Advanced Econometrics I

Jürgen Meinecke

Lecture 7 of 12

Research School of Economics, Australian National University

Roadmap

Instrumental Variables Estimation

- Invalid Instruments
- Weak Instruments

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: $\mathrm{E}Z_i = 0$ and $\mathrm{E}Z_i^2 = 1$

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

What happens when $\mathrm{E}(X_iZ_i) = 0$ 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 $\xi_1 = \xi_0 + \rho\xi_2$:

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

(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 $\hat{\beta}^{\mathsf{IV}}$-based t statistic under invalid instruments?

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

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

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

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

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

therefore

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

Notice

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

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

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

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:

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

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

Instrumental Variables Estimation

- Invalid Instruments
- Weak Instruments

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

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

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

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

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

Let's turn to the IV estimator, remember

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

We start by looking at

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_i X_i = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_i^2 \pi + \frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_i v_i$$
$$= \frac{1}{N}\sum_{i=1}^{N} Z_i^2 \tau + \frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_i v_i$$
$$\xrightarrow{\text{d}} \tau + \xi_2$$

and recall

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

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

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

What happens to the t test based on $\hat{\beta}^{\mathsf{IV}}$ 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 $\hat{\beta}^{\mathsf{IV}}$ under homoskedasticity

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

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

therefore

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

Notice

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

It follows that

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

And for the t statistic:

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

Copy and paste last line from previous slide:

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

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)