

Ec533: Labour Economics for Research Students

Weak instruments

Jörn-Steffen Pischke

LSE

October 6, 2010

IV regression with weak instruments

Bound, Jaeger, and Baker (1995) pointed out that the quarter of birth instruments explain only a tiny proportion of the variation in schooling. This leads to two distinct problems:

- The 2SLS estimator with weak instruments is biased in small samples.
- Any inconsistency from a small violation of the exclusion restriction gets magnified by weak instruments.

Start with the 2SLS small sample bias. To get an intuition for this situation, look at the simplest formulation of the IV problem:

$$y_i = \beta x_i + \eta_i \quad (\text{structural equation})$$

$$x_i = \pi_1 z_i + \zeta_i \quad (\text{first stage})$$

The small sample behavior with one instrument

- If an instrument is basically irrelevant then $\pi_1 \approx 0$. Recall

$$\beta_{IV} = \frac{\text{cov}(y_i, z_i)}{\text{cov}(x_i, z_i)}$$

but

$$\text{cov}(x_i, z_i) = \text{cov}(\pi_1 z_i + \zeta_i, z_i) = \pi_1 \sigma_Z^2$$

So if $\pi_1 = 0$,

$$\text{cov}(x_i, z_i) = 0$$

and the IV estimator doesn't exist.

- Even when π_1 is truly zero, in any finite sample the sample analogue to $\text{cov}(x_i, z_i)$ will not be exactly zero. But this is of little comfort as the sampling variation in $\text{cov}(x_i, z_i)$ is not helpful to estimate β .

The 2SLS bias with many instruments

With multiple instruments the first stage is:

$$x = Z\pi + \zeta.$$

OLS estimates are biased because η_i is correlated with ζ_i . The instruments Z_i are uncorrelated with ζ_i by construction and uncorrelated with η_i by assumption.

The 2SLS estimator is

$$\hat{\beta}_{2SLS} = (x'P_Zx)^{-1} x'P_Zy = \beta + (x'P_Zx)^{-1} x'P_Z\eta$$

where $P_Z = Z(Z'Z)^{-1}Z'$ is the projection matrix that produces fitted values from a regression of x on Z . Substitute the first stage for x in $x'P_Z\eta$ to get

$$\begin{aligned}\hat{\beta}_{2SLS} - \beta &= (x'P_Zx)^{-1} (\pi'Z' + \zeta') P_Z\eta \\ &= (x'P_Zx)^{-1} \pi'Z'\eta + (x'P_Zx)^{-1} \zeta' P_Z\eta\end{aligned}$$

Group asymptotics

The expectation of this expression is hard to evaluate because the expectation operator does not pass through the inverse $(x'P_Zx)^{-1}$, a nonlinear function.

- Trick: group asymptotics. Still use an asymptotic argument but let the number of instruments grow at the same rate as the sample size. This “keeps the instruments weak.”
- Group asymptotics gives us something like an expectation, it essentially says that we can take these expectations through non-linear functions anyway:

$$E[\widehat{\beta}_{2SLS} - \beta] \approx (E[x'P_Zx])^{-1} E[\pi'Z'\eta] + (E[x'P_Zx])^{-1} E[\zeta'P_Z\eta].$$

This approximation is much better than the usual first-order asymptotic approximation invoked in large-sample theory, so it gives us a good measure of the finite-sample behavior of the 2SLS estimator.

The 2SLS bias with many instruments

Remember the instruments Z_i are uncorrelated with ζ_i and η_i . Therefore $E[\pi' Z' \eta] = 0$, and we have

$$\begin{aligned} E[\widehat{\beta}_{2SLS} - \beta] &\approx (E[x' P_Z x])^{-1} E[\pi' Z' \eta] + (E[x' P_Z x])^{-1} E[\zeta' P_Z \eta] \\ &= (E[x' P_Z x])^{-1} E[\zeta' P_Z \eta]. \end{aligned}$$

Substitute in the first stage again.

$$E[\widehat{\beta}_{2SLS} - \beta] \approx (E[(\pi' Z' + \zeta') P_Z (Z\pi + \zeta)])^{-1} E[\zeta' P_Z \eta].$$

