Online appendix
Dynastic Human Capital, Inequality and Intergenerational Mobility

Adrian Adermon   Mikael Lindahl   Mårten Palme

October, 2020

A Econometric appendix

This appendix provides a set of results on the behavior of sums of coefficients in latent variable models. All results are derived under the assumptions that the population coefficients and the correlation between latent variables are non-negative. Section A.1 specifies our econometric model and assumptions.

In Section A.2, we study OLS estimates from a model with errors in variables, and establish the following results:

1. In a regression model where all variables are measured with error, the individual coefficient estimates are biased in unknown directions.
2. The sum of coefficients from such a regression provides a lower bound on the sum of coefficients in the true latent variable model.
3. The negative bias in the sum of coefficients can be reduced by adding more proxies to the regression.

Results 1 and 2 hold under very general assumptions; result 3 requires uncorrelated measurement errors, but mostly holds more generally in simulations.

Section A.3 shows that when there is measurement error, sums of coefficients are less biased than R^2. This result only relies on using standardized variables.

Finally, Section A.4 shows that an instrumental variables estimator can be viewed as an *upper* bound on the sum of coefficients. This result relies on the assumption of uncorrelated measurement errors.

A.1 Model

Assume an observed outcome variable y is generated by the model

$$y = \sum_{k=1}^{K} \beta_k x_k^* + \varepsilon, \tag{A.1}$$

where x_k^* are unobserved latent variables.

We observe M proxy variables $x_k = x_k^* + u_k$, which have been standardized so that $\text{Var}(x_k) = 1$. We allow the measurement errors to have different variances, $\text{Var}(u_k) = \lambda_k$, and denote the covariances among latent variables and measurement errors, as $\text{Cov}(x_k^*, x_{k'}^*) = \rho_{k,k'}$ and $\text{Cov}(u_k, u_{k'}) = \sigma_{k,k'}$, respectively.

We generally think of y as a child outcome, and x_k as the outcomes of relatives in the parental generation, with subscript $k = 1$ representing the parents, $k = 2$ the aunts and uncles, and so on to successively more distant relatives. However, to facilitate exposition, we use numerical indices througout this section.

In addition, we make the following assumptions:

Assumption 1 (Classical measurement errors).

$$\text{Cov}(u_k, x^*_{k'}) = 0, \forall k, k',$$
$$\text{Cov}(u_k, y) = 0, \forall k.$$

It follows from Assumption 1 that $\text{Var}(x^*_k) = 1 - \lambda_k$.

Assumption 2 (Non-negative parameters). $\beta_k \geq 0$, $\forall k$.

Assumption 3 (Positively correlated latent variables). $\rho_{k,k'} \geq 0$, $\forall k \neq k'$.

Initially, we also maintain

Assumption 4 (Uncorrelated measurement errors). $\sigma_{k,k'} = 0$,

although we relax this assumption in different ways below.

Finally, the following two assumptions are required when we allow for negatively correlated measurement errors:

Assumption 5 (Non-increasing coefficients). $\beta_{k+1} \leq \beta_k$.

Assumption 6 (Non-increasing measurement errors). $\lambda_{k+1} \leq \lambda_k$.

Assumptions 5 and 6 are reasonable in our application: more distant relatives are likely to have a smaller impact on children; and since outcomes for more distant relatives are averaged over more individuals, measurement errors are likely to be smaller.

We regress the outcome variable on the set of proxy variables:

$$y = \sum_{k=1}^{M} bx_k + \nu, \tag{A.2}$$

where we allow the number of proxies to be either fewer, more, or the same as the number of latent variables in the population regression. For simplicity, in what follows all variables are expressed as deviations from their means.

A.2 Coefficient sums in latent variable models

This section discusses some properties of the sum of coefficient estimates from Eq. (A.2).

Proposition 1. *Single coefficient estimates from the proxy regression are biased in unknown directions:*

$$\text{plim}\, \hat{b}_k \lesseqgtr \beta_k.$$

For simplicity, let $K = M = 2$, and $\text{Cov}(x_1^*, x_2^*) = \rho$. Then the coefficient estimates are[1]

$$\text{plim}\, \hat{b}_1 = \beta_1 - \frac{\beta_1\lambda_1 - \rho\beta_2\lambda_2}{1-\rho^2},$$
$$\text{plim}\, \hat{b}_2 = \beta_2 - \frac{\beta_2\lambda_2 - \rho\beta_1\lambda_1}{1-\rho^2}.$$

The signs of the bias terms depend on the relative size of the parameters, measurement error variances, and the covariance.[2]

Proposition 2. *Under weak assumptions, the sum of coefficients from the proxy regression is a lower bound on the sum of parameters in the population regression:*

$$\text{plim} \sum_{k=1}^{M} \hat{b}_k \leq \sum_{k=1}^{K} \beta_k.$$

We discuss this result in several steps. First, we analyze the simple two-variable case in order to build intuition. We then derive more general expressions under somewhat restrictive assumptions. Finally, we relax these assumptions in a Monte Carlo simulation.

Two variables If we only use the first proxy, $y = bx_1 + \nu$, it is easy to show that[3]

$$\text{plim}\, \hat{b} = \beta_1(1-\lambda_1) + \beta_2\rho \leq \beta_1 + \beta_2.$$

If we instead use both proxies and take the sum of coefficients, we get[4]

$$\text{plim}\left(\hat{b}_1 + \hat{b}_2\right) = \beta_1 + \beta_2 - \frac{1}{1+\rho}(\beta_1\lambda_1 + \beta_2\lambda_2) \leq \beta_1 + \beta_2.$$

In the special case where $\beta_2 = 0$, this simplifies to

$$\text{plim}\left(\hat{b}_1 + \hat{b}_2\right) = \beta_1\left(1 - \frac{\lambda_1}{1+\rho}\right) \leq \beta_1.$$

If we relax Assumption 4 and allow the measurement errors to be correlated, the coefficient sum becomes[5]

$$\text{plim}\left(\hat{b}_1 + \hat{b}_2\right) = \beta_1 + \beta_2 - \frac{1}{1+\rho+\sigma}(\beta_1(\lambda_1+\sigma) + \beta_2(\lambda_2+\sigma)).$$

This is still a lower bound either if $\sigma \geq 0$, or if $\sigma < 0$ and Assumptions 5 and 6 hold.[6]

1. This follows from setting $M = K = 2$ in Eq. (A.10). See also, e.g., Maddala (1992, pp. 456–457).
2. This result holds for more than two proxies; see Eqs. (A.10) and (A.14).
3. To see this, set $M = 1$ and $K = 2$ in Eq. (A.16)
4. Setting $M = K = 2$ in Eq. (A.12).
5. Set $M = K = 2$ in Eq. (A.11).
6. Start with the numerator of the bias term. By the Cauchy-Schwarz inequality, $|\sigma| \leq \sqrt{\lambda_1\lambda_2}$. Assume the worst-case scenario, $-\sigma = \sqrt{\lambda_1\lambda_2}$. Then $\beta_1(\lambda_1+\sigma) + \beta_2(\lambda_2+\sigma) = \left(\beta_1\sqrt{\lambda_1} - \beta_2\sqrt{\lambda_2}\right)\left(\sqrt{\lambda_1} - \sqrt{\lambda_2}\right)$, which is positive by Assumptions 5 and 6. The denominator is always positive since $|\sigma| \leq 1$. From this, it follows that the bias term is always negative.

Many variables We now turn to the more general case, where we allow for arbitrary numbers of proxies and latent variables. We also allow the measurement errors to be positively correlated across variables. In order for analytical results to be feasible, however, we make the admittedly strong assumptions that all covariances between latent variables and measurement errors are constant. We relax these assumptions below, where we present from a Monte Carlo simulation.

Assumption 7 (Constant positive covariances between latent variables)**.**

$$\rho_{k,k'} = \rho \geq 0, \, \forall k \neq k'.$$

Assumption 8 (Constant positive covariances between measurement errors)**.**

$$\sigma_{k,k'} = \sigma \geq 0, \, \forall k \neq k'.$$

If $M \geq K$, the sum of coefficients takes the form[7]

$$\text{plim} \sum_{k=1}^{M} \hat{b}_k = \sum_{k=1}^{K} \beta_k - \frac{1}{1 + (M-1)(\rho + \sigma)} \left(\sum_{k=1}^{K} \lambda_k \beta_k + (M-1)\sigma \sum_{k=1}^{K} \beta_k \right).$$

If instead $M < K$, the sum of coefficients takes the form

$$\begin{aligned} \text{plim} \sum_{k=1}^{M} \hat{b}_k &= \sum_{k=1}^{K} \beta_k \\ &- \frac{1}{1 + (M-1)(\rho + \sigma)} \left(\sum_{k=1}^{M} \lambda_k \beta_k + (1-\rho) \sum_{k=M+1}^{K} \beta_k + (M-1)\sigma \sum_{k=1}^{K} \beta_k \right). \end{aligned}$$

In both cases, the second term is always negative.

Monte Carlo simulations Assumptions 7 and 8 are unrealistic. For the simulations, we relax them and allow for a completely arbitrary correlation structure among the latent variables and the measurement errors, respectively, with the only restriction being that all correlations are positive.

We generate data according to Eq. (A.1), where

$$\begin{aligned} \beta_k, \lambda_k &\sim U(0,1), \\ \varepsilon &\sim \mathcal{N}(0,1), \\ \mathbf{X}^* &\sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}^*), \\ \mathbf{U} &\sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}), \end{aligned}$$

7. See Section A.5 for the full derivations of these equations.

where $\mathbf{\Sigma}^*$ is the covariance matrix for the latent variables, with diagonal entries $\Sigma^*_{kk} = 1 - \lambda_k$ and randomly drawn positive off-diagonal entries. The measurement errors have covariance matrix $\mathbf{\Sigma}$, with diagonal entries $\Sigma_{kk} = \lambda_k$ and arbitrary non-negative off-diagonal entries.[8]

We perform $r = 1,000$ simulations, with $n = 500,000$ observations in each replication. We arbitrarily set $K = 5$ and $M = 10$, and estimate one OLS regression for each $m \in \{1, 2, \ldots, M\}$.

Figure A.1 shows the results from these simulations. The horizontal axes show the true sum of parameters, and the vertical axes the sum of OLS estimates. The black points show estimates with only one proxy for 5 latent variables; the orange points show estimates with 5 proxies for 5 latent variables; and the blue dots show estimates with 10 proxies for 5 latent variables. Panel A shows results for uncorrelated measurement errors, while the errors are allowed to be positively correlated in Panel B.

