mitnik
Center on Poverty and Inequality
Stanford University

Sociological Methodology, Vol. 50, Issue 1, 2020.

CONTENTS

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]:

$$\ddot{\beta_1} = \frac{\lambda_1}{\eta_1}\beta_1 + \frac{\lambda_1}{\eta_1}\varrho_2 \frac{SD(L)}{SD(\ln X)} \left[\frac{1 - [Corr(\ln X, L)]^2}{Corr(\ln X, L) + \frac{Cov(Q, L)}{\eta_1 SD(\ln X)\, SD(L)}} \right] - \frac{\lambda_1}{\eta_1}\varrho_2 \frac{Cov(\ln X, L)}{Var(\ln X)} \frac{Cov(Q, L)}{Cov(\ln S, L)} + \frac{Cov(V, L)}{Cov(\ln S, L)} - \frac{\lambda_1}{\eta_1}\varrho_1 \frac{Cov(Q, L)}{Cov(\ln S, L)}$$

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

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

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

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

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

IV estimation of the IGE of the expectation

I derive here Equations [26], [27] and [28] and the sign of $\partial E\left(X^{\ddot{\alpha}_1} \ln X\right) \left[E\left(X^{\ddot{\alpha}_1}\right)\right]^{-1} / \partial \ddot{\alpha}_1$. 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:

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

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

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 $\exp(\alpha_0) = E(Y)/E(X^{\alpha_1})$ 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:

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

$$E(X^{\pi_1 \ddot{\alpha}_1} L) = E\left([\exp(\ln X)]^{\pi_1 \ddot{\alpha}_1} L\right)$$
$$\approx 0 + 0.5\;0\;\;Var(\ln X) + 0.5\;0\;Var(L) + \exp(\pi_1 \ddot{\alpha}_1)\,\pi_1 \ddot{\alpha}_1\;Cov(\ln X\,, L)$$
$$\approx \exp(\pi_1 \ddot{\alpha}_1)\;\;\pi_1 \ddot{\alpha}_1\;Cov(\ln X\,, L)$$

$$E\left(X^{\pi_1 \ddot{\alpha}_1} [\exp(P)]^{\ddot{\alpha}_1}\right) = \;E\left([\exp(\ln X)]^{\pi_1 \ddot{\alpha}_1} [\exp(P)]^{\ddot{\alpha}_1}\right)$$
$$\approx \exp(\pi_1 \ddot{\alpha}_1) + \;0.5\;[\pi_1 \ddot{\alpha}_1]^2 \exp(\pi_1 \ddot{\alpha}_1)\,Var(\ln X) + 0.5\;[\ddot{\alpha}_1]^2 \exp(\pi_1 \ddot{\alpha}_1)\;\;Var(P)$$
$$+\;[\ddot{\alpha}_1]^2 \pi_1 \exp(\pi_1 \ddot{\alpha}_1)\;\;Cov(\ln X\,, P)$$

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

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

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

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

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

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

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

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

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

This is the result shown in Equation [28].

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

$$\frac{\partial E\big(X^{\ddot{\alpha}_1}\ln X\big)\big[E\big(X^{\ddot{\alpha}_1}\big)\big]^{-1}}{\partial\ddot{\alpha}_1} = \frac{E\big(X^{\ddot{\alpha}_1}\big)\dfrac{dE\big(X^{\ddot{\alpha}_1}\ln X\big)}{d\ddot{\alpha}_1} - E\big(X^{\ddot{\alpha}_1}\ln X\big)\dfrac{dE\big(X^{\ddot{\alpha}_1}\big)}{d\ddot{\alpha}_1}}{[E(X^{\ddot{\alpha}_1})]^2}$$

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

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

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:

