“A Course in Applied Econometrics”
Lecture 15

Weak Instruments and Many Instruments

Guido Imbens

IRP Lectures, UW Madison, August 2008

Outline

1. Introduction
2. Motivation
3. Weak Instruments
4. Many (Weak) Instruments

1. Introduction

Standard normal asymptotic approximation to sampling distribution of IV, TSLS, and LIML estimators relies on non-zero correlation between instruments and endogenous regressors.

If correlation is close to zero, these approximations are not accurate even in fairly large samples.

In the just identified case TSLS/LIML confidence intervals will still be fairly wide in most cases, even if not valid, unless degree of endogeneity is very high. If concerned with this, alternative confidence intervals are available that are valid uniformly. No better estimators available.

In the case with large degree of overidentification TSLS has poor properties: considerable bias towards OLS, and substantial underestimation of standard errors.

LIML is much better in terms of bias, but its standard error is not correct. A simple multiplicative adjustment to conventional LIML standard errors based on Bekker asymptotics or random effects likelihood works well.

Overall: use LIML, with Bekker-adjusted standard errors.

2.A Motivation : Angrist-Krueger

AK were interested in estimating the returns to years of education. Their basic specification is:

$$Y_i = \alpha + \beta \cdot E_i + \varepsilon_i,$$

where Y_i is log (yearly) earnings and E_i is years of education.

In an attempt to address the endogeneity problem AK exploit variation in schooling levels that arise from differential impacts of compulsory schooling laws by quarter of birth and use quarter of birth as an instrument. This leads to IV estimate (using only 1st and 4th quarter data):

$$\hat{\beta} = \frac{\overline{Y}_4 - \overline{Y}_1}{\overline{E}_4 - \overline{E}_1} = 0.089 \quad (0.011)$$

2.B AK with Many Instruments

AK also present estimates based on additional instruments. They take the basic 3 qob dummies and interact them with 50 state and 9 year of birth dummies.

Here (following Chamberlain and Imbens) we interact the single binary instrument with state times year of birth dummies to get 500 instruments. Also including the state times year of birth dummies as exogenous covariates leads to the following model:

$$Y_i = X_i'\beta + \varepsilon_i, \qquad \mathbb{E}[Z_i \cdot \varepsilon_i] = 0,$$

where X_i is the 501-dimensional vector with the 500 state/year dummies and years of education, and Z_i is the vector with 500 state/year dummies and the 500 state/year dummies multiplying the indicator for the fourth quarter of birth.

The TSLS estimator for β is

$$\hat{\beta}_{\text{TSLS}} = 0.073 \quad (0.008)$$

suggesting the extra instruments improve the standard errors a little bit.

However, LIML estimator tells a somewhat different story,

$$\hat{\beta}_{\text{LIML}} = 0.095 \quad (0.017)$$

with an increase in the standard error.

1.C Bound-Jaeger-Baker Critique

BJB suggest that despite the large (census) samples used by AK asymptotic normal approximations may be very poor because the instruments are only very weakly correlated with the endogenous regressor.

The most striking evidence for this is based on the following calculation. Take the AK data and re-calculate their estimates after replacing the actual quarter of birth dummies by random indicators with the same marginal distribution.

In principle this means that the standard (gaussian) large sample approximations for TSLS and LIML are invalid since they rely on non-zero correlations between the instruments and the endogenous regressor.

	Single Instr		500 Instruments			
			TSLS		LIML	
Real QOB	0.089	(0.011)	0.073	(0.008)	0.095	(0.017)
Random QOB	-1.96	(18.12)	0.059	(0.009)	-0.330	(0.100)

With many random instruments the results are troubling. Although the instrument contains no information, the results suggest that the instruments can be used to infer precisely what the returns to education are.

