NBER Summer Institute

What's New in Econometrics – Time Series

Lecture 3

July 14, 2008

Weak Instruments, Weak Identification,

and Many Instruments, Part I

Outline

Lecture 3

1) What is weak identification, and why do we care?
2) Classical IV regression I: Setup and asymptotics
3) Classical IV regression II: Detection of weak instruments
4) Classical IV regression III: hypothesis tests and confidence intervals

Lecture 4

5) Classical IV regression IV: Estimation
6) GMM I: Setup and asymptotics
7) GMM II: Detection of weak identification
8) GMM III: Hypothesis tests and confidence intervals
9) GMM IV: Estimation
10) Many instruments

1) What is weak identification, and why do we care?

<u>1a) Four examples</u>

Example #1 (cross-section IV): Angrist-Kreuger (1991),

What are the returns to education?

$Y = \log(\text{earnings})$

$X = \text{years of education}$

$Z = \text{quarter of birth}; \ k = \#\text{IVs} = 3 \text{ binary variables or up to } 178$

(interacted with YoB, state-of-birth)

$n = 329{,}509$

A-K results: $\hat{\beta}^{TSLS} = .081 \ (SE = .011)$

Then came Bound, Jaeger, and Baker (1995)…

$\Rightarrow$ The problem is that **Z** (once you include all the interactions) **is weakly correlated with** X

Example #2 (time-series IV): Estimating the elasticity of intertemporal substitution, linearized Euler equation

e.g. Campbell (2003), *Handbook of Economics of Finance*

Δc_{t+1} = consumption growth, t to $t+1$

$r_{i,t+1}$ = return on i^{th} asset, t to $t+1$

linearized Euler equation moment condition:
$$E_t(\Delta c_{t+1} - \tau_i - \psi r_{i,t+1}) = 0$$

Resulting IV estimating equation:
$$E[(\Delta c_{t+1} - \tau_i - \psi r_{i,t+1})Z_t] = 0$$

(this ignores temporal aggregation concerns)

EIS estimating equations:

$$\Delta c_{t+1} = \tau_i + \psi r_{i,t+1} + u_{i,t+1} \qquad \text{(a)}$$

or
$$r_{i,t+1} = \mu_i + (1/\psi)\Delta c_{t+1} + \eta_{i,t+1} \qquad \text{(b)}$$

Under homoskedasticity, standard estimation is by the TSLS estimator in (a) or by the inverse of the TSLS estimator in (b).

Findings in literature (e.g. Campbell (2003), US data):

- regression (a): 95% TSLS CI for ψ is (-.14, .28)
- regression (b): 95% TSLS CI for $1/\psi$ is (-.73, 2.14)

What is going on?

Reverse regression:

$$r_{i,t+1} = \mu_i + (1/\psi)\Delta c_{t+1} + \eta$$

Can you forecast Δc_{t+1} using Z_t?

$\Rightarrow$ **Z_t is weakly correlated with Δc_{t+1}**

Example #3 (linear GMM): New Keynesian Phillips Curve

e.g. Gali and Gertler (1999), where x_t = labor share; see survey by Kleibergen and Mavroeidis (2008). Hybrid NKPC with shock η_t:

$$\pi_t = \lambda x_t + \gamma_f E_t \pi_{t+1} + \gamma_b \pi_{t-1} + \eta_t$$

Rational expectations: $\quad E_{t-1}(\pi_t - \lambda x_t - \gamma_f \pi_{t+1} - \gamma_b \pi_{t-1}) = 0$

GMM moment condition: $\quad E[(\pi_t - \gamma_f \pi_{t+1} - \gamma_b \pi_{t-1} - \lambda x_t)Z_t] = 0$

Instruments: $\quad Z_t = \{\pi_{t-1}, x_{t-1}, \pi_{t-2}, x_{t-2}, \ldots\}$ (GG: 23 total)

Issues:

- Z_t needs to predict π_{t+1} − beyond π_{t-1} (included regressor)
- But predicting inflation is really hard! Atkeson-Ohanian (2001) found that, over 1985-1999 quarterly sample, it is difficult to outperform a year-over-year random walk forecast at the 4-quarter horizon

Example #4 (nonlinear GMM): Estimating the elasticity of intertemporal substitution, nonlinear Euler equation

With CRRA preferences, in standard GMM notation,

$$h(Y_t, \theta) = \delta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} \overset{G \times 1}{R_{t+1}} - \iota_G$$

where R_{t+1} is a $G \times 1$ vector of asset returns and ι_G is the G-vector of 1's. GMM moment conditions (Hansen-Singleton (1982)):

$$E[h(Y_t, \theta) \otimes Z_t] = 0 \text{ where } Z_t = \Delta c_t, R_t, \text{ etc.}$$

$\Rightarrow$ Same problem as example 2, but now nonlinear: **Z_t must predict consumption growth (and stock returns) using past data**

1b) Identification and Weak Identification

Let θ be a parameter vector, Θ be the parameter space, and $\boldsymbol{Y}$ be the vector of data.

Observational equivalence and identification

Cowles terminology, e.g. Rothenberg (1973):

- Two values of θ, θ_1 and θ_2, are **observationally equivalent** if they give rise to the same distribution of the data (i.e. $f_Y(\boldsymbol{Y};\theta_1) = f_Y(\boldsymbol{Y};\theta_2)$) for all $\boldsymbol{Y}$)

- θ is **identified at the point θ_0** if there does not exist $\theta \in \Theta$ such that (θ_0, θ) are observationally equivalent

- θ is (globally) **identified** if θ is identified at all $\theta_0 \in \Theta$

Identification – objective function definition

Let $S(\theta; Y)$ denote an objective function.

- Two values of θ, θ_1 and θ_2, are **observationally equivalent using S** if $S(\theta_1; Y) = S(\theta_2; Y)$ for all Y

- θ is **identified at the point θ_0 using S** if there does not exist $\theta' \in \Theta$ such that (θ_0, θ') are observationally equivalent

- θ is (globally) **identified using S** if θ is identified at all $\theta_0 \in \Theta$

If S is the likelihood function, this is the same as the previous definition

Identification and consistency

Identification does not imply consistent estimability. Consider the regression model, with $T \to \infty$:

$$Y_t = \beta_0 D_t + \beta_1(1 - D_t) + u_t, \text{ where } D_t = \begin{cases} 1, & t = 1,...,10 \\ 0, & t = 11,...,T \end{cases}$$

Both β_0 and β_1 are identified, but only β_1 is consistently estimable.

Working definition of weak identification

We will say that θ is **weakly identified** if the distributions of GMM or IV estimators and test statistics are not well approximated by their standard asymptotic normal or chi-squared limits because of limited information in the data.

- Departures from standard asymptotics are what matters in practice
- The source of the failures is limited information, not (for example) heavy tailed distributions, near-unit roots, unmodeled breaks, etc.
- We will focus on large samples - the source of the failure is not small-sample problems in a conventional sense. In fact most available tools for weak instruments have large-sample justifications. This is not a theory of finite sample inference (although it is closely related, at least in the linear model.)
- Throughout, we assume instrument exogeneity – weak identification is about instrument relevance, not instrument exogeneity

Two ways to think about weak identification

1) First order condition (estimating equation) interpretation:

   - In linear IV, instruments are (collectively) weakly correlated with the included endogenous regressor – Z doesn't predict the included endogenous regressor

   - In GMM, Z doesn't predict the deviation of the moment "residual" from the error term using the true value

2) Objective function interpretation:

   - The objective function is not well approximated by a (possibly local) quadratic with curvature matrix that is (i) nonrandom and (ii) does not depend on θ

<u>Some special cases:</u>

- Special cases we will come back to
    - θ is unidentified
    - Some elements of θ are strongly identified, some are weakly identified
- A special cases we won't come back to
    - θ is ***partially identified***, i.e. some elements of θ are identified and the rest are not identified
- Not a special case
    - θ is ***set identified***, i.e. the true value of θ is identified only up to a set within Θ. Weak identification and set identification could be married in theory, but they haven't been.
    - Inference when there is set identification is a hot topic in econometric theory. Set identification will come up in Lecture 7.

