2SLS (IV)

$\hat{\beta}_{OLS}$ if X - stochastic

consistent if $\operatorname{cov}(X, \varepsilon) = 0$

unbiased if $X \perp \varepsilon$

$$\hat{\beta} = (X'X)^{-1} X'y = (X'X)^{-1} X'(X\beta + \varepsilon) =$$

$$= \beta + \underbrace{(X'X)^{-1} X'\varepsilon}_{\text{bias}}$$

$$E(\hat{\beta}) = \beta + E((X'X)^{-1} X'\varepsilon)$$

deterministic

$$(X'X)^{-1} X' \underset{\substack{\shortparallel \\ 0}}{E(\varepsilon)}$$

stochastic

$= 0$

if $X \perp \varepsilon$

$\operatorname{cov}(\varepsilon, X) = 0$
is not enough

$$E(\hat{\beta} | X) = \beta + E((X'X)^{-1} X'\varepsilon | X) =$$

$$= \beta + (X'X)^{-1} X' \underset{\substack{\shortparallel \\ 0}}{E(\varepsilon | X)} = \beta$$

IV

X – endogenous variable, z – instrumental var.

1 step:

$$\hat{X}_i = \underbrace{\hat{\theta}_1 + \hat{\theta}_2 \cdot z_i}$$

① exogenous

$cov(z, \varepsilon) = 0$

2 step:

$$y_i = \hat{\beta}_1 + \hat{\beta}_2 \cdot \hat{X}_i + \hat{\varepsilon}_i$$

② relevance

$cov(z, x) \neq 0$

Problem 1.

$$\hat{\beta}_2 = \frac{\hat{cov}(\hat{x}, y)}{\hat{var}(\hat{x})} = \frac{\hat{cov}(\hat{\theta}_1 + \hat{\theta}_2 z, y)}{\hat{var}(\hat{\theta}_1 + \hat{\theta}_2 z)} =$$

$$\left\{ \begin{array}{l} \hat{x} = \hat{\theta}_1 + \hat{\theta}_2 z \\ \hat{\theta}_2 = \dfrac{\hat{cov}(z, x)}{\hat{var}(z)} \end{array} \right\}$$

$$= \frac{\cancel{\hat{\theta}_2}\, \hat{cov}(z, y)}{(\hat{\theta}_2)^{\cancel{2}} \cdot \hat{var}(z)} = \frac{\hat{cov}(z, y)}{\dfrac{\hat{cov}(z, x)}{\cancel{\hat{var}(z)}} \cdot \cancel{\hat{var}(z)}} = \frac{\hat{cov}(z, y)}{\hat{cov}(z, x)}$$

$$\frac{\hat{cov}(\overset{z}{x}, y)}{\hat{cov}(\underset{z}{x}, x)}$$

$$\operatorname{plim} \hat{\beta}_{TSLS} = \operatorname{plim}_{n\to\infty} \frac{\widehat{\operatorname{cov}}(z,y)}{\widehat{\operatorname{cov}}(z,X)} = \frac{\operatorname{cov}(z_i, \beta_1 + \beta_2 \cdot X_i + \varepsilon_i)}{\operatorname{cov}(z_i, X_i)} =$$

$$= \beta_2 + \frac{\operatorname{cov}(z_i, \varepsilon_i)^{=0}}{\underbrace{\operatorname{cov}(z_i, X_i)}_{}} \neq 0$$

(exogenous)

(relevance)

$$se(\hat{\beta}_2^{IV}) = \sqrt{\underbrace{\frac{s^2}{\sum(X_i - \bar{X})^2}}_{Var(\hat{\beta}_2^{OLS})} \cdot \frac{1}{\widehat{\operatorname{corr}}^2(X,z)}}$$

$$s^2 = \frac{\sum e_i^2}{n-2}$$

$$e_i = y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i$$

$x_1, \dots, x_p$ - endogenous regr.

$w_1, \dots, w_r$ - exogeneus regr.

$z_1, \dots, z_m$ - instruments

1) $\hat{x}_1 \mid z_1, \dots, z_m, w_1, \dots, w_r$

$\vdots$

$\hat{x}_p \mid z_1, \dots, z_m, w_1, \dots, w_r$

2) $y \mid \hat{x}_1, \dots, \hat{x}_p, w_1, \dots, w_r$

$m < p$ => under identified

$\hat{x}_1 = \hat{\alpha}_1 + \hat{\alpha}_2 \cdot z_i$

$\hat{x}_2 = \hat{\beta}_1 + \hat{\beta}_2 \cdot z_i$

=> perf. multicollinearity

$m = p$ => exactly identified

(IV)

$$\hat{\beta}_{IV} = (Z'X)^{-1} Z'y$$

$m > p$ => over identified

(2SLS)

Testing:

1) Relevance

2) Exogeneity (Sargan's test)

3) $\hat{\beta}_{OLS}$ vs $\hat{\beta}_{IV}$ (Hausman test)

H_0: Z - weak

H_a: Z - relevant instrument (strong)

1) $x \mid z_1, \dots, z_m, w_1, \dots, w_r$

$\hookrightarrow$ $\underline{F > 10}$ $\Rightarrow$ relevance $\checkmark$

if $F < 10$ $\Rightarrow$ weak instruments

2) Sargan test $m > p$

H_0: z - exogenous

H_a: z - endogenous

$\hat{\varepsilon}_1 \mid z_1, \dots, z_m, w_1, \dots, w_r \Rightarrow F$

$\uparrow$ 2^{nd} step of 2SLS

$$y = m \cdot F \overset{H_0}{\sim} \chi^2_{m-p}$$

3) Hausman test:

H_0: $\hat{\beta}_{OLS}$ - consistent $\Rightarrow$ $\hat{\beta}_{OLS}$ (more efficient)

H_a: $\hat{\beta}_{OLS}$ - inconsistent $\Rightarrow$ $\hat{\beta}_{2SLS}$ (consistent)

$$(\hat{\beta}_{2SLS} - \hat{\beta}_{OLS})'(Var(\hat{\beta}_{2SLS}) - Var(\hat{\beta}_{OLS}))^{-1}(\hat{\beta}_{2SLS} - \hat{\beta}_{OLS}) \overset{H_0}{\sim} \chi^2_k$$

k - # regr. on 2^{nd} step