ECMA 31100: Intro to Empirical Analysis II

Weak Instruments

Joe Hardwick

University of Chicago

Winter 2022

Weak Instruments

- Consider the model

$$y = x'\beta + u; \quad \mathrm{E}(zu) = 0,$$

where $y \in \mathbb{R}$, $x \in \mathbb{R}^{k+1}$, $z \in \mathbb{R}^{l+1}$ and $u \in \mathbb{R}$ is unobserved.
- Validity condition $\mathrm{E}(zu) = 0$ implies there exists a β such that:

$$\mathrm{E}(zy) = \mathrm{E}\left(zx'\right)\beta.$$

- Rank condition $\mathrm{E}(zx')$ has rank $k+1$ implies the solution is unique.

Weak Instruments

- Question: What if $\mathrm{E}(zx')$ has full rank, but its columns are 'almost' linearly dependent?
- Special case: Scalar endogenous variable, scalar excluded instrument z, no covariates:

$$y = \beta_0 + \beta_1 d + u; \quad \mathrm{E}(zu) = 0;$$
$$d = \pi_0 + \pi_1 z + v; \quad \mathrm{E}(zv) = 0.$$

Second equation is the 'first stage regression' in an IV approach. Identification requires

$$\pi_1 = \frac{Cov(d, z)}{Var(z)} \neq 0.$$

What are the properties of the IV estimator if $Cov(d, z) \approx 0$?

Weak Instruments

- Our current asymptotic approximations do not distinguish between $Cov(d,z) \approx 0$ and $Cov(d,z) >> 0$ or $Cov(d,z) << 0$, except to say that the standard errors may be larger if this correlation is weak.
- In our special case, and assuming $\mathrm{E}(u|z) = 0$; $\mathrm{E}\left(u^2|z\right) = \sigma^2$:

$$\sqrt{n}\left(\hat{\beta}_{1,IV} - \beta\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma^2}{Corr(d,z)^2\, Var(d)}\right).$$

- If in fact there is no endogeneity issue and $\mathrm{E}(u|d) = 0$, $\mathrm{E}\left(u^2|d\right) = \sigma^2$, the OLS estimate is more efficient:

$$\sqrt{n}\left(\hat{\beta}_{1,OLS} - \beta_1\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma^2}{Var(d)}\right).$$

Weak Instruments

- There is no issue with the theory involving the asymptotic approximation: We get larger standard errors with weaker correlation between d, z.
- However, in finite samples the true distribution of the IV estimator may not resemble its asymptotic counterpart.
- We will come up with a different asymptotic approximation which (hopefully) better represents the finite sample distribution of $\hat{\beta}_{IV}$ and which nests the existing approximation in case our instruments are not weak ($Corr\,(d, z) \not\approx 0$).

Weak Instruments: Example

- For now suppose that

$$y = \beta x + u; \quad \mathrm{E}\left(zu\right) = 0$$
$$x = \pi z + v; \quad \mathrm{E}\left(zv\right) = 0$$

where all random variables are scalar. The IV estimator is

$$\hat{\beta}_{IV} = \frac{\frac{1}{n}\sum_{i=1}^{n} z_i y_i}{\frac{1}{n}\sum_{i=1}^{n} z_i x_i} = \beta + \frac{\frac{1}{n}\sum_{i=1}^{n} z_i u_i}{\frac{1}{n}\sum_{i=1}^{n} z_i x_i}.$$

- Goal is to see what happens when $\pi \approx 0$, so plug in x_i:

Weak Instruments: Example

- Note that

$$\sqrt{n}\left(\hat{\beta}_{IV} - \beta\right) = \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i u_i}{\left(\frac{1}{n} \sum_{i=1}^{n} z_i^2\right) \cdot \pi + \frac{1}{n} \sum_{i=1}^{n} z_i v_i}.$$

- First let's see what happens when $\pi = 0$ (rank condition fails):

$$\sqrt{n}\left(\hat{\beta}_{IV} - \beta\right) = \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i u_i}{\frac{1}{n} \sum_{i=1}^{n} z_i v_i}.$$

- Numerator should converge in distribution, but denominator converges in probability to 0! Can't apply Slutsky.

