Advanced Econometrics I

Jürgen Meinecke

Lecture 6 of 12

Research School of Economics, Australian National University

Roadmap

We combined the structural and first stage equations like so:

$$\begin{aligned} Y_i &= X_i'\beta + e_i \\ &= (\pi' Z_i + v_i)'\beta + e_i \\ &= Z_i'\lambda + w_i, \end{aligned}$$

with $\lambda := \pi\beta$ and $w_i := v_i'\beta + e_i$

Recall the two reduced form projection coefficients

- regressing Y_i on Z_i results in $\lambda = \mathsf{E}(Z_i Z_i')^{-1}\mathsf{E}(Z_i Y_i)$
- regressing X_i on Z_i results in $\pi = \mathsf{E}(Z_i Z_i')^{-1}\mathsf{E}(Z_i X_i')$

Let's recall their dimensions

- $\dim \lambda = L \times 1$
- $\dim \pi = L \times K$

We learned that the projection coefficients λ and π are identified because they are explicit functions of population moments

This means we can uniquely estimate them

Practically we treat them as if they were known to us (because we have faith in uniquely estimating them via analog principle)

In contrast, identification of β is not so easy because $\lambda = \pi\beta$ is a system of L equations for K unknowns

Linear algebra tells you that there

- are no solutions or infinitely many solutions if $L < K$
- is hope for unique solution only if $L \geq K$

So let's only consider $L \geq K$

Two sub-cases here

- $L = K$
  then $\dim \pi = K \times K$ and if it is invertible then
  $$\beta = \pi^{-1}\lambda = \mathsf{E}(Z_i X_i')^{-1}\mathsf{E}(Z_i Y_i)$$
  This solution for β motivates the IV estimator
- $L > K$ then we cannot simply invert, but we can do this:
  $$\pi\beta = \lambda \quad \Leftrightarrow \quad \pi'\pi\beta = \pi'\lambda$$
  and therefore
  $$\beta = (\pi'\pi)^{-1}\pi'\lambda$$
  But $(\hat{\pi}'\hat{\pi})^{-1}\hat{\pi}'\hat{\lambda}$ is not the 2SLS estimator

The 2SLS estimator has a different motivation

Again looking at our structural equation and plugging in the first stage

$$\begin{aligned} Y_i &= X_i'\beta + e_i \\ &= (\pi' Z_i + v_i)'\beta + e_i \\ &= Z_i'\pi\beta + (v_i'\beta + e_i) \\ &= Z_i'\pi\beta + w_i \end{aligned}$$

If you knew π you could define $\tilde{Z}_i' = Z_i'\pi$ and write

$$Y_i = \tilde{Z}_i'\beta + w_i,$$

where $\text{E}(\tilde{Z}_i w_i) = 0$

Clearly, OLS would work fine here

Notice that $\dim \tilde{Z}_i = \dim X_i = K \times 1$

The corresponding matrix $\tilde{Z} := Z\pi$ with $\dim \tilde{Z} = \dim X = N \times K$

The OLS estimator is

$$
\begin{aligned}
\hat{\beta}^{\text{i2SLS}} &:= \left(\tilde{Z}'\tilde{Z}\right)^{-1} \tilde{Z}'Y \\
&= \left(\pi'Z'Z\pi\right)^{-1} \pi'Z'Y
\end{aligned}
$$

This OLS estimator is *infeasible* because we don't know π

But we can turn it into a feasible estimator by plugging in the consistent estimator $\hat{\pi} := (Z'Z)^{-1}Z'X$

And this is indeed what the 2SLS estimator does

Definition (Two Stage Least Squares (2SLS) Estimator)

$$\begin{aligned}\hat{\beta}^{\text{2SLS}} &:= \left(\hat{\pi}' Z' Z \hat{\pi}\right)^{-1} \hat{\pi}' Z' Y \\ &= \left(X' Z (Z' Z)^{-1} Z' X\right)^{-1} X' Z (Z' Z)^{-1} Z' Y\end{aligned}$$

In summation notation:

$$\begin{aligned}\hat{\beta}^{\text{2SLS}} = &\left[\left(\sum_{i=1}^{N} X_i Z_i'\right)\left(\sum_{i=1}^{N} Z_i Z_i'\right)^{-1}\left(\sum_{i=1}^{N} Z_i X_i'\right)\right]^{-1} \times \\ &\left(\sum_{i=1}^{N} X_i Z_i'\right)\left(\sum_{i=1}^{N} Z_i Z_i'\right)^{-1}\left(\sum_{i=1}^{N} Z_i Y_i\right)\end{aligned}$$