Additional preparatory comments

- The literature has differing degrees of maturity and completion:
  - o Testing and confidence intervals in classical (cross-sectional) IV regression model with a single included endogenous regressor: a mature area in which the first order problems are solved
  - o Estimation in general nonlinear GMM – little is known
- These lectures focus on:
  - o explaining how weak identification arises at a general level;
  - o providing practical tools and advice ("state of the art")
  - o providing references to the most recent literature (untested methods)
- Literature reviews:
  - o Stock, Yogo, Wright (2002), Hahn and Hausman (2003), Dufour (2003) (all dated but idiosyncratic and therefore interesting)
  - o Andrews and Stock (2007) (comprehensive but technical)

Outline

1) What is weak identification, and why do we care?
2) **Classical IV regression I: Setup and asymptotics**
3) Classical IV regression II: Detection of weak instruments
4) Classical IV regression III: hypothesis tests and confidence intervals
5) Classical IV regression IV: Estimation
6) GMM I: Setup and asymptotics
7) GMM II: Detection of weak identification
8) GMM III: Hypothesis tests and confidence intervals
9) GMM IV: Estimation
10) Many instruments

2) Classical IV regression I: Setup and asymptotics

Classical IV regression model & notation

Equation of interest:	$y_t = Y_t\beta + u_t, \ m = \dim(Y_t)$
k exogenous instruments Z_t:	$E(u_t Z_t) = 0, \ k = \dim(Z_t)$
Auxiliary equations:	$Y_t = \Pi' Z_t + v_t, \ \text{corr}(u_t, v_t) = \rho \ (\text{vector})$
Sampling assumption	(y_t, Y_t, Z_t) are i.i.d.

Equations in matrix form:
$$\mathbf{y} = \mathbf{Y}\beta + \mathbf{u}$$
$$\mathbf{Y} = \mathbf{Z}\Pi + \mathbf{v}$$

Comments:

- Included exogenous regressors have been omitted without loss of generality (all variables can be taken to be residuals from projections on included exogenous regressors)
- Auxiliary is just the projection of Y on Z

IV regression with one *Y* and a single irrelevant instrument

$$\hat{\beta}^{TSLS} = \frac{\mathbf{Z'y}}{\mathbf{Z'Y}} = \frac{\mathbf{Z'(Y\beta + u)}}{\mathbf{Z'Y}} = \beta + \frac{\mathbf{Z'u}}{\mathbf{Z'Y}}$$

If *Z* is irrelevant (as in Bound et. al. (1995)), then $\mathbf{Y} = \mathbf{Z}\Pi + \mathbf{v} = \mathbf{v}$, so

$$\hat{\beta}^{TSLS} - \beta = \frac{\mathbf{Z'u}}{\mathbf{Z'v}} = \frac{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}Z_t u_t}{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}Z_t v_t} \xrightarrow{d} \frac{z_u}{z_v}, \text{ where } \begin{pmatrix} z_u \\ z_v \end{pmatrix} \sim N\left(0, \begin{bmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{bmatrix} \right)$$

Comments:

- $\hat{\beta}^{TSLS}$ isn't consistent (nor should it be!)
- Distribution of $\hat{\beta}^{TSLS}$ is Cauchy-like (ratio of correlated normals)

- The distribution of $\hat{\beta}^{TSLS}$ is a ***mixture of normals with nonzero mean***: write $z_u = \delta z_v + \eta$, $\eta \perp z$, where $\delta = \sigma_{uv}/\sigma_v^2$. Then

$$\frac{z_u}{z_v} = \frac{\delta z_v + \eta}{z_v} = \delta + \frac{\eta}{z_v}, \text{ and } \frac{\eta}{z_v}\Big|z_v \sim N(0, \frac{\sigma_\eta^2}{z_v^2})$$

so **the asymptotic distribution** of $\hat{\beta}^{TSLS} - \beta_0$ is the mixture of normals,

$$\hat{\beta}^{TSLS} - (\beta_0 + \delta) \xrightarrow{d} \int N(0, \frac{\sigma_\eta^2}{z_v^2}) f_{z_v}(z_v) dz_v \text{ (1 irrelevant instrument)}$$

- heavy tails (mixture is based on inverse chi-squared)
- center of distribution of $\hat{\beta}^{TSLS}$ is $\beta_0 + \delta$. But

$$\hat{\beta}^{OLS} - \beta_0 = \frac{\mathbf{X'u}/n}{\mathbf{X'X}/n} = \frac{\mathbf{v'u}/n}{\mathbf{v'v}/n} \xrightarrow{p} \frac{\sigma_{uv}}{\sigma_v^2} = \delta, \text{ so } plim(\hat{\beta}^{OLS}) = \beta_0 + \delta$$

Thus $\hat{\beta}^{TSLS}$ is centered around $plim(\hat{\beta}^{OLS})$

This is one end of the spectrum; the usual normal approximation is the other. If instruments are weak the distribution is somewhere in between…

<u>TSLS with possibly weak instruments, 1 included endogenous regressor</u>

Suppose that $\mathbf{Z}$ is fixed and $\mathbf{u}$, $\mathbf{v}$ are normally distributed. Then the sample size enters the distribution of $\hat{\beta}^{TSLS}$ only through the **concentration parameter** μ^2, where

$$\mu^2 = \Pi'\mathbf{Z}'\mathbf{Z}\Pi/\sigma_v^2 \text{ (concentration parameter).}$$

Calculation due to Rothenberg (1984). Write $\hat{\beta}^{TSLS} = \dfrac{\mathbf{Y}'P_{\mathbf{Z}}\mathbf{y}}{\mathbf{Y}'P_{\mathbf{Z}}\mathbf{Y}}$, where

$P_{\mathbf{Z}} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$. Then

$$\mu(\hat{\beta}^{TSLS} - \beta) = \left(\frac{\sigma_u^2}{\sigma_v^2}\right)^{1/2} \frac{\zeta_u + (S_{vu}/\mu)}{1 + (2\zeta_v/\mu) + (S_{vv}/\mu^2)},$$

where $\zeta_u = \Pi'\mathbf{Z}'\mathbf{u}/(\sigma_u^2\Pi'\mathbf{Z}'\mathbf{Z}\Pi)^{1/2}$, $\zeta_v = \Pi'\mathbf{Z}'\mathbf{v}/(\sigma_v^2\Pi\mathbf{Z}'\mathbf{Z}\Pi)^{1/2}$,

$S_{vu} = \mathbf{v}'P_{\mathbf{Z}}\mathbf{u}/(\sigma_u^2\sigma_v^2)^{1/2}$, and $S_{vv} = \mathbf{v}'P_{\mathbf{Z}}\mathbf{v}/\sigma_v^2$

TSLS– estimating equation approach, ctd

$$\mu(\hat{\beta}^{TSLS} - \beta) = \left(\frac{\sigma_u^2}{\sigma_v^2}\right)^{1/2} \frac{\zeta_u + (S_{vu}/\mu)}{1 + (2\zeta_v/\mu) + (S_{vv}/\mu^2)},$$

With fixed instruments and normal errors, the distributions of ζ_u, ζ_v, S_{vu}, and S_{vv} do not depend on the sample size - the sample size enters the distribution of the TSLS estimator only through μ^2