The simulations confirm our analytical results. Three features of the results are worth noting — first, the sum of estimated coefficients is always below the true sum of parameters, validating the lower bound result. Second, the lower bound is tighter when more proxies are used.[9] Third, the bias increases when measurement errors are correlated.

Panel C of Fig. A.1 shows simulations where we relax the assumption of positive correlations, and allow for a completely unrestricted correlation matrix between measurement errors, including negative correlations. These results show that the sum of estimates is no longer a sharp lower bound. However, the estimated sum of coefficients only exceeds the true parameter sum in rare cases, and only when there are more proxies than latent variables. Furthermore, the sum of coefficients never exceeds the true sum more than marginally.

Finally, Panel D shows results from a simulation where we allow unrestricted correlations between measurement errors, but impose Assumptions 5 and 6, so that both the true parameters and the measurement error variances are non-increasing with the order of the variable. As for the cases considered in Panels A and B, the sum of coefficients is a lower bound.

Proposition 3. *The bias in the sum of coefficients shrinks when more proxies are added to the regression.*

Many variables Maintaining Assumptions 7 and 8, let $\hat{b}^{sum}_M$ be the sum of coefficients from a regression with M proxies, and $\hat{b}^{sum}_{M+1}$ the sum of coefficients

8. The covariance matrices are generated in the following steps. First, we use the algorithm in Joe (2006) to draw a random correlation matrix from a uniform distribution of all positive definite correlation matrices. To impose positive correlations, we then replace each entry by its absolute value, and apply the algorithm in Higham (2002) to find the neares positive definite matrix. Finally, we multiply element i, j of the correlation matrix by $\sqrt{(1-\lambda_i)(1-\lambda_j)}$ (for the latent variables) or $\sqrt{\lambda_i \lambda_j}$ (for the measurement errors) to get the appropriate covariance matrix.

9. A partial exception is when measurement errors are positively correlated (Panel B). Then additional proxies do not seem to add information after $M = K$.

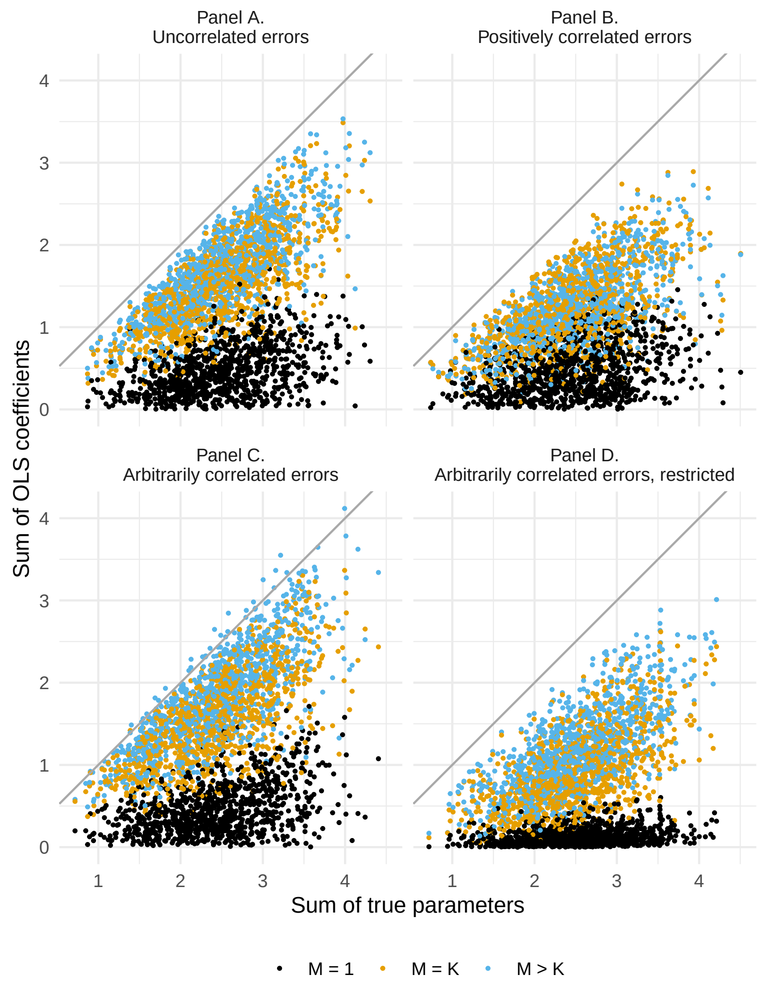

Figure A.1: Sum of coefficients as lower bound, simulation results from a regression with $M+1$ proxies. If $M \geq K$, the change in the sum of coefficients when adding a proxy to the regression is

$$\hat{b}^{sum}_{M+1} - \hat{b}^{sum}_{M} = \frac{\rho \sum_{k=1}^{K} \lambda_k \beta_k - \sigma \sum_{k=1}^{K} (1-\lambda_k)\beta_k}{(1+M(\rho+\sigma))(1+(M-1)(\rho+\sigma))}. \tag{A.3}$$

If instead $M < K$, the change in the sum of coefficients is

$$\begin{aligned}\hat{b}^{sum}_{M+1} - \hat{b}^{sum}_{M} &= \frac{1}{(1+M(\rho+\sigma))(1+(M-1)(\rho+\sigma))} \\ &\quad \times \Bigg[(\rho+\sigma)\sum_{k=1}^{M} \lambda_k \beta_k + (\rho+\sigma)(1-\rho)\sum_{k=M+2}^{K} \beta_k - \sigma \sum_{k=1}^{K} \beta_k \\ &\quad + \left[(1+M(\rho+\sigma))(1-\rho) - (1+(M-1)(\rho+\sigma))\lambda_{M+1}\right]\beta_{M+1}\Bigg]. \end{aligned} \tag{A.4}$$

If $\sigma = 0$, Eq. (A.3) is positive, and Eq. (A.4) is positive under Assumption 6.[10] Furthermore, it follows from Eq. (A.12) that the estimator is consistent as $M \to \infty$, since the bias term disappears in the limit.

If $\sigma > 0$, the proposition holds when σ is small relative to the other model parameters, but it is difficult to characterize the precise conditions in a meaningful way.

Monte Carlo simulation Using the same simulations as for Proposition 2 above, we study how the estimator behaves as we sequentially add more proxies to the regression. Figure A.2 shows the distributions of changes in the sum of coefficient for each additional proxy variable, as a share of the true parameter sum (points are medians). For example, the left-most distribution plot in each panel shows how the sum of coefficients changes when we move from a regression with only one proxy to one with two proxies. In all cases, the true underlying regression has $K = 5$ variables.

As long as the added proxies correspond to latent variables in the population regression ($M \leq K$), adding more proxies essentially always improves the estimator. Once we have one proxy for each latent variable, adding more proxies can cause the sum of coefficients to increase or decrease ($M > K$). With uncorrelated measurement errors, further proxies almost always tend to increase the sum (Panel A), while this is not the case when errors are allowed to be correlated (Panels B–D).

10. If $\sigma = 0$, the expression is positive if $(1+M\rho)(1-\rho) - (1+(M-1)\rho)\lambda_{M+1}$ is non-negative. This term can be rewritten as $(1+M\rho)[(1-\lambda_{M+1})-\rho]+\rho\lambda_{M+1}$, which is positive if $(1-\lambda_{M+1})-\rho$ is non-negative. Notice that $1-\lambda_{M+1} = \mathrm{Var}(x^*_{M+1})$, and remember that by the Cauchy-Schwartz inequality, $|\mathrm{Cov}(X,Y)| \leq \sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}$. Since $\rho = \mathrm{Cov}(x^*_k, x^*_{k'})$, $\forall k \neq k'$, and furthermore $\rho > 0$, it follows that $\rho \leq \sqrt{(1-\lambda_{M+1})(1-\lambda_k)}, \forall k$. By Assumption 6, $1-\lambda_1 x^*_1$ has smaller or equal variance as x^*_{M+1}, so that $1-\lambda_{M+1} \geq \sqrt{(1-\lambda_{M+1})(1-\lambda_1)}$, and therefore $\rho \leq 1-\lambda_{M+1}$.

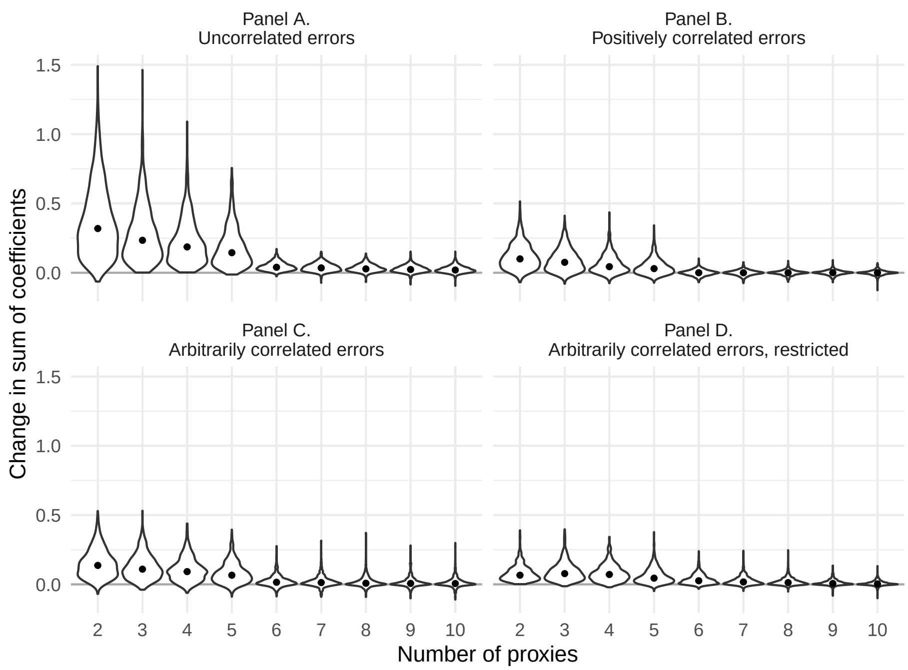

Figure A.2: Bias reduction as number of proxies increases, simulation results

Furthermore, in Panels A–B, each added proxy on average contributes less than the previous ones. It is also clear that correlated measurement errors slows down the rate of convergence to the true parameter sum.