1.D Simulations with a Single Instrument

10,000 artificial data sets, all of size 160,000, designed to mimic the AK data. In each of these data sets half the units have quarter of birth (denoted by Q_i) equal to 0 and 1 respectively.

$$\begin{pmatrix} \nu_i \\ \eta_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.446 & \rho \cdot \sqrt{0.446} \cdot \sqrt{10.071} \\ \rho \cdot \sqrt{0.446} \cdot \sqrt{10.071} & 10.071 \end{pmatrix} \right).$$

The correlation between the reduced form residuals in the AK data is $\rho = 0.318$.

$$E_i = 12.688 + 0.151 \cdot Q_i + \eta_i,$$

$$Y_i = 5.892 + 0.014 \cdot Q_i + \nu_i.$$

Now we calculate the IV estimator and its standard error, using either the actual qob variable or a random qob variable as the instrument.

We are interested in the size of tests of the null that coefficient on years of education is equal to $0.089 = 0.014/0.151$.

We base the test on the t-statistic. Thus we reject the null if the ratio of the point estimate minus 0.089 and the standard error is greater than 1.96 in absolute value.

We repeat this for 12 different values of the reduced form error correlation. In Table 3 we report the coverage rate and the median and 0.10 quantile of the width of the estimated 95% confidence intervals.

Table 3: Coverage Rates of Conv. TSLS CI by Degree of Endogeneity

ρ	0.0	0.4	0.6	0.8	0.9	0.95	0.99
implied OLS	0.00	0.08	0.13	0.17	0.19	0.20	0.21
Real QOB							
Cov rate	0.95	0.95	0.96	0.95	0.95	0.95	0.95
Med Width 95% CI	0.09	0.08	0.07	0.06	0.05	0.05	0.05
0.10 quant Width	0.08	0.07	0.06	0.05	0.04	0.04	0.04
Random QOB							
Cov rate	0.99	1.00	1.00	0.98	0.92	0.82	0.53
Med Width 95% CI	1.82	1.66	1.45	1.09	0.79	0.57	0.26
0.10 quant Width	0.55	0.51	0.42	0.33	0.24	0.17	0.08

In this example, unless the reduced form correlations are very high, e.g., at least 0.95, with irrelevant instruments the conventional confidence intervals are wide and have good coverage.

The amount of endogeneity that would be required for the conventional confidence intervals to be misleading is higher than one typically encounters in cross-section settings.

Put differently, although formally conventional confidence intervals are not valid uniformly over the parameter space (e.g., Dufour, 1997), the subsets of the parameter space where results are substantively misleading may be of limited interest.

This in contrast to the case with many weak instruments where especially TSLS can be misleading in empirically relevant settings.

3.A Single Weak Instrument

$$Y_i = \beta_0 + \beta_1 \cdot X_i + \varepsilon_i,$$

$$X_i = \pi_0 + \pi_1 \cdot Z_i + \eta_i,$$

with $(\varepsilon_i, \eta_i) \perp\!\!\!\perp Z_i$, and jointly normal with covariance matrix Σ. The reduced form for the first equation is

$$Y_i = \alpha_0 + \alpha_1 \cdot Z_i + \nu_i,$$

where the parameter of interest is $\beta_1 = \alpha_1/\pi_1$. Let

$$\Omega = \mathbb{E}\left[\begin{pmatrix} \nu_i \\ \eta_i \end{pmatrix} \cdot \begin{pmatrix} \nu_i \\ \eta_i \end{pmatrix}'\right], \quad \text{and} \quad \Sigma = \mathbb{E}\left[\begin{pmatrix} \varepsilon_i \\ \eta_i \end{pmatrix} \cdot \begin{pmatrix} \varepsilon_i \\ \eta_i \end{pmatrix}'\right],$$

Standard IV estimator:

$$\hat{\beta}_1^{\text{IV}} = \frac{\frac{1}{N}\sum_{i=1}^N \left(Y_i - \overline{Y}\right)\left(Z_i - \overline{Z}\right)}{\frac{1}{N}\sum_{i=1}^N \left(X_i - \overline{X}\right)\left(Z_i - \overline{Z}\right)},$$