- μ^2 plays the role usually played by n
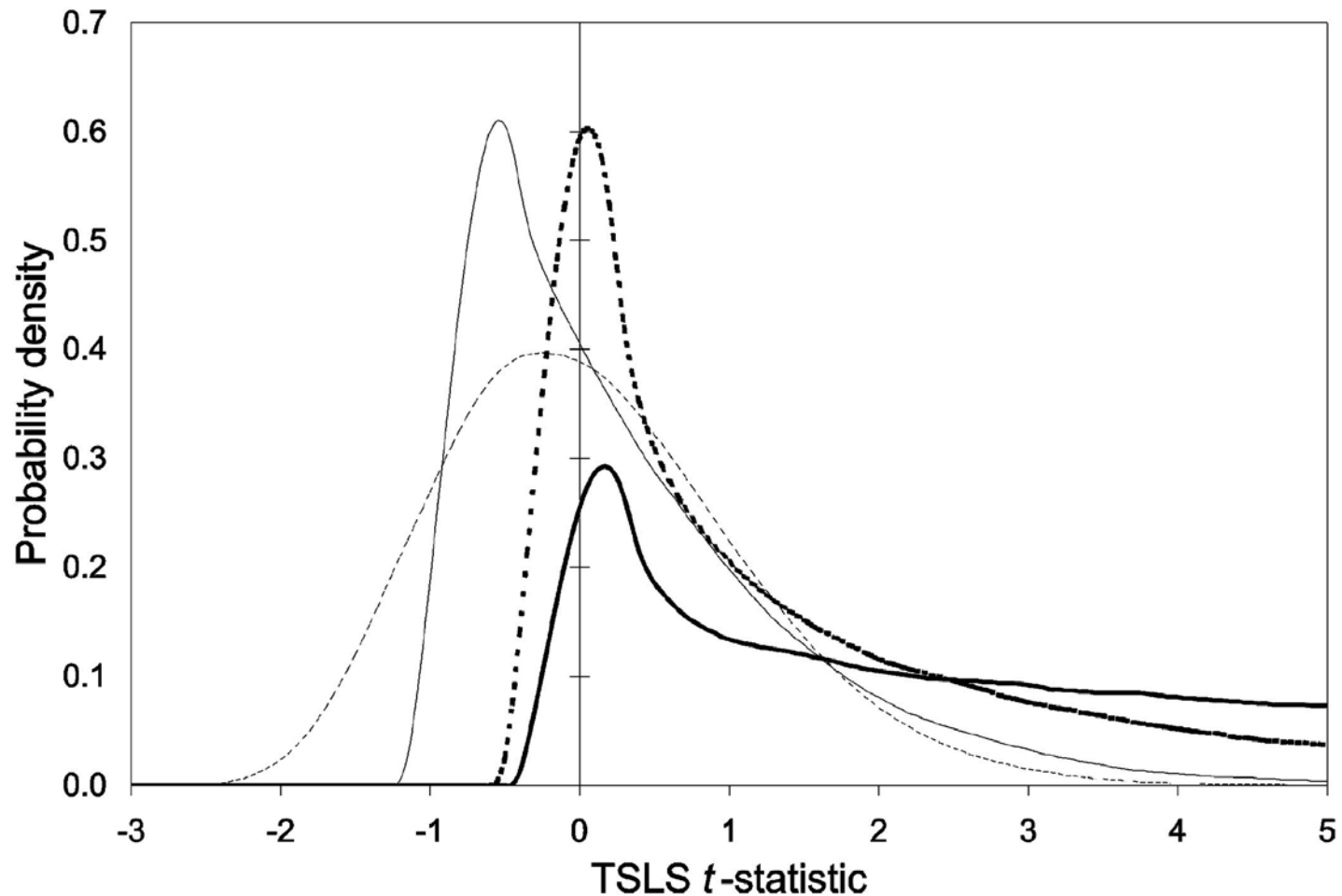- As $\mu^2 \to \infty$, the usual asymptotic approximation obtains:

$$\text{as } \mu^2 \to \infty, \ \mu(\hat{\beta}^{TSLS} - \beta) \xrightarrow{d} N(0, \sigma_u^2/\sigma_v^2)$$

  (the σ_v^2 terms in μ and limiting variance cancel)

- for small values of μ^2, the distribution is nonstandard

How important are these deviations from normality quantitatively?

Nelson-Startz (1990a,b) plots of the distribution of the TSLS t-statistic:

Dark line = irrelevant instruments; dashed light line = strong instruments;
intermediate cases: weak instruments

Four approaches to computing distributions of IV statistics with weak IVs

The goal: a distribution theory that is tractable; provides good approximations uniformly in μ^2; and can be used to compare procedures

1. Finite sample theory?

- large literature in 70s and 80s under the strong assumptions that $\mathbf{Z}$ is fixed (strictly exogenous) and (u_t, v_t) are i.i.d. normal
- literature died – distributions aren't tractable, results aren't useful

2. Edgeworth expansions?

- expand dist^n in orders of $T^{-1/2}$ – requires consistent estimability
- work poorly when instruments are very weak (Rothenberg (1984))

3. Bootstrap and subsampling?

- Neither work uniformly (irrelevant to weak to strong) in general
- We return to these later (recent interesting literature)

4. Weak instrument asymptotics

Adopt nesting that makes the concentration parameter tend to a constant as the sample size increases; that is, model F as not increasing with the sample size.

This is accomplished by setting $\Pi = C/\sqrt{T}$

- This is the Pitman drift for obtaining the local power function of the first-stage F.

- Under this nesting, $F \xrightarrow{d}$ noncentral χ^2_k/k with noncentrality parameter μ^2/k (so $F = O_p(1)$)

- Letting the parameter depend on the sample size is a common ways to obtain good approximations – e.g. local to unit roots (Bobkoski 1983, Cavanagh 1985, Chan and Wei 1987, and Phillips 1987)

Weak IV asymptotics for TSLS estimator, 1 included endogenous vble:

$$\hat{\beta}^{TSLS} - \beta_0 = (\mathbf{Y}'P_{\mathbf{Z}}\mathbf{u})/(\mathbf{Y}'P_{\mathbf{Z}}\mathbf{Y})$$

Now

$$\mathbf{Y}'P_{\mathbf{Z}}\mathbf{Y} = \left(\frac{(\mathbf{Z}\Pi + \mathbf{v})'\mathbf{Z}}{\sqrt{T}}\right)\left(\frac{\mathbf{Z}'\mathbf{Z}}{T}\right)^{-1}\left(\frac{\mathbf{Z}'(\mathbf{Z}\Pi + \mathbf{v})}{\sqrt{T}}\right)$$

$$= \left(\frac{\Pi\mathbf{Z}'\mathbf{Z}}{\sqrt{T}} + \frac{\mathbf{v}'\mathbf{Z}}{\sqrt{T}}\right)\left(\frac{\mathbf{Z}'\mathbf{Z}}{T}\right)^{-1/2'}\left(\frac{\mathbf{Z}'\mathbf{Z}}{T}\right)^{-1/2}\left(\frac{\mathbf{Z}'\mathbf{Z}\Pi}{\sqrt{T}} + \frac{\mathbf{Z}'\mathbf{v}}{\sqrt{T}}\right)$$

$$= \left[C'\left(\frac{\mathbf{Z}'\mathbf{Z}}{T}\right)^{1/2} + \frac{\mathbf{v}'\mathbf{Z}}{\sqrt{T}}\left(\frac{\mathbf{Z}'\mathbf{Z}}{T}\right)^{-1/2'}\right]\left[\left(\frac{\mathbf{Z}'\mathbf{Z}}{T}\right)^{1/2'}C + \left(\frac{\mathbf{Z}'\mathbf{Z}}{T}\right)^{-1/2}\frac{\mathbf{Z}'\mathbf{v}}{\sqrt{T}}\right]$$

$$\xrightarrow{d} (\lambda + z_v)'\,(\lambda + z_v),$$

where

$$\lambda = C'Q_{ZZ}^{1/2}, \; Q_{ZZ} = EZ_tZ_t', \text{ and } \begin{pmatrix} z_u \\ z_v \end{pmatrix} \sim N\left(0, \begin{bmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{bmatrix}\right)$$