Weak Instruments: Example

- Need to find a different rate of convergence. (Imagine example on slide 4 with $Corr(x, z) = 0$).
- Find that
$$\hat{\beta}_{IV} - \beta = \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i u_i}{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i v_i}.$$
Let's use the multivariate CLT and the continuous mapping theorem:

Weak Instruments: Example

- By the multivariate CLT:

$$\begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i u_i \\ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i v_i \end{pmatrix} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \begin{pmatrix} z_i u_i \\ z_i v_i \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, \begin{pmatrix} \mathrm{E}\left(u^2 z^2\right) & \mathrm{E}\left(z^2 uv\right) \\ \mathrm{E}\left(z^2 uv\right) & \mathrm{E}\left(z^2 v^2\right) \end{pmatrix}\right)$$

so by the continuous mapping theorem, since the limit distribution of $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i v_i$ is normal and equals zero with probability zero,

$$\hat{\beta}_{IV} - \beta = \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i u_i}{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i v_i} \xrightarrow{d} \text{Ratio of (correlated) normal RVs}$$

Weak Instruments: Example

- Now let

$$\begin{pmatrix} A \\ B \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} \mathrm{E}\left(u^2 z^2\right) & \mathrm{E}\left(z^2 uv\right) \\ \mathrm{E}\left(z^2 uv\right) & \mathrm{E}\left(z^2 v^2\right) \end{pmatrix}\right),$$

and recall that jointly normal random variables have linear conditional means. Since both have mean zero:

$$\mathrm{E}\left(A|B=b\right) = \rho \cdot \sqrt{\frac{\mathrm{E}\left(u^2 z^2\right)}{\mathrm{E}\left(z^2 v^2\right)}} \cdot b = \frac{\mathrm{E}\left(z^2 uv\right)}{\mathrm{E}\left(z^2 v^2\right)} \cdot b,$$

where

$$\rho = Corr\left(A, B\right) = \frac{\mathrm{E}\left(z^2 uv\right)}{\sqrt{\mathrm{E}\left(u^2 z^2\right)\mathrm{E}\left(z^2 v^2\right)}}.$$

Weak Instruments: Example

- We can therefore write

$$A = \rho \cdot \sqrt{\frac{\mathrm{E}\left(u^2 z^2\right)}{\mathrm{E}\left(z^2 v^2\right)}} \cdot B + U; \quad \mathrm{E}\left(U|B\right) = 0.$$

  Since linear transformations of the multivariate normals are multivariate normal:

$$\begin{pmatrix} U \\ B \end{pmatrix} = \begin{pmatrix} A - \rho \cdot \sqrt{\frac{\mathrm{E}(u^2 z^2)}{\mathrm{E}(z^2 v^2)}} \cdot B \\ B \end{pmatrix}$$

  is multivariate normal also.

- In fact, U and B are independent, since they are uncorrelated and bivariate normal.

Weak Instruments: Example

- Conclude that

$$\hat{\beta}_{IV} - \beta \xrightarrow{d} \rho \cdot \sqrt{\frac{\mathrm{E}\left(u^2 z^2\right)}{\mathrm{E}\left(z^2 v^2\right)}} + \frac{U}{B}.$$

- The standard Cauchy distribution can be represented as the ratio of two independent standard normals:
- If $X \sim \mathit{Cauchy}(0, 1)$, X is symmetric about zero and has the same distribution as

$$\frac{N_1}{N_2},$$

where N_1 is independent of N_2 and $N_1 \sim \mathcal{N}\left(0, 1\right), N_2 \sim \mathcal{N}\left(0, 1\right)$.
- Therefore, U/B is a scaled standard cauchy distribution.

Weak Instruments: Example

- If we make some homoskedasticity assumptions, we can relate this behaviour to that of the OLS estimator.
- Assume:

$$\mathrm{E}\left(v^2|z\right) = \sigma_v^2; \quad \mathrm{E}\left(uv|z\right) = \sigma_{uv}.$$

Then:

$$\rho \cdot \sqrt{\frac{\mathrm{E}\left(u^2 z^2\right)}{\mathrm{E}\left(z^2 v^2\right)}} = \frac{\mathrm{E}\left(z^2 uv\right)}{\mathrm{E}\left(z^2 v^2\right)} = \frac{\sigma_{uv}}{\sigma_v^2} = \frac{\mathrm{E}\left(uv\right)}{\mathrm{E}\left(v^2\right)}.$$

Weak Instruments: Example

- Finally note that when $\pi = 0$, $x = \pi z + v = v$, so:

$$\hat{\beta}_{IV} \xrightarrow{d} \beta + \frac{\mathrm{E}(xu)}{\mathrm{E}(x^2)} + \frac{U}{B}.$$

Now recall that

$$\hat{\beta}_{OLS} = \beta + \frac{\frac{1}{n}\sum_{i=1}^{n} x_i u_i}{\frac{1}{n}\sum_{i=1}^{n} x_i^2} \xrightarrow{p} \beta + \frac{\mathrm{E}(xu)}{\mathrm{E}(x^2)},$$

so the IV estimator is (asymptotically) centered around the probability limit of the OLS estimator.

