Instrumental Variables

Kosuke Imai

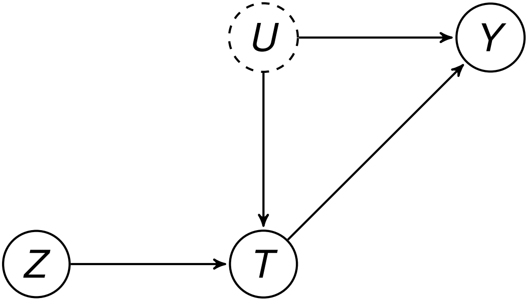
Harvard University

STAT186/GOV2002 CAUSAL INFERENCE

Fall 2018

Instrumental Variables

- From randomized encouragement design to general instrumental variabes approach:



- Instruments in the nature $\rightsquigarrow$ natural experiments
  1. random assignment of Z
  2. no direct effect of Z on Y

Classical Instrumental Variables Estimator

- Linear model (in matrix notation):

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon \quad \text{where } \mathbb{E}(\epsilon) = \mathbf{0}_n \text{ and } \mathbf{X} \text{ is } n \times K$$

- Endogeneity:

$$\mathbb{E}(\epsilon_i \mid \mathbf{X}) \neq 0$$

- Instruments $\mathbf{Z}$ is $n \times L$
  1. Exogeneity: $\mathbb{E}(\epsilon_i \mid \mathbf{Z}) = 0$
  2. Exclusion restriction: Z_i does not belong to the outcome model
  3. Rank condition: $\mathbf{Z}^\top\mathbf{X}$ and $\mathbf{Z}^\top\mathbf{Z}$ have full rank
- Experimental setting:
  - $\mathbf{X}_i =$ the treatment and pre-treatment covariates
  - $\mathbf{Z}_i =$ 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 $\mathcal{S}(\mathbf{Z})$): $\mathbf{P_Z} = \mathbf{Z}(\mathbf{Z}^\top\mathbf{Z})^{-1}\mathbf{Z}^\top$
- "Purge" endogeneity: $\widehat{\mathbf{X}} = \mathbf{P_Z X}$
- Since $\mathbf{P_Z} = \mathbf{P_Z^\top}$ and $\mathbf{P_Z P_Z} = \mathbf{P_Z}$, we have

$$
\begin{aligned}
\hat{\beta}_{\text{IV}} &= (\widehat{\mathbf{X}}^\top\widehat{\mathbf{X}})^{-1}\widehat{\mathbf{X}}^\top\mathbf{Y} \\
&= (\mathbf{X}^\top\mathbf{Z}(\mathbf{Z}^\top\mathbf{Z})^{-1}\mathbf{Z}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Z}(\mathbf{Z}^\top\mathbf{Z})^{-1}\mathbf{Z}^\top\mathbf{Y}
\end{aligned}
$$

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

Asymptotic Inference

- Estimation error:

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_{\mathrm{IV}} - \boldsymbol{\beta} &= (\mathbf{X}^\top \mathbf{Z}(\mathbf{Z}^\top \mathbf{Z})^{-1}\mathbf{Z}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{Z}(\mathbf{Z}^\top \mathbf{Z})^{-1}\mathbf{Z}^\top \boldsymbol{\epsilon} \\
&= \left\{\left(\frac{1}{n}\sum_{i=1}^n \mathbf{X}_i \mathbf{Z}_i^\top\right)\left(\frac{1}{n}\sum_{i=1}^n \mathbf{Z}_i \mathbf{Z}_i^\top\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n \mathbf{Z}_i \mathbf{X}_i^\top\right)\right\}^{-1} \\
&\quad \times \left(\frac{1}{n}\sum_{i=1}^n \mathbf{X}_i \mathbf{Z}_i^\top\right)\left(\frac{1}{n}\sum_{i=1}^n \mathbf{Z}_i \mathbf{Z}_i^\top\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n \mathbf{Z}_i \epsilon_i\right)
\end{aligned}
$$

- Thus, $\hat{\boldsymbol{\beta}}_{\mathrm{IV}} \xrightarrow{p} \boldsymbol{\beta}$
- Under the homoskedasticity, $\mathbb{V}(\boldsymbol{\epsilon} \mid \mathbf{Z}) = \sigma^2 \mathbf{I}_n$:

$$
\sqrt{n}(\hat{\boldsymbol{\beta}}_{\mathrm{IV}} - \boldsymbol{\beta}) \overset{d}{\rightsquigarrow} \mathcal{N}(0,\ \sigma^2[\mathbb{E}(\mathbf{X}_i \mathbf{Z}_i^\top)\{\mathbb{E}(\mathbf{Z}_i \mathbf{Z}_i^\top)\}^{-1}\mathbb{E}(\mathbf{Z}_i \mathbf{X}_i^\top)]^{-1})
$$