Three different interpretations of $\hat{\beta}^{\text{2SLS}}$

Recall $P_Z := Z(Z'Z)^{-1}Z'$ is the symmetric and idempotent projection matrix

Then $\hat{X} := P_Z X$ is the projection of X on Z

It follows

$$
\begin{aligned}
\hat{\beta}^{\text{2SLS}} &= \left(X'Z(Z'Z)^{-1}Z'X\right)^{-1} X'Z(Z'Z)^{-1}Z'Y \\
&= (X'P_Z X)^{-1} X'P_Z Y && (1) \\
&= ((P_Z X)'X)^{-1} (P_Z X)'Y \\
&= \left(\hat{X}'X\right)^{-1} \hat{X}'Y && (2) \\
&= (X'P_Z P_Z X)^{-1} X'P_Z Y \\
&= ((P_Z X)'(P_Z X))^{-1} (P_Z X)'Y \\
&= \left(\hat{X}'\hat{X}\right)^{-1} \hat{X}'Y && (3)
\end{aligned}
$$

Equation (1) is the most common matrix representation of $\hat{\beta}^{\text{2SLS}}$ in textbooks and lecture notes

Equation (2) presents the 2SLS estimator as an IV estimator, it has the same structure as $\hat{\beta}^{\text{IV}}$ with $\hat{X}$ used in place of Z

Equation (3) presents the 2SLS estimator as an OLS estimator of Y on $\hat{X}$

The third interpretation justifies label ‘two stage least squares’:

(1) regress X on Z, obtain $\hat{\pi} = (Z'Z)^{-1}Z'X$ and $\hat{X} = Z\hat{\pi} = P_Z X$
(2) regress Y on $\hat{X}$ and obtain $\hat{\beta}^{\text{2SLS}} = (\hat{X}'\hat{X})^{-1}\hat{X}'Y$

Roadmap

Proposition (Consistency of $\hat{\beta}^{\text{2SLS}}$)

$\hat{\beta}^{2SLS} = \beta + o_p(1).$

Some definitions needed for asymptotic variance:
Let $C_{XZ} = \mathrm{E}(X_i Z_i')$, and $C_{ZZ} = \mathrm{E}(Z_i Z_i')$, and $C_{ZX} = \mathrm{E}(Z_i X_i')$.

Proposition (Asymptotic Distribution of $\hat{\beta}^{\text{2SLS}}$)

$$\sqrt{N}(\hat{\beta}^{2SLS} - \beta) \xrightarrow{d} N(0, \Omega)$$

where
$\Omega = (C_{XZ} C_{ZZ}^{-1} C_{ZX})^{-1} C_{XZ} C_{ZZ}^{-1} E(e_i^2 Z_i Z_i') C_{ZZ}^{-1} C_{ZX} (C_{XZ} C_{ZZ}^{-1} C_{ZX})^{-1}$

Corollary

Under homoskedasticity, $\Omega = \sigma_e^2 (C_{XZ} C_{ZZ}^{-1} C_{ZX})^{-1}$.

Consistent estimators for the asymptotic covariances are readily obtained by using the analogy principle

So replace population moments by sample moments, because

$$\sum_{i=1}^{N} X_i Z_i' / N = C_{XZ} + o_p(1)$$

$$\sum_{i=1}^{N} Z_i Z_i' / N = C_{ZZ} + o_p(1)$$

$$\sum_{i=1}^{N} Z_i X_i' / N = C_{ZX} + o_p(1)$$

$$\sum_{i=1}^{N} Z_i Z_i' \hat{e}_i^2 / N = \mathrm{E}(Z_i Z_i' e_i^2) + o_p(1)$$

where $\hat{e}_i := Y_i - X_i' \hat{\beta}^{\text{2SLS}}$

The resulting covariance matrix estimator will be consistent

Roadmap

What is the expected value of $\hat{\beta}^{2SLS}$?

It's quite complicated to work this out

There exists a somewhat sobering result that offers some guidance

Lemma (Existence of Moments of 2SLS (Kinal))

Let (X, Y, Z) be jointly normal. The r-th moment of $\hat{\beta}^{2SLS}$ only exists for integers $r < L_2 - K_2 + 1$.

To obtain this result, Kinal had to impose the strong restriction of normality (which is almost certainly far from the truth)

Applying this to a common case in practice:
$K_2 = L_2 = 1 \Rightarrow$ no expected value

Only over-identified models can hope to have $\mathrm{E}(\hat{\beta}^{2SLS}) < \infty$

Although the 2SLS estimator is consistent, it is biased

Where does this bias come from?

Recall the infeasible 2SLS estimator:

$$\hat{\beta}^{\text{i2SLS}} = (\pi' Z' Z \pi)^{-1} \pi' Z' Y$$

We can't use i2SLS because we don't know π

Brainwave: use $\hat{\pi}$ instead, and obtain

$$\hat{\beta}^{\text{2SLS}} = (\hat{\pi}' Z' Z \hat{\pi})^{-1} \hat{\pi}' Z' Y$$

Seems like a good analogy principle solution, however using $\hat{\pi}$ in place of π is the source of the bias of 2SLS
(even though $\hat{\pi}$ is a trusty and consistent estimator for π)

Usually we don't make a big deal if an estimator has a little bias, but the bias in the 2SLS setting can get out of control quickly

