Advanced Econometrics I

Jürgen Meinecke

Lecture 6 of 12

Research School of Economics, Australian National University

Roadmap

Instrumental Variables Estimation

- Two Stage Least Squares (2SLS) Estimator
- Large Sample Properties of 2SLS Estimator
- Bias of 2SLS Estimator

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

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

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

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

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

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

Instrumental Variables Estimation

- Two Stage Least Squares (2SLS) Estimator
- Large Sample Properties of 2SLS Estimator
- Bias of 2SLS Estimator

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

The resulting covariance matrix estimator will be consistent

Roadmap

Instrumental Variables Estimation

- Two Stage Least Squares (2SLS) Estimator
- Large Sample Properties of 2SLS Estimator
- Bias of 2SLS Estimator

Although the 2SLS estimator is consistent, it is biased

Where does this bias come from?

Recall the infeasible 2SLS estimator:

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

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

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

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

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
- $\Xi := \sum_{i=1}^{N} Z_i Z_i' / N$

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)

In the scalar case, by definition:

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

Rearranging results in

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

Let's dissect numerator and denominator

Turning first to the numerator

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

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

These can be used in numerator and denominator

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

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, in that case $\mathrm{E}(v_i w_i) = 0$ implies that v_i and w_i are statistically independent

Copy and paste from previous slide:

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

Looking at first two terms:

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

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

where I have used the lemma: if $P \sim \mathsf{N}(0, I_N)$ then $P'QP \sim \chi^2(\text{tr }(Q))$

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

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

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

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

What's the expected value of the third term?

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

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

Bottom line 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

Have you noticed at which stage the bias has entered?

Now 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

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

$$\begin{aligned}
\frac{2}{N}\pi'Z'v &\sim \mathsf{N}\left(0, \frac{4\sigma_v^2}{N^2}\pi'Z'Z\pi\right) \\
&= \frac{2}{\sqrt{N}}\mathsf{N}\left(0, \sigma_v^2\pi'\Xi\pi\right) \\
&= \frac{2}{\sqrt{N}}\mathsf{O}_p(1) = \mathsf{O}_p\left(\frac{1}{\sqrt{N}}\right) \\
\frac{1}{N}v'P_Zv &\sim \frac{1}{N}\sigma_v^2\chi^2(L) = \frac{1}{N}\sigma_v^2\mathsf{O}_p(1) = \mathsf{O}_p\left(\frac{1}{N}\right) = \mathsf{o}_p\left(\frac{1}{\sqrt{N}}\right)
\end{aligned}$$

Bottom line:

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

Putting things together and applying an asymptotic approximation from Hahn and Kuersteiner (2002)

For small r, they use: $\frac{1}{\pi'\Xi\pi+r} \approx \frac{1}{\pi'\Xi\pi} - \frac{1}{(\pi'\Xi\pi)^2} \cdot r$

We use: $\frac{1}{\pi'\Xi\pi+r} \approx \frac{1}{\pi'\Xi\pi}$

Then

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

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

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

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

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

$$\mathrm{E}\left(\hat{\beta}^{\mathrm{2SLS}} - \beta\right) \approx \frac{1}{N}\frac{L}{\pi'\Xi\pi}\sigma_{ev} = \frac{L}{\pi' Z' Z\pi}\sigma_{ev} = \frac{L}{\mu^2}\frac{\sigma_{ev}}{\sigma_v^2},$$

where $\mu^2 := \pi' Z' Z\pi/\sigma_v^2$ is the *concentration parameter*

Aside: Hahn and Kuersteiner obtain $\mathrm{E}\left(\hat{\beta}^{\mathrm{2SLS}} - \beta\right) \approx \frac{L-2}{\mu^2}\frac{\sigma_{ev}}{\sigma_v^2}$

Notice: $\mathrm{Var}\,(X) \approx \pi' Z' Z\pi + \sigma_v^2$

So the concentration parameter is the proportion of the variation in X that is captured by the instruments

The concentration parameter is a measure of strength of the instruments

If instruments are weak, in the sense of $\mu^2 \approx 0$, then we suspect a large bias for the 2SLS estimator

We will pursue this further, both analytically and computationally