Note that $E[\pi' Z' \zeta] = 0$, so we get no cross-terms:

$$E[\widehat{\beta}_{2SLS} - \beta] \approx [E(\pi' Z' Z\pi) + E(\zeta' P_Z \zeta)]^{-1} E(\zeta' P_Z \eta).$$

The 2SLS bias with many instruments

Matrix algebra trick: $\zeta' P_Z \zeta$ is a scalar, therefore equal to its trace; the trace is a linear operator which passes through expectations and is invariant to cyclic permutations; finally, the trace of P_Z , an idempotent matrix, is equal to its rank, Q . Using these facts

$$\begin{aligned} E(\zeta' P_Z \zeta) &= E[\text{tr}(\zeta' P_Z \zeta)] \\ &= E[\text{tr}(P_Z \zeta \zeta')] \\ &= \text{tr}(P_Z E[\zeta \zeta']) \\ &= \text{tr}(P_Z \sigma_\zeta^2 I) \\ &= \sigma_\zeta^2 \text{tr}(P_Z) \\ &= \sigma_\zeta^2 Q, \end{aligned}$$

where we have assumed that ζ_i is homoskedastic. Similarly, applying the trace trick to $\zeta' P_Z \eta$ shows that this term is equal to $\sigma_\eta \zeta Q$.

The 2SLS bias with many instruments

Now we have

$$\begin{aligned} E[\widehat{\beta}_{2SLS} - \beta] &\approx \sigma_{\eta\xi} Q \left[E(\pi' Z' Z \pi) + \sigma_{\xi}^2 Q \right]^{-1} \\ &= \frac{\sigma_{\eta\xi}}{\sigma_{\xi}^2} \left[\frac{E(\pi' Z' Z \pi) / Q}{\sigma_{\xi}^2} + 1 \right]^{-1}. \end{aligned}$$

Note that

$$F = \frac{E(\pi' Z' Z \pi) / Q}{\sigma_{\xi}^2}$$

is the population F -statistic for the joint significance of all regressors in the first stage regression and hence

$$E[\widehat{\beta}_{2SLS} - \beta] \approx \frac{\sigma_{\eta\xi}}{\sigma_{\xi}^2} \frac{1}{F + 1}.$$

The 2SLS bias with a zero first stage

Suppose the first stage coefficients π are truly zero. Then $F = 0$.
Furthermore

$$\sigma_x^2 = \sigma_{\xi}^2.$$

Hence

$$E[\hat{\beta}_{2SLS} - \beta] \approx \frac{\sigma_{\eta\xi}}{\sigma_x^2}$$

but this is just the bias in the OLS estimator, since

$$\begin{aligned}\beta_{OLS} &= \frac{\text{cov}(y_i, x_i)}{\text{var}(x_i)} = \frac{\text{cov}(\beta x_i + \eta_i, x_i)}{\text{var}(x_i)} \\ &= \beta + \frac{\text{cov}(\eta_i, x_i)}{\text{var}(x_i)} = \frac{\sigma_{\eta\xi}}{\sigma_x^2}\end{aligned}$$

since $\text{cov}(\eta_i, x_i) = \sigma_{\eta\xi}$ if $\pi = 0$.

2SLS is biased towards OLS with weak instruments

Where does this come from?

- If π is truly zero, then any variation in \hat{x}_i in the sample just comes from ζ_i . So, the variation in \hat{x}_i is no different from the variation in x_i , and hence OLS and 2SLS have to estimate the same quantity on average.
- If π is not truly zero but F is small, then 2SLS will be biased towards OLS.

What does the weak instrument bias depend on?

The weak instrument bias tends to get worse as we add more (weak) instruments. To see this consider

$$F = \frac{E(\pi' Z' Z \pi) / Q}{\sigma_{\xi}^2}$$

Suppose you have some existing instruments, and you add new ones with no additional exploratory power. I.e. the π coefficients on the additional instruments will be zero.

- $\pi' Z' Z \pi$ will remain the same as before adding more instruments.
- Since the first stage regression is unchanged by the additional instruments, σ_{ξ}^2 will also remain the same.
- Q will go up.