A.3 Relation between sum of coefficients and R^2

The coefficient of determination is defined as

$$R^2 = \frac{\text{Var}(\hat{y})}{\text{Var}(y)}.$$

Consider a univariate regression model,

$$y = bx + \varepsilon,$$

with standardized variables ($\text{Var}(y) = \text{Var}(x) = 1$). This regression has $R^2 = \text{Var}(\hat{y}) = \text{Var}(\hat{b}x) = \hat{b}^2$, which is simply the square of the OLS estimator.

In a multivariate regression with K regressors, the coefficient of determination is

$$R^2 = \text{Var}(\hat{y}) = \text{Var}\left(\sum_{k=1}^{K} \hat{b}_k x_k\right) = \sum_{k=1}^{K} \hat{b}_k^2 + \sum_{k \neq s} \hat{b}_k \hat{b}_s \,\text{Cov}(x_k, x_s).$$

For comparison, we can calculate the square of the sum of coefficients:

$$\left(\sum_{k=1}^{K} \hat{b}_k\right)^2 = \sum_{k=1}^{K} \hat{b}_k^2 + \sum_{k \neq s} \hat{b}_k \hat{b}_s.$$

With standardized variables, the Cauchy-Schwarz inequality guarantees that $|\text{Cov}(x_k, x_s)| \leq 1$ always holds. It follows that

$$R^2 \leq \left(\sum_{k=1}^{K} \hat{b}_k\right)^2.$$

In words, the coefficient of determination is always (weakly) smaller than the squared sum of coefficients. This is a general results that only relies on using standardized variables. In a situation where the sum of coefficients is a lower bound on a true parameter of interest, this means that R^2 has a (weakly) *larger* downward bias than the sum of coefficients once they are transformed to a comparable scale.

A.4 IV estimation in latent variable models

We now turn to instrumental variables estimation. Assume that the true model is Eq. (A.1), and we use all proxies $\{x_k : 2 \leq k \leq K\}$ as instruments for x_1 in a two-stage least squares regression. This amounts to instrumenting parental outcomes with outcomes of more distant relatives.

Proposition 4. *The 2SLS estimator provides an upper bound on the sum of population parameters.*

Two variables Assume the true model is $y = \beta_1 x_1^* + \beta_2 x_2^* + \varepsilon$, and we estimate the regression $y = bx_1 + \nu$, using x_2 as an instrument for x_1.

The IV estimator is

$$\hat{b}_{IV} = \frac{\text{Cov}(x_2, y)}{\text{Cov}(x_2, x_1)} = \frac{\beta_1 \rho + \beta_2 (1 - \lambda_2)}{\rho + \sigma}.$$

If measurement errors are uncorrelated, this simplifies to $\beta_1 + \beta_2 \frac{1-\lambda_2}{\rho}$. Under Assumption 6, the IV estimator is an upper bound on the true sum of coefficients: $\hat{b}_{IV} \geq \beta_1 + \beta_2$.[11]

When the measurement errors are allowed to be correlated with each other, this result no longer holds.

Many variables Under Assumptions 4, 7 and 8, the 2SLS estimator has probability limit

$$\text{plim}\, \hat{b}_{IV} = \beta_1 + \sum_{k=2}^{K} \frac{(1 - \lambda_k) + (K - 2)\rho}{(K - 1)\rho} \beta_k. \tag{A.5}$$

This is an upper bound on the true coefficient sum under Assumptions 2 and 6.[12]

Proposition 5. *If the true model is $AR(1)$, 2SLS provides a consistent estimate of the true parameter.*

This follows trivially from that fact that, if $\beta_k = 0, k \in \{2, \ldots, K\}$, the IV exclusion restriction is satisfied.

Monte Carlo simulation Finally, Fig. A.3 presents simulations similar to those discussed in Section A.2. All simulations maintain Assumption 4.[13] The horizontal axis shows the sum of coefficients in the true model, while the vertical axis shows the IV estimates. In all three models, we use x_2 as an instrument for x_1.

Panel A shows estimates from a model where $\beta_2 = 0$, so that the exclusion restriction holds. As expected, the estimates cluster around the 45 degree line, reflecting that the estimator is consistent.

Panel B shows what happens when we allow $\beta_2 \geq 0$, so that the exclusion restriction does not hold. Now the estimates spread out much more, and the IV tends to overestimate the sum of coefficients. It is worth noting that, although there is a substantial number of simulation draws in which the IV underestimates the true coefficient sum, the size of the bias is generally quite small.

11. By the Cauchy-Schwarz inequality, $\rho \leq \sqrt{(1-\lambda_1)(1-\lambda_2)}$. Under Assumption 6, this implies that $1 - \lambda_2 \geq \rho$, and the result follows.

12. See Section A.5.3 for the proof.

13. In order to avoid IV regressions with extremely weak first stages, we have imposed two additional restrictions: first, the variance of the measurement errors, λ_k, is not allowed to go above 0.8; and second, the correlations between the latent variables is required to be at least 0.5.

Finally, Panel C shows estimates with an endogenous instrument, but where we impose that the measurement error is smaller in x_2 than in x_1 (Assumption 6). Now the IV estimator gives a strict upper bound for the true coefficient sum.

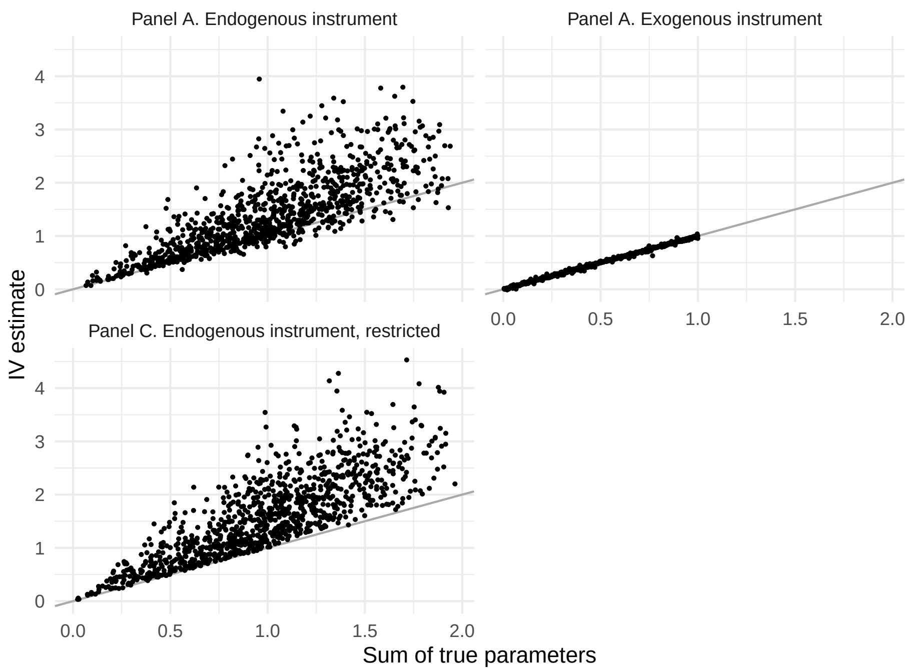

Figure A.3: Instrumental variables as upper bound, simulation results

A.5 Derivations

In this section, we derive expressions when m and k are unrestricted positive integers. We maintain Assumptions 7 and 8. All probability limits are taken as $N \to \infty$.

We can write the observed regression model as $\mathbf{y} = \mathbf{X}\mathbf{b} + \boldsymbol{\nu}$, where $\mathbf{b}$ is the $M \times 1$ vector of coefficients to be estimated, $\mathbf{X}$ is the $N \times M$ matrix of observed proxies, and $\boldsymbol{\nu}$ is a vector of regression errors. The proxies can be written as

$$\mathbf{X} = \mathbf{X}^* + \mathbf{U},$$

where $\mathbf{X}^*$ is the matrix of latent variables, and $\mathbf{U}$ is the matrix of measurement errors.

The vector of OLS estimates is

$$\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{X}^{*\prime}\mathbf{X}^* + \mathbf{X}^{*\prime}\mathbf{U} + \mathbf{U}'\mathbf{X}^* + \mathbf{U}'\mathbf{U})^{-1}(\mathbf{X}^{*\prime}\mathbf{y} + \mathbf{U}'\mathbf{y}). \quad \text{(A.6)}$$

Applying the continuous mapping theorem, and assuming finite first- and second-order moments, it is easy to show that

$$N^{-1}\mathbf{X}^{*\prime}\mathbf{X}^* \xrightarrow{p} \begin{bmatrix} 1-\lambda_1 & \cdots & \rho \\ \vdots & \ddots & \vdots \\ \rho & \cdots & 1-\lambda_M \end{bmatrix},$$

$$N^{-1}\mathbf{U}'\mathbf{U} \xrightarrow{p} \begin{bmatrix} \lambda_1 & \cdots & \sigma \\ \vdots & \ddots & \vdots \\ \sigma & \cdots & \lambda_M \end{bmatrix}.$$

Furthermore, by Assumption 1, $N^{-1}\mathbf{X}^{*\prime}\mathbf{U} \xrightarrow{p} \mathbf{0}$ and $N^{-1}\mathbf{U}\mathbf{X}^{*\prime} \xrightarrow{p} \mathbf{0}$.

It follows that

$$N^{-1}(\mathbf{X}^{*\prime}\mathbf{X}^* + \mathbf{X}^{*\prime}\mathbf{U} + \mathbf{U}'\mathbf{X}^* + \mathbf{U}'\mathbf{U}) \xrightarrow{p} \left[(1-\rho-\sigma)\mathbf{I}_M + (\rho+\sigma)\mathbf{1}_M\mathbf{1}_M'\right],$$

It follows that the inverse term in Eq. (A.6) converges to

$$N(\mathbf{X}^{*\prime}\mathbf{X}^* + \mathbf{X}^{*\prime}\mathbf{U} + \mathbf{U}'\mathbf{X}^* + \mathbf{U}'\mathbf{U})^{-1} \xrightarrow{p} \left[(1-\rho-\sigma)\mathbf{I}_M + (\rho+\sigma)\mathbf{1}_M\mathbf{1}_M'\right]^{-1}.$$

To find the inverse, we use the Sherman-Morrison formula (e.g., Eq. (2) in Bartlett 1951):

$$(\mathbf{A} + \mathbf{u}\mathbf{v}')^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}'\mathbf{A}^{-1}}{1 + \mathbf{v}'\mathbf{A}^{-1}\mathbf{u}}, \quad \text{(A.7)}$$

to get

$$N(\mathbf{X}^{*\prime}\mathbf{X}^* + \mathbf{X}^{*\prime}\mathbf{U} + \mathbf{U}'\mathbf{X}^* + \mathbf{U}'\mathbf{U})^{-1} \xrightarrow{p} \left(\frac{1}{1-\rho-\sigma}\mathbf{I}_M - \frac{\rho+\sigma}{(1-\rho-\sigma)(1+(M-1)(\rho+\sigma))}\mathbf{J}_M\right). \quad \text{(A.8)}$$

We treat the two cases $M \geq K$ and $M < K$ separately.