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

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

$$= E\left(X^{\ddot{\alpha}_{1}{}_1} \dot{X}^{\ddot{\alpha}_1} \ln \dot{X} \ \ln \dot{X} - X^{\ddot{\alpha}_1} \ln X \, \dot{X}^{\ddot{\alpha}_1} \ln \dot{X} - \dot{X}^{\ddot{\alpha}_1} \ln \dot{X} \ X^{\ddot{\alpha}_{1}{}_1} \ln X + \dot{X}^{\ddot{\alpha}_1} X^{\ddot{\alpha}_1} \ln X \ln X\right)$$

$$= E\left(X^{\ddot{\alpha}_1}\right) E\left(\dot{X}^{\ddot{\alpha}_1} \ln \dot{X} \ln \dot{X}\right) - E\left(X^{\ddot{\alpha}_1} \ln X\right) E\left(\dot{X}^{\ddot{\alpha}_1} \ln \dot{X}\right) + E\left(\dot{X}^{\ddot{\alpha}_1}\right) E\left(X^{\ddot{\alpha}_1} \ln X \ \ln X\right) - E\left(\dot{X}^{\ddot{\alpha}_1} \ln \dot{X}\right) E\left(X^{\ddot{\alpha}_1} \ln X\right)$$

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

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").

Like in the previous appendix, 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:

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

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:

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

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

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

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:

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

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, Equation [30] of the GEIVE-IV model reduces to:

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

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:

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

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 $\partial E\left(X^{\ddot{\alpha}_1}\ln X\right)\left[E\left(X^{\ddot{\alpha}_1}\right)\right]^{-1}/\partial\ddot{\alpha}_1 > 0$ (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 $Var(\ln X)Cov(L,\Psi)/Cov(L,\ln X)$. 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. Constructing confidence intervals for partially identified IGEs: Additional details

In the main text I pointed out that, when constructing confidence intervals for the partially identified IGEs, we need to consider the uncertainty due to partial identification as well as the uncertainty regarding the estimated bounds. I considered two contexts. In the first, simpler, context, there is only one estimate of the upper bound. As I explained, asymptotically the noncoverage risk is effectively one-sided, which suggests using $100\ (1-2\alpha)$ % confidence intervals for the identified sets as $100\ (1-\alpha)$ % confidence intervals for the IGEs. I indicated, however, that the resulting confidence intervals are problematic. Indeed, 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:

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

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

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 the standard error operator; $\Phi(.)$ denotes the CDF of the standard normal distribution; and the superscripts identify estimators. As $c_{\beta_1}(\alpha)$ and $c_{\alpha_1}(\alpha)$ are inversely related to RW_{β_1} and RW_{α_1}, respectively, the coverage probabilities of the confidence intervals converge to their nominal values uniformly across values of the width of the identified sets. As I indicated in the main text, 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.

In the second context, the upper bound is estimated more than once with different sets of instruments, which creates more serious complications. 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. I indicated in the main text that, 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 (Nevo and Rosen 2012: Section IV). However, as this estimator is based on multiple estimators, each with its own distribution, the construction of confidence intervals needs to consider the sampling uncertainty about all upper-bound estimates.

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 2S estimator. This can be achieved by using a procedure proposed by Chernozhukov, Lee and Rosen (2013), which they dubbed "adaptive inequality selection" (AIS).[2] Skipping this step, i.e., keeping all upper-bound estimates when carrying out the procedures described in later steps, produces valid but generally conservative confidence intervals.

(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.[3]

(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.

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

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

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

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.[4] 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.[5]

Although the 2S estimator of the upper bound is consistent, it is not asymptotically unbiased—in fact, as I indicated in the main text, no estimator involving minimization or maximization can be unbiased (Hirano and Porter 2012). Chernozhukov, Lee and Rosen's (2013) alternative three-step (3S) estimator 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. 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 $p = 1 - \Phi(\ln N \;\Delta)\; \alpha$. Confidence intervals are then computed exactly or almost exactly as I described above.[6]

As I indicated in the main text, 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."

D. 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).[7] 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 Forthcoming). 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] Nevo and Rosen (2012: 666) used a different procedure, which has been superseded by AIS.

[3] 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.

[4] 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.

[5] These are the confidence intervals reported in the paper's empirical section (in the analyses in which the upper bound is estimated only once).

[6] 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 did in the paper's empirical section (where it made almost no difference).

[7] 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

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

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

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

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.

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

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

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

Mitnik, Pablo and David Grusky. Forthcoming. "The Intergenerational Elasticity of What? The Case for Redefining the Workhorse Measure of Economic Mobility." *Sociological Methodology*.

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