Weak Instrument Asymptotics

- Our goal is to more accurately reflect the behaviour of the IV estimator in finite samples.
- We know how to approximate its distribution when $\pi = 0$ and when π is 'far' from zero.
- How to model π being 'close' to zero?
- This will depend on the sample size. As $n \to \infty$, $\pi = 0.00001$ will provide $\sqrt{n}\left(\hat{\beta}_{IV} - \beta\right) \xrightarrow{d} \mathcal{N}(0, V)$, but we don't expect the finite sample distribution of the IV estimator to behave this way.

Questions?

Nelson and Startz (1990)

- Consider the DGP

$$
\begin{aligned}
y &= \beta x + u \\
x &= \epsilon + \lambda^{-1} u \\
z &= \nu + \gamma \epsilon
\end{aligned}
$$

where $\beta = 0$, $\lambda = 1$, $u \sim \mathcal{N}(0, 1)$ and

$$
\frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} \nu_i \\ \epsilon_i \end{pmatrix} \begin{pmatrix} \nu_i & , \epsilon_i \end{pmatrix} \to \begin{pmatrix} \sigma_\nu^2 & 0 \\ 0 & \sigma_\epsilon^2 \end{pmatrix}.
$$

Nelson and Startz (1990)

- Derive finite sample distribution of $\hat{\beta}_{IV}$ conditional on ν, ϵ:

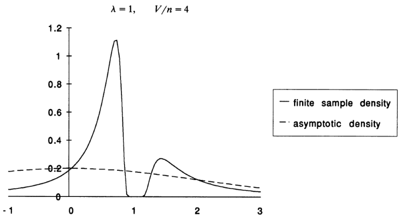


FIGURE 2.—Finite sample and asymptotic density functions for instrumental variables.

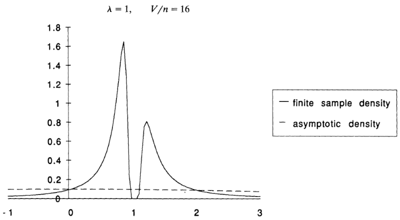


FIGURE 3.—Finite sample and asymptotic density functions for instrumental variables.

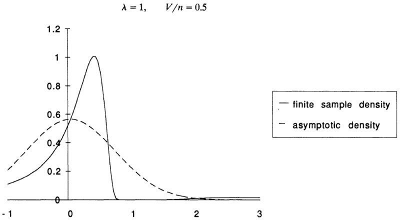


FIGURE 4.—Finite sample and asymptotic density functions for instrumental variables.

Pitman Drift

- Pitman drift allows for an unknown parameter in the model to vary with n to produce asymptotic approximations that hopefully better reflect finite sample statistical properties.
- To fix ideas: Consider testing $H_0 : \mu \leq c$ vs. $H_1 : \mu > c$ with a sample $\{X_i\}_{i=1}^n$ drawn from $\mathcal{N}(\mu, 1)$, $\mu \in \mathbb{R}$ at significance level α.
- Typical test statistic:

$$T_n = \sqrt{n}\left(\bar{X}_n - c\right).$$

- Rejection rule: Reject iff $T_n \geq z_{1-\alpha}$, where $z_{1-\alpha}$ is the $1-\alpha$ quantile of a standard normal distribution.

