Identification of Causal Effects Using Instrumental Variables

Stat 232 Presentation

Mingxun Wang   Mengyao Huang

April 21, 2023

Outline

- Introduction
- Structural Equation Models
- Causal estimands with IV
- Sensitivity of the IV assumptions
- The effect of military service on civilian mortality

Introduction

- "Treatment assignment" does not necessarily equal to "being treated"
- Motivate us to extend the potential outcomes:
  - Experimental assignment of subject i: z_i
    $z_i = 1$: subject is assigned to the treatment group
    $z_i = 0$: subject is assigned to the control group
  - Actual treatment of subject i: $d_i(z)$
  - Compliers: $d_i(z = 1) = 1, d_i(z = 0) = 0$
  - Never Takers: $d_i(z = 0) = 0, d_i(z = 0) = 0$
  - Always Takers: $d_i(z = 1) = 1, d_i(z = 0) = 1$
  - Defiers: $d_i(z = 1) = 0, d_i(z = 0) = 1$
- For our own convenience, just denote them as C, NT, AT and D.

Introduction

- Intent to treat of z_i on d_i and Intent to treat of z_i on Y_i

$$
\begin{aligned}
ITT_{i,D} &\equiv d_i(1) - d_i(0) \\
ITT_D &\equiv E[d_i(1)] - E[d_i(0)] \\
ITT_{i,Y} &\equiv Y_i(z=1, d(1)) - Y_i(z=0, d(0)) \\
ITT_Y &\equiv \frac{1}{N}\sum_{i=1}^{N}[Y_i(z=1, d(1)) - Y_i(z=0, d(0))] \\
&= E[Y_i(z=1, d(1))] - E[Y_i(z=0, d(0))] \\
&= E[ITT_{i,Y}]
\end{aligned}
$$

- For experiments with 100% compliance, treatment assignment is the same as treatment status, so the ITT_Y is the same as the ATE.

Structural Equation Models

- Structural equation model is an alternative framework to potential outcomes for thinking about causation. The other one is Prof. Pearl's DAG framework.
- Stochastic models in which each equation represents a causal link, rather than a mere empirical association.

$$Y_i = \beta_0 + \beta_1 D_i + \epsilon_i$$
$$D_i^* = \alpha_0 + \alpha_1 Z_i + v_i$$

$$D_i = \begin{cases} 1, & D_i^* > 0, \\ 0, & D_i^* \leq 0. \end{cases}$$

- Assumptions (instrument variable):

$$E[Z_i \epsilon_i] = 0, E[Z_i v_i] = 0$$
$$Cov(D_i, Z_i) \neq 0$$

IV estimator

- Although, we are just briefly mentioning it, but it's important so we give it a separate page.

$$
\begin{aligned}
\hat{\beta}^{IV} &= \frac{\hat{Cov}(Y_i, Z_i)}{\hat{Cov}(D_i, Z_i)} \\
&= \frac{\sum_i Y_i Z_i / \sum_i Z_i - \sum Y_i(1 - Z_i) / \sum_i (1 - Z_i)}{\sum_i D_i Z_i / \sum Z_i - \sum_i D_i(1 - Z_i) / \sum_i (1 - Z_i)}
\end{aligned}
$$

Wold estimator $\hat{\beta}_{IV} = \dfrac{E[Y_i|Z_i=1] - E[Y_i|Z_i=0]}{E[D_i|Z_i=1] - E[D_i|Z_i=0]}$ $\qquad \hat{\beta}_{IV} = \dfrac{Cov(Z,Y)}{Cov(X,Z)}$

$$\begin{tikzcd} Z \arrow[dr] & U \arrow[d] \arrow[dr] & \\ & D \arrow[r, shift left] & Y \arrow[l, shift left] \end{tikzcd}$$

$$= \frac{E[Y_i|Z_i=1] - E[Y_i|Z_i=0]}{P(D_i=1|Z_i=1) - P(D_i=1|Z_i=0)}$$

We can further show that IV can be obtained by 2SLS.