Let's investigate

Let's look at a toy model

$$Y_i = X_i\beta + e_i$$
$$X_i = Z_i'\pi + v_i,$$

where X_i is a scalar and $\dim Z_i = L \geq 1$

Let $(e_i, v_i) \sim \mathrm{N}(0, \Sigma)$
(that is, we assume an exact bivariate normal distribution)

2SLS estimation makes sense here because $\mathrm{E}(e_i X_i) \neq 0$

To make life easier, let's pretend that

- Z_i
- $\hat{\Theta} = \sum_{i=1}^{N} Z_i Z_i'/N$

are non-stochastic (we treat them as constants)

We work with a simple toy model and make many simplifying assumptions (otherwise the math becomes even more tedious)

Let's start by showing that infeasible 2SLS is unbiased

Recall

$$\hat{\beta}^{\text{i2SLS}} = \frac{\pi' Z' Y}{\pi' Z' Z \pi}$$

Therefore

$$\sqrt{N}\left(\hat{\beta}^{\text{i2SLS}} - \beta\right) = \frac{\frac{1}{\sqrt{N}} \pi' Z' e}{\frac{1}{N} \pi' Z' Z \pi}$$

Let's take a closer look at the numerator

For our toy model we can obtain an exact distribution:

$$\frac{1}{\sqrt{N}} \pi' Z' e \sim \mathrm{N}\left(0, \frac{\sigma_e^2}{N} \pi' Z' Z \pi\right) = \mathrm{N}\left(0, \sigma_e^2 \pi' \hat{\Theta} \pi\right)$$

Therefore, infeasible 2SLS is unbiased

For 2SLS things are not so simple

In the scalar case, by definition:

$$\hat{\beta}^{\text{2SLS}} = \frac{\hat{\pi}'Z'Y}{\hat{\pi}'Z'Z\hat{\pi}} = \frac{X'Z(Z'Z)^{-1}Z'Y}{X'Z(Z'Z)^{-1}Z'X}$$

Rearranging results in

$$\sqrt{N}\left(\hat{\beta}^{\text{2SLS}} - \beta\right) = \frac{\frac{1}{\sqrt{N}}X'Z(Z'Z)^{-1}Z'e}{\frac{1}{N}X'Z(Z'Z)^{-1}Z'X}$$

Both numerator and denominator are more complicated than for the infeasible case

Let's dissect them

We will make the following substitution: given $X = Z\pi + v$

- $Z'X = Z'Z\pi + Z'v$
- $X'Z = \pi'Z'Z + v'Z$

Turning first to the numerator

$$
\begin{aligned}
\frac{1}{\sqrt{N}} X'Z(Z'Z)^{-1}Z'e &= \frac{1}{\sqrt{N}}\pi'Z'e + \frac{1}{\sqrt{N}} v'P_Z e \\
&= \frac{1}{\sqrt{N}}\pi'Z'e + \frac{1}{\sqrt{N}}\frac{\sigma_{ev}}{\sigma_v^2} v'P_Z v + \frac{1}{\sqrt{N}} v'P_Z w
\end{aligned}
$$

where I use the projection $e_i = \frac{\sigma_{ev}}{\sigma_v^2} v_i + w_i$ with $\mathrm{E}(v_i w_i) = 0$

Because both e_i and v_i are normal, it follows that w_i is normal

Moreover, for the normal distribution the zero covariance $\mathrm{E}(v_i w_i) = 0$
implies that v_i and w_i are statistically independent
(that's a special feature of the normal distribution)

For infeasible 2SLS the numerator only consisted of $\frac{1}{\sqrt{N}}\pi'Z'e$

Copy and paste from previous slide:

$$\frac{1}{\sqrt{N}}X'Z(Z'Z)^{-1}Z'e = \frac{1}{\sqrt{N}}\pi'Z'e + \frac{1}{\sqrt{N}}\frac{\sigma_{ev}}{\sigma_v^2}v'P_Z v + \frac{1}{\sqrt{N}}v'P_Z w$$

Looking at first two terms:

$$\frac{1}{\sqrt{N}}\pi'Z'e \sim \mathrm{N}\left(0, \frac{\sigma_e^2}{N}\pi'Z'Z\pi\right) = \mathrm{N}\left(0, \sigma_e^2\pi'\hat{\Theta}\pi\right)$$

$$\frac{1}{\sqrt{N}}\frac{\sigma_{ev}}{\sigma_v^2}v'P_Z v = \frac{1}{\sqrt{N}}\sigma_{ev}\frac{v'}{\sigma_v}P_Z\frac{v}{\sigma_v} \sim \frac{1}{\sqrt{N}}\sigma_{ev}\chi^2(\mathrm{tr}\ P_Z) \sim \frac{1}{\sqrt{N}}\sigma_{ev}\chi^2(L)$$

Notice that $\frac{v}{\sigma_v} \sim N(0, I_N)$, then using the lemma
if $P \sim \mathrm{N}(0, I_N)$ then $P'QP \sim \chi^2(\mathrm{tr}\ (Q))$

Having worked out the distributions of these two terms, we can consider their expected values