A.5.1 More proxies than latent variables

When $M \geq K$, the last term in Eq. (A.6) is

$$N^{-1}(\mathbf{X}^{*\prime}\mathbf{y} + \mathbf{U}'\mathbf{y}) \xrightarrow{p} \begin{bmatrix} \text{Cov}(x_1^*, y) \\ \vdots \\ \text{Cov}(x_M^*, y) \end{bmatrix} = \begin{bmatrix} \text{Cov}(x_1^*, \sum_{k=1}^K \beta_k x_k^*) \\ \vdots \\ \text{Cov}(x_M^*, \sum_{k=1}^K \beta_k x_k^*) \end{bmatrix} + o_p(N)$$
$$= \begin{bmatrix} (1-\lambda_1)\beta_1 + \rho \sum_{k\neq 1}^K \beta_k \\ \vdots \\ (1-\lambda_k)\beta_k + \rho \sum_{k\neq k}^K \beta_k \\ \rho \sum_1^K \beta_k \\ \vdots \\ \rho \sum_1^K \beta_k \end{bmatrix}. \tag{A.9}$$

Substituting Eqs. (A.8) and (A.9) in Eq. (A.6) we have

$$\text{plim}\,\hat{\mathbf{b}} = \frac{1}{1-\rho-\sigma} \begin{bmatrix} (1-\lambda_1)\beta_1 + \rho \sum_{k\neq 1}^K \beta_k \\ \vdots \\ (1-\lambda_k)\beta_k + \rho \sum_{k\neq k}^K \beta_k \\ \rho \sum_1^K \beta_k \\ \vdots \\ \rho \sum_1^K \beta_k \end{bmatrix}$$
$$- \frac{\rho+\sigma}{(1-\rho-\sigma)(1+(M-1)(\rho+\sigma))} \begin{bmatrix} \sum_{k=1}^K (1-\lambda_k)\beta_k + (M-1)\rho \sum_{k=1}^K \beta_k \\ \vdots \\ \sum_{k=1}^K (1-\lambda_k)\beta_k + (M-1)\rho \sum_{k=1}^K \beta_k \end{bmatrix}.$$

The coefficient estimate for proxy $k \leq M$ is

$$\begin{aligned} \text{plim}\,\hat{b}_k &= \frac{1-\rho}{1-\rho-\sigma}\beta_k - \frac{1}{1-\rho-\sigma}\lambda_k\beta_k \\ &+ \left(\frac{\rho}{1-\rho-\sigma} - \frac{(\rho+\sigma)(1+(M-1)\rho)}{(1-\rho-\sigma)(1+(M-1)(\rho+\sigma))}\right)\sum_{k=1}^K \beta_k \\ &+ \frac{\rho+\sigma}{(1-\rho-\sigma)(1+(M-1)(\rho+\sigma))}\sum_{k=1}^K \lambda_k\beta_k. \end{aligned}$$

When $\sigma = 0$, this simplifies to

$$\text{plim}\,\hat{b}_k = \beta_k - \frac{1}{1-\rho}\lambda_k\beta_k + \frac{\rho}{(1-\rho)(1+(M-1)\rho)}\sum_{k=1}^K \lambda_k\beta_k. \tag{A.10}$$

The sum of coefficients is

$$\operatorname{plim} \sum_{k=1}^{M} \hat{b}_k$$
$$= \sum_{k=1}^{K} \beta_k - \frac{1}{1+(M-1)(\rho+\sigma)} \left(\sum_{k=1}^{K} \lambda_k \beta_k + (M-1)\sigma \sum_{k=1}^{K} \beta_k \right). \quad \text{(A.11)}$$

In the special case where $\sigma = 0$, this simplifies to

$$\operatorname{plim} \sum_{k=1}^{M} \hat{b}_k = \sum_{k=1}^{K} \beta_k - \frac{1}{1+(M-1)\rho} \sum_{k=1}^{K} \lambda_k \beta_k. \quad \text{(A.12)}$$

A.5.2 Fewer proxies than latent variables

When $M < K$, the second term of Eq. (A.6) can be rewritten as

$$N^{-1}(\mathbf{X}^{*\prime}\mathbf{y} + \mathbf{U}'\mathbf{y}) \xrightarrow{p} \begin{bmatrix} (1-\lambda_1)\beta_1 + \rho \sum_{k \neq 1}^{K} \beta_k \\ \vdots \\ (1-\lambda_M)\beta_M + \rho \sum_{k \neq M}^{K} \beta_k \end{bmatrix}. \quad \text{(A.13)}$$

Substitute Eqs. (A.8) and (A.13) in Eq. (A.6):

$$\begin{aligned} \operatorname{plim} \hat{\mathbf{b}} =& \frac{1}{1-\rho-\sigma} \begin{bmatrix} (1-\lambda_1)\beta_1 + \rho \sum_{k \neq 1}^{K} \beta_k \\ \vdots \\ (1-\lambda_M)\beta_M + \rho \sum_{k \neq M}^{K} \beta_k \end{bmatrix} \\ & - \frac{\rho+\sigma}{(1-\rho-\sigma)(1+(K-1)(\rho+\sigma))} \\ & \times \begin{bmatrix} (1-\rho)\sum_{k=1}^{M} \beta_k + M\rho \sum_{k=1}^{K} \beta_k - \sum_{k=1}^{M} \lambda_k \beta_k \\ \vdots \\ (1-\rho)\sum_{k=1}^{M} \beta_k + M\rho \sum_{k=1}^{K} \beta_k - \sum_{k=1}^{M} \lambda_k \beta_k \end{bmatrix}. \end{aligned}$$

A single coefficient is

$$\begin{aligned} \operatorname{plim} \hat{b}_k =& \frac{1}{1-\rho-\sigma} \left((1-\lambda_k)\beta_k + \rho \sum_{k=1}^{K} \beta_k - \rho\beta_k \right) \\ & - \frac{\rho+\sigma}{(1-\rho-\sigma)(1+(K-1)(\rho+\sigma))} \\ & \left((1-\rho) \sum_{k=1}^{M} \beta_k + M\rho \sum_{k=1}^{K} \beta_k - \sum_{k=1}^{M} \lambda_k \beta_k \right), \end{aligned}$$

which simplifies to

$$\begin{aligned}\operatorname{plim}\hat{b}_k =& \frac{1}{1-\rho}\left((1-\lambda_k)\beta_k + \rho\sum_{k=1}^{K}\beta_k - \rho\beta_k\right)\\ &- \frac{\rho}{(1-\rho)(1+(K-1)\rho)}\\ &\left((1-\rho)\sum_{k=1}^{M}\beta_k + M\rho\sum_{k=1}^{K}\beta_k - \sum_{k=1}^{M}\lambda_k\beta_k\right)\end{aligned} \tag{A.14}$$

when $\sigma = 0$.

The sum of coefficients is

$$\begin{aligned}\operatorname{plim}\sum_{k=1}^{M}\hat{b}_k = \sum_{k=1}^{K}\beta_k - \frac{1}{1+(M-1)(\rho+\sigma)}\\ \times\left(\sum_{k=1}^{M}\lambda_k\beta_k + (1-\rho)\sum_{k=M+1}^{K}\beta_k + (M-1)\sigma\sum_{k=1}^{K}\beta_k\right).\end{aligned} \tag{A.15}$$

Setting $\sigma = 0$, this simplifies to

$$\operatorname{plim}\sum_{k=1}^{M}\hat{b}_k = \sum_{k=1}^{K}\beta_k - \frac{\sum_{k=1}^{M}\lambda_k\beta_k + (1-\rho)\sum_{k=M+1}^{K}\beta_k}{1+(M-1)\rho}. \tag{A.16}$$

A.5.3 Instrumental variables estimator

This section shows the derivation of Eq. (A.5).

Assume the true model is Eq. (A.1), and we use all proxies x_k, $2 \le k \le K$, as instruments for the first proxy, x_1, in a two-stage least squares regression. The matrix of instruments is

$$\begin{aligned}\mathbf{Z} &= \mathbf{Z}^* + \mathbf{V},\\ \mathbf{Z}^* &= \begin{bmatrix}\mathbf{x}_2^* & \cdots & \mathbf{x}_K^*\end{bmatrix},\\ \mathbf{V} &= \begin{bmatrix}\mathbf{u}_2 & \cdots & \mathbf{u}_K\end{bmatrix}.\end{aligned}$$

The 2SLS estimator is

$$\hat{b}_{IV} = (\mathbf{x}_1'\mathbf{P}_Z\mathbf{x}_1)^{-1}\mathbf{x}_1'\mathbf{P}_Z\mathbf{y}, \tag{A.17}$$

where

$$\begin{aligned}\mathbf{P}_Z &= \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\\ &= (\mathbf{Z}^* + \mathbf{V})(\mathbf{Z}^{*\prime}\mathbf{Z}^* + \mathbf{Z}^{*\prime}\mathbf{V} + \mathbf{V}'\mathbf{Z}^* + \mathbf{V}'\mathbf{V})^{-1}(\mathbf{Z}^{*\prime} + \mathbf{V}')\end{aligned} \tag{A.18}$$

The inner term in Eq. (A.18) is essentially the same as in Eq. (A.8), so we have

$$N(\mathbf{Z}^{*\prime}\mathbf{Z}^* + \mathbf{Z}^{*\prime}\mathbf{V} + \mathbf{V}'\mathbf{Z}^* + \mathbf{V}'\mathbf{V})^{-1} \xrightarrow{p} \left(\frac{1}{1-\rho}\mathbf{I}_{K-1} - \frac{\rho}{(1-\rho)(1+(K-2)\rho)}\mathbf{J}_{K-1}\right).$$

Notice that

$$N^{-2}\mathbf{u}'\mathbf{A}\mathbf{B}'\mathbf{v} = N^{-2}\sum_{k=1}^{K}\sum_{i=1}^{N} u_i a_{i,k} \sum_{l=1}^{N} v_l b_{l,k} \xrightarrow{p} \sum_{k=1}^{K} \operatorname{Cov}(u, a_k)\operatorname{Cov}(v, b_k), \quad \text{(A.19)}$$

and

$$N^{-2}\mathbf{u}'\mathbf{A}\mathbf{J}_K\mathbf{B}'\mathbf{v} = N^{-2}\left(\sum_{k=1}^{K}\sum_{i=1}^{N} u_i a_{i,k}\right)\left(\sum_{s=1}^{K}\sum_{l=1}^{N} v_l b_{l,s}\right) \xrightarrow{p} \sum_{k=1}^{K} \operatorname{Cov}(u, a_k)\sum_{s=1}^{K}\operatorname{Cov}(v, b_s). \quad \text{(A.20)}$$

Furthermore, by Assumption 1,

$$N^{-1}\mathbf{u}_1'\mathbf{A} \xrightarrow{p} \mathbf{0},$$
$$N^{-1}\mathbf{A}'\mathbf{u}_1 \xrightarrow{p} \mathbf{0},$$

where $\mathbf{A}$ is $\mathbf{Z}^*$ or $\mathbf{V}$.