Concentration parameter:

$$\lambda = \pi_1^2 \cdot \sum_{i=1}^N \left(Z_i - \overline{Z}\right)^2 / \sigma_\eta^2.$$

Normal approximations for numerator and denominator are accurate:

$$\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^N \left(Y_i - \overline{Y}\right)\left(Z_i - \overline{Z}\right) - \text{Cov}(Y_i, Z_i)\right) \approx \mathcal{N}\left(0, V(Y_i \cdot Z_i)\right),$$

$$\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^N \left(X_i - \overline{X}\right)\left(Z_i - \overline{Z}\right) - \text{Cov}(X_i, Z_I)\right) \approx \mathcal{N}\left(0, V(X_i \cdot Z_i)\right).$$

If $\pi_1 \neq 0$, as the sample size gets large, then the ratio will eventually be well approximated by a normal distribution as well.

However, if $\text{Cov}(X_i, Z_i) \approx 0$, the ratio may be better approximated by a Cauchy distribution, as the ratio of two normals centered close to zero.

3.B Staiger-Stock Asymptotics and Uniformity

Staiger and Stock investigate the distribution of the standard IV estimator under an alternative asymptotic approximation.

The standard asymptotics (strong instrument asymptotics in the SS terminology) is based on fixed parameters and the sample size getting large.

In their alternative asymptotic sequence SS model π_1 as a function of the sample size, $\pi_{1N} = c/\sqrt{N}$, so that the concentration parameter converges to a constant:

$$\lambda \longrightarrow c^2 \cdot V(Z_i).$$

SS then compare coverage properties of various confidence intervals under this (weak instrument) asymptotic sequence.

The importance of the SS approach is in demonstrating for any sample size there are values of the nuisance parameters such that the actual coverage is substantially away from the nominal coverage.

More recently the issue has therefore been reformulated as requiring confidence intervals to have asymptotically the correct coverage probabilities uniformly in the parameter space. See for a discussion from this perspective Mikusheva.

Note that there cannot exist estimators that are consistent for β^* uniformly in the parameter space since if $\pi_1 = 0$, there are no consistent estimators for β_1. However, for testing there are generally confidence intervals that are uniformly valid, but they are not of the conventional form, that is, a point estimate plus or minus a constant times a standard error.

3.C Anderson-Rubin Confidence Intervals

Let the instrument $\tilde{Z}_i = Z_i - \overline{Z}$ be measured in deviations from its mean. Then define the statistic

$$S(\beta_1) = \frac{1}{N}\sum_{i=1}^{N} \tilde{Z}_i \cdot (Y_i - \beta_1 \cdot X_i)\,.$$

Then, under the null hypothesis that $\beta_1 = \beta_1^*$, and conditional on the instruments, the statistic $\sqrt{N}\cdot S(\beta_1^*)$ has an exact normal distribution

$$\sqrt{N} \cdot S(\beta_1^*) \sim \mathcal{N}\left(0, \sum_{i=1}^{N} \tilde{Z}_i^2 \cdot \sigma_\varepsilon^2\right).$$

Anderson and Rubin (1949) propose basing tests for the null hypothesis

$$H_0: \quad \beta_1 = \beta_1^0, \qquad \text{against the alternative hypothesis} \quad H_a: \quad \beta_1 \neq \beta_1^0$$

on this idea, through the statistic

$$\mathrm{AR}\left(\beta_1^0\right) = \frac{N \cdot S(\beta_1^0)^2}{\sum_{i=1}^N \tilde{Z}_i^2} \cdot \left(\begin{pmatrix} 1 & -\beta_1^0 \end{pmatrix} \Omega \begin{pmatrix} 1 \\ -\beta_1^0 \end{pmatrix}\right)^{-1}.$$

A confidence interval can be based on this test statistic by inverting it:

$$\mathrm{CI}_{0.95}^{\beta_1} = \{\beta_1 \,|\mathrm{AR}(\beta_1) \leq 3.84\}$$

This interval can be equal to the whole real line.