Residuals and Robust Standard Error

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

$$\widehat{\mathbb{V}(\hat{\beta}_{\text{IV}})} = \hat{\sigma}^2 \{\mathbf{X}^\top \mathbf{Z}(\mathbf{Z}^\top \mathbf{Z})^{-1}\mathbf{Z}^\top \mathbf{X}\}^{-1} = \hat{\sigma}^2 (\widehat{\mathbf{X}}^\top \widehat{\mathbf{X}})^{-1}$$

- Sandwich heteroskedasticity consistent estimator:

$$\begin{aligned}
\text{bread} &= \{\mathbf{X}^\top \mathbf{Z}(\mathbf{Z}^\top \mathbf{Z})^{-1}\mathbf{Z}^\top \mathbf{X}\}^{-1}\mathbf{X}^\top \mathbf{Z}(\mathbf{Z}^\top \mathbf{Z})^{-1} \\
\text{meat} &= \mathbf{Z}^\top \text{diag}(\hat{\epsilon}_i^2)\mathbf{Z} \left(= \sum_{i=1}^n \hat{\epsilon}_i^2 \mathbf{Z}_i \mathbf{Z}_i^\top \right) \\
\text{bread meat bread}^\top &= (\widehat{\mathbf{X}}^\top \widehat{\mathbf{X}})^{-1} \widehat{\mathbf{X}}^\top \text{diag}(\hat{\epsilon}_i^2) \widehat{\mathbf{X}} (\widehat{\mathbf{X}}^\top \widehat{\mathbf{X}})^{-1}
\end{aligned}$$

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

Multi-valued Treatment (Angirst and Imbens. 1995. *J. Am. Stat. Assoc*)

- Two stage least squares regression:

$$\begin{aligned} Y_i &= \alpha + \beta T_i + \eta_i, \\ T_i &= \delta + \gamma Z_i + \epsilon_i \end{aligned}$$

- Binary encouragement and binary treatment,
  - $\hat{\beta} = \widehat{\text{CATE}}$ (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

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

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

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

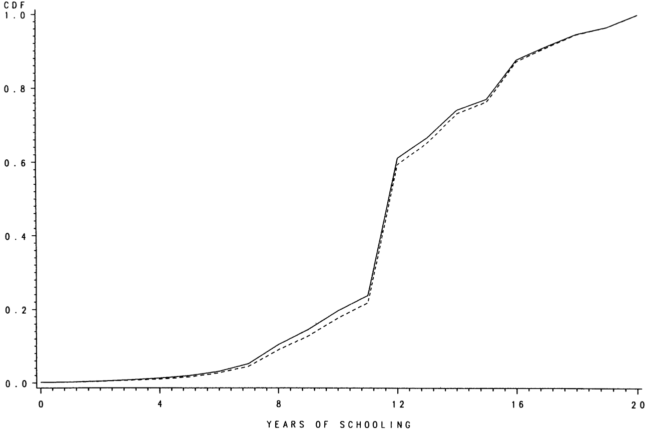


Figure 2. Schooling CDF by Quarter of Birth (Men Born 1930–1939; Data From the 1980 Census). Quarter of birth: ——, first; – – –, fourth.

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:

$$\begin{aligned} Y_i &= \alpha + \beta T_i + \epsilon_i, \\ T_i &= \underbrace{\gamma}_{\approx 0} Z_i + \eta_i, \quad \text{where } \mathbb{E}(\epsilon_i \mid Z_i) \;=\; \mathbb{E}(\eta_i \mid Z_i) \;=\; 0 \end{aligned}$$

where (ϵ_i, η_i) follows a bivariate normal with mean zero. Then,

$$\hat{\beta}_{\text{iv}} - \beta \;\approx\; \frac{\sum_{i=1}^{n} \epsilon_i Z_i}{\sum_{i=1}^{n} \eta_i Z_i} \;\overset{d}{\rightsquigarrow}\; \text{Corr}(\epsilon_i, \eta_i)\sqrt{\frac{\mathbb{V}(\epsilon_i)}{\mathbb{V}(\eta_i)}} + \underbrace{W_i}_{\textit{Cauchy}}$$

- Asymptotic analysis for weak instruments

Simulated Instruments (Bound et al. 1995. *J. Am. Stat. Assoc*)

- 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 $= 427582$

Randomization Inference (Imbens and Rosenbaum. 2005. *J. R. Stat. Soc. A.*)