Using these results, it's easy to see that

$$N^{-2}\mathbf{x}_1^{*\prime}\mathbf{Z}^*\mathbf{Z}^{*\prime}\mathbf{x}_1^* \xrightarrow{p} (K-1)\rho^2,$$
$$N^{-2}\mathbf{x}_1^{*\prime}\mathbf{Z}^*\mathbf{J}_{K-1}\mathbf{Z}^{*\prime}\mathbf{x}_1^* \xrightarrow{p} (K-1)^2\rho^2,$$
$$N^{-2}\mathbf{x}_1^{*\prime}\mathbf{V}\mathbf{Z}^{*\prime}\mathbf{x}_1^* = N^{-2}\mathbf{x}_1^{*\prime}\mathbf{Z}^*\mathbf{V}'\mathbf{x}_1^* \xrightarrow{p} \sum_{k=2}^{K}\operatorname{Cov}(x_1^*, u_k)\operatorname{Cov}(x_1^*, x_k^*) = 0,$$
$$N^{-2}\mathbf{x}_1^{*\prime}\mathbf{V}\mathbf{J}_{K-1}\mathbf{Z}^{*\prime}\mathbf{x}_1^* = N^{-2}\mathbf{x}_1^{*\prime}\mathbf{Z}^*\mathbf{J}_{k-1}\mathbf{V}'\mathbf{x}_1^* \xrightarrow{p} \left(\sum_{k=2}^{K}\operatorname{Cov}(x_1^*, u_k)\right)\left(\sum_{k=2}^{K}\operatorname{Cov}(x_1^*, x_k^*)\right) = 0,$$
$$N^{-2}\mathbf{x}_1^{*\prime}\mathbf{V}\mathbf{V}'\mathbf{x}_1^* \xrightarrow{p} \sum_{k=2}^{K}\operatorname{Cov}(x_1^*, u_k)^2 = 0,$$
$$N^{-2}\mathbf{x}_1^{*\prime}\mathbf{V}\mathbf{J}_{K-1}\mathbf{V}'\mathbf{x}_1^* \xrightarrow{p} \left(\sum_{k=2}^{K}\operatorname{Cov}(u_1, u_k)\right)\left(\sum_{k=2}^{K}\operatorname{Cov}(u_1, u_k)\right) = 0.$$

Applying these results and the continuous mapping theorem, it can be shown that

$$N^{-1}\mathbf{x}_1'\mathbf{P}_Z\mathbf{x}_1 \xrightarrow{p} \frac{(K-1)\rho^2}{1+(K-2)\rho}.$$

For the second term of Eq. (A.17), first note that

$$\begin{aligned}
\sum_{k=2}^{K}\mathrm{Cov}(y, x_k^*) &= \sum_{k=2}^{K}\mathrm{Cov}\left(\sum_{s=1}^{K}\beta_s x_s^*, x_k^*\right) \\
&= \sum_{k=2}^{K}\sum_{s=1}^{K}\beta_s\,\mathrm{Cov}\left(x_s^*, x_k^*\right) \\
&= (K-1)\rho\beta_1 + \sum_{k=2}^{K}(1-\lambda_k)\beta_k + (K-2)\rho\sum_{k=2}^{K}\beta_k.
\end{aligned}$$

Using this and Eqs. (A.19) and (A.20), we have

$$\begin{aligned}
N^{-2}\mathbf{x}_1^{*\prime}\mathbf{Z}^*\mathbf{Z}^{*\prime}\mathbf{y} &\xrightarrow{p} \sum_{k=2}^{K}\mathrm{Cov}(x_1^*, x_k^*)\,\mathrm{Cov}(y, x_k^*) \\
&= \rho\left((K-1)\rho\beta_1 + \sum_{k=2}^{K}(1-\lambda_k)\beta_k + (K-2)\rho\sum_{k=2}^{K}\beta_k\right), \\
N^{-2}\mathbf{x}_1^{*\prime}\mathbf{Z}^*\mathbf{J}_{K-1}\mathbf{Z}^{*\prime}\mathbf{y} &\xrightarrow{p} \sum_{k=2}^{K}\mathrm{Cov}(x_1^*, x_k^*)\sum_{s=2}^{K}\mathrm{Cov}(y, x_s^*) \\
&= \rho(K-1) \\
&\quad\times\left((K-1)\rho\beta_1 + \sum_{k=2}^{K}(1-\lambda_k)\beta_k + \rho(K-2)\sum_{k=2}^{K}\beta_k\right), \\
N^{-2}\mathbf{x}_1^{*\prime}\mathbf{V}\mathbf{Z}^{*\prime}\mathbf{y}_1 &= N^{-2}\mathbf{x}_1^{*\prime}\mathbf{Z}^*\mathbf{V}'\mathbf{y}_1 \xrightarrow{p} 0, \\
N^{-2}\mathbf{x}_1^{*\prime}\mathbf{V}\mathbf{J}_{K-1}\mathbf{Z}^{*\prime}\mathbf{y}_1 &= N^{-2}\mathbf{x}_1^{*\prime}\mathbf{Z}^*\mathbf{J}_{K-1}\mathbf{V}'\mathbf{y}_1 \xrightarrow{p} 0, \\
N^{-2}\mathbf{x}_1^{*\prime}\mathbf{V}\mathbf{V}'\mathbf{y}_1 &\xrightarrow{p} 0, \\
N^{-2}\mathbf{x}_1^{*\prime}\mathbf{V}\mathbf{J}_{K-1}\mathbf{V}'\mathbf{y}_1 &\xrightarrow{p} 0.
\end{aligned}$$

This in turn yields

$$\begin{aligned}
&N^{-1}\mathbf{x}_1'\mathbf{P}_Z\mathbf{y}_1 \\
&\qquad\xrightarrow{p} \frac{\rho}{1+(K-2)\rho}\left((K-1)\rho\beta_1 + \sum_{k=2}^{K}(1-\lambda_k)\beta_k + (K-2)\rho\sum_{k=2}^{K}\beta_k\right).
\end{aligned}$$

Putting it all together, the 2SLS estimator converges to

$$\begin{aligned}
\operatorname{plim} \hat{b}_{IV} &= \left(\frac{(K-1)\rho^2}{1+(K-2)\rho}\right)^{-1} \frac{\rho}{1+(K-2)\rho} \\
&\times \left((K-1)\rho\beta_1 + \sum_{k=2}^{K}(1-\lambda_k)\beta_k + (K-2)\rho\sum_{k=2}^{K}\beta_k\right) \\
&= \beta_1 + \sum_{k=2}^{K}\frac{(1-\lambda_k)+(K-2)\rho}{(K-1)\rho}\beta_k.
\end{aligned} \tag{A.21}$$

First, assume the true model is AR(1) — i.e., $\beta_k = 0$, $\forall 2 \leq k \leq K$. Then the exclusion restriction holds, and the IV estimator is unbiased ($\operatorname{plim} \hat{b}_{IV} = \beta_1$).

Second, assume instead the extended model is true, so that $\beta_k > 0, \forall k$. Then the exclusion restriction fails, and the IV estimator is upwards biased. Furthermore, the Cauchy-Schwarz inequality states that $\rho \leq \sqrt{(1-\lambda_k)(1-\lambda_s)}, \forall k, s$. This implies that $(1-\lambda_k) \geq \rho$ for all but the largest λ_k. From Assumption 6, it then follows that

$$\sum_{k=2}^{K}\frac{(1-\lambda_k)+(K-2)\rho}{(K-1)\rho}\beta_k \geq \sum_{k=2}^{K}\beta_k,$$

and thus

$$\operatorname{plim} \hat{b}_{IV} \geq \sum_{k=1}^{K}\beta_k.$$

This means that the IV estimate is an upper bound on the sum of parameters in the true model (and the bound becomes exact if the exclusion restriction holds).

References

Bartlett, M. S. 1951. "An Inverse Matrix Adjustment Arising in Discriminant Analysis." *The Annals of Mathematical Statistics* 22 (1): 107–111.

Higham, Nicholas J. 2002. "Computing the nearest correlation matrix—a problem from finance." *IMA Journal of Numerical Analysis* 22 (3): 329–343.

Joe, Harry. 2006. "Generating random correlation matrices based on partial correlations." *Journal of Multivariate Analysis* 97 (10): 2177–2189.

Maddala, G.S. 1992. *Introduction to Econometrics.* Second edition. Macmillan Publishing Company.

B Variable definitions

This appendix provides additional details on variable definitions and construction.

B.1 Children

The grade point average (GPA) is constructed from the national grade 9 registers, using grades in all compulsory subjects. The original scores were percentile ranked by birth cohort, in order to take changes in the grading over time into account.

Data on highest attained educational level is obtained from Swedish education registers, available since 1985. The years of schooling variable for the child generation is defined as follows: nine for compulsory schooling (*Grundskola*), 11 for short high school, 12 for long high school, 14 for short university, 15.5 for long university and 19 for a PhD. We use the latest educational register available, which is for 2009. If education for the individual is missing in 2009, we use 2008, and so forth.

B.2 Parents and other ancestor generations

Data on highest attained educational level is obtained from Swedish education registers, available since 1985. In addition, information is used from the Census 1970. The years of schooling variable for the parental (and other ancestor) generations is defined as follows: seven for (old) primary school (*Folkskola*), nine for (new) compulsory schooling (*Grundskola*), 9.5 for (old) post-primary school (*Realskola*), 11 for short high school, 12 for long high school, 14 for short university, 15.5 for long university and 19 for a PhD. We use information from the latest educational register available.