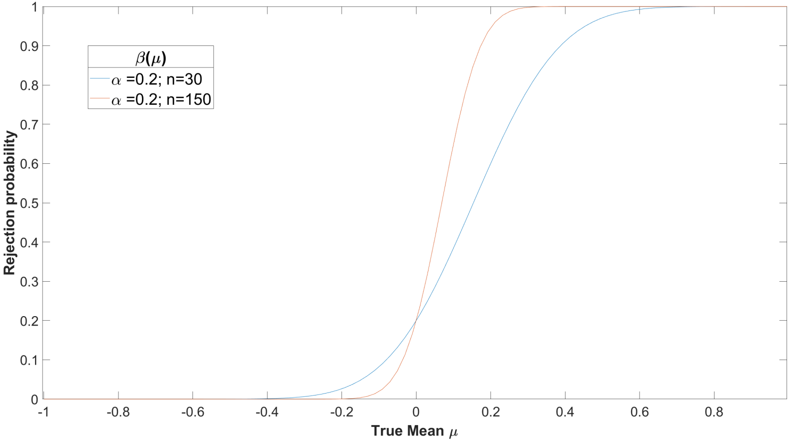
Pitman Drift

- The power function for this test is defined as

$$
\begin{aligned}
\beta_n(\mu) &= \mathrm{P}_\mu\left(T_n \geq z_{1-\alpha}\right) \\
&= \mathrm{P}_\mu\left(\sqrt{n}\left(\bar{X}_n - \mu\right) + \sqrt{n}\left(\mu - c\right) \geq z_{1-\alpha}\right) \\
&= 1 - \Phi\left(z_{1-\alpha} - \sqrt{n}\left(\mu - c\right)\right).
\end{aligned}
$$

The power is increasing in n when $\mu > c$. We plot this function for $\alpha = 0.2, n = 30, 150$:

Pitman Drift



Pitman Drift

- As $n \to \infty$, the power function converges to zero at all points $\mu < 0$, to 1 at all points $\mu > 0$ and to α at $\mu = 0$.
- Many tests would do the same, so when the distribution of X is unknown, how are we supposed to compare their power based on the asymptotic distribution of the test statistic?
- Idea: Let μ get closer to c as $n \to \infty$ to approximate what's going on in a finite sample.
- Test $\mu \leq c$ vs. a "sequence of local alternatives" $\mu_n = c + \frac{k}{\sqrt{n}}$ for some $k > 0$.

Pitman Drift

- Under this sequence, the power function becomes

$$
\begin{aligned}
\beta_n(\mu_n) &= \mathrm{P}_{\mu_n}\left(T_n \geq z_{1-\alpha}\right) \\
&= \mathrm{P}_{\mu_n}\left(\sqrt{n}\left(\bar{X}_n - \mu_n\right) + \sqrt{n}\left(\mu_n - c\right) \geq z_{1-\alpha}\right) \\
&= 1 - \Phi\left(z_{1-\alpha} - \sqrt{n} \cdot \frac{k}{\sqrt{n}}\right) \\
&= 1 - \Phi\left(z_{1-\alpha} - k\right),
\end{aligned}
$$

which no longer depends on n! When X is drawn from a normal distribution we can compute the power exactly, but what if we only know

$$\sqrt{n}\left(\bar{X}_n - \mu_n\right) \xrightarrow{d} \mathcal{N}(0,1)?$$

Pitman Drift

- For a standard alternative $\mu > c$,

$$\begin{aligned}
&\mathrm{P}_{\mu}\left(\sqrt{n}\left(\bar{X}_n - c\right) \geq z_{1-\alpha}\right) \\
&= \mathrm{P}_{\mu}\left(\sqrt{n}\left(\bar{X}_n - \mathrm{E}\left(X\right)\right) \geq z_{1-\alpha} - \sqrt{n}\left(\mathrm{E}\left(X\right) - c\right)\right) \\
&\to 1,
\end{aligned}$$

but under a sequence of local alternatives $\mu_n = c + \frac{k}{\sqrt{n}}$:

$$\begin{aligned}
&\mathrm{P}_{\mu_n}\left(\sqrt{n}\left(\bar{X}_n - c\right) \geq z_{1-\alpha}\right) \\
&= \mathrm{P}_{\mu_n}\left(\sqrt{n}\left(\bar{X}_n - \mu_n\right) \geq z_{1-\alpha} - \sqrt{n} \cdot \frac{k}{\sqrt{n}}\right) \\
&\to \mathrm{P}\left(\mathcal{N}\left(0, 1\right) \geq z_{1-\alpha} - k\right) \\
&= 1 - \Phi\left(z_{1-\alpha} - k\right).
\end{aligned}$$

Pitman Drift

- In this regard we can approximate the finite sample power of the test for non-normal distributions. We can also compare this 'asymptotic local power' to that of competing tests of the same hypothesis.