Similarly,

$$\mathbf{Y}' P_{\mathbf{Z}} \mathbf{u} = \left(\frac{(\mathbf{Z}\Pi + \mathbf{v})'\mathbf{Z}}{\sqrt{T}} \right) \left(\frac{\mathbf{Z}'\mathbf{Z}}{T} \right)^{-1} \left(\frac{\mathbf{Z}'\mathbf{u})}{\sqrt{T}} \right)$$

$$= \left(C' \frac{\mathbf{Z}'\mathbf{Z}}{T} + \frac{\mathbf{v}'\mathbf{Z}}{\sqrt{T}} \right) \left(\frac{\mathbf{Z}'\mathbf{Z}}{T} \right)^{-1} \left(\frac{\mathbf{Z}'\mathbf{u}}{\sqrt{T}} \right)$$

$$\overset{d}{\to} (\lambda + z_v)' z_u,$$

so

$$\hat{\beta}^{TSLS} - \beta_0 \overset{d}{\to} \frac{(\lambda + z_v)' z_u}{(\lambda + z_v)'(\lambda + z_v)}$$

- Under weak instrument asymptotics, $\mu^2 \overset{p}{\to} C'Q_{ZZ}C/\sigma_v^2 = \lambda'\lambda/\sigma_v^2$

- Unidentified special case: $\hat{\beta}^{TSLS} - \beta_0 \overset{d}{\to} \dfrac{z_v' z_u}{z_v' z_v}$ (obtained earlier)

- Strong identification: $\sqrt{\lambda'\lambda}\,(\hat{\beta}^{TSLS} - \beta_0) \overset{d}{\to} N(0,\sigma_u^2)$ (standard limit)

<u>Summary of weak IV asymptotic results:</u>

- Resulting asymptotic distributions are the same as in the exact normal classical model with fixed Z – but with **known** covariance matrices.

- IV estimators are not consistent (and are biased) under this nesting

- IV estimators are nonnormal ($\hat{\beta}^{TSLS}$ has mixture of normals with nonzero mean, where mean $\propto k/\mu^2$)

- Test statistics (including the J-test of overidentifying restrictions) do not have normal or chi-squared distributions

- Conventional confidence intervals do not have correct coverage (coverage can be driven to zero)

- Provide good approximations to sampling distributions uniformly in μ^2 for T moderate or greater (say, 100+ observations).

- Remember, μ^2 is unknown – so these distributions can't be used directly in practice to obtain a "corrected" distribution….

Outline

3) Classical IV regression II: Detection of weak instruments

Bound et. al. revisited

- $n = 329{,}509$ (it is μ^2, or μ^2/k, not sample size that matters!)
- for $K = 3$ (quarter of birth only), $F = 30.53$,
    - Recall that $E(F) = 1 + \mu^2/k$
    - Estimate of μ^2/k is 29.53
    - Estimate μ^2 as $k(F-1) = 3 \times (30.53-1) = 88.6$
- for $K = 178$ (all interactions), $F = 1.869$
    - Estimate of $\mu^2 = 178 \times (1.869-1) = 154.7$
    - Estimate of μ^2/k is 0.869
- We will see that numerical work suggests that
    - $\mu^2/k = 29.53$: strong instruments
    - $\mu^2/k = 0.869$: very weak instruments

How weak is weak? Need a cutoff value for μ^2

The basic idea is to compare F to some cutoff. But how should that cutoff be chosen? In general, this depends on the statistic you are using (different statistics have different sensitivities to μ^2). TSLS is among the worst (most sensitive) – and is also most frequently used. So, it is reasonable to develop a cutoff for F assuming use of TSLS.

Various procedures:
- First stage $F > 10$ rule of thumb
- Stock-Yogo (2005a) bias method
- Stock-Yogo (2005a) size method
- Hahn-Hausman (2003) test
- Other methods (R^2, partial R^2, Shea (1997), etc.)

TSLS bias cutoff method (Stock-Yogo (2005a))

Let $\mu^2_{10\%bias}$ be the value of μ^2 such that, if $\mu^2 \leq \mu^2_{10\%bias}$, the maximum bias of TSLS will be no more than 10% of the bias (inconsistency) of OLS. Stock-Yogo (2005a): decision rule of the form:

$$\text{if } F \begin{pmatrix} \leq \\ > \end{pmatrix} \kappa_{.10}(k), \text{ conclude that instruments are } \begin{pmatrix} \text{weak} \\ \text{strong} \end{pmatrix}$$

where F is the first stage F-statistic* and $\kappa_{.10}(k)$ is chosen so that P($F > \kappa_{.10}(k)$; $\mu^2 = \mu^2_{10\%bias}$) = .05 (so that the rule acts like a 5% significance test at the boundary value $\mu^2 = \mu^2_{10\%bias}$).

*$F = F$-statistic testing the hypothesis that the coefficients on $Z_t = 0$ in the regression of Y_t on Z_t, W_t, and a constant, where W_t = the exogenous regressors included in the equation of interest.

TSLS bias cutoff method (Stock-Yogo (2005a)), ctd

Some background:

The relative squared normalized bias of TSLS to OLS is,

$$B_n^2 = \frac{(E\hat{\beta}^{\mathrm{IV}} - \beta)' \Sigma_{YY} (E\hat{\beta}^{\mathrm{IV}} - \beta)}{(E\hat{\beta}^{\mathrm{OLS}} - \beta)' \Sigma_{YY} (E\hat{\beta}^{\mathrm{OLS}} - \beta)}$$

The square root of the maximal relative squared asymptotic bias is:

$$B^{max} = \max_{\rho: \, 0 < \rho'\rho \leq 1} \lim_{n \to \infty} |B_n|, \text{ where } \rho = \mathrm{corr}(u_t, v_t)$$

This maximization problem is a ratio of quadratic forms so it turns into a (generalized) eigenvalue problem; algebra reveals that the solution to this eigenvalues problem depends only on μ^2/k and k; this yields the cutoff μ_{bias}^2.