Using $\mathrm{E}(\chi^2(L)) = L$, it follows that

$$\mathrm{E}\left(\frac{1}{\sqrt{N}}\pi'Z'e\right) = 0$$

$$\mathrm{E}\left(\frac{1}{\sqrt{N}}\frac{\sigma_{ev}}{\sigma_v^2}v'P_Z v\right) = \frac{1}{\sqrt{N}}L\sigma_{ev}$$

What’s the expected value of the third term?

$$\mathsf{E}\left(\frac{1}{\sqrt{N}}v'P_Z w\right) = \frac{1}{\sqrt{N}}\mathsf{E}(v')P_Z\mathsf{E}(w) = 0$$

Why? Because v and w are independent rvs with zero mean

It follows, for the entire numerator:

$$\mathsf{E}\left(\frac{1}{\sqrt{N}}X'Z(Z'Z)^{-1}Z'e\right) = \frac{1}{\sqrt{N}}L\sigma_{ev}$$

Ideally, this should be zero

Recall our earlier substitutions

- $Z'X = Z'Z\pi + Z'v$
- $X'Z = \pi'Z'Z + v'Z$

Now applying to the denominator:

$$\frac{1}{N}X'Z(Z'Z)^{-1}Z'X = \frac{1}{N}\pi'Z'Z\pi + \frac{2}{N}\pi'Z'v + \frac{1}{N}v'Z(Z'Z)^{-1}Z'v$$

Looking at the individual terms

$$\frac{1}{N}\pi'Z'Z\pi = \pi'\hat{\Theta}\pi = O(1)$$

$$\frac{2}{N}\pi'Z'v \sim \mathrm{N}\left(0, \frac{4\sigma_v^2}{N^2}\pi'Z'Z\pi\right) = \mathrm{N}\left(0, \frac{4\sigma_v^2}{N}\pi'\hat{\Theta}\pi\right) = \frac{2}{\sqrt{N}}\mathrm{N}\left(0, \sigma_v^2\pi'\hat{\Theta}\pi\right)$$

$$= \frac{2}{\sqrt{N}}\mathrm{O}_p(1) = \mathrm{O}_p\left(\frac{1}{\sqrt{N}}\right)$$

$$\frac{1}{N}v'P_Zv \sim \frac{1}{N}\sigma_v^2\chi^2(L) = \frac{1}{N}\mathrm{O}_p(1) = \mathrm{O}_p\left(\frac{1}{N}\right)$$

Bottom line: a decent approximation for the denominator is

$$\frac{1}{N}X'Z(Z'Z)^{-1}Z'X \approx \pi'\hat{\Theta}\pi$$

Putting things together and applying an asymptotic approximation from Hahn and Hausman (2005)

For small r, they use: $\frac{1}{\pi'\hat{\Theta}\pi+r} \approx \frac{1}{\pi'\hat{\Theta}\pi}$

Then

$$\sqrt{N}(\hat{\beta}^{\text{2SLS}} - \beta) \approx \frac{\frac{1}{\sqrt{N}}\pi'Z'e + \frac{1}{\sqrt{N}}v'P_Z e}{\pi'\hat{\Theta}\pi}$$

Big picture: We want the study the expected value of $\hat{\beta}^{\text{2SLS}}$

We have done all the hard work, now we can derive the expected value of the rhs

$$\text{E}\left(\sqrt{N}(\hat{\beta}^{\text{2SLS}} - \beta)\right) \approx \text{E}\left(\frac{\frac{1}{\sqrt{N}}\pi'Z'e + \frac{1}{\sqrt{N}}v'P_Z e}{\pi'\hat{\Theta}\pi}\right) = \frac{1}{\sqrt{N}}\frac{L}{\pi'\hat{\Theta}\pi}\sigma_{ev}$$

We have successfully approximated the bias of the 2SLS estimator:

$$\mathsf{E}\left(\hat{\beta}^{\text{2SLS}} - \beta\right) \approx \frac{1}{N}\frac{L}{\pi'\hat{\Theta}\pi}\sigma_{ev} \approx \frac{1}{N}\frac{L}{\pi'\Theta\pi}\sigma_{ev},$$

where $\Theta = \mathsf{E}(Z_i Z_i')$

This is the result from Hahn and Hausman (2005)

Let's get a 'feeling' for what's going on

Following Hahn and Hausman, we make further simplifications and recall a few basic concepts

Recall that earlier $(e_i, v_i) \sim \mathsf{N}(0, \Sigma)$

Now, $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$

(this amounts to a normalization that is wlog)

Notice that this implies $\sigma_{ev} = \rho$

Do you remember R^2 from undergrad metrics? Refresher!