Weak Instruments

- The relevant approximation for a weak instrument is to model the first stage coefficient as a sequence

$$\pi_n = \frac{\pi}{\sqrt{n}} \to 0.$$

and write

$$x_i = \pi_n z_i + v_i.$$

Plug this in to the formula for $\hat{\beta}_{IV} - \beta$:

$$\hat{\beta}_{IV} - \beta = \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i u_i}{\left(\sqrt{n}\pi_n\right)\left(\frac{1}{n} \sum_{i=1}^{n} z_i^2\right) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i v_i}.$$

Weak Instruments

- Now have

$$\left(\sqrt{n}\pi_n\right)\left(\frac{1}{n}\sum_{i=1}^{n} z_i^2\right) = \pi \cdot \frac{1}{n}\sum_{i=1}^{n} z_i^2 \xrightarrow{p} \pi \mathrm{E}\left(z^2\right).$$

- Apply the multivariate CLT and CMT once again to conclude

$$\hat{\beta}_{IV} - \beta \xrightarrow{d} \frac{A}{\pi \mathrm{E}\left(z^2\right) + B},$$

where

$$\begin{pmatrix} A \\ B \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} \mathrm{E}\left(u^2 z^2\right) & \mathrm{E}\left(z^2 uv\right) \\ \mathrm{E}\left(z^2 uv\right) & \mathrm{E}\left(z^2 v^2\right) \end{pmatrix}\right).$$

Weak Instruments

- Now consider the standard asymptotic argument:

$$\sqrt{n}\left(\hat{\beta}_{IV}-\beta\right) = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i u_i}{\left(\frac{1}{n}\sum_{i=1}^{n} z_i^2\right)\cdot\pi + \frac{1}{n}\sum_{i=1}^{n} z_i v_i}$$
$$\xrightarrow{d} \frac{A}{\pi\mathrm{E}\left(z^2\right)+0}.$$

This is captured in the weak instrument asymptotics for large π:

$$\sqrt{n}\left(\hat{\beta}_{IV}-\beta\right) \overset{d}{\approx} \frac{A}{\pi_n\mathrm{E}\left(z^2\right)+B/\sqrt{n}}.$$

Weak instrument problem

- Impose homoskedasticity: $\mathrm{E}\left(uv|z\right) = \sigma_{uv}$, $\mathrm{E}\left(u^2|z\right) = \sigma_u^2$ and $\mathrm{E}\left(v^2|z\right) = \sigma_v^2$.
- Under weak instrument asymptotics:

$$\sqrt{n}\left(\hat{\pi}_{OLS} - \pi_n\right) = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i v_i}{\frac{1}{n}\sum_{i=1}^{n} z_i^2} \xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma_v^2}{\mathrm{E}\left(z^2\right)}\right),$$

and so

$$\sqrt{n}\hat{\pi}_{OLS} \xrightarrow{d} \pi + \mathcal{N}\left(0, \frac{\sigma_v^2}{\mathrm{E}\left(z^2\right)}\right).$$

- Can't estimate $\pi = \sqrt{n}\pi_n$ consistently under weak instrument asymptotics.

Weak instrument problem

- Nevertheless, we can test hypotheses about π using the first stage F statistic. To do this we find the asymptotic distribution of the first stage F-statistic under the sequence of first stage coefficients $\pi_n = \frac{\pi}{\sqrt{n}}$.
- Consider a test of linear restriction $\pi = 0$:

$$T_n = \frac{n\left(\hat{\pi}_{OLS} - 0\right)^2}{\frac{\hat{\sigma}_v^2}{\frac{1}{n}\sum_{i=1}^n z_i^2}}; \quad \hat{\sigma}_v^2 = \frac{1}{n-1}\sum_{i=1}^n \left(x_i - \hat{\pi}_{OLS} z_i\right)^2.$$

F-Stat under weak instrument asymptotics

- Under the sequence π_n we verify $\hat{\sigma}_v^2$ is consistent:

$$\begin{aligned}
\hat{\sigma}_v^2 &= \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \hat{\pi}_{OLS} z_i\right)^2 \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\left(\left(\pi_n - \hat{\pi}_{OLS}\right) z_i + v_i\right)^2 \\
&= \frac{1}{n-1}\sum_{i=1}^{n} v_i^2 + o_p\left(1\right) \xrightarrow{p} \sigma_v^2.
\end{aligned}$$