Value of cutoff μ^2_{bias}/k to ensure indicated maximal bias (Stock-Yogo, 2005)

Boundary of weak instrument set ($n = 1$)
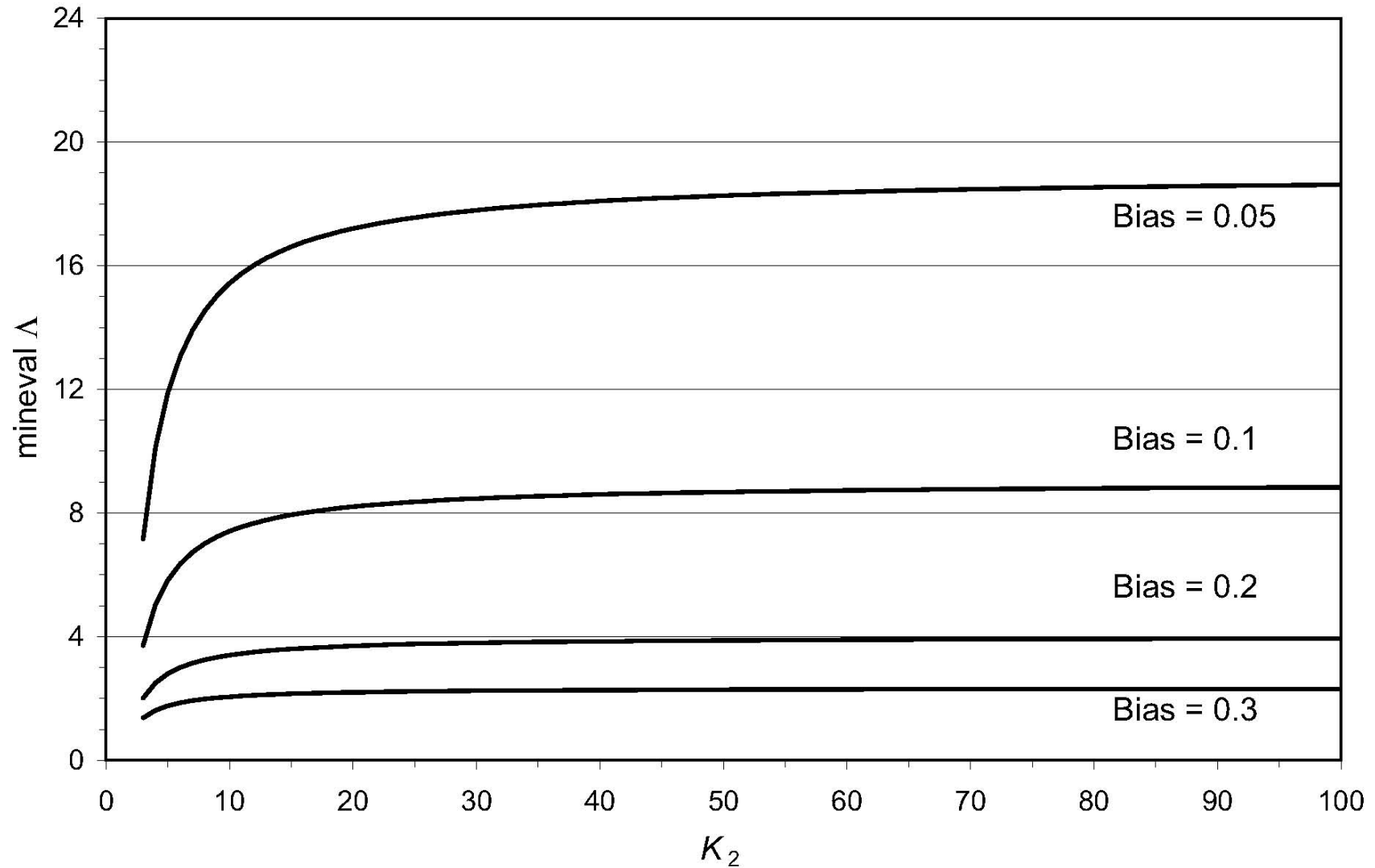
Critical values

One included endogenous regressor

The 5% critical value of the test is the 95% percentile value of the noncentral χ_k^2/k distribution, with noncentrality parameter μ_{bias}^2/k

Multiple included endogenous regressors

The Cragg-Donald (1993) statistic is:

$$g_{min} = \text{mineval}(G_T), \text{ where } G_T = \hat{\Sigma}_{VV}^{-1/2}\prime \mathbf{Y}'P_{\mathbf{Z}}\mathbf{Y}\hat{\Sigma}_{VV}^{-1/2}/k,$$

- G_T is essentially a matrix first stage F statistic
- Critical values are given in Stock-Yogo (2005a)

Software

STATA (ivreg2),…

5% critical value of F to ensure indicated maximal bias (Stock-Yogo, 2005a)

Critical value at 5% significance ($n = 1$)
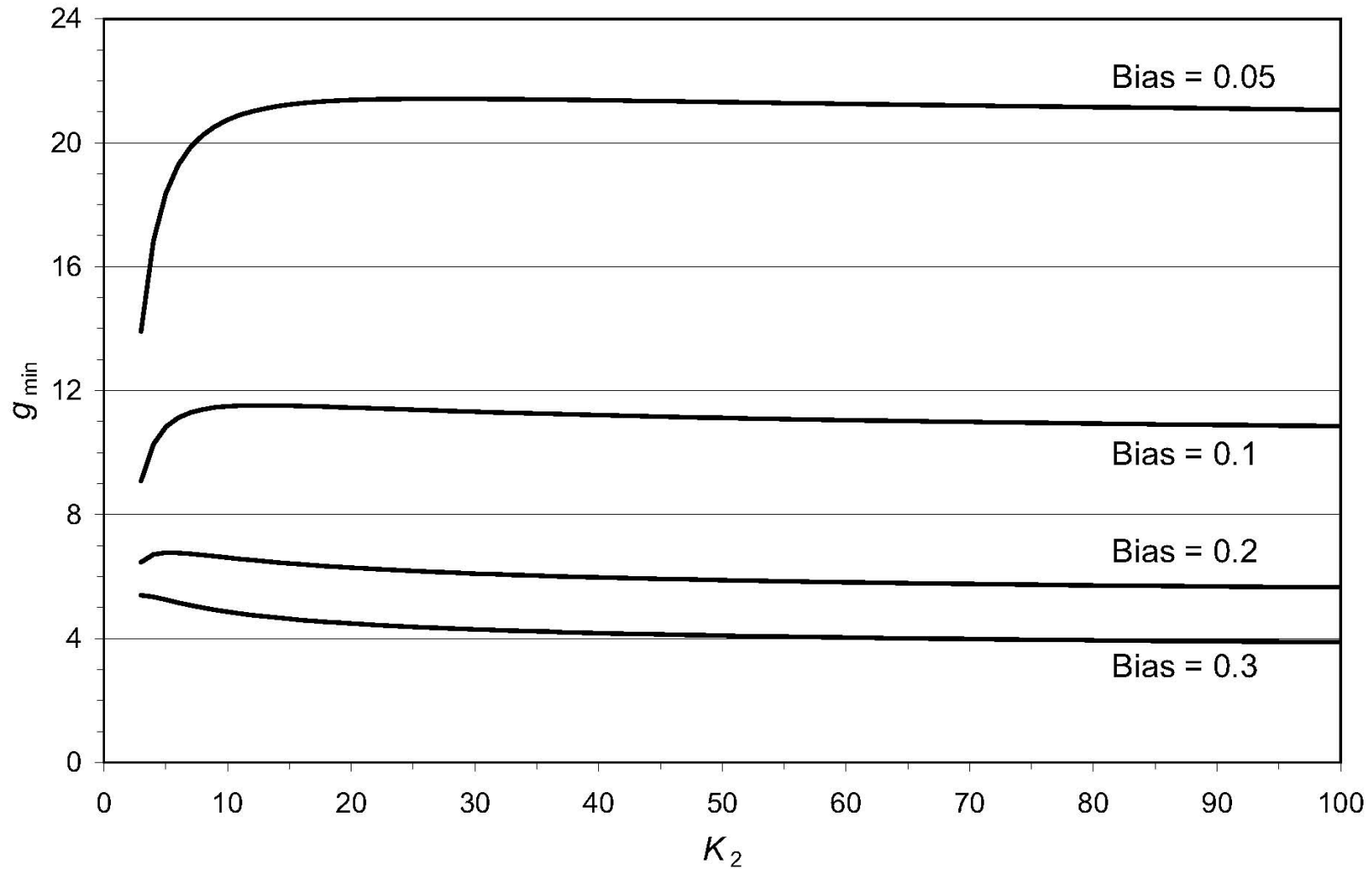
To ensure 10% maximal bias, need $F < 11.52$; F < 10 is a rule of thumb