- Constant additive treatment effect model for the QoB example:

$$Y_i(t) \;=\; Y_i(0) + \beta \cdot t \quad \text{for } t = 0, 1, \ldots$$

- Randomization test:
  1. Null hypothesis: $H_0 : \beta = \beta_0$
  2. Test statistic: $S_i = f(Y_i - \beta_0 T_i, Z_i)$
  3. Assume $Z_i \sim \text{Bernoulli}(\overline{Z}_n)$ to obtain the reference distribution
- Application:
  - $S_i \;=\; \sum_{i=1}^{N} Z_i \cdot \text{rank}(Y_i - \beta_0 T_i)$
  - 95% confidence intervals:
    - 1920 cohorts: $[0.036, 0.106]$, $[0.028, 0.115]$ (Wald)
    - 1930 cohorts: $[0.049, 0.122]$, $[0.055, 0.149]$ (Wald)
  - Simulation (rejection rates of 0.05 level tests with 1000 simulations):
    - 1920 cohorts: 0.048, 0.001 (Wald)
    - 1930 cohorts: 0.051, 0.004 (Wald)

Violations of IV Assumptions

1. Violation of exclusion restriction:

$$\text{bias} \quad = \quad \text{ITT}_{\text{noncomplier}} \times \frac{\Pr(\text{noncomplier})}{\Pr(\text{complier})}$$

   - Weak encouragement (instruments)
   - Direct effects of encouragement; failure of randomization, alternative causal paths

2. Violation of monotonicity:

$$\text{bias} \quad = \quad \frac{\{\text{CATE} + \text{ITT}_{\text{defier}}\}\Pr(\text{defier})}{\Pr(\text{complier}) - \Pr(\text{defier})}$$

   - Proportion of defiers
   - Heterogeneity of causal effects

Bounding the Average Treatment Effect

(Manski. (1990). *Am. Econ. Rev.*)

- Instrumental variable estimator does not point-identify the ATE
- Partial identification (Manski. 1995. *Identification Problems in the Social Sciences*. Harvard UP)
- Consider a binary outcome with the randomized encouragement and exclusion restriction:

$$\begin{aligned}
\Pr(Y_i(1)=1) &= \Pr(Y_i(1)=1 \mid Z_i=1) \\
&= \mu_{11}\pi_1 + \Pr(Y_i(1)=1 \mid D_i=0, Z_i=1)(1-\pi_1) \\
\Pr(Y_i(0)=1) &= \mu_{00}(1-\pi_0) + \Pr(Y_i(0)=1 \mid D_i=1, Z_i=0)\pi_0
\end{aligned}$$

  where $\mu_{dz} = \Pr(Y_i=1 \mid D_i=d, Z_i=z)$, $\pi_z = \Pr(D_i=1 \mid Z_i=z)$
- Bounds on the ATE:

$$\mu_{11}\pi_1 - \mu_{00}(1-\pi_0) - \pi_0 \ \leq\ \tau\ \leq\ \mu_{11}\pi_1 - \mu_{00}(1-\pi_0) + 1 - \pi_1$$

  where the width equals $1-(\pi_1-\pi_0)$

Sharp Bounds (Balke and Pearl. 1997. *J. Am. Stat. Assoc*)

- The previous bounds are not sharp:
  - only consider four latent types based on mapping from Z to D
  - there are four additional mappings from D to Y
  - a total of 16 latent types: $(D_i(1), D_i(0), Y_i(1), Y_i(0))$
  - equal to Manki's bounds under monotonicity
- Linear programming problem:

$$\text{maximize/minimize} \sum_u \Pr(Y_i(d) = 1 \mid U_i = u) \Pr(U_i = u)$$

subject to

$$\begin{aligned} & \Pr(Y_i = y, D_i = d \mid Z_i = z) \\ = \;& \Pr(Y_i(d) = y \mid D_i = d, U_i = u) \Pr(D_i = d \mid Z_i = z, U_i = u) \\ & \quad \Pr(Z_i = z) \Pr(U_i = u) \end{aligned}$$

- A general strategy for a discrete potential outcome case

Revisiting the Habitual Voting Example

- Effect of voting in 2006 election on the turnout in the 2008 election: est $= 0.128$, s.e. $= 0.022$
- Potential bias of estimated CATE due to exclusion restriction:

$$\text{ITT}_{noncomplier} \times \frac{1 - 0.083}{0.083} \;=\; 11.05 \times \text{ITT}_{noncomplier}$$

- Inference for the Average Treatment Effect
  - exclusion restriction + monotonicity: $[-0.315,\ 0.602]$
  - exclusion restriction alone: $[-0.315,\ 0.602]$

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 and randomization inference
- ATE vs. CATE $\rightsquigarrow$ partial identification, method of bounds

- Suggested readings:
    - IMBENS AND RUBIN. Chapter 25
    - ANGRIST AND PISCHKE. Chapter 4