We also have:

$$\frac{1}{n}\sum_{i=1}^{n} z_i^2 \xrightarrow{p} \mathrm{E}\left(z^2\right).$$

F-Stat under weak instrument asymptotics

- Therefore

$$T_n = \frac{n\left(\hat{\pi}_{OLS} - 0\right)^2}{\frac{\hat{\sigma}_v^2}{\frac{1}{n}\sum_{i=1}^n z_i^2}} \xrightarrow{d} \left(\frac{\pi\sqrt{\mathrm{E}\left(z^2\right)}}{\sigma_v} + \mathcal{N}\left(0,1\right)\right)^2.$$

The concentration parameter is defined as:

$$\mu := \frac{\pi\sqrt{\mathrm{E}\left(z^2\right)}}{\sigma_v},$$

and equals π divided by the asymptotic standard error of $\hat{\pi}_{OLS}$.

F-Stat under weak instrument asymptotics

- The concentration parameter is a normalized measure of the strength of the instrument, and

$$\left(\frac{\pi\sqrt{\mathrm{E}\left(z^2\right)}}{\sigma_v}+\mathcal{N}\left(0,1\right)\right)^2\overset{d}{=}\chi_1^2\left(\mu^2\right).$$

- The non-centrality parameter depends on μ which indexes the sequence of 'local alternatives' . $\mu=0$ means the instrument fails the rank condition and the first stage F statistic converges in distribution to a (central) χ_1^2.
- An instrument is 'weak' if μ is 'too close' to zero.

Towards a definition of 'weak' instruments

- How should we choose a suitable μ? Recall that when $\mu = 0$ we found (under homoskedasticity)

$$\hat{\beta}_{IV} \xrightarrow{d} \text{plim}\left(\hat{\beta}_{OLS}\right) + \text{scaled standard Cauchy.}$$

When $\mu \neq 0$ the IV/OLS estimators satisfy (under homoskedasticity)

$$\hat{\beta}_{IV} - \beta \xrightarrow{d} \frac{A}{\pi \mathrm{E}\left(z^2\right) + B};$$

$$\hat{\beta}_{OLS} - \beta \xrightarrow{p} \frac{\mathrm{E}\left(uv\right)}{\mathrm{E}\left(v^2\right)} = \frac{\sigma_{uv}}{\sigma_v^2}.$$

where

$$\begin{pmatrix} A \\ B \end{pmatrix} \sim \mathcal{N}\left(0, \mathrm{E}\left(z^2\right)\begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix}\right).$$

'Bias' of the IV estimator

- We might try to take the expectation of the RHS:

$$"\mathrm{E}\left(\frac{A}{\pi \mathrm{E}\left(z^2\right)+B}\right)",$$

  which unfortunately doesn't exist, and compare it to $\mathrm{E}(uv)/\mathrm{E}\left(v^2\right)$. In the standard asymptotics, the IV estimator was asymptotically centered at β. When $\mu = 0$, the IV estimator is centered at the probability limit of the OLS estimator.

- So we can define an instrument to be 'weak' if

$$"\mathrm{E}\left(\frac{A}{\pi \mathrm{E}\left(z^2\right)+B}\right)"$$

  is large relative to the inconsistency in the OLS estimator $\mathrm{E}(uv)/\mathrm{E}\left(v^2\right)$.

'Bias' of the IV estimator

- We use the Nagar bias, which computes the mean of a related random variable that does have a finite expectation. This alternative hopefully captures the relevant features of the limit distribution that make it 'far' from zero.
- Write:

$$
\begin{aligned}
\mathrm{E}\left(\frac{A}{\pi \mathrm{E}\left(z^2\right) + B} \,\middle|\, B = b \right) &= \frac{\sigma_{uv}}{\sigma_v^2} \cdot \frac{b}{\pi \mathrm{E}\left(z^2\right) + b}. \\
&= \frac{\sigma_{uv}}{\sigma_v^2} \cdot \frac{b}{\pi \mathrm{E}\left(z^2\right)} \left[\frac{1}{1 + b\left[\pi \mathrm{E}\left(z^2\right)\right]^{-1}} \right] \\
&\approx \frac{\sigma_{uv}}{\sigma_v^2} \cdot \frac{b}{\pi \mathrm{E}\left(z^2\right)} \left[1 - \frac{b}{\pi \mathrm{E}\left(z^2\right)} \right].
\end{aligned}
$$