5% critical values for Weak IV test statistic g_{min}, for 10% maximal TSLS Bias (Stock-Yogo (2005), Table 1) $m = \dim(Y_t)$

k	m = 1	m = 2	m = 3
3	9.08	–	–
4	10.27	7.56	–
5	10.83	8.78	6.61
6	11.12	9.48	7.77
7	11.29	9.92	8.50
8	11.39	10.22	9.01
9	11.46	10.43	9.37
10	11.49	10.58	9.64
15	11.51	10.93	10.33
20	11.45	11.03	10.60
25	11.38	11.06	10.71
30	11.32	11.05	10.77

Other methods for detecting weak instruments

Stock-Yogo (2005a) size method

- Instead of controlling bias, control the size of a Wald test of $\beta = \beta_0$
- Less frequently used
- Not really relevant (any more) since fully robust methods for testing exist

Hahn-Hausman (2003) test

1)   Idea is to test the null of strong instruments, under which the TSLS estimator, and the inverse of the TSLS estimator from the "reverse" regression, should be the same (recall the Campbell linearized CCAPM results)

Other methods for detecting weak instruments, ctd.

2)   Unfortunately the HH test is not consistent against weak instruments (power of 5% level test depends on parameters, is typically $\approx$ 15-20% (Hausman, Stock, Yogo (2005))

Examination of R^2, partial R^2, or adjusted R^2

- None of these are a good idea, more precisely, what needs to be large is the concentration parameter, not the R^2. An $R^2 = .10$ is small if $T = 50$ but is large if $T = 5000$.
- The first-stage R^2 is especially uninformative if the first stage regression has included exogenous regressors (W's) because it is the marginal explanatory content of the Z's, given the W's, that matters.

Outline

1) What is weak identification, and why do we care?
2) Classical IV regression I: Setup and asymptotics
3) Classical IV regression II: Detection of weak instruments
4) **Classical IV regression III: hypothesis tests and confidence intervals**
5) Classical IV regression IV: Estimation
6) GMM I: Setup and asymptotics
7) GMM II: Detection of weak identification
8) GMM III: Hypothesis tests and confidence intervals
9) GMM IV: Estimation
10) Many instruments

4) Classical IV regression III:
Hypothesis tests and confidence intervals

There are two approaches to improving inference (providing tools):

Fully robust methods:

- Inference that is valid for any value of the concentration parameter, including zero, at least if the sample size is large, under weak instrument asymptotics
  - o For tests: asymptotically correct size (and good power!)
  - o For confidence intervals: asymptotically correct coverage rates
  - o For estimators: asymptotically unbiased (or median-unbiased)

Partially robust methdos:

- Methods are less sensitive to weak instruments than TSLS – e.g. bias is "small" for a "large" range of μ^2

Fully Robust Testing

- The TSLS t-statistic has a distribution that depends on μ^2, which is unknown

- Approach #1: use a statistic whose distribution depends on μ^2, but use a "worst case" conservative critical value
  - This is unattractive – substantial power loss

- Approach #2: use a statistic whose distribution does not depend on μ^2 (two such statistics are known)

- Approach #3: use statistics whose distribution depends on μ^2, but compute the critical values as a function of another statistic that is sufficient for μ^2 under the null hypothesis.
  - Both approaches 2 and 3 have advantages and disadvantages – we discuss both

Approach #2: Tests that are valid unconditionally

(that is, the distribution of the test statistic does not depend on μ^2)

The Anderson-Rubin (1949) test

Consider $H_0: \beta = \beta_0$ in

$$\mathbf{y} = \mathbf{Y}\beta + \mathbf{u},$$
$$\mathbf{Y} = \mathbf{Z}\Pi + \mathbf{v}$$

The Anderson-Rubin (1949) statistic is the F-statistic in the regression of $\mathbf{y} - \mathbf{Y}\beta_0$ on $\mathbf{Z}$.

$$\text{AR}(\beta_0) = \frac{(\mathbf{y} - \mathbf{Y}\beta_0)' P_{\mathbf{Z}} (\mathbf{y} - \mathbf{Y}\beta_0) / k}{(\mathbf{y} - \mathbf{Y}\beta_0)' M_{\mathbf{Z}} (\mathbf{y} - \mathbf{Y}\beta_0) / (T - k)}$$

$$AR(\beta_0) = \frac{(\mathbf{y} - \mathbf{Y}\beta_0)' P_{\mathbf{Z}} (\mathbf{y} - \mathbf{Y}\beta_0) / k}{(\mathbf{y} - \mathbf{Y}\beta_0)' M_{\mathbf{Z}} (\mathbf{y} - \mathbf{Y}\beta_0) / (T - k)}$$

Comments

- $AR(\hat{\beta}^{TSLS}) =$ the J-statistic

- Null distribution doesn't depend on μ^2:

  Under the null, $\mathbf{y} - \mathbf{Y}\beta_0 = \mathbf{u}$, so

  $$AR = \frac{\mathbf{u}' P_{\mathbf{Z}} \mathbf{u} / k}{\mathbf{u}' M_{\mathbf{Z}} \mathbf{u} / (T - k)} \sim F_{k, n-k} \qquad \text{if } u_t \text{ is normal}$$

  $$AR \xrightarrow{d} \chi_k^2 / k \qquad \text{if } u_t \text{ is i.i.d. and } Z_t u_t \text{ has 2 moments (CLT)}$$

- The distribution of AR under the alternative depends on μ^2 – more information, more power (of course)

The AR statistic if there are included endogenous regressors

Let $\mathbf{W}$ denote the matrix of observations on included exogenous regressors, so the structural equation and first stage regression are,

$$\mathbf{y} = \mathbf{Y}\beta + \mathbf{W}\gamma + \mathbf{u}$$

$$\mathbf{Y} = \mathbf{Z}\Pi + \mathbf{W}\Pi_W + \mathbf{v}$$

Then the AR statistic is the F-statistic testing the hypothesis that the coefficients on $\mathbf{Z}$ are zero in the regression of $\mathbf{y} - \mathbf{Y}\beta_0$ on $\mathbf{Z}$ and $\mathbf{W}$.

Advantages and disadvantages of AR

Advantages

- Easy to use – entirely regression based
- Uses standard F critical values
- Works for $m > 1$ (general dimension of Z)

Disadvantages

- Difficult to interpret: rejection arises for two reasons: β_0 is false *or* Z is endogenous
- Power loss relative to other tests (we shall see)
- Is not efficient if instruments are strong – under strong instruments, not as powerful as TSLS Wald test (power loss because $AR(\beta_0)$ has k degrees of freedom)

Kleibergen's (2002) LM test

Kleibergen developed an LM test that has a null distribution that is χ_1^2 - doesn't depend on μ^2.