3.D Anderson-Rubin with K instruments

The reduced form is

$$X_i = \pi_0 + \pi_1' Z_i + \eta_i,$$

$S\left(\beta_1^0\right)$ is now normally distributed vector.
AR statistic with associated confidence interval:

$$\text{AR}\left(\beta_1^0\right) = N \cdot S\left(\beta_1^0\right)' \left(\sum_{i=1}^{N} \tilde{Z}_i \cdot \tilde{Z}_i'\right)^{-1} S\left(\beta_1^0\right) \cdot \left(\begin{pmatrix} 1 & -\beta_1^0 \end{pmatrix} \Omega \begin{pmatrix} 1 \\ -\beta_1^0 \end{pmatrix}\right.$$

$$\text{CI}_{0.95}^{\beta_1} = \left\{\beta_1 \middle| \text{AR}(\beta_1) \leq \mathcal{X}_{0.95}^2(K)\right\},$$

The problem is that this confidence interval can be empty because it simultaneously tests validity of instruments.

3.E Kleibergen Test

Kleibergen modfies AR statistic through

$$\tilde{S}\left(\beta_1^0\right) = \frac{1}{N} \sum_{i=1}^{N} \left(\tilde{Z}_i' \hat{\pi}_1(\beta_1^0)\right) \cdot \left(Y_i - \beta_1^0 \cdot X_i\right),$$

where $\hat{\pi}$ is the maximum likelihood estimator for π_1 under the restriction $\beta_1 = \beta_1^0$. The test is then based on the statistic

$$K\left(\beta_1^0\right) = \frac{N \cdot \tilde{S}(\beta_1^0)^2}{\sum_{i=1}^{N} \tilde{Z}_i^2} \cdot \left(\begin{pmatrix} 1 & -\beta_1^0 \end{pmatrix} \Omega \begin{pmatrix} 1 \\ -\beta_1^0 \end{pmatrix}\right)^{-1}.$$

This has an approximate chi-squared distribution, and can be used to construct a confidence interval.

3.F Moreira's Similar Tests

Moreira (2003) proposes a method for adjusting the critical values that applies to a number of tests, including the Kleibergen test. His idea is to focus on *similar* tests, test that have the same rejection probability for all values of the nuisance parameter (the π) by adjusting critical values (instead of using quantiles from the chi-squared distribution).

The way to adjust the critical values is to consider the distribution of a statistic such as the Kleibergen statistic conditional on a complete sufficient statistic for the nuisance parameter. In this setting a complete sufficient statistic is readily available in the form of the maximum likelihood estimator under the null, $\hat{\pi}_1(\beta_1^0)$.

Moreira's preferred test is based on the likelihood ratio. Let

$$LR\left(\beta_1^0\right) = 2 \cdot \left(L\left(\hat{\beta}_1, \hat{\pi}\right) - L\left(\beta_1^0, \hat{\pi}(\beta_1^0)\right)\right),$$

be the likelihood ratio.

Then let $c_{LR}(p, 0.95)$, be the 0.95 quantile of the distribution of $LR(\beta_1^0)$ under the null hypothesis, conditional on $\hat{\pi}(\beta_1^0) = p$. The proposed test is to reject the null hypothesis at the 5% level if

$$LR\left(\beta_1^0\right) > c_{LR}(\hat{\pi}(\beta_1^0), 0.95),$$

where conventional test would use critical values from a chi-squared distribution with a single degree of freedom. The critical values are tabulated for low values of K.

This test can then be converted to construct a 95% confidence intervals.

3.G Conditioning on the First Stage

These confidence intervals are asymptotically valid irrespective of the strength of the first stage (the value of π_1). However, they are not valid if one first inspects the first stage, and conditional on the strength of that, decides to proceed.

