Tutorial: Two Stage Least Squares for IV

Feedback Form: `https://tiny.cc/hammadfeedback`

Hammad Shaikh

Instrumental Variable Review

$$
\begin{tikzcd}
Z \arrow[r] & X \arrow[rr] & & Y \\
& & \varepsilon \arrow[ul] \arrow[ur] &
\end{tikzcd}
\Rightarrow
\begin{cases}
X \to Y \\
X \leftarrow \varepsilon \to Y
\end{cases}
$$

- $Y = \beta_0 + \beta_1 X + \epsilon$

$$\hat{\beta}_{IV} = \widehat{Cov}(Z,Y)/\widehat{Cov}(Z,X) \xrightarrow{p} \beta_1$$

- Endogenity problem: $Cov(X, \epsilon) \neq 0$

↳ Indirect path $X \leftarrow \varepsilon \rightarrow Y$

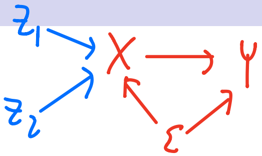
↳ OLS is biased

- Solution: IV Z so that $i)\underbrace{Cov(Z, X) \neq 0}_{\text{Relevance}}$ and $ii)\underbrace{Cov(Z, \epsilon) = 0}_{\text{Exogenity}}$

Z random $\Rightarrow Z \perp \varepsilon$

iii) Exclusion: $Z \underset{\text{not directly}}{\not\longrightarrow} Y$

Multiple Instrumental Variables



- Consider Z_1 and Z_2 as valid IVs for X

i) $Cov(Z_j, \varepsilon) = 0$, ii) $Cov(Z_j, X) \neq 0$, $j = 1, 2$

- Idea: Combine Z_1 and Z_2 into one valid IV for X

$$\tilde{Z} = w_1 Z_1 + w_2 Z_2$$

$$\hookrightarrow \begin{cases} Cov(\tilde{Z}, \varepsilon) = 0 \\ Cov(\tilde{Z}, X) \neq 0 \end{cases}$$

Two Stage Least Squares

$\boxed{2SLS}$

$$\begin{tikzcd} Z_1 \arrow[dr] & & & \varepsilon \arrow[dll] \arrow[dr] & \\ & X \arrow[r] & \hat{X} \arrow[rr] & & Y \\ Z_2 \arrow[ur] & & & & \end{tikzcd}$$

$$\begin{tikzcd} \hat{X} \arrow[r] \arrow[d, phantom, "\Updownarrow"] & Y \\ X \arrow[r] & Y \end{tikzcd}$$

- First stage: impact of IVs Z_1 and Z_2 on X

  $\hookrightarrow X = \alpha_0 + \alpha_1 Z_1 + \alpha_2 Z_2 + u$

  $\hookrightarrow \hat{X} = \hat{\alpha}_0 + \hat{\alpha}_1 Z_1 + \hat{\alpha}_2 Z_2$

- Second stage: Use exogenous variation in X that is induced by the IVs as a determinant of Y

  $\hookrightarrow Y = \beta_0 + \beta_1 \hat{X} + \varepsilon$

  $\hookrightarrow \hat{\beta}_1^{2SLS}$

Two Stage Least Squares with Covariates

- $Y = \beta_0 + \beta_1 X + \beta_2 X^* + \epsilon$

$X^* =$ exog. Controls

$\hookrightarrow X^* \perp \varepsilon$

$\hookrightarrow \mathrm{Cov}(X, \varepsilon) \neq 0 \Rightarrow$ OLS is biased for β_1

- Let Z_1 and Z_2 be valid IVs for X after controlling for X^*

Exog. $Z_1, Z_2 \perp \varepsilon \mid X^*$

- What are the first and second stages?

① $X = \alpha_0 + \alpha_1 Z_1 + \alpha_2 Z_2 + \alpha_3 X^* + u$

$\hookrightarrow \hat{X} = \hat{\alpha}_0 + \hat{\alpha}_1 Z_1 + \hat{\alpha}_2 Z_2 + \hat{\alpha}_3 X^*$

② $Y = \beta_0 + \beta_1 \hat{X} + \beta_2 X^* + \varepsilon \Rightarrow \hat{\beta}_1^{2SLS}$

ε

Practice Problem (Dec, 2016 Exam)

Innovation of printing presses on economic growth

$X \rightarrow Y$, Wealth, Emp $\rightarrow X$, Wealth, Emp $\rightarrow Y$

- $$\underbrace{popgrowth_i}_{Y} = \beta_0 + \overbrace{\beta_1 lnpop_i + \beta_2 \underbrace{print_i}_{X} + \beta_3 port_i}^{X^*} + U_i$$

  $Cov(X, u) \neq 0$ → OLS is biased

- $\overset{Z}{dist_i}$ is distance between city i and Mainz, Germany (printing presses invented)

- Why may $dist_i$ be a valid IV for $print_i$?

① $Cov(Z, X) \neq 0$

② $Cov(Z, \varepsilon) = 0$

③ $Z \not\rightarrow Y$ not directly

Practice Problem Continued (Dec, 2016 Exam)

$X^* = \{lnpop, port\}$

- $\overbrace{popgrowth_i}^{Y} = \beta_0 + \beta_1 lnpop_i + \beta_2 \underbrace{print_i}_{X} + \beta_3 port_i + U_i$

- $\overbrace{dist_i}^{Z}$ is distance between city i and Mainz, Germany (where printing presses were invented)

① $X = \alpha_0 + \alpha_1 Z + \alpha_2 \ln pop + \alpha_3 port + \varepsilon$

↳ $\hat{X} = \hat{\alpha}_0 + \hat{\alpha}_1 Z + \hat{\alpha}_2 \ln pop + \hat{\alpha}_3 port$

- How to apply 2SLS with $dist_i$ as a valid IV for $print_i$?

② $Y = \beta_0 + \beta_1 \ln pop + \beta_2 \hat{X} + \beta_3 port + U$

↳ $\hat{\beta}_2^{2SLS}$