Given the reduced form $X_i = Z_i'\pi + v_i$, define

$$\text{TSS} := \sigma_x^2$$

$$\text{ESS} := \text{Var}\,(Z_i'\pi) = \pi'\Theta\pi$$

$$\text{RSS} := \sigma_v^2 = 1$$

(these are the definitions based on the population moments)

Recall from undergrad that $R^2 := \text{ESS}/\text{TSS}$
(the proportion of the variance of X_i that is explained by Z_i)

Algebraic facts: $R^2 = \frac{1-\text{RSS}}{\text{TSS}}$, or equivalently $\text{TSS} = \frac{\text{RSS}}{1-R^2}$

It follows that $\text{TSS} = 1/(1-R^2)$

Lastly, let $F := N \cdot R^2/(1-R^2)$
(proportion of the variance of X_i that is explained by Z_i
divided by
proportion of the variance of X_i that is explained by v_i)

Let's fiddle around with our bias formula

We get the following results for the 2SLS bias

$$
\begin{aligned}
\mathsf{E}\left(\hat{\beta}^{\mathsf{2SLS}} - \beta\right) &\approx \frac{1}{N}\frac{L \cdot \sigma_{ev}}{\pi'\Theta\pi} \\
&= \frac{1}{N}\frac{L \cdot \rho}{\pi'\Theta\pi} \\
&= \frac{1}{N}\frac{L \cdot \rho}{\mathsf{ESS}} \qquad (1) \\
&= \frac{1}{N}\frac{L \cdot \rho}{R^2 \cdot \mathsf{TSS}} \\
&= \frac{1}{N}\frac{L \cdot \rho \cdot (1 - R^2)}{R^2} \qquad (2) \\
&= \frac{L \cdot \rho}{F} \qquad (3)
\end{aligned}
$$

Let's interpret these

In each case, suppose that ρ is nonzero, implying that there is indeed endogeneity present

The three equations illustrate ways in which the bias could blow up

They all concern the first stage regression of X_i on Z_i

Bias could blow up if

1. ESS ≈ 0
2. $R^2 \approx 0$
3. first stage F statistic is zero

These are all equivalent ways of saying:
the instruments don't explain the endogenous variable well enough

Of course, asymptotically, the bias is zero

But the problem that we point out here confronts researchers who, in practice, deal with finite samples

We will pursue this further, both analytically and computationally

Roadmap

Consider the simple scalar model

$$Y_i = X_i\beta + e_i$$
$$X_i = Z_i\pi + v_i$$

In other words: $K_1 = 0$, $K_2 = L_2 = L = 1$

Let's make life easy: $\mathsf{E}Z_i = 0$ and $\mathsf{E}Z_i^2 = 1$

Then $\pi = \mathsf{Cov}(X_i, Z_i)/\mathsf{Var}\,(Z_i) = \mathsf{E}(X_i Z_i)/\mathsf{E}(Z_i^2) = \mathsf{E}(X_i Z_i)$

What happens when $\mathsf{E}(X_i Z_i) = 0$ so that $\pi = 0$?

In that case, the first stage equation simplifies to $X_i = v_i$

Let's label this case *invalid instrument*

Using Z_i as an IV doesn't make sense because it isn't one

Let's further assume, for simplicity,

$$\text{Var}\left(\begin{pmatrix} e_i \\ v_i \end{pmatrix} | Z_i\right) = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

Endogeneity, of course, implies $\rho \neq 0$

Let's say, you recognize that Z_i isn't really an IV and you decide to resort to OLS instead

$$\hat{\beta}^{\text{OLS}} - \beta = \frac{\sum_{i=1}^{N} X_i e_i}{\sum_{i=1}^{N} X_i^2} = \frac{N^{-1} \sum_{i=1}^{N} v_i e_i}{N^{-1} \sum_{i=1}^{N} v_i^2} \xrightarrow{\text{p}} \frac{\text{E}(v_i e_i)}{\text{E}(v_i^2)} = \rho \neq 0$$

So $\hat{\beta}^{\text{OLS}}$ is not consistent, which we knew already

Can the instrument help, although it is invalid?

And if it doesn't help, could the instrument do any harm? (spoiler alert: Yes!)

$$\hat{\beta}^{\mathsf{IV}} - \beta = \frac{N^{-1}\sum_{i=1}^{N} Z_i e_i}{N^{-1}\sum_{i=1}^{N} X_i Z_i} \xrightarrow{\mathsf{p}} \frac{\mathsf{E}(Z_i e_i)}{\mathsf{E}(X_i Z_i)} = \frac{0}{0},$$

which is indeterminate

Notice that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\begin{pmatrix} Z_i e_i \\ Z_i v_i \end{pmatrix} \xrightarrow{\mathsf{d}} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} \sim \mathsf{N}\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

Notice that Var $(Z_i e_i) = \mathsf{E}(Z_i^2 e_i^2) = \mathsf{E}(Z_i^2 \mathsf{E}(e_i^2 | Z_i)) = 1$
(and similarly for Var $(Z_i v_i)$)

Here Cov$(\xi_1, \xi_2) = \mathsf{E}(\xi_1 \xi_2) = \rho$

Then define $\xi_0 := \xi_1 - \rho\xi_2$

This makes Cov$(\xi_0, \xi_2) = \mathsf{E}(\xi_0 \xi_2) = 0$,
meaning ξ_0 and ξ_2 are independent
(joint normal and zero covariance implies independence)