The income measure we use is calculated as the sum of gross labor earnings, income from businesses, and unemployment benefits. Data is from the IoT-register. Average log income is calculated in the following way: we use income data for all available years for each individual between ages 30 and 60; we take logs and residualize by adjusting for both birth cohort and income year fixed effects; we then take the average of the residuals for each individual. Lastly, we take averages among members within the dynasty category

The Cambridge Social Interaction and Stratification (CAMSIS) measure of social distance uses occupations of spouses to create an index (0–100) of *Social stratification.* The basic idea is that individuals who are similar in terms of social status are more likely to marry each other.[1] While there are many occupation-based social classifications, the CAMSIS scale has two advantages

1. The CAMSIS score is constructed by analyzing a frequency cross-table of husbands' and wives' occupations. This table maps out the space of social distances, and from this, it is possible to locate each occupation along an index of social status or stratification. We use the Swedish CAMSIS scale based on data for 2001–2007 prepared by Erik Bihagen and Paul Lambert, available at http://www.camsis.stir.ac.uk/Data/Sweden90.html.

for our purposes—first, unlike categorical classifications of social class schemes (e.g., Erikson, Goldthorpe, and Portocarero 1979), it is continuous; second, unlike the Socio-Economic Index of occupational status (ISEI) and similar measures (Ganzeboom, De Graaf, and Treiman 1992), it does not rely on income or education in its construction. Hence, the CAMSIS scale provides independent information beyond that contained in our schooling and income variables.[2]

References

Erikson, Robert, John H. Goldthorpe, and Lucienne Portocarero. 1979. "Intergenerational Class Mobility in Three Western European Societies: England, France and Sweden." *The British Journal of Sociology* 30 (4): 415–441.

Ganzeboom, Harry B. G., Paul M. De Graaf, and Donald J. Treiman. 1992. "A Standard International Socio-Economic Index of Occupational Status." *Social Science Research* 21 (1): 1–56.

Lambert, Paul S., and Erik Bihagen. 2014. "Using Occupation-Based Social Classifications." *Work, Employment & Society* 28 (3): 481–494.

2. Lambert and Bihagen (2014) compare a large set of occupation-based social classifications, showing that most measures tend to be relatively highly correlated with each other, and that CAMSIS performs relatively well in predicting unemployment and health.

C Additional results for adoptees analysis

C.1 Descriptive statistics

To increase the probability of meeting assumptions 1, 2, and 4 in Section 5.4, we have restricted the sample to international adoptees and to those adopted before their first birthday.[1] We show summary statistics for the sample of adoptees in Table C.1. If we compare these figures to those in the population (Table D.1), we see that adoptive parents are on average more educated, have higher income, and score higher on the social stratification index, but that the children's GPA is very similar. The adoptive parents are on average also born earlier, whereas the adopted children are similarly aged to the population of children.

Table C.1: Summary statistics, adoptees sample

	Years of schooling	Log income (residualized)	Social stratification	Obs./ child	Birth year
	Child generation				
Child (GPA)	46.14				1988.54
	(26.42)				(3.90)
	Parental generation				
Parents	12.12	.12	51.70	1.99	1956.48
	(2.01)	(.36)	(9.83)	(.08)	(3.54)
Aunts and uncles	11.94	.02	49.21	4.24	1958.50
	(1.56)	(.34)	(8.52)	(2.18)	(5.15)
Spouses of aunts/uncles	12.07	.05	49.83	3.56	1958.16
	(1.67)	(.34)	(9.17)	(1.85)	(5.50)
Parents' cousins	12.29	-.01	47.00	7.42	1966.25
	(1.35)	(.38)	(7.65)	(5.63)	(4.05)
Spouses of parents' cousins	12.24	.03	46.78	5.61	1964.84
	(1.35)	(.37)	(8.32)	(4.36)	(4.34)
Siblings of spouses of aunts/uncles	11.86	-.04	48.83	6.99	1957.96
	(1.59)	(.36)	(7.85)	(4.85)	(6.94)

Note: Cells show means with standard deviations in parentheses. N = 903 observations. "Obs./child" shows number of observations with non-missing data on all variables. The first row shows grade point average for the child in the "years of schooling" column.

1. The adoption age is calculated as the difference between the immigration date and the birth date. Both dates are obtained from Swedish administrative registers.

C.2 Tests of quasi-randomization of adoptees

Following the previous literature[2] we investigate the quasi-randomization assumption for foreign-born adoptees by regressing variables determined prior to adoption on measures for the parents and the other dynasty categories (in our case years of schooling, income, and social stratification). The results from these tests are shown in Table C.2.[3] The pre-determined outcome variables are child gender (Panel A), and age of adoption in months (Panel B). We show results without and with controls for region-of-birth fixed effects. The magnitudes of estimates are in all cases extremely small: one year of additional parental schooling (about 0.5 SD) is associated with one-tenth of a month (3 days) lower adoption age; and with the probability of the child being a girl by at most 2 percent.[4] We conclude that we cannot reject that international adoptees in Sweden (adopted at infancy) during this time are in effect quasi-randomly assigned to their adopting parents.[5]

C.3 Estimation results using matched samples

Table C.3 shows estimates of our main model for a sample of biological children that has been matched to be similar to the adoptees sample with regard to the distribution across birth cohorts for children and parents. Table C.4 shows sample balance before and after matching.

2. See Sacerdote 2007; Fagereng, Mogstad, and Rønning 2018; Holmlund, Lindahl, and Plug 2011; Lundborg, Nordin, and Rooth 2018.

3. To facilitate comparison with our other results, the variables are standardized using the means and standard deviations from the full population sample.

4. For older adoptees (not adopted within 12 months) we do see evidence of systematic placement, likely because those who wanted to adopt quicker could do so by adopting older children, which probably also meant that these adoption families were of higher SES on average.

5. The excess demand for *infant* adoptees is probably what made the adoptions conditionally quasi-random in Sweden, since it is very costly for a family to decline a child in terms of waiting time. Holmlund, Lindahl, and Plug (2008, 2011) and Lundborg, Nordin, and Rooth (2018) investigate quasi-randomness for children adopted by Swedish parents and born abroad mainly during the 1970s. Both of these papers find some evidence of selection using tests similar to ours, but conclude that magnitudes of the estimates are very small.

Table C.2: Test of quasi-randomization of adopted children

	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Female child indicator					
Sum of coefficients	-.023	-.019	-.010	-.010	-.004	-.002
	(.014)	(.018)	(.019)	(.022)	(.023)	(.025)
	Panel B: Female child indicator, region F.E.					
Sum of coefficients	-.008	-.002	.012	.013	.019	.020
	(.015)	(.019)	(.020)	(.022)	(.024)	(.025)
	Panel C: Adoption age in months					
Sum of coefficients	-.051	-.071	-.081	-.066	-.096	-.100
	(.080)	(.100)	(.110)	(.127)	(.134)	(.139)
	Panel D: Adoption age in months, region F.E.					
Sum of coefficients	-.096	-.177	-.218	-.199	-.234	-.249
	(.076)	(.093)	(.105)	(.120)	(.126)	(.131)
Parents	✓	✓	✓	✓	✓	✓
Aunts and uncles		✓	✓	✓	✓	✓
Spouses of aunts/uncles			✓	✓	✓	✓
Parents' cousins				✓	✓	✓
Spouses of parents' cousins					✓	✓
Siblings of spouses of aunts/uncles						✓

Note: Each cell shows the sum of coefficients from a separate regression on years of schooling for the indicated relatives. N = 903 observations. Data is restricted to foreign-born adoptees, with an age at adoption of at most 12 months. Dependent variable is an indicator for female child in Panels A–B, and child's age at adoption in months in Panels C–D. All regressions include linear and quadratic controls for average years of birth for each included type of relative and birth year indicators for the children, and Panels B and D include fixed effects for region-of-birth. Robust standard errors in parentheses.

Table C.3: Biological children, matched to adoptees sample

	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Main estimates					
Parents	.345	.288	.284	.277	.276	.275
	(.003)	(.003)	(.003)	(.003)	(.003)	(.003)
Aunts and uncles		.127	.112	.106	.105	.102
		(.003)	(.004)	(.004)	(.004)	(.004)
Spouses of aunts/uncles			.033	.030	.030	.021
			(.004)	(.003)	(.003)	(.004)
Parents' cousins				.049	.042	.041
				(.003)	(.003)	(.003)
Spouses of parents' cousins					.015	.015
					(.003)	(.003)
Siblings of spouses of aunts/uncles						.025
						(.003)
Sum of coefficients	.345	.415	.428	.462	.468	.479
	(.003)	(.003)	(.004)	(.004)	(.004)	(.005)
R^2	.159	.174	.175	.177	.177	.178
	Panel B: Lubotsky-Wittenberg estimates					
LW estimates	.448	.512	.526	.557	.563	.570
	(.002)	(.002)	(.002)	(.002)	(.003)	(.003)

Note: Each column shows results from a separate regression of child's grade point average on parental generation outcomes. N = 278,277 observations. The sample has been matched to the adoptees sample using exact matching on year of birth and gender of the child, and average year of birth for the parents rounded to the nearest integer. Parental generation variable is years of schooling in Panel A, and the LW index of years of schooling, log income, and social stratification in Panel B. Robust standard errors in parentheses

Table C.4: Balance before and after matching to adoptees sample

	Birth year			Years of schooling			Log income			Social stratification		
	Adopt.	Δmain	Δmatch	Adopt.	Δmain	Δmatch	Adopt.	Δmain	Δmatch	Adopt.	Δmain	Δmatch
Child	1988.54	.46	.00									
Parents	1956.48	-4.41	-.04	12.12	.51	.31	.12	.16	.16	51.70	4.93	2.25
Parents' siblings	1958.50	-3.21	-.02	11.94	.28	.24	.02	.08	.06	49.21	2.48	1.14
Spouses of aunts/uncles	1958.16	-2.81	-.04	12.07	.30	.26	.05	.06	.05	49.83	2.63	1.13
Parents' cousins	1966.25	-1.32	-.46	12.29	.15	.12	-.01	.01	-.00	47.00	.97	.54
Spouses of parents' cousins	1964.84	-1.03	-.39	12.24	.11	.08	.03	.01	-.00	46.78	.61	.17
Siblings of spouses of aunts/uncles	1957.96	-2.31	.06	11.86	.19	.21	-.04	.01	-.00	48.83	1.52	.67

Note: Each set of columns show averages in the adoptees sample; the difference between the average for the adoptees and the average in the main sample; and the difference between the adoptees average and the average in the matched sample. The matched sample has been matched to the adoptees sample using exact matching on year of birth and gender of the child, and average year of birth for the parents rounded to the nearest integer.

