Lecture 7: Instrumental Variables and Difference in Differences

Hammad Shaikh

Part 1: Annotations until slide 4

Endogenity Problem

- Regressor is exogenous if it is independent with error term

  ↳ $X \perp \varepsilon$

- Regressor is endogenous if it is correlated with the error term
  - Biases regression parameter estimates

  $Corr(X, \varepsilon) \neq 0 \Rightarrow E[\hat{\beta}] \neq \beta$

- Different types of endogenity:
  - Omitted variable bias

    ε contains unobs. variables related to X & Y
  - Measurement error

    Obs. $\tilde{X} = X + \varepsilon$ (X unobserved)
  - Reverse Causality

    want $X \to Y$, but have $Y \to X$

Remedies for Endogenity in Observational Data

- Multiple Regression (Control for confounding variables)
- Regression Discontinuity
- Instrumental Variable
- Differences in Differences

↳ Panel data

Instrumental Variable Application

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

- Outcome Y = earnings, policy variable $X = \mathrm{I}(\text{college grad})$

- 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

  ↳ ε related to X, $E[\varepsilon | X=1] \neq E[\varepsilon | X=0]$

- Suppose $Z = \mathrm{I}(\text{college aid})$ is randomly given to HS students

$\uparrow Z \rightarrow \uparrow X \rightarrow \uparrow Y$, $Z \perp \varepsilon$

Instrumental Variable Intuition

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

- Solution: Use only exogenous variation in X for estimation

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

Instrumental Variable Intuition

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

- IV Estimate: $\widehat{\beta}_{IV} = \frac{\triangle Y_{X_z}}{\triangle X_z}$
  - $\widehat{b}_1$ has a causal interpretation if $Z \perp\!\!\!\perp \epsilon$ and $Corr(Z, X) \neq 0$

Instrumental Variables Framework

- Suppose $Corr(X, \epsilon) \neq 0$, endogeneity problem
- A instrumental variable Z satisfies:
  - $Corr(Z, X) \neq 0$, that is Z related to X
  - Z doesn't directly effect outcome Y
  - $Z \perp\!\!\!\perp \epsilon$, Z is randomly assigned
- $\widehat{\beta}_{IV}$ estimates Local Average Treatment Effect (LATE)
  - ATE for those who comply with instrument

Interaction Term in Regression

- Recall: $Y_i = \beta_0 + \beta_1 I(\text{i college grad}) + \epsilon_i$
  - $\widehat{b}_1 = \overline{Y}_{college} - \overline{Y}_{HS}$

- Question: Do returns to college education differ by gender?

- Answer: Include interaction term
  $Y_i = \alpha_0 + \alpha_1 College_i + \alpha_2 Male_i + \alpha_3 College_i \times Male_i + \epsilon_i$

- $\widehat{a}_1 = \overline{Y}_{college,female} - \overline{Y}_{HS,female}$

- $\widehat{a}_3 = (\overline{Y}_{college,male} - \overline{Y}_{HS,male}) - (\overline{Y}_{college,female} - \overline{Y}_{HS,female})$

Differences in Differences Application

- Question: What is the causal effect of schools receiving funding on students achievement?

- Data: School test scores in 2009 to 2010 in Ontario

- Background: Suppose that Toronto received school funding in 2010 but other cities did not

Differences in Differences Application

Table: Average School Performance

Year/City	Toronto	Other cities	Difference
2009	65	60	5
2010	75	65	10
Difference	10	5	5

Assume: Achievement trends are same in Toronto and other cities

Differences in Differences Intuition

- Control group never receives intervention
- Treatment group receives intervention for $t > t^*$
- Common Trends Assumption: Control and treatment group have same outcome trends over time
- Use trend of the control group to construct the counterfactual for treatment group for $t > t^*$

Differences in Differences Framework

- Requires time variation (before and after) and treatment variation (control and treatment)

- Treatment indicator: $T_i = \text{I(i in treatment group)}$

- Diff - Diff Regression:
  $Y_{it} = \beta_0 + \beta_1 I(t > t^*) + \beta_2 T_i + \beta_3 I(t > t^*) \times T_i + \epsilon_{it}$

- $\widehat{b}_3$ has causal interpretation if "common trends" satisfied
  - Estimates Average Treatment Effect on Treated (ATT)

Causal Inference Summary

- Experiments are ideal
  - Estimates ATE
- Multiple regression is starting point
  - Estimates ATE
- Regression discontinuity
  - Estimates ATE at cutoff
- Instrumental Variables
  - Estimates LATE
- Difference in Differences
  - Estimates ATT