Let’s take another look now, plugging in $\xi_1 = \xi_0 + \rho\xi_2$:

$$\hat{\beta}^{\mathrm{IV}} - \beta = \frac{\frac{1}{\sqrt{N}} \sum_{i=1}^{N} Z_i e_i}{\frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i Z_i} \xrightarrow{\mathrm{d}} \frac{\xi_1}{\xi_2} = \rho + \frac{\xi_0}{\xi_2}$$

(and applying the continuous mapping theorem: the limiting distribution of the ratio is the ratio of the limiting distributions)

The ratio of two independently normally distributed rvs with zero mean results in a Cauchy distributed random variable that is centered at zero

The Cauchy distribution is nasty

- although it is centered at zero it has infinite mean
- its median is zero
- it has thick tails (outliers)

We've learned that using $\hat{\beta}^{\text{IV}}$ when Z_i isn't a valid IV results in an estimator $\hat{\beta}^{\text{IV}}$ that

- does not converge in probability
- instead converges to a Cauchy distribution
- has a median of $\beta + \rho$

Let's say, you ignore all that and use an IV based t test anyway

What will happen?

What happens to the $\hat{\beta}^{\text{IV}}$-based t statistic under invalid instruments?

Recall the generic t statistic that is based on an estimator $\hat{\beta}$:

$$t_{\hat{\beta}}(\beta) = \frac{\hat{\beta} - \beta}{\text{se}(\hat{\beta})}$$

Let's make our lives easy and consider the standard error of $\hat{\beta}^{\text{IV}}$ under homoskedasticity

The estimator of the asymptotic variance for $\hat{\beta}^{\text{IV}}$ is

$$\text{Var}\left(\hat{\beta}^{\text{IV}} | Z_i\right) = \hat{\sigma}_e^2 \frac{\sum_{i=1}^{N} Z_i^2}{\left(\sum_{i=1}^{N} X_i Z_i\right)^2}$$

therefore

$$\text{se}(\hat{\beta}^{\text{IV}}) = \frac{\sqrt{\hat{\sigma}_e^2 \sum_{i=1}^{N} Z_i^2}}{\sum_{i=1}^{N} X_i Z_i}$$

Notice

$$
\begin{aligned}
\hat{\sigma}_e^2 &= N^{-1}\sum_{i=1}^{N}(Y_i - X_i\hat{\beta}^{\mathsf{IV}})^2 = N^{-1}\sum_{i=1}^{N}\left(X_i(\beta - \hat{\beta}^{\mathsf{IV}}) + e_i\right)^2 \\
&= \left(\sum_{i=1}^{N} e_i^2/N\right) - 2(\hat{\beta}^{\mathsf{IV}} - \beta)\left(\sum_{i=1}^{N} X_i e_i/N\right) + (\hat{\beta}^{\mathsf{IV}} - \beta)^2\left(\sum_{i=1}^{N} X_i^2/N\right) \\
&\xrightarrow{\mathrm{d}} 1 - 2\rho\frac{\xi_1}{\xi_2} + \left(\frac{\xi_1}{\xi_2}\right)^2
\end{aligned}
$$

It follows for the standard error (using con't mapping theorem):

$$
\mathrm{se}(\hat{\beta}^{\mathsf{IV}}) = \frac{\sqrt{\hat{\sigma}_e^2\sum_{i=1}^{N} Z_i^2}}{\sum_{i=1}^{N} X_i Z_i} = \frac{\sqrt{\hat{\sigma}_e^2\frac{1}{N}\sum_{i=1}^{N} Z_i^2}}{\frac{1}{\sqrt{N}}\sum_{i=1}^{N} X_i Z_i} \xrightarrow{\mathrm{d}} \frac{\sqrt{1 - 2\rho\frac{\xi_1}{\xi_2} + \left(\frac{\xi_1}{\xi_2}\right)^2}}{\xi_2}
$$

And for the t statistics:

$$
t_{\hat{\beta}^{\mathsf{IV}}}(\beta) = \frac{\hat{\beta}^{\mathsf{IV}} - \beta}{\mathrm{se}(\hat{\beta}^{\mathsf{IV}})} \xrightarrow{\mathrm{d}} \frac{\xi_1/\xi_2}{\frac{\sqrt{1-2\rho\frac{\xi_1}{\xi_2}+\left(\frac{\xi_1}{\xi_2}\right)^2}}{\xi_2}} = \frac{\xi_1}{\sqrt{1 - 2\rho\frac{\xi_1}{\xi_2} + \left(\frac{\xi_1}{\xi_2}\right)^2}}
$$

(Note: the numerator is slightly different from Hansen)

Copy and paste last line from previous slide:

$$t_{\hat{\beta}^{\mathrm{IV}}}(\beta) \xrightarrow{\mathrm{d}} \frac{\xi_1}{\sqrt{1 - 2\rho\frac{\xi_1}{\xi_2} + \left(\frac{\xi_1}{\xi_2}\right)^2}} =: S(\rho)$$

What does this mean?