As a result, F will go down, and the 2SLS bias will get worse.

Summing up on the bias

With weak instruments

- 2SLS is biased towards OLS.
- The bias will tend to be worse when there are many overidentifying restrictions (many instruments compared to endogenous regressors).
- Just identified IV is approximately unbiased (or less biased) even with weak instruments (although it is not possible to see this from the bias formula).
- Estimated standard errors of 2SLS and IV estimators may be too small.

- There are alternative estimators, which have better small sample properties than 2SLS with weak instruments. One such estimator is *LIML* (limited information maximum likelihood).
- LIML is a linear combination of the OLS and 2SLS estimate (with the weights depending on the data), and the weights happen to be such that they (approximately) eliminate the 2SLS bias.

A Monte Carlo Experiment

Simulate data from the following model

$$y_i = \beta x_i + \eta_i$$
$$x_i = \sum_{j=1}^Q \pi_j z_{ij} + \tilde{\zeta}_i$$

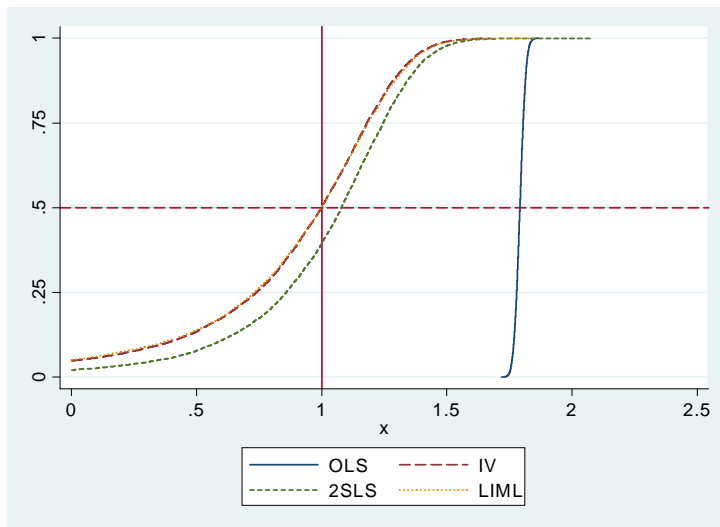
with $\beta = 1$, $\pi_1 = 0.1$, $\pi_j = 0$ for $j = 2, \dots, Q$,

$$\begin{pmatrix} \eta_i \\ \tilde{\zeta}_i \end{pmatrix} \Big| Z \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix} \right),$$

where the z_{ij} are independent, normally distributed random variables with mean zero and unit variance. The sample size is 1000.

