Instrumental Variables

Kosuke Imai

Harvard University

Stat186/Gov2002 Causal Inference

Fall 2019
From randomized encouragement design to general instrumental variables approach:

- Instruments in the nature $\leadsto$ natural experiments
  1. random assignment of $Z$
  2. no direct effect of $Z$ on $Y$
Classical Instrumental Variables Estimator

- **Linear model (in matrix notation):**
  \[ Y = X\beta + \epsilon \quad \text{where } \mathbb{E}(\epsilon) = 0_n \text{ and } X \text{ is } n \times K \]

- **Endogeneity:**
  \[ \mathbb{E}(\epsilon_i | X) \neq 0 \]

- **Instruments** \( Z \) is \( n \times L \)
  1. **Exogeneity:** \( \mathbb{E}(\epsilon_i | Z) = 0 \)
  2. **Exclusion restriction:** \( Z_i \) does not belong to the outcome model
  3. **Rank condition:** \( Z^\top X \) and \( Z^\top Z \) have full rank

- **Experimental setting:**
  - \( X_i \) = the treatment and pre-treatment covariates
  - \( Z_i \) = the randomized encouragement and pre-treatment covariates

- **Identification**
  1. \( K = L \): just-identified
  2. \( K < L \): over-identified
  3. \( K > L \): under-identified
Geometry of Instrumental Variables

- Projection matrix (onto $S(Z)$): $P_Z = Z(Z^T Z)^{-1} Z^T$
- “Purge” endogeneity: $\hat{X} = P_Z X$
- Since $P_Z = P_Z^T$ and $P_Z P_Z = P_Z$, we have

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

- Two stage least squares:
  1. Regress $X$ on $Z$ and obtain the fitted values $\hat{X}$
  2. Regress $Y$ on $\hat{X}$
- We do not assume the linearity of $X$ in $Z$
Asymptotic Inference

- Estimation error:
  \[ \hat{\beta}_{IV} - \beta = (X^\top Z(Z^\top Z)^{-1}Z^\top X)^{-1}X^\top Z(Z^\top Z)^{-1}Z^\top \epsilon \]
  \[ = \left\{ \left( \frac{1}{n} \sum_{i=1}^{n} x_i z_i^\top \right) \left( \frac{1}{n} \sum_{i=1}^{n} z_i z_i^\top \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^{n} z_i x_i^\top \right) \right\}^{-1} \]
  \[ \times \left( \frac{1}{n} \sum_{i=1}^{n} x_i z_i^\top \right) \left( \frac{1}{n} \sum_{i=1}^{n} z_i z_i^\top \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^{n} z_i \epsilon_i \right) \]

- Thus, \( \hat{\beta}_{IV} \xrightarrow{p} \beta \)
- Under the homoskedasticity, \( \mathbb{V}(\epsilon \mid Z) = \sigma^2 I_n \):
  \[ \sqrt{n}(\hat{\beta}_{IV} - \beta) \xrightarrow{d} \mathcal{N}(0, \sigma^2 [\mathbb{E}(X_i z_i^\top) \{\mathbb{E}(Z_i Z_i^\top)\}^{-1} \mathbb{E}(Z_i X_i^\top)]^{-1}) \]
Residuals and Robust Standard Error

- $\hat{\epsilon} = Y - X\hat{\beta}_{IV}$ and not $\hat{\epsilon} \neq Y - X\hat{\beta}_{IV}$
- Under homoskedasticity: $\hat{\sigma}^2 = \frac{\|\hat{\epsilon}\|^2}{n-K} \xrightarrow{p} \sigma^2$

$$\sqrt{V(\hat{\beta}_{IV})} = \hat{\sigma}^2 \{X^TZ(Z^TZ)^{-1}Z^TX\}^{-1} = \hat{\sigma}^2 (X^T\hat{X})^{-1}$$

- Sandwich heteroskedasticity consistent estimator:

  bread $= \{X^TZ(Z^TZ)^{-1}Z^TX\}^{-1}X^TZ(Z^TZ)^{-1}$

  meat $= Z^T\text{diag}(\hat{\epsilon}_i^2)Z \left( = \sum_{i=1}^{n} \hat{\epsilon}_i^2 Z_iZ_i^T \right)$

  bread meat

bread meat bread$^\top = (X^T\hat{X})^{-1}X^T\text{diag}(\hat{\epsilon}_i^2)\hat{X}(X^T\hat{X})^{-1}$

- Robust standard errors for clustering, auto-correlation, etc.