The t statistic does NOT converge to a normal distribution

So we can't simply compare it to the ± 1.96 cutoffs

The asymptotic distribution of t depends on ρ, a parameter that we don't know and cannot estimate

- ρ is the degree of endogeneity

To get more intuition about what's going on, let's send ρ to 1 which is the worst possible case of endogeneity

The closer $\rho \to 1$, the more ξ_1 and ξ_2 will resemble each other

Weird things will happen in the limit case as $\rho \to 1$:

- $\xi_1 \xrightarrow{\text{p}} \xi_2$
- $\hat{\sigma}_e^2 \xrightarrow{\text{p}} 0$
- $\text{se}(\hat{\beta}^{\text{IV}}) \xrightarrow{\text{p}} 0$
- $S(\rho) \to \infty$
- and ultimately the t statistic converges in probability to ∞

That can't be good

It means, that you are mechanically rejecting H_0 irrespective of the true value of β

Hansen puts it nicely in his book:
...users may incorrectly interpret estimates as precise, despite the fact that they are useless.

Put slightly differently:

- the t statistic based on $\hat{\beta}^{\text{IV}}$ when instruments are invalid is deceivingly optimistic
- it tends to be large suggesting a nonzero coefficient
- irrespective of the true value of β
- the large t statistic is merely an artifact of the breakdown of the asymptotic normal distribution

In the case $\pi = 0$, perhaps better to use OLS instead of IV?

Problem: in applications you don't usually know that $\pi = 0$

Anyway, maybe the case $\pi = 0$ is too extreme and produces problems that are too dramatic

Let's study a case that is less extreme and therefore, maybe, less dramatic: $\pi \neq 0$ but $\pi \approx 0$ (so-called *weak instruments*)

Roadmap

We have seen that $\pi = 0$ (*invalid instruments*) leads to a breakdown of statistical inference for the IV estimator

Now let's look at: $\pi \neq 0$ but $\pi \approx 0$

What I'm trying to say here:
π is not equal to zero but it is close to zero or *local to zero*

We will use the same setup as in the *invalid instrument* case (one endogenous regressor and one instrument)

Technically, local to zero is generated by letting $\pi = N^{-1/2}\tau$ where $\tau \neq 0$

Where does this come from? You could guess that, once you plug this into an asymptotic expansion, it delivers a useful rate of convergence

Reminder of the setup

$$Y_i = X_i\beta + e_i$$
$$X_i = Z_i\pi + v_i$$

In other words: $K_1 = 0$, $K_2 = L_2 = L = 1$

We still assume that $\mathrm{E}Z_i = 0$ and $\mathrm{E}Z_i^2 = 1$

Recall that $\pi = \mathrm{E}(X_iZ_i)/\mathrm{E}(Z_i^2) = \mathrm{E}(X_iZ_i)$

What happens when $\mathrm{E}(X_iZ_i) \approx 0$ so that $\pi \approx 0$?

Let's label this case *weak instrument*

To make life easy, let's assume

$$\mathrm{Var}\left(\begin{pmatrix} e_i \\ v_i \end{pmatrix} | Z_i\right) = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

Endogeneity, of course, implies $\rho \neq 0$

Let's again first look at the OLS estimator

$$\hat{\beta}^{\text{OLS}} - \beta = \frac{\sum_{i=1}^{N} X_i e_i}{\sum_{i=1}^{N} X_i^2}$$

$$= \frac{N^{-1} \sum_{i=1}^{N} (N^{-1/2} \tau Z_i + v_i) e_i}{N^{-1} \sum_{i=1}^{N} (N^{-1/2} \tau Z_i + v_i)^2}$$

$$\xrightarrow{\text{p}} \frac{\mathsf{E}(v_i e_i)}{\mathsf{E}(v_i^2)} = \rho \neq 0$$

which is the same as before when $\pi = 0$

Let's turn to the IV estimator, remember

$$\hat{\beta}^{\text{IV}} - \beta = \frac{\sum_{i=1}^{N} Z_i e_i}{\sum_{i=1}^{N} Z_i X_i}$$

We start by looking at

$$\begin{aligned}\frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_i X_i &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_i^2 \pi + \frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_i v_i \\ &= \frac{1}{N}\sum_{i=1}^{N} Z_i^2 \tau + \frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_i v_i \\ &\xrightarrow{\mathrm{d}} \tau + \xi_2\end{aligned}$$

and recall

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\begin{pmatrix} Z_i e_i \\ Z_i v_i \end{pmatrix} \xrightarrow{\mathrm{d}} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} \sim \mathrm{N}\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right), \text{therefore}$$

$$\hat{\beta}^{\mathrm{IV}} - \beta = \frac{\frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_i e_i}{\frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_i X_i} \xrightarrow{\mathrm{d}} \frac{\xi_1}{\tau + \xi_2}$$

Again: $\hat{\beta}^{\mathrm{IV}}$ is inconsistent with non-normal asymptotic distribution

What happens to the t test based on $\hat{\beta}^{\text{IV}}$ under weak identification?