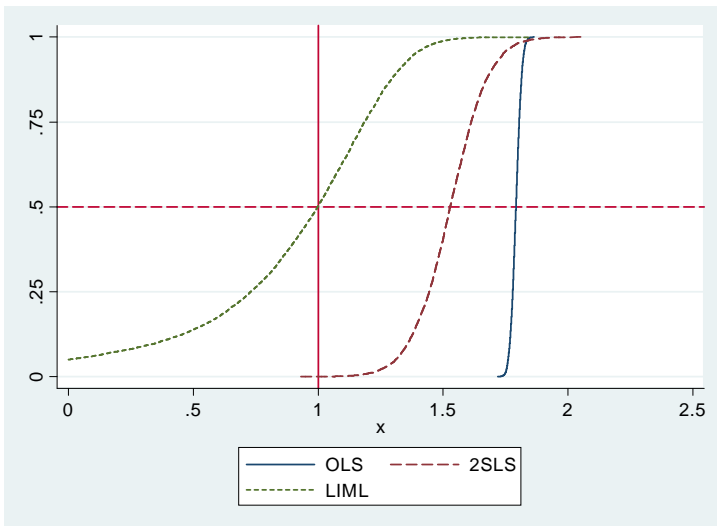
Monte Carlo Results

2SLS, LIML: 2 instruments, IV: one instrument



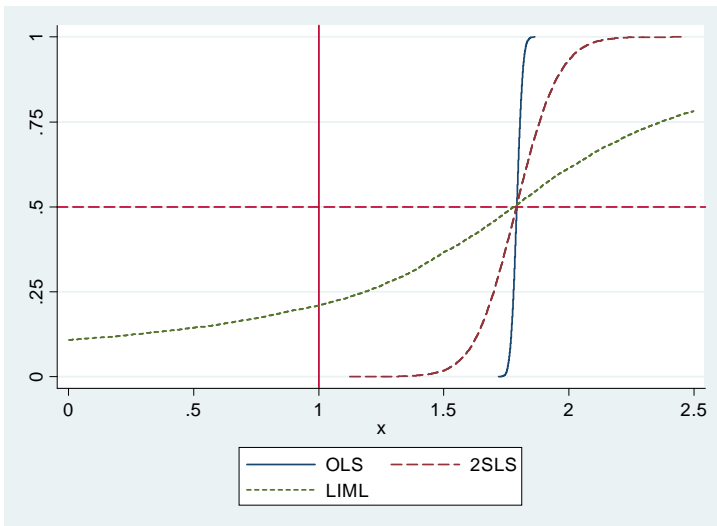
Monte Carlo Results

20 instruments



Monte Carlo Results

20 garbage instruments



What should you do in practice?

- Report the first stage and think about whether it makes sense. Are the magnitude and sign as you would expect?
- Report the F -statistic on the excluded instruments. The bigger this is, the better. F s above 10 to 20 are considered relatively safe, lower F s put you in the danger zone.
- Pick your best single instrument and report just-identified estimates using this one only. Just-identified IV is approximately median-unbiased.
- Check over-identified 2SLS estimates with LIML. If the LIML estimates are very different, or standard errors are much bigger, worry.
- Look at the coefficients, t -statistics, and F -statistics for excluded instruments in the reduced-form regression of dependent variables on instruments. The reduced-form estimates are just OLS, so they are unbiased. If the relationship you expect is not the reduced form, it's probably not there.

Violations of the exclusion restriction

- In general, we do not believe that any of the assumptions we are making in statistics are literally true. But we typically proceed if we think our assumptions are “pretty good.”
- Is “pretty good” enough for the exclusion restriction? Suppose we have a candidate instrument Z_i for a regressor D_i . Write

$$\begin{aligned}y_i &= \alpha + \rho D_i + \gamma Z_i + e_i \\D_i &= \pi_0 + \pi_1 Z_i + \xi_i.\end{aligned}$$

The exclusion restriction amounts to the assumption $\gamma = 0$.

Weak instruments and the exclusion restriction

What happens if $\gamma = \gamma_0 \neq 0$?

$$\begin{aligned}\rho_{IV} &= \frac{\text{cov}(\alpha + \rho D_i + \gamma_0 Z_i + e_i, Z_i)}{\text{cov}(D_i, Z_i)} \\ &= \frac{\rho \text{cov}(D_i, Z_i) + \gamma_0 \text{var}(Z_i)}{\text{cov}(D_i, Z_i)} \\ &= \rho + \gamma_0 \frac{\text{var}(Z_i)}{\text{cov}(D_i, Z_i)} = \rho + \frac{\gamma_0}{\pi_1}\end{aligned}$$

- The IV estimate of ρ is biased, the bias is γ_0 / π_1 .
- The bias is larger in absolute value the smaller π_1 . I.e. the bias is worse with weak instruments.
- With π_1 very small, a very small violation of the exclusion restriction can lead to a large (asymptotic) bias.

Can violation of the exclusion restriction explain the AK91 results?

The bias is equal to γ_0/π_1 . The first stage coefficient with only the quarter 4 instrument is 0.09. This means if

- $\gamma_0 = 0.001$ the bias is 0.011.
- $\gamma_0 = 0.008$ the bias is 0.088.
- Bound, Jaeger, and Baker report that family income between first and other quarter births differs by about 2%. Intergenerational correlations are about 0.4, so this suggests off-spring earnings may differ by about 0.008. This would explain the entire return to schooling estimate!