'Bias' of the IV estimator

- We use the first order Taylor expansion to write

$$
\begin{aligned}
\text{"}\mathrm{E}\left(\frac{A}{\pi \mathrm{E}\left(z^2\right)+B}\right)\text{"} &\approx \mathrm{E}\left(\frac{\sigma_{uv}}{\sigma_v^2}\cdot\frac{B}{\pi \mathrm{E}\left(z^2\right)}\left[1-\frac{B}{\pi \mathrm{E}\left(z^2\right)}\right]\right)\\
&= \frac{\sigma_{uv}}{\sigma_v^2}\cdot\frac{1}{\pi \mathrm{E}\left(z^2\right)}\cdot\left[\mathrm{E}\left(B\right)-\frac{\mathrm{E}\left(B^2\right)}{\pi \mathrm{E}\left(z^2\right)}\right]\\
&= -\frac{\sigma_{uv}}{\sigma_v^2}\cdot\frac{1}{\pi \mathrm{E}\left(z^2\right)}\cdot\frac{\mathrm{E}\left(z^2\right)\sigma_v^2}{\pi \mathrm{E}\left(z^2\right)}\\
&= -\frac{\sigma_{uv}}{\pi^2 \mathrm{E}\left(z^2\right)}\\
&= -\frac{\sigma_{uv}}{\sigma_v^2}\cdot\frac{\sigma_v^2}{\pi^2 \mathrm{E}\left(z^2\right)} = -\frac{\sigma_{uv}}{\sigma_v^2}\cdot\frac{1}{\mu^2}
\end{aligned}
$$

'Bias' of the IV estimator

- Since $\frac{\sigma_{uv}}{\sigma_v^2}$ is the inconsistency in the OLS estimate, we have

$$\left| \frac{\text{"}\textit{bias}\text{"}\left(\hat{\beta}_{IV}\right)}{\text{plim}\left(\hat{\beta}_{OLS} - \beta\right)} \right| = \frac{1}{\mu^2}.$$

- As μ increases, the bias of the IV estimator relative to OLS decreases.
- We may define instruments to be 'weak' if the absolute relative bias of the IV estimator exceeds, say, 5%:

$$\frac{1}{\mu^2} \geq 0.05 \implies \mu \leq \sqrt{20}.$$

Testing for weak instruments

- Given a cutoff for μ, we can test

$$H_0 : \mu \leq \sqrt{20} \text{ vs.} H_1 : \mu > \sqrt{20}$$

at significance level α using the weak instruments asymptotic distribution of the F-stat. The critical value should be chosen to ensure that the probability of rejection under H_0 does not exceed $\alpha = 0.05$. We have

$$\chi_1^2(20, 0.95) = 37.42.$$

- Often recommended to use F-stat greater than 10 as a reasonable cutoff, but this depends on the definition of a weak instrument.

Asymptotic distribution of $\hat{\beta}_{IV}$

- Recall under homoskedasticity

$$\hat{\beta}_{IV} - \beta \xrightarrow{d} \frac{A}{\pi \mathrm{E}\left(z^2\right) + B},$$

where

$$\begin{aligned}\begin{pmatrix} A \\ B \end{pmatrix} &\sim \mathcal{N}\left(0, \mathrm{E}\left(z^2\right)\begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix}\right) \\ &\overset{d}{=} \sigma_v \sqrt{\mathrm{E}\left(z^2\right)} \cdot \mathcal{N}\left(0, \begin{pmatrix} \sigma_u^2/\sigma_v^2 & \sigma_{uv}/\sigma_v^2 \\ \sigma_{uv}/\sigma_v^2 & 1 \end{pmatrix}\right),\end{aligned}$$