Advantages

- Fairly easy to implement
- Is efficient if instruments are strong

Disadvantages

- Has very strange power properties (we shall see)
- Its power is dominated by the conditional likelihood ratio test

Approach #3: Conditional tests

Conditional tests have rejection rate 5% for all points under the null (β_0, μ^2) ("similar tests")

Recall your first semester probability and statistics course…

- Let S be a statistic with a distribution that depends on θ
- Let T be a sufficient statistic for θ
- Then the distribution of $S|T$ does not depend on θ

Here (Moreira (2003)):

- LR will be a statistic testing $\beta = \beta_0$ (LR is "S" in notation above)
- Q_T will be sufficient for μ^2 under the null (Q_T is "T")
- Thus the distribution of $LR|\,Q_T$ does not depend on μ^2 under the null
- Thus valid inference can be conducted using the quantiles of $LR|\,Q_T$ — that is, critical values that are a function of Q_T

Moreira's (2003) conditional likelihood ratio (CLR) test

$$LR = \max_\beta \text{log-likelihood}(\beta) - \text{log-likelihood}(\beta_0)$$

After lots of algebra, this becomes:

$$LR = \tfrac{1}{2}\{\hat{Q}_S - \hat{Q}_T + [(\hat{Q}_S - \hat{Q}_T)^2 + 4\hat{Q}_{ST}^2]^{1/2}\}$$

where

$$\hat{Q} = \begin{bmatrix} \hat{Q}_S & \hat{Q}_{ST} \\ \hat{Q}_{ST} & \hat{Q}_T \end{bmatrix} = \hat{J}_0{'}\hat{\Omega}^{-1/2}\mathbf{Y}^{+\prime}P_{\mathbf{Z}}\mathbf{Y}^{+}\hat{\Omega}^{-1/2\prime}\hat{J}_0$$

$$\hat{\Omega} = \mathbf{Y}^{+\prime}M_{\mathbf{Z}}\mathbf{Y}^{+}/(T{-}k), \quad \mathbf{Y}^{+} = (\mathbf{y} \ \ \mathbf{Y})$$

$$\hat{J}_0 = \begin{bmatrix} \dfrac{\hat{\Omega}^{1/2\prime}b_0}{\sqrt{b_0{'}\hat{\Omega}b_0}} & \dfrac{\hat{\Omega}^{-1/2}a_0}{\sqrt{a_0{'}\hat{\Omega}^{-1}a_0}} \end{bmatrix}, \quad b_0 = \begin{pmatrix} 1 \\ -\beta_0 \end{pmatrix} \quad a_0 = \begin{pmatrix} \beta_0 \\ 1 \end{pmatrix}.$$

CLR test, ctd.

Implementation:

- Q_T is sufficient for μ^2 (under weak instrument asymptotics)
- The distribution of $LR|Q_T$ does not depend on μ^2
- LR proc exists in STATA (condivreg), GAUSS
- STATA (condivreg), Gauss code for computing LR and conditional p-values exists

<u>Advantages and disadvantages of the CLR test</u>

Advantages

- More powerful than AR or LM
- In fact, effectively uniformly most powerful among valid tests that are invariant to rotations of the instruments (Andrews, Moreira, Stock (2006) – among similar tests; Andrews, Moreira, Stock (2008) – among nonsimilar tests)
- Implemented in software (STATA,…)

Disadvantages

- More complicated to explain and write down
- Only developed (so far) for a single included endogenous regressor
- As written, the software requires homoskedastic errors; extensions to heteroskedasticity and serial correlation have been developed but are not in common statistical software

Web link:

Some power comparisons: AR, LM, CLR

Local link:

Some power comparisons: AR, LM, CLR

Full results on power for various tests are in Andrews, Moreira, Stock (2006) and on Stock's Harvard Econ Web site

Confidence Intervals

(a) A 95% confidence set is a function of the data contains the true value in 95% of all samples

(b) A 95% confidence set is constructed as the set of values that cannot be rejected as true by a test with 5% significance level

Usually (b) leads to constructing confidence sets as the set of β_0 for which

$$-1.96 < \frac{\hat{\beta} - \beta_0}{SE(\hat{\beta})} < 1.96. \text{ Inverting this } t\text{-statistic yields } \hat{\beta} \pm 1.96 SE(\hat{\beta})$$

- This won't work for TSLS $- t^{TSLS}$ isn't normal (the critical values of t^{TSLS} depend on μ^2)
- Dufour (1997) impossibility result for weak instruments: unbounded intervals must occur with positive probability.
- However, you can compute a valid, fully robust confidence interval by inverting a fully robust test!

(1) Inversion of AR test: AR Confidence Intervals

$$95\% \text{ CI} = \{\beta_0: \text{AR}(\beta_0) < F_{k,T-k;.05}\}$$

Computational issues:

- For $m = 1$, this entails solving a quadratic equation:

$$\text{AR}(\beta_0) = \frac{(\mathbf{y} - \mathbf{Y}\beta_0)' P_{\mathbf{Z}} (\mathbf{y} - \mathbf{Y}\beta_0) / k}{(\mathbf{y} - \mathbf{Y}\beta_0)' M_{\mathbf{Z}} (\mathbf{y} - \mathbf{Y}\beta_0) / (T - k)} < F_{k,T-k;.05}$$

- For $m > 1$, solution can be done by grid search or using methods in Dufour and Taamouti (2005)

- Sets for a single coefficient can be computed by projecting the larger set onto the space of the single coefficient (see Dufour and Taamouti (2005)), also see recent work by Kleibergen (2008)

AR confidence intervals, ctd.

$$95\% \text{ CI} = \{\beta_0: \text{AR}(\beta_0) < F_{k,T-k;.05}\}$$

Four possibilities:

- a single bounded confidence interval
- a single unbounded confidence interval
- a disjoint pair of confidence intervals
- an empty interval

Note:

- Difficult to interpret
- Intervals aren't efficient (AR test isn't efficient) under strong instruments

(2) Inversion of CLR test: CLR Confidence Intervals

$$95\% \ CI = \{\beta_0 : LR(\beta_0) < cv_{.05}(Q_T)\}$$

where $cv_{.05}(Q_T) = 5\%$ conditional critical value

Comments:

- Efficient GAUSS and STATA (condivreg) software
- Will contain the LIML estimator (Mikusheva (2005))
- Has certain optimality properties: nearly uniformly most accurate invariant; also minimum expected length in polar coordinates (Mikusheva (2005))
- Only available for $m = 1$

Example #2: Consumption CAPM and the EIS

Yogo (2004)

Δc_{t+1} = consumption growth, t to $t+1$

$r_{i,t+1}$ = return on i^{th} asset, t to $t+1$

Moment conditions: $\qquad E_t(\Delta c_{t+1} - \tau_i - \psi r_{i,t+1}) = 0$

EIS estimating equations: $\qquad \Delta c_{t+1} = \tau_i + \psi r_{i,t+1} + u_{i,t+1}$

or $\qquad\qquad\qquad\qquad\quad r_{i,t+1} = \mu_i + (1/\psi)\Delta c_{t+1} + \eta_{i,t+1}$

Under homoskedasticity, standard estimation is by TSLS or by the inverse of the TSLS estimator (remember Hahn-Hausman (2003) test?); but with weak instruments, the normalization matters

First stage *F*-statistics for EIS (Yogo (2004)):

THE REVIEW OF ECONOMICS AND STATISTICS

TABLE 1.—TEST FOR WEAK INSTRUMENTS

Country	Sample Period	Variable	F	p-Value			
				TSLS Bias	TSLS Size	Fuller-k	LIML
USA	1947.3–1998.4	Δc	2.93	0.93	1.00	0.53	0.37
		r_f	15.53	0.00	0.66	0.00	0.00
		r_e	2.88	0.93	1.00	0.54	0.39
AUL	1970.3–1998.4	Δc	1.79	0.99	1.00	0.81	0.69
		r_f	21.81	0.00	0.14	0.00	0.00
		r_e	1.82	0.99	1.00	0.80	0.68
CAN	1970.3–1999.1	Δc	3.03	0.92	1.00	0.50	0.35
		r_f	15.37	0.00	0.67	0.00	0.00
		r_e	2.51	0.96	1.00	0.64	0.48
FR	1970.3–1998.3	Δc	0.17	1.00	1.00	1.00	1.00
		r_f	38.43	0.00	0.00	0.00	0.00
		r_e	3.09	0.91	1.00	0.49	0.34
GER	1979.1–1998.3	Δc	0.83	1.00	1.00	0.97	0.93
		r_f	17.66	0.00	0.45	0.00	0.00
		r_e	0.69	1.00	1.00	0.98	0.95
ITA	1971.4–1998.1	Δc	0.73	1.00	1.00	0.98	0.95

