Tutorial: Instrumental Variables

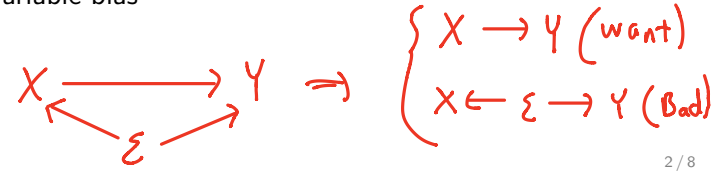
Feedback Form: `https://tiny.cc/hammadfeedback`
Video on IV: `https://youtu.be/CBIE4uwWKwo`

Hammad Shaikh

Endogenity Problem

Ex. Educ → Earnings
Parental Educ → Educ, Parental Educ → Earnings

- Regressor is exogenous if it is independent with error term
  - $X \perp \varepsilon \Rightarrow Cov(X, \varepsilon) = 0$
- Regressor is endogenous if it is correlated with the error term
  - Biases regression parameter estimates
  - $Cov(X, \varepsilon) \neq 0 \Rightarrow E(\hat{\beta}) \neq \beta$ (Bias)
- Omitted variable bias



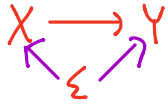
Remedies for Endogenity in Observational Data

OMVB: $X \to Y$, $\varepsilon \to X$, $\varepsilon \to Y$ $\Rightarrow$ RCT $\Rightarrow$ $X \to Y$, $\varepsilon \to Y$

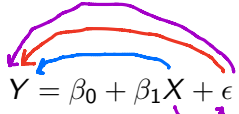
- Randomized Control Trial (RCT)
  - ↳ X random assigned $\Rightarrow X \perp \varepsilon$
- Multiple Regression (Control for confounding variables)
  - ↳ $X \perp \varepsilon \mid \tilde{X}, \hat{X}$ are controls $\Rightarrow$ $\tilde{X} \to Y$, $X \to Y$, $\varepsilon \to Y$
- Instrumental Variable
  - ↳ $Y = \beta_0 + \beta_1 X + \beta_2 \tilde{X} + \varepsilon$

Instrumental Variable Application

- Question: What is the causal effect from going to college on earnings?



- Outcome Y = earnings, policy variable X = I(college grad)



⇒ indirect $X \to \varepsilon \to Y$

- Regression: $Y = \beta_0 + \beta_1 X + \epsilon$
  - Selection bias: individuals that choose to attend college are different from the ones that don't attend

↳ $Cov(X, \varepsilon) \neq 0 \Rightarrow E(\varepsilon | X=0) \neq E(\varepsilon | X=1)$

- Suppose Z = I(college aid) is randomly given to HS students

$\uparrow Z \Rightarrow \uparrow X \to \uparrow Y$, $Z \perp \varepsilon$

Instrumental Variable Intuition

- Endogenity problem: $\triangle X$ implies $\triangle Y = \triangle Y_X + \triangle Y_\epsilon$ (bad)
  - Occurs because $\triangle X \iff \triangle \epsilon \implies \triangle Y_\epsilon$

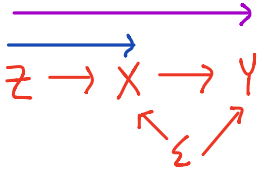
$\uparrow$ Educ $\Rightarrow$ $\uparrow$ Motivation $\Rightarrow$ $\uparrow$ Earnings

- Solution: Use only exogenous variation in X for estimation

$$Y = \beta_0 + \beta_1 X + \varepsilon \implies \frac{dy}{dx} = \beta_1 + \frac{d\varepsilon}{dx}$$

$\frac{d\varepsilon}{dx} \neq 0$ since $Cov(X, \varepsilon) \neq 0$

- Suppose $Z \perp\!\!\!\perp \epsilon$ and Z is related to X



$$\beta_1 = \frac{Cov(Y, Z)}{Cov(X, Z)}$$

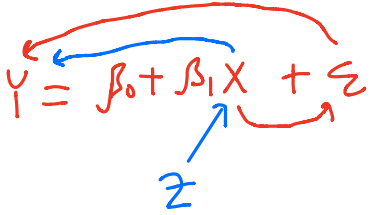
Instrumental Variable Intuition

- Z effects Y only through X: $\triangle Z \implies \triangle X_z \implies \triangle Y_{X_z}$

$$Y = \beta_0 + \beta_1 X(Z) + \varepsilon \Rightarrow \frac{dy}{dz} = \beta_1 \frac{dx}{dz} + \underbrace{\frac{d\varepsilon}{dz}}_{0}$$

- IV Estimate: $\widehat{\beta}_{IV} = \frac{\triangle Y_{X_z}}{\triangle X_z}$

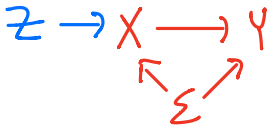
$$\hookrightarrow \beta_1 = \frac{dy/dz}{dx/dz}$$



Instrumental Variables Framework

$Y = \beta_0 + \beta_1 X + \varepsilon, \quad Cov(X, \varepsilon) \neq 0$

- Suppose $Corr(X, \epsilon) \neq 0$, endogeneity problem
  - ↳ Bias & inconsistent slope estimates
- A instrumental variable Z satisfies:
  - $Corr(Z, X) \neq 0$, that is Z related to X (Relevance) [Testable]
  - Z doesn't directly effect outcome Y (Exclusion) } not testable
  - $Z \perp\!\!\!\perp \epsilon$, Z is randomly assigned (Exogenity) }



Ex. Y = Final grade
X = # tutorials
Z = Randomly assigned to get cupcake per tutorial

IV Problem (Dec 2017, Exam)

Private Sch → Test Score ← Teacher IQ

- $Score_i = \beta_0 + \beta_1 YrsPrivSch_i + u_i$

Yrs want to private school

- Why is OLS estimation not appropriate?

↳ OMVB, $Cov(u, YrsPrivsch) \neq 0 \Rightarrow$ biased estimates

- Suppose students were randomly assigned vouchers which allow them to attend private school for $SelYrs_i$ at a discount

# of Yrs discount at private school

- Why is $\underbrace{SelYrs_i}_{Z}$ a valid IV for $\underbrace{YrsPrivSch_i}_{X}$?

(1) Relevance: $Cov(Z, X) \neq 0$

(2) Exclusion: Z not directly effect Y

(3) Exog: $Z \perp \varepsilon$