so

$$\hat{\beta}_{IV} - \beta \xrightarrow{d} \frac{A'}{\mu + B'}.$$

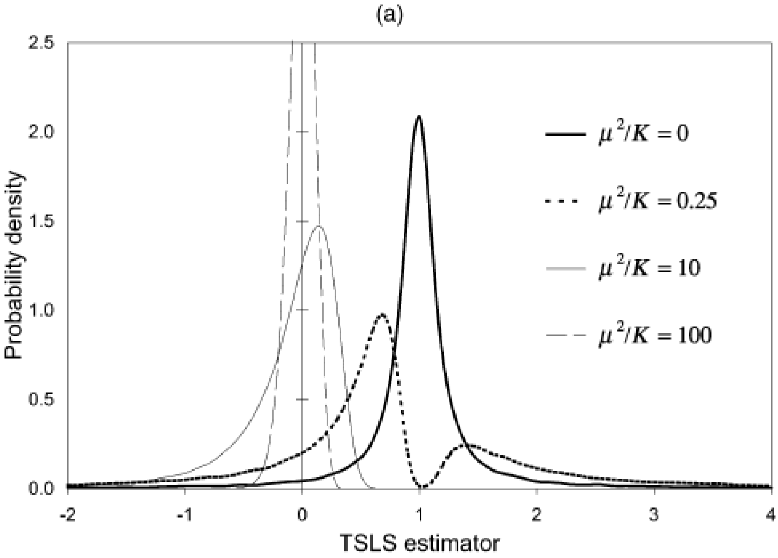
Stock,Wright and Yogo (2002)

- Stock, Wright and Yogo (2002) simulate the finite sample distribution of $\hat{\beta}_{IV}$:

$$y = \beta x + u;$$
$$x = \pi z + v;$$
$$\begin{pmatrix} u \\ v \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} 1 & 0.99 \\ 0.99 & 1 \end{pmatrix}\right),$$

and $\beta = 0$. Instruments non-random. In this case $\hat{\beta}_{OLS} \xrightarrow{p} \frac{\sigma_{uv}}{\sigma_v^2} = 0.99$. $\mu = \pi \left(\sum_{i=1}^{n} z_i^2\right)^{1/2}$. $K = \dim(z) = 1$.

Stock, Wright and Yogo (2002)



Conducting inference on β

- Two-step methods that check the first stage F-statistic and then use $\hat{\beta}_{IV}$ to conduct inference on β are often unreliable, and it's generally difficult to come up with a reasonable value for the F-statistic.
- Rule of thumb F-stat≥ 10 reasonable for absolute relative bias of 10% and for the specific case of three or more instruments (where the mean actually exists, rather than being approximated by a nagar bias). Stock and Yogo (2005) show the critical value lies between 9 and 12 for number of instruments between 3 and 30.
- In joint normal errors/fixed instruments case, $\hat{\beta}_{IV}$ has number of moments equal to the number of excluded instruments - number of endogenous variables (Kinal (1980)).

Conducting inference on β

- Other methods of defining weak instruments exist, for example actual size of t-test should be 'close' to specified significance level.
- With several endogenous variables, the bias of the 2SLS estimate becomes the norm of the vector of biases for the coefficients on each of the endogenous variables.
- Under heteroskedasticity, Montiel Olea and Pflueger (2013) suggest a modification to the F-stat, but it's no longer the case that $\hat{\beta}_{IV}$ is centered at the prob. limit of OLS. See Pflueger and Wang (2015) for a Stata implementation.

Conducting inference on β

- Instead we use a test that is robust to weak instruments, called Anderson-Rubin test. Suppose $y = x'\beta + u$.
- Idea: $\mathrm{E}(zu) = 0$ implies $\mathrm{E}(z(y - x'\beta)) = 0$, so under $H_0 : \beta = \beta_0$,

$$\mathrm{E}(zu(\beta_0)) = 0,$$

where $u(\beta_0) = y - x'\beta_0$. We can write this equivalently as

$$\mathrm{E}\left(zz'\right)^{-1} \mathrm{E}(zu(\beta_0)) = 0,$$

and this is the vector of coefficients of a regression of $u(\beta_0)$ on z.

Conducting inference on β

- Reject if coefficient estimates are significantly different from zero.
- Substitute $y = x'\beta + u$ to give

$$\begin{aligned} \mathrm{E}\left(zz'\right)^{-1} \mathrm{E}\left(zu\left(\beta_0\right)\right) &= \mathrm{E}\left(zz'\right)^{-1} \mathrm{E}\left(zx'\left(\beta - \beta_0\right) + zu\right) \\ &= \mathrm{E}\left(zz'\right)^{-1} \mathrm{E}\left(zx'\right)\left(\beta - \beta_0\right). \end{aligned}$$

  which equals zero iff $\beta = \beta_0$.
- Therefore, should expect such a test to have power against any alternative $\beta \neq \beta_0$.