Specifically, if in practice one first inspects the first stage, and decide to abandon the project if the first stage F-statistic is less than some fixed value, and otherwise proceed by calculating confidence interval, the large sample coverage probabilities would not be the nominal ones.

Chioda and Jansson propose a confidence interval that is valid conditional on the strength of the first stage. A caveat is that this involves loss of information, and thus the Chioda-Jansson confidence intervals are wider than confidence intervals that are not valid conditional on the first stage.

4.A Many (Weak) Instruments

In this section we discuss the case with many weak instruments. The problem is both the bias in the standard estimators, and the misleadingly small standard errors based on conventional procedures, leading to poor coverage rates for standard confidence intervals in many situations.

Resampling methods such as bootstrapping do not solve these problems.

The literature has taken a number of approaches. Part of the literature has focused on alternative confidence intervals analogues to the single instrument case. In addition a variety of new point estimators have been proposed.

Generally LIML still does well, but standard errors need to be adjusted.

4.B Bekker Asymptotics

Bekker (1995) derives large sample approximations for TSLS and LIML based on sequences where the number of instruments increases proportionally to the sample size.

He shows that TSLS is not consistent in that case.

LIML is consistent, but the conventional LIML standard errors are not valid. Bekker then provides LIML standard errors that are valid under this asymptotic sequence. Even with relatively small numbers of instruments the differences between the Bekker and conventional asymptotics can be substantial.

Bekker correction, single endogenous regressor:

$$Y_i = \beta_1' X_{1i} + \beta_2' X_{2i} + \varepsilon_i = \beta' X_i + \varepsilon_i,$$

$$X_{1i} = \pi_1' Z_{1i} + \pi_2' X_{2i} + \eta_i = \pi' Z_i + \eta_i.$$

Define the matrices $\mathbf{P}_Z$ and $\mathbf{M}_Z$ as:

$$\mathbf{P}_Z = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}', \quad \mathbf{M}_Z = \mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'.$$

Let σ^2 be the variance of ε_i, with consistent estimator $\hat{\sigma}^2$. The standard TSLS variance is

$$V_{tsls} = \hat{\sigma}^2 \cdot (\mathbf{X}\mathbf{P}_Z\mathbf{X})^{-1}.$$

Under the standard, fixed number of instrument asymptotics, the asymptotic variance for LIML is identical to that for TSLS, and so in principle we can use the same estimator. In practice researchers typically estimate the variance for LIML as

$$V_{liml} = \hat{\sigma}^2 \cdot \left(\mathbf{X P}_Z \mathbf{X} - \hat{\lambda} \cdot \mathbf{X}'\mathbf{M}_Z \mathbf{X}\right)^{-1}.$$

To get Bekker's correction, we need a little more notation. Define

$$\Omega = \begin{pmatrix} \mathbf{Y} & \mathbf{X} \end{pmatrix} \mathbf{P}_Z \begin{pmatrix} \mathbf{Y} & \mathbf{X} \end{pmatrix} / N = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{12}' & \Omega_{22} \end{pmatrix},$$

$$\Omega_{11} = \mathbf{Y P}_Z \mathbf{Y}/N, \quad \Omega_{12} = \mathbf{Y P}_Z \mathbf{X}/N, \quad \text{and} \quad \Omega_{22} = \mathbf{X P}_Z \mathbf{X}/N.$$

$$\mathbf{A} = N \cdot \frac{\Omega_{12}'\Omega_{12} - \Omega_{22}\beta\Omega_{12} - \Omega_{12}'\beta'\Omega_{22} + \Omega_{22}\beta\beta'\Omega_{22}}{\Omega_{11} - 2\Omega_{12}\beta + \beta'\Omega_{22}\beta}.$$