- Two stage least squares regression:
  \[
  Y_i = \alpha + \beta T_i + \eta_i, \\
  T_i = \delta + \gamma Z_i + \epsilon_i
  \]

- Binary encouragement and binary treatment,
  \[
  \hat{\beta} = \frac{\text{Cov}(Y_i, Z_i)}{\text{Cov}(T_i, Z_i)} \xrightarrow{p} \text{CATE (no covariate)} \\
  \hat{\beta} \xrightarrow{p} \text{CATE (with covariates)}
  \]

- Binary encouragement multi-valued treatment
- Monotonicity: \( T_i(1) \geq T_i(0) \)
- Exclusion restriction: \( Y_i(1, t) = Y_i(0, t) \) for each \( t = 0, 1, \ldots, K \)
Estimator

\[ \hat{\beta}_{TSLS} \xrightarrow{p} \frac{\text{Cov}(Y_i, Z_i)}{\text{Cov}(T_i, Z_i)} = \frac{\mathbb{E}(Y_i(1) - Y_i(0))}{\mathbb{E}(T_i(1) - T_i(0))} \]

\[ = \sum_{k=0}^{K} \sum_{j=k+1}^{K} w_{jk} \mathbb{E} \left( \frac{Y_i(1) - Y_i(0)}{j - k} \left| T_i(1) = j, T_i(0) = k \right. \right) \]

where \( w_{jk} \) is the weight, which sums up to one, defined as,

\[ w_{jk} = \frac{(j - k) \Pr(T_i(1) = j, T_i(0) = k)}{\sum_{k'=0}^{K} \sum_{j'=k'+1}^{K} (j' - k') \Pr(T_i(1) = j', T_i(0) = k')} . \]

Easy interpretation under the constant additive effect assumption for every complier type

Assume encouragement induces at most only one additional dose

Then, \( w_k = \Pr(T_i(1) = k, T_i(0) = k - 1) \)
Quarter of Birth  (Angrist and Krueger. 1991. Q. J. Econ.)

- Instrument for educational attainment to address “ability bias”
  - Outcome: men’s log weekly earnings in 1980
  - Compulsory education law in US: students must attend school until they reach age 16
  - Those born in the third or fourth quarter typically finish tenth grade before reaching age 16
  - Instrument at most decreases years of education by one year

- Weak instrument: first quarter vs. 2nd to 4th quarter
  - 1920s cohorts: est. = −0.126, s.e. (HC) = 0.016, corr = −0.016
  - 1930s cohorts: est. = −0.109, s.e. (HC) = 0.013, corr = −0.014

- Wald estimates:
  - 1920 cohorts: est. = 0.072, s.e. (HC) = 0.022
  - 1930 cohorts: est. = 0.102, s.e. (HC) = 0.024

- OLS estimates:
  - 1920 cohorts: est. = 0.080, s.e. (HC) = 0.0004
  - 1930 cohorts: est. = 0.071, s.e. (HC) = 0.0004
CDFs for First and Fourth Quarter of Birth

Figure 2. Schooling CDF by Quarter of Birth (Men Born 1930–1939; Data From the 1980 Census). Quarter of birth: ——, first; – – –, fourth.
Analysis of Weak Instruments

- Recall the Wald estimator:
  \[ \hat{\beta}_{\text{IV}} = \frac{\text{Cov}(Y_i, Z_i)}{\text{Cov}(T_i, Z_i)} \]

- \( \hat{\beta}_{\text{IV}} \) does not exist if the instrument is irrelevant

- Consider the following model:
  \[
  Y_i = \alpha + \beta T_i + \epsilon_i, \\
  T_i = \gamma Z_i + \eta_i, \quad \text{where } E(\epsilon_i | Z_i) = E(\eta_i | Z_i) = 0
  \]
  where \((\epsilon_i, \eta_i)\) follows a bivariate normal with mean zero. Then,
  \[
  \hat{\beta}_{\text{IV}} - \beta \approx \frac{\sum_{i=1}^{n} \epsilon_i Z_i}{\sum_{i=1}^{n} \eta_i Z_i} \quad \text{d} \quad \text{Corr}(\epsilon_i, \eta_i) \sqrt{\frac{\text{V}(\epsilon_i)}{\text{V}(\eta_i)}} + \frac{W_i}{\text{Cauchy}}
  \]

- Rich literature on asymptotic analysis for weak instruments
Simulated Instruments \cite{Bound1995JASA}

- **Simulation exercise:**
  1. Simulate $Z_i$ from Bernoulli with success probability equal to its empirical estimate
  2. Compute the Wald estimate as before

- **1920s cohorts:**
  - Estimates: min $= -694.718$, 1st Qu. $= -0.093$, median $= 0.0876$, 3rd Qu. $= 0.260$, max $= 36.236$
  - Std. Errors: min $= 0.057$, 1st Qu. $= 0.185$, median $= 0.393$, 3rd Qu. $= 1.467$, max $= 4865.657$

- **1930s cohorts:**
  - Estimates: min $= -36.223$, 1st Qu. $= -0.117$, median $= 0.078$, 3rd Qu. $= 0.284$, max $= 202.667$
  - Std. Errors: min $= 0.064$, 1st Qu. $= 0.197$, median $= 0.421$, 3rd Qu. $= 1.814$, max $= 427582$
Summary

- Instrumental variables as a general strategy for coping with selection bias
  - randomization of instruments
  - monotonicity
  - exclusion restriction

- Extensions to multi-valued treatment

- Weak instruments

Readings:
- IMBENS AND RUBIN. Chapter 25
- ANGRIST AND PISCHKE. Chapter 4