Bound et al. 1995: Table 1

	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV
Coefficient	.063 (.000)	.142 (.033)	.063 (.000)	.081 (.016)	.063 (.000)	.060 (.029)
F (excluded instruments)		13.486		4.747		1.613
Partial R ² (excluded instruments, ×100)		.012		.043		.014
F (overidentification)		.932		.775		.725
<i>Age Control Variables</i>						
Age, Age ²	x	x			x	x
9 Year of birth dummies			x	x	x	x
<i>Excluded Instruments</i>						
Quarter of birth		x		x		x
Quarter of birth × year of birth				x		x
Number of excluded instruments		3		30		28

NOTE: Calculated from the 5% Public-Use Sample of the 1980 U.S. Census for men born 1930–1939. Sample size is 329,509. All specifications include Race (1 = black), SMSA (1 = central city), Married (1 = married, living with spouse), and 8 Regional dummies as control variables. F (first stage) and partial R² are for the instruments in the first stage of IV estimation. F (overidentification) is that suggested by Basman (1960).

Bound et al. 1995: Table 2

	(1) OLS	(2) IV	(3) OLS	(4) IV
Coefficient	.063 (.000)	.083 (.009)	.063 (.000)	.081 (.011)
F (excluded instruments)		2.428		1.869
Partial R ² (excluded instruments, ×100)		.133		.101
F (overidentification)		.919		.917
<i>Age Control Variables</i>				
Age, Age ²			x	x
9 Year of birth dummies	x	x	x	x
<i>Excluded Instruments</i>				
Quarter of birth		x		x
Quarter of birth × year of birth		x		x
Quarter of birth × state of birth		x		x
Number of excluded instruments		180		178

NOTE: Calculated from the 5% Public-Use Sample of the 1980 U.S. Census for men born 1930–1939. Sample size is 329,509. All specifications include Race (1 = black), SMSA (1 = central city), Married (1 = married, living with spouse), 8 Regional dummies, and 50 State of Birth dummies as control variables. F (first stage) and partial R² are for the instruments in the first stage of IV estimation. F (overidentification) is that suggested by Basman (1960).

Bound et al. 1995: Table 3

<i>Table 3. Estimated Effect of Completed Years of Education on Men's Log Weekly Earnings, Using Simulated Quarter of Birth (500 replications)</i>					
<i>Table (column)</i>	<i>*</i>	<i>1 (4)</i>	<i>1 (6)</i>	<i>2 (2)</i>	<i>2 (4)</i>
<i>Estimated Coefficient</i>					
Mean		.062	.061	.060	.060
Standard deviation of mean		.038	.039	.015	.015
5th percentile		-.001	-.002	.034	.035
Median		.061	.061	.060	.060
95th percentile		.119	.127	.083	.082
<i>Estimated Standard Error</i>					
Mean		.037	.039	.015	.015

NOTE: Calculated from the 5% Public-Use Sample of the 1980 U.S. Census for men born 1930-1939. Sample size is 329,509.

TABLE 4.6.2
Alternative IV estimates of the economic returns to schooling

	(1)	(2)	(3)	(4)	(5)	(6)
2SLS	.105 (.020)	.435 (.450)	.089 (.016)	.076 (.029)	.093 (.009)	.091 (.011)
LIML	.106 (.020)	.539 (.627)	.093 (.018)	.081 (.041)	.106 (.012)	.110 (.015)
<i>F</i> -statistic (excluded instruments)	32.27	.42	4.91	1.61	2.58	1.97
<i>Controls</i>						
Year of birth	✓	✓	✓	✓	✓	✓
State of birth					✓	✓
Age, age squared		✓		✓		✓
<i>Excluded instruments</i>						
Quarter-of-birth dummies	✓	✓				
Quarter of birth*year of birth			✓	✓	✓	✓
Quarter of birth*state of birth					✓	✓
Number of excluded instruments	3	2	30	28	180	178

Notes: The table compares 2SLS and LIML estimates using alternative sets of instruments and controls. The age and age squared variables measure age in quarters. The OLS estimate corresponding to the models reported in columns 1–4 is .071; the OLS estimate corresponding to the models reported in columns 5 and 6 is .067. Data are from the Angrist and Krueger (1991) 1980 census sample. The sample size is 329,509. Standard errors are reported in parentheses.