References

Fagereng, Andreas, Magne Mogstad, and Marte Rønning. 2018. *Why Do Wealthy Parents Have Wealthy Children?* Working Paper 6955. CESifo.

Holmlund, Helena, Mikael Lindahl, and Erik Plug. 2008. *The Causal Effect of Parent's Schooling on Children's Schooling: A Comparison of Estimation Methods.* Discussion Paper 3630. IZA.

———. 2011. "The Causal Effect of Parent's Schooling on Children's Schooling: A Comparison of Estimation Methods." *Journal of Economic Literature* 49 (4): 615–651.

Lundborg, Petter, Martin Nordin, and Dan Olof Rooth. 2018. "The Intergenerational Transmission of Human Capital: The Role of Skills and Health." *Journal of Population Economics* 31 (4): 1035–1065.

Sacerdote, Bruce. 2007. "How Large Are the Effects from Changes in Family Environment? A Study of Korean American Adoptees." *The Quarterly Journal of Economics* 122 (1): 119–157.

D Additional tables and figures

Table D.1: Summary statistics

	Years of schooling	Log income (residualized)	Social stratification	Obs./ child	Birth year
	Child generation, N = 541,459				
Child (GPA)	46.71				1988.07
	(27.96)				(4.78)
	Parental generation, N = 541,459				
Parents	11.61	-.04	46.78	2.00	1960.89
	(1.69)	(.49)	(9.75)	(.05)	(4.64)
Aunts and uncles	11.66	-.05	46.73	4.50	1961.71
	(1.45)	(.40)	(8.38)	(2.25)	(5.70)
Spouses of aunts/uncles	11.77	-.01	47.20	3.60	1960.97
	(1.50)	(.39)	(9.07)	(1.87)	(6.03)
Parents' cousins	12.14	-.02	46.03	10.11	1967.57
	(1.18)	(.34)	(7.14)	(7.39)	(3.87)
Spouses of parents' cousins	12.13	.02	46.16	7.34	1965.87
	(1.23)	(.33)	(7.83)	(5.58)	(4.15)
Siblings of spouses of aunts/uncles	11.67	-.05	47.31	7.16	1960.27
	(1.41)	(.36)	(7.88)	(5.14)	(7.01)
	Grandparent generation, N = 539,493				
Grandparents	9.27	-.16	45.63	3.87	1934.18
	(1.70)	(.38)	(7.66)	(.42)	(5.55)
Parents' aunts/uncles	9.86	-.14	46.68	5.30	1942.01
	(1.80)	(.38)	(8.11)	(3.63)	(4.17)
	Great grandparent generation, N = 337,265				
Great grandparents	7.74	-.18	40.84	3.54	1913.88
	(1.36)	(.58)	(8.11)	(1.66)	(3.70)

Note: Cells show means with standard deviations in parentheses. "Obs./child" shows number of observations with non-missing data on all variables. The first row shows grade point average for the child in the "years of schooling" column.

Table D.2: Summary statistics, years of schooling sample

	Years of schooling	Log income (residualized)	Social stratification	Obs./ child	Birth year
	Child generation				
Child	12.31				1979.74
	(1.96)				(2.86)
	Parental generation				
Parents	10.98	-.13	46.19	1.99	1955.75
	(1.59)	(.49)	(9.15)	(.07)	(3.29)
Aunts and uncles	11.17	-.11	46.08	5.06	1958.09
	(1.35)	(.38)	(7.53)	(2.56)	(5.09)
Spouses of aunts/uncles	11.34	-.05	46.98	3.93	1957.25
	(1.44)	(.37)	(8.26)	(2.06)	(5.62)
Parents' cousins	11.92	-.04	45.72	8.34	1966.39
	(1.22)	(.37)	(7.39)	(6.48)	(3.96)
Spouses of parents' cousins	11.85	-.01	46.02	5.30	1963.94
	(1.32)	(.37)	(8.34)	(4.19)	(4.24)
Siblings of spouses of aunts/uncles	11.33	-.08	47.01	8.23	1957.16
	(1.37)	(.34)	(7.23)	(5.88)	(6.68)

Note: Cells show means with standard deviations in parentheses. N = 91,243 observations. "Obs./child" shows number of observations with non-missing data on all variables.

Table D.3: Correlation matrix, parental generation outcomes

		Parents			Aunts and uncles			Siblings of aunts/uncles			Parents' cousins			Spouses of parents' cousins			Siblings of spouses of aunts/uncles	
		Sch.	Inc.	S.S.	Sch.	Inc.	S.S.	Sch.	Inc.	S.S.	Sch.	Inc.	SS	Sch.	Inc.	SS	Inc.	SS
Parents	Inc.	.300																
	S.S.	.519	.308															
Aunts and uncles	Sch.	.466	.210	.323														
	Inc.	.211	.265	.179	.313													
	S.S.	.304	.162	.304	.495	.259												
Siblings of aunts/uncles	Sch.	.345	.163	.254	.508	.228	.282											
	Inc.	.152	.177	.147	.226	.268	.195	.277										
	S.S.	.230	.127	.234	.280	.184	.221	.476	.241									
Parents' cousins	Sch.	.247	.126	.171	.251	.126	.157	.187	.089	.117								
	Inc.	.109	.108	.087	.109	.112	.078	.084	.075	.063	.326							
	S.S.	.140	.071	.130	.136	.071	.121	.109	.060	.094	.410	.216						
Spouses of parents' cousins	Sch.	.198	.101	.138	.199	.099	.126	.154	.074	.097	.531	.248	.243					
	Inc.	.089	.082	.075	.089	.078	.066	.071	.059	.057	.235	.274	.171	.274				
	S.S.	.106	.060	.102	.102	.059	.091	.085	.048	.078	.235	.162	.175	.408	.211			
Siblings of spouses of aunts/uncles	Sch.	.244	.123	.178	.356	.160	.208	.458	.200	.270	.141	.062	.080	.118	.053	.064		
	Inc.	.119	.102	.104	.162	.149	.125	.195	.213	.140	.073	.053	.045	.058	.043	.035	.324	
	S.S.	.169	.093	.165	.222	.117	.219	.266	.145	.260	.093	.046	.074	.075	.040	.059	.488	.265

Note: Correlation matrix between years of schooling (Sch.), log income (Inc.), and social stratification (S.S.) across categories of relatives. N = 541,459 observations.

Table D.4: Horizontal GPA-schooling regressions, only biological relatives

	(1)	(2)	(3)
Parents	.361	.301	.291
	(.001)	(.001)	(.001)
Aunts and uncles		.138	.127
		(.001)	(.001)
Parents' cousins			.064
			(.001)
Sum of coefficients	.361	.439	.483
	(.001)	(.001)	(.002)
R^2	.151	.166	.170

Note: Each column shows results from a separate regression of child's grade point average on parental generation years of schooling. N = 541,459 observations. Robust standard errors in parentheses.

Table D.5: Years of schooling sample

	Years of schooling						GPA					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Panel A: Main estimates											
Parents	.285	.243	.239	.234	.234	.234	.287	.243	.239	.234	.234	.233
	(.003)	(.003)	(.003)	(.003)	(.003)	(.003)	(.003)	(.003)	(.003)	(.003)	(.003)	(.003)
Aunts and uncles		.115	.100	.095	.095	.093		.119	.100	.094	.094	.091
		(.004)	(.004)	(.004)	(.004)	(.004)		(.004)	(.004)	(.004)	(.004)	(.004)
Spouses of aunts/uncles			.036	.035	.035	.031			.044	.042	.042	.036
			(.004)	(.004)	(.004)	(.004)			(.004)	(.004)	(.004)	(.004)
Parents' cousins				.046	.047	.047				.052	.048	.048
				(.003)	(.004)	(.004)				(.003)	(.004)	(.004)
Spouses of parents' cousins					-.002	-.002					.008	.008
					(.004)	(.004)					(.004)	(.004)
Siblings of spouses of aunts/uncles						.011						.020
						(.004)						(.004)
Sum of coefficients	.285	.358	.376	.410	.409	.414	.287	.362	.384	.422	.426	.435
	(.003)	(.004)	(.004)	(.005)	(.005)	(.005)	(.003)	(.004)	(.004)	(.005)	(.005)	(.005)
R^2	.102	.114	.115	.117	.117	.117	.104	.117	.118	.121	.121	.121
	Panel B: Lubotsky-Wittenberg estimates											
LW estimates	.390	.457	.469	.502	.499	.500	.408	.477	.494	.529	.531	.536
	(.004)	(.004)	(.005)	(.005)	(.005)	(.006)	(.004)	(.004)	(.005)	(.005)	(.005)	(.006)
R^2	.131	.141	.141	.143	.143	.144	.138	.149	.150	.152	.152	.152

Note: Each column shows results from a separate regression of child's years of schooling (columns 1–6) or grade point average (columns 7–12) on parental generation outcomes, using only the subsample with non-missing years of schooling for the child. N = 91,243 observations. Parental generation variable is years of schooling in Panel A, and the LW index of years of schooling, log income, and social stratification in Panel B. Each parental generation outcome is the average across all members of the given category of relatives. All variables have been normalized to have mean zero and standard deviation one. Robust standard errors in parentheses.

Table D.6: Region fixed effects regressions

	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: 2,706 Mother's parish F.E., N = 540,968					
Sum of coefficients	.354	.435	.456	.497	.504	.515
	(.001)	(.002)	(.002)	(.002)	(.002)	(.002)
	Panel B: 10,913 SAMS region F.E., N = 493,774					
Sum of coefficients	.339	.419	.439	.479	.486	.496
	(.001)	(.002)	(.002)	(.002)	(.002)	(.002)
	Panel C: 26,868 School-by-year F.E., N = 541,459					
Sum of coefficients	.345	.425	.446	.485	.493	.503
	(.001)	(.002)	(.002)	(.002)	(.002)	(.002)
Parents	✓	✓	✓	✓	✓	✓
Aunts and uncles		✓	✓	✓	✓	✓
Spouses of aunts/uncles			✓	✓	✓	✓
Parents' cousins				✓	✓	✓
Spouses of parents' cousins					✓	✓
Siblings of spouses of aunts/uncles						✓

Note: Each cell shows the sum of coefficients from a separate fixed effects regression. Panel A controls for mothers' (or fathers' when mother is missing) parish of residence within 10 years before or after the child's birth. Panel B controls for child's SAM region of residence at ages 16–17. Panel C controls for year-specific school fixed effects for when the child was in 9th grade. All standard errors are clustered on the fixed effects level.