Various estimates of the EIS, forward and backward

TABLE 2.—ESTIMATES OF THE EIS USING THE INTEREST RATE

Country	Sample Period	$1/\psi$			ψ		
		TSLS	Fuller-k	LIML	TSLS	Fuller-k	LIML
USA	1947.3–1998.4	0.68 (0.48)	3.30 (3.20)	34.11 (112.50)	0.06 (0.09)	0.03 (0.10)	0.03 (0.10)
AUL	1970.3–1998.4	0.50 (0.48)	2.37 (2.45)	30.03 (107.71)	0.05 (0.11)	0.04 (0.12)	0.03 (0.12)
CAN	1970.3–1999.1	−1.04 (0.39)	−2.40 (1.13)	−2.98 (1.54)	−0.30 (0.16)	−0.33 (0.17)	−0.34 (0.17)
FR	1970.3–1998.3	−3.12 (3.75)	−1.83 (1.72)	−12.38 (29.61)	−0.08 (0.19)	−0.08 (0.19)	−0.08 (0.19)
GER	1979.1–1998.3	−1.05 (0.62)	−1.38 (0.90)	−2.29 (1.87)	−0.42 (0.35)	−0.43 (0.35)	−0.44 (0.36)
ITA	1971.4–1998.1	−3.34 (1.98)	−5.82 (4.47)	−14.81 (18.55)	−0.07 (0.08)	−0.07 (0.08)	−0.07 (0.08)
JAP	1970.3–1998.4	−0.18 (0.43)	−0.86 (1.23)	−21.56 (106.53)	−0.04 (0.21)	−0.04 (0.23)	−0.05 (0.23)

AR, LM, and CLR confidence intervals for ψ:

TABLE 3.—WEAK-INSTRUMENT-ROBUST CONFIDENCE INTERVALS FOR THE
EIS USING THE INTEREST RATE

Country	Sample Period	AR	LM	Cond. LR
USA	1947.3–1998.4	$\varnothing$	[−0.21, 0.23]	[−0.19, 0.22]
AUL	1970.3–1998.4	[−0.16, 0.21]	[−0.22, 13.74]	[−0.22, 0.27]
CAN	1970.3–1999.1	[−0.54, −0.14]	[−0.73, 14.15]	[−0.71, 0.00]
FR	1970.3–1998.3	[−0.68, 0.53]	[−0.47, 0.31]	[−0.48, 0.33]
GER	1979.1–1998.3	[−1.57, 0.54]	[−1.21, 0.26]	[−1.23, 0.28]
ITA	1971.4–1998.1	[−0.29, 0.18]	[−0.24, 0.11]	[−0.24, 0.12]
JAP	1970.3–1998.4	[−0.60, 0.49]	[−∞, ∞]	[−0.56, 0.45]
NTH	1977.3–1998.4	[−0.91, 0.64]	[−∞, ∞]	[−0.76, 0.48]
SWD	1970.3–1999.2	[−0.30, 0.29]	[−∞, ∞]	[−0.22, 0.21]
SWT	1976.2–1998.4	[−1.69, 0.37]	[−1.19, 0.07]	[−1.22, 0.09]
UK	1970.3–1999.1	[0.04, 0.28]	[−∞, ∞]	[−0.12, 0.43]
USA	1970.3–1998.4	$\varnothing$	[−∞, ∞]	[−0.23, 0.23]
SWD	1921–1994	[−0.30, 0.40]	[−∞, ∞]	[−0.25, 0.35]
UK	1921–1994	[−0.05, 0.88]	[0.01, 0.70]	[0.01, 0.70]
USA	1891–1995	[−0.49, 0.46]	[−∞, ∞]	[−∞, ∞]

The table reports 95% confidence intervals for the EIS, constructed from AR, LM, and conditional LR tests. $\varnothing$ indicates an empty confidence interval. The instruments are twice lagged nominal interest rate, inflation, consumption growth, and log dividend-price ratio.

What about stock returns – should they "work"?

TABLE 5.—WEAK-INSTRUMENT-ROBUST CONFIDENCE INTERVALS FOR THE
EIS USING THE STOCK RETURN

Country	Sample Period	AR	LM	Cond. LR
USA	1947.3–1998.4	[−0.21, −0.02]	[−∞, ∞]	[−∞, ∞]
AUL	1970.3–1998.4	[−∞, ∞]	[−∞, ∞]	[−∞, ∞]
CAN	1970.3–1999.1	[0.02, 4.03]	[0.05, 0.35]	[0.04, 0.41]
FR	1970.3–1998.3	[−0.28, 0.20]	[−∞, ∞]	[−0.16, 0.11]
GER	1979.1–1998.3	[−∞, ∞]	[−∞, ∞]	[−∞, ∞]
ITA	1971.4–1998.1	[−∞, ∞]	[−∞, ∞]	[−∞, ∞]
JAP	1970.3–1998.4	[−0.05, 0.32]	[−1.01, 0.20]	[−0.02, 0.21]
NTH	1977.3–1998.4	[−∞, ∞]	[−∞, ∞]	[−∞, ∞]
SWD	1970.3–1999.2	[−∞, ∞]	[−∞, ∞]	[−∞, ∞]
SWT	1976.2–1998.4	[−∞, ∞]	[−∞, ∞]	[−∞, ∞]
UK	1970.3–1999.1	[−0.51, −0.02]	[−∞, ∞]	[−∞, ∞]
USA	1970.3–1998.4	[−∞, ∞]	[−∞, ∞]	[−∞, ∞]
SWD	1921–1994	[−∞, ∞]	[−∞, ∞]	[−∞, ∞]
UK	1921–1994	[−0.04, 0.10]	[−∞, ∞]	[−0.10, 0.14]
USA	1891–1995	[−∞, ∞]	[−∞, ∞]	[−∞, ∞]

See notes to table 3.

Summary: EIS and CCAPM

- *a-priori* reason for thinking that instruments are weak
- Empirical pathologies evident from TSLS: strongly different estimators for different instrument lists, reverse and forward estimators strikingly different (HH intuition)
- First-stage F's confirm that instruments are weak
- Point estimation is difficult but LIML and Fuller are more reliable than TSLS
- AR and CLR confidence intervals are reliable and give similar answers: not much precision, but EIS appears to be small (<1)
- Yogo (2004) is a template for how to proceed (1 endogenous regressor)

Extensions to >1 included endogenous regressor

- Usually the extension to higher dimensions is easy – standard normal t-ratios, chi-squared F-tests, etc.

- That extension all rests on normality. Once normality of estimators and chi-squared tests are gone, the extensions are not easy.

- CLR exists in theory, but unsolved computational issues because the conditioning statistic has dimension $m(m+1)/2$ (Kleibergen (2007))

- Can test joint hypothesis $H_0: \beta = \beta_0$ using the AR statistic:

$$AR(\beta_0) = \frac{(\mathbf{y} - \mathbf{Y}\beta_0)' P_{\mathbf{Z}} (\mathbf{y} - \mathbf{Y}\beta_0) / k}{(\mathbf{y} - \mathbf{Y}\beta_0)' M_{\mathbf{Z}} (\mathbf{y} - \mathbf{Y}\beta_0) / (T - k)}$$

under $H_0,$ $AR \xrightarrow{d} \chi_k^2 / k$