Recall the generic t statistic that is based on an estimator $\hat{\beta}$:

$$t_{\hat{\beta}}(\beta) = \frac{\hat{\beta} - \beta}{\text{se}(\hat{\beta})}$$

Let's make our lives easy and consider the standard error of $\hat{\beta}^{\text{IV}}$ under homoskedasticity

The estimator of the asymptotic variance for $\hat{\beta}^{\text{IV}}$ is

$$\text{Var}\left(\hat{\beta}^{\text{IV}} | Z_i\right) = \hat{\sigma}_e^2 \frac{\sum_{i=1}^N Z_i^2}{(\sum_{i=1}^N X_i Z_i)^2}$$

therefore

$$\text{se}(\hat{\beta}^{\text{IV}}) = \hat{\sigma}_e \frac{\sqrt{\sum_{i=1}^N Z_i^2}}{\sum_{i=1}^N X_i Z_i}$$

Notice

$$\hat{\sigma}_e^2 = N^{-1}\sum_{i=1}^{N}(Y_i - X_i\hat{\beta}^{\mathsf{IV}})^2 = N^{-1}\sum_{i=1}^{N}\left(X_i(\beta - \hat{\beta}^{\mathsf{IV}}) + e_i\right)^2$$

$$= \left(\sum_{i=1}^{N} e_i^2/N\right) - 2(\hat{\beta}^{\mathsf{IV}} - \beta)\left(\sum_{i=1}^{N} X_i e_i/N\right) + (\hat{\beta}^{\mathsf{IV}} - \beta)^2\left(\sum_{i=1}^{N} X_i^2/N\right)$$

$$\xrightarrow{\mathrm{d}} 1 - 2\rho\frac{\xi_1}{\tau + \xi_2} + \left(\frac{\xi_1}{\tau + \xi_2}\right)^2$$

It follows that

$$\mathrm{se}(\hat{\beta}^{\mathsf{IV}}) = \frac{\sqrt{\hat{\sigma}_e^2 \sum_{i=1}^{N} Z_i^2}}{\sum_{i=1}^{N} X_i Z_i} = \frac{\sqrt{\hat{\sigma}_e^2 \frac{1}{N}\sum_{i=1}^{N} Z_i^2}}{\frac{1}{\sqrt{N}}\sum_{i=1}^{N} X_i Z_i} \xrightarrow{\mathrm{d}} \frac{\sqrt{1 - 2\rho\frac{\xi_1}{\tau+\xi_2} + \left(\frac{\xi_1}{\tau+\xi_2}\right)^2}}{\tau + \xi_2}$$

And for the t statistic:

$$t_{\hat{\beta}^{\mathsf{IV}}}(\beta) = \frac{\hat{\beta}^{\mathsf{IV}} - \beta}{\mathrm{se}(\hat{\beta}^{\mathsf{IV}})} \xrightarrow{\mathrm{d}} \frac{\xi_1/(\tau + \xi_2)}{\frac{\sqrt{1-2\rho\frac{\xi_1}{\tau+\xi_2} + \left(\frac{\xi_1}{\tau+\xi_2}\right)^2}}{\tau+\xi_2}} = \frac{\xi_1}{\sqrt{1 - 2\rho\frac{\xi_1}{\tau+\xi_2} + \left(\frac{\xi_1}{\tau+\xi_2}\right)^2}}$$

Copy and paste last line from previous slide:

$$t_{\hat{\beta}^{\text{IV}}}(\beta) \xrightarrow{\text{d}} \frac{\xi_1}{\sqrt{1 - 2\rho\frac{\xi_1}{\tau+\xi_2} + \left(\frac{\xi_1}{\tau+\xi_2}\right)^2}} =: S(\rho, \tau)$$

What does this mean?

The t statistic does NOT converge to a normal distribution

So we can't simply compare it to the ± 1.96 cutoffs

The asymptotic distribution of t depends on ρ and τ, two parameters that we don't know and cannot estimate

- ρ is the degree of endogeneity
- τ is the strength of the instrument

To get more intuition about what's going on, let's set $\rho = 1$ which is the worst possible case of endogeneity

Then $\xi_1 = \xi_2$ and the t statistic collapses to

$$S(1, \tau) = \xi_1 + \frac{\xi_1^2}{\tau},$$

Recall that $\xi_1 \sim \mathrm{N}(0, 1)$ and $\xi_1^2 \sim \chi_1^2$

So $S(1, \tau)$ is a mixture of a $\mathrm{N}(0, 1)$ and a χ_1^2 distribution

The degree of the mixture is controlled by the value of τ

- if τ is very large, then $S(1, \tau)$ will be close to $\mathrm{N}(0, 1)$ (strong instrument case)
- if τ is very small, then the χ_1^2 dominates and distorts away from normality (weak instrument case)
- in the extreme we get $\lim_{\tau \to 0} S(1, \tau) = \infty$ (that's a terrible result: very weak instruments will yield misleadingly large t statistics suggesting significant β regardless of the truth)