Table D.7: Lubotsky-Wittenberg regressions

	(1)	(2)	(3)	(4)	(5)	(6)
Parents	.465	.397	.392	.382	.382	.381
	(.001)	(.002)	(.002)	(.002)	(.002)	(.002)
Aunts and uncles		.137	.120	.111	.111	.109
		(.002)	(.002)	(.002)	(.002)	(.002)
Spouses of aunts/uncles			.038	.035	.035	.029
			(.002)	(.002)	(.002)	(.002)
Parents' cousins				.056	.049	.048
				(.001)	(.002)	(.002)
Spouses of parents' cousins					.014	.014
					(.002)	(.002)
Siblings of spouses of aunts/uncles						.016
						(.002)
Sum of LW coefficients	.465	.534	.550	.585	.590	.597
	(.001)	(.002)	(.002)	(.002)	(.002)	(.002)
Sum of underlying coefficients	.524	.607	.626	.666	.671	.678
	(.002)	(.002)	(.002)	(.002)	(.003)	(.003)
R^2	.186	.198	.199	.202	.202	.202

Note: Each column shows Lubotsky-Wittenberg weighted sums of coefficients from a separate regression of child's grade point average on parental generation years of schooling, log income, and social stratification. N = 541,459 observations. *Sum of LW coefficients* shows the sum of the Lubotsky-Wittenberg coefficients across dynasty categories, while *Sum of underlying coefficients* shows the sum of the individual coefficients from the underlying regression of child GPA on all three parental generation outcomes for each category of relatives. Robust standard errors in parentheses.

Table D.8: Horizontal GPA-income regressions

	(1)	(2)	(3)	(4)	(5)	(6)
Parents	.262	.233	.228	.224	.223	.222
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
Aunts and uncles		.111	.100	.096	.095	.092
		(.001)	(.001)	(.001)	(.001)	(.001)
Spouses of aunts/uncles			.046	.044	.044	.038
			(.001)	(.001)	(.001)	(.001)
Parents' cousins				.052	.045	.045
				(.001)	(.001)	(.001)
Spouses of parents' cousins					.024	.024
					(.001)	(.001)
Siblings of spouses of aunts/uncles						.034
						(.001)
Sum of coefficients	.262	.344	.374	.416	.432	.455
	(.001)	(.002)	(.002)	(.002)	(.002)	(.002)
R^2	.096	.108	.110	.113	.113	.115

Note: Each column shows results from a separate regression of child's grade point average on parental generation log income. N = 541,459 observations. Each parental generation outcome is the average across all members of the given category of relatives. All variables have been normalized to have mean zero and standard deviation one. Robust standard errors in parentheses.

Table D.9: Horizontal GPA-social stratification regressions

	(1)	(2)	(3)	(4)	(5)	(6)
Parents	.282	.250	.237	.234	.232	.231
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
Aunts and uncles		.115	.105	.101	.100	.096
		(.001)	(.001)	(.001)	(.001)	(.001)
Spouses of aunts/uncles			.072	.070	.069	.063
			(.001)	(.001)	(.001)	(.001)
Parents' cousins				.041	.037	.036
				(.001)	(.001)	(.001)
Spouses of parents' cousins					.029	.028
					(.001)	(.001)
Siblings of spouses of aunts/uncles						.028
						(.001)
Sum of coefficients	.282	.365	.415	.446	.467	.483
	(.001)	(.002)	(.002)	(.002)	(.002)	(.002)
R^2	.103	.116	.121	.122	.123	.124

Note: Each column shows results from a separate regression of child's grade point average on parental generation social stratification index. N = 541,459 observations. Each parental generation outcome is the average across all members of the given category of relatives. All variables have been normalized to have mean zero and standard deviation one. Robust standard errors in parentheses.

Table D.10: Instrumental variables estimates

	(1)	(2)	(3)	(4)	(5)
	Panel A: IV first stage				
Aunts and uncles	.434	.369	.346	.344	.338
	(.001)	(.002)	(.002)	(.002)	(.002)
Spouses of aunts/uncles		.131	.123	.122	.107
		(.001)	(.001)	(.001)	(.002)
Parents' cousins			.119	.099	.098
			(.001)	(.001)	(.001)
Spouses of parents' cousins				.040	.040
				(.001)	(.001)
Siblings of spouses of aunts/uncles					.040
					(.001)
R^2	.245	.258	.271	.272	.273
	Panel B: IV second stage				
Parents	.599	.599	.612	.613	.613
	(.003)	(.003)	(.003)	(.003)	(.003)
	Panel C: IV-LW second stage				
Parents	.654	.654	.664	.664	.664
	(.003)	(.003)	(.003)	(.003)	(.003)

Note: Each column shows results from a separate regression. N = 541,459 observations. Panel A shows first stage regressions of parents' average years of schooling on average years of schooling of the other relatives, while Panel B shows 2SLS estimates of child's GPA on parents' years of schooling, instrumented with the other relatives' years of schooling. Panel C shows Lubotsky-Wittenberg coefficients based on 2SLS estimates of child's GPA on parents' years of schooling, log income, and social stratification, using the corresponding variables for the other relatives as instruments. Robust standard errors in parentheses.

Table D.11: IV estimates, distant relatives

	(1)	(2)	(3)	(4)	(5)
	Panel A: IV first stage				
Aunts and uncles	.338				
	(.002)				
Spouses of aunts/uncles	.107	.241			
	(.002)	(.002)			
Parents' cousins	.098	.139	.161		
	(.001)	(.002)	(.002)		
Spouses of parents' cousins	.040	.055	.065	.148	
	(.001)	(.002)	(.002)	(.001)	
Siblings of spouses of aunts/uncles	.040	.087	.190	.202	.217
	(.001)	(.001)	(.001)	(.001)	(.001)
R^2	.273	.195	.150	.132	.110
	Panel B: IV second stage				
Parents	.613	.634	.662	.645	.620
	(.003)	(.004)	(.005)	(.005)	(.006)
	Panel C: IV-LW second stage				
Parents	.664	.668	.685	.667	.643
	(.003)	(.004)	(.005)	(.005)	(.006)

Note: Each column shows results from a separate regression. N = 541,459 observations. Panel A shows first stage regressions of parents' average years of schooling on average years of schooling of the other relatives, while Panel B shows 2SLS estimates of child's GPA on parents' years of schooling, instrumented with the other relatives' years of schooling. Panel C shows Lubotsky-Wittenberg coefficients based on 2SLS estimates of child's GPA on parents' years of schooling, log income, and social stratification, using the corresponding variables for the other relatives as instruments. Robust standard errors in parentheses.

Table D.12: Multigenerational Lubotsky-Wittenberg estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Vertical regressions				Horizontal coefficient sums			
Parental generation	.465		.429	.442	.597		.570	.587
	(.001)		(.002)	(.002)	(.002)		(.003)	(.004)
Grandparental generation		.252	.068	.073		.292	.018	.037
		(.001)	(.002)	(.002)		(.002)	(.002)	(.003)
Great grandparents				.003				-.008
				(.002)				(.002)
Observations	539,493	539,493	539,493	337,265	539,493	539,493	539,493	337,265
R^2	.186	.085	.192	.187	.202	.091	.205	.199

Note: Each column shows results from a separate regression of child's grade point average on ancestor LW indices of years of schooling, log income, and social stratification. In columns 5–8, each table entry shows the sum of coefficients for all available types of relatives in each generation. Robust standard errors in parentheses.

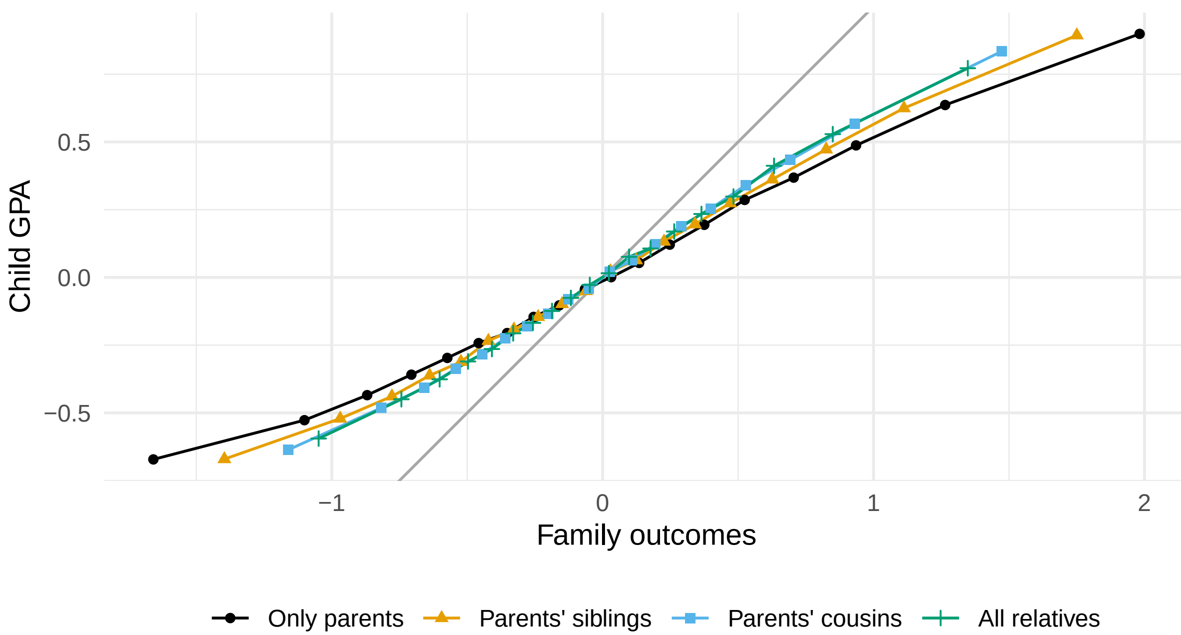

Figure D.1: Lubotsky-Wittenberg regressions

Note: This figure is constructed in the same way as Figure 1, with the difference that the parental generation variable is the Lubotsky-Wittenberg index of years of schooling, log income, and social stratification.