Then:

$$V_{bekker} = \hat{\sigma}^2 \cdot \left(\mathbf{X P}_Z \mathbf{X} - \hat{\lambda} \cdot \mathbf{X}'\mathbf{M}_Z \mathbf{X}\right)^{-1}$$

$$\times \left(\mathbf{X P}_Z \mathbf{X} - \lambda \cdot \mathbf{A}\right) \cdot \left(\mathbf{X P}_Z \mathbf{X} - \hat{\lambda} \cdot \mathbf{X}'\mathbf{M}_Z \mathbf{X}\right)^{-1}.$$

Recommended in practice

4.C Random Effects Estimators

Chamberlain and Imbens propose a random effects quasi maximum likelihood (REQML) estimator. They propose modelling the first stage coefficients π_k, for $k = 1, \ldots, K$, in the regression

$$X_i = \pi_0 + \pi_1' Z_i + \eta_i = \pi_0 + \sum_{k=1}^{K} \pi_k \cdot Z_{ik} + \eta_i,$$

(after normalizing the instruments to have mean zero and unit variance,) as independent draws from a normal $\mathcal{N}(\mu_\pi, \sigma_\pi^2)$ distribution.

Assuming also joint normality for (ε_i, η_i), one can derive the likelihood function

$$\mathcal{L}(\beta_0, \beta_1, \pi_0, \mu_\pi, \sigma_\pi^2, \Omega).$$

In contrast to the likelihood function in terms of the original parameters $(\beta_0, \beta_1, \pi_0, \pi_1, \Omega)$, this likelihood function depends on a small set of parameters, and a quadratic approximation to its logarithms is more likely to be accurate.

CI discuss some connections between the REQML estimator and LIML and TSLS in the context of this parametric set up. First they show that in large samples, with a large number of instruments, the TSLS estimator corresponds to the restricted maximum likelihood estimator where the variance of the first stage coefficients is fixed at a large number, or $\sigma_\pi^2 = \infty$:

$$\hat{\beta}_{\text{TSLS}} \approx \arg \max_{\beta_0,\beta_1,\pi_0,\mu_\pi} = L(\beta_0, \beta_1, \pi_0, \mu_\pi, \sigma_\pi^2 = \infty, \Omega).$$

From a Bayesian perspective, TSLS corresponds approximately to the posterior mode given a flat prior on all the parameters, and thus puts a large amount of prior mass on values of the parameter space where the instruments are jointly powerful.

In the special case where we fix $\mu_\pi = 0$, and Ω is known, and the random effects specification applies to all instruments, CI show that the REQML estimator is identical to LIML.

However, like the Bekker asymptotics, the REQML calculations suggests that the standard LIML variance is too small: the variance of the REQML estimator is approximately equal to the standard LIML variance times

$$1 + \sigma_\pi^{-2} \cdot \left(\begin{pmatrix} 1 \\ \beta_1 \end{pmatrix}' \Omega^{-1} \begin{pmatrix} 1 \\ \beta_1 \end{pmatrix} \right)^{-1}.$$

This is similar to the Bekker adjustment.

4.D Choosing the Number of Instruments

Donald and Newey (2001) consider the problem of choosing a subset of an infinite sequence of instruments.

They assume the instruments are ordered, so that the choice is the number of instruments to use.

The criterion they focus on is based on an estimable approximation to the expected squared error. version of this leads to approximately the same expected squared error as using the infeasible criterion.

Although in its current form not straightforward to implement, this is a very promising approach that can apply to many related problems such as generalized method of moments settings with many moments.

4.E Flores' Simulations

In one of the more extensive simulation studies Flores-Lagunes reports results comparing TSLS, LIML, Fuller, Bias corrected versions of TSLS, LIML and Fuller, a Jacknife version of TSLS (Hahn, Hausman and Kuersteiner), and the REQML estimator, in settings with 100 and 500 observations, and 5 and 30 instruments for the single endogenous variable. Does not include LIML with Bekker standard errors.

He looks at median bias, median absolute error, inter decile range, coverage rates.

He concludes that "our evidence indicates that the random-effects quasi-maximum likelihood estimator outperforms alternative estimators in terms of median point estimates and coverage rates."