

Lecture 57: 2SLS standard errors; an over-identified example; 2SLS, IV and ILS under exact identification

Econometrics ■ Unit IV: Simultaneous equations models

Kapil Kumar

Department of Statistics
Chaudhary Charan Singh University, Meerut

B.Sc. (Hons) Statistics, Semester V
Academic Session 2026–2027

Previous lecture

Lecture 56: Two-stage least squares: the method, derivation and consistency.

Today

- Two-stage least squares
 - Standard errors
 - 2SLS, IV and ILS under exact identification

The estimated covariance matrix is

$$\widehat{\text{Var}}(\hat{\theta}_{2SLS}) = s^2(D'P_X D)^{-1}, \quad s^2 = \frac{\hat{u}'\hat{u}}{n - m - k},$$

$$\hat{u} = y_1 - D\hat{\theta}_{2SLS} = y_1 - Y_1\hat{\beta} - X_1\hat{\gamma}.$$

Important

- The residuals $\hat{u}$ use the *original* Y_1 , not $\hat{Y}_1$.
- The residuals printed by the second-stage OLS regression, $y_1 - \hat{Y}_1\hat{\beta} - X_1\hat{\gamma}$, differ from $\hat{u}$ by $(Y_1 - \hat{Y}_1)\hat{\beta}$, and do not estimate the structural disturbance.
- So the standard errors printed by a manual second-stage regression must be corrected.

Example 8.3 (An over-identified 2SLS calculation)

One endogenous regressor x , an intercept, and two excluded exogenous variables z_1, z_2 ; $n = 20$. All variables are centred, so the intercept estimate is zero and we work with $Z = [z_1 \ z_2]$. Suppose

$$Z'Z = \begin{pmatrix} 10 & 0 \\ 0 & 10 \end{pmatrix}, \quad Z'x = \begin{pmatrix} 10 \\ 5 \end{pmatrix}, \quad Z'y = \begin{pmatrix} 20 \\ 5 \end{pmatrix},$$

$$x'x = 16, \quad x'y = 28, \quad y'y = 100.$$

Stage 1. $\hat{\pi} = (Z'Z)^{-1}Z'x = (1, 0.5)'$, so $\hat{x} = z_1 + 0.5z_2$.

Example 8.3 (cont.)

Stage 2 quantities. Using $(Z'Z)^{-1} = 0.1I$,

$$\hat{x}'\hat{x} = x'P_Zx = (Z'x)'(Z'Z)^{-1}(Z'x) = 0.1(10^2 + 5^2) = 12.5,$$

$$\hat{x}'y = x'P_Zy = 0.1\{10(20) + 5(5)\} = 22.5.$$

Estimates.

$$\hat{\beta}_{2SLS} = 22.5/12.5 = 1.8,$$

while OLS gives $x'y/x'x = 28/16 = 1.75$.

Example 8.3 (cont.)

Standard error. Structural residual sum of squares:

$$\begin{aligned}\hat{u}'\hat{u} &= (y - 1.8x)'(y - 1.8x) = y'y - 2(1.8)x'y + 1.8^2x'x \\ &= 100 - 100.8 + 51.84 = 51.04.\end{aligned}$$

With $p = 2$ coefficients (intercept and slope), $s^2 = 51.04/18 = 2.8356$ and

$$se(\hat{\beta}_{2SLS}) = \sqrt{2.8356/12.5} = 0.4763.$$

Using second-stage residuals instead would give

$$y'y - 2(1.8)(22.5) + 1.8^2(12.5) = 100 - 81 + 40.5 = 59.5,$$

which is wrong for this purpose.

If the equation is exactly identified, the number of exogenous variables K equals the number of coefficients $m + k$, so $X'D$ is a square matrix. If it is nonsingular, then by $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$,

$$\begin{aligned}\hat{\theta}_{2SLS} &= \{D'X(X'X)^{-1}X'D\}^{-1}D'X(X'X)^{-1}X'y_1 \\ &= (X'D)^{-1}(X'X)(D'X)^{-1}D'X(X'X)^{-1}X'y_1 \\ &= (X'D)^{-1}(X'X)(X'X)^{-1}X'y_1 = (X'D)^{-1}X'y_1.\end{aligned}$$

- This is the IV estimator of Chapter 6 with the exogenous variables X as instruments.
- It solves $X'(y_1 - D\hat{\theta}) = 0$, the same equations that ILS solves.
- Hence, *for an exactly identified equation, ILS, IV and 2SLS give identical estimates.*
- For an over-identified equation, 2SLS combines all the instruments in a single estimate.

Practice

- Exercises 8.3 and 8.4 (Exercises and Solutions booklet).

Next lecture

Lecture 58: LIML: the least variance ratio and its minimisation.