Assume $Y = \beta_0 + \beta_1 X_i + u_i$

$X_i = \pi_0 + \pi_1 Z_i + v_i$

Rule of thumb: First stage F-statistic >10

Two steps : ① Regress X on Z

② Regress Y on $\hat{X}$

$$\hat{\pi}_1 = \frac{Cov(X,Z)}{Var(Z)}$$

$$\hat{\beta}_{2SLS} = \frac{Cov(Y,\hat{X})}{Var(\hat{X})}$$

From the formula, we can use reduced form / first stage.

$$= \frac{Cov(Y, \hat{\pi}_0 + \hat{\pi}_1 Z_i)}{Var(\hat{\pi}_0 + \hat{\pi}_1 Z_i)}$$

$$= \frac{\hat{\pi}_1 Cov(Y,Z)}{\hat{\pi}_1^2 Var(Z)}$$

$$= \frac{Cov(Y,Z)}{\hat{\pi}_1 Var(Z)}$$

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

$$= \hat{\beta}_{IV}$$

Crucial Assumptions in Causal Inference

- SUTVA
  If $Z_i = Z_i^{'}$, then $D_i(\boldsymbol{Z}) = D_i(\boldsymbol{Z'})$
  If $Z_i = Z_i^{'}$ and $D_i = D_i^{'}$, then $Y_i(\boldsymbol{Z}, \boldsymbol{D}) = Y_i(\boldsymbol{Z}', \boldsymbol{D}')$
  Potential outcomes for each person i are unrelated to the treatment status of other individuals.
- Random Assignment

$$P(\boldsymbol{Z} = \boldsymbol{c}) = P(\boldsymbol{Z} = \boldsymbol{c}') \ \forall \boldsymbol{c}, \boldsymbol{c}' \ \ s.t. \ 1^T \boldsymbol{c} = 1^T \boldsymbol{c}'$$

Sample Analogue of ITT

- Given the SUTVA and random assignment, we can obtain the unbiased estimate for the ITTs.

$$
\begin{aligned}
IT\hat{T}_Y &= \frac{\sum_i Y_i Z_i}{\sum_i Z_i} - \frac{\sum_i Y_i(1-Z_i)}{\sum_i (1-Z_i)} \\
&= \frac{(1/N)\sum_i Y_i Z_i - (1/N)\sum_i Y_i (1/N)\sum_i Z_i}{(1/N)\sum_i Z_i Z_i - (1/N)\sum_i Z_i (1/N)\sum_i Z_i}
\end{aligned}
$$

$$
\begin{aligned}
IT\hat{T}_D &= \frac{\sum_i D_i Z_i}{\sum_i Z_i} - \frac{\sum_i D_i(1-Z_i)}{\sum_i (1-Z_i)} \\
&= \frac{(1/N)\sum_i D_i Z_i - (1/N)\sum_i Y_i (1/N)\sum_i Z_i}{(1/N)\sum_i Z_i Z_i - (1/N)\sum_i Z_i (1/N)\sum_i Z_i}
\end{aligned}
$$

- The ratio of the two ITT (ITT_Y / ITT_D) equals to the conventional instrumental variable estimator in the previous slide.

Crucial Assumptions in NON-COMPLIANCE

- Exclusion Restriction
  - $\boldsymbol{Y}(\boldsymbol{Z}, \boldsymbol{D}) = \boldsymbol{Y}(\boldsymbol{Z}', \boldsymbol{D})$ for all $\boldsymbol{Z}$,$\boldsymbol{Z}'$ and for all $\boldsymbol{D}$
  - The assumption implies that $Y_i(1, d) = Y_i(0, d)$ for $d = 0, 1$
  - Any effect of Z on Y must be via an effect of Z on D
- Using the assumption, we can define our new potential outcomes as:

$$\boldsymbol{Y}(\boldsymbol{Z}, \boldsymbol{D}) = \boldsymbol{Y}(\boldsymbol{Z}', \boldsymbol{D}) \;\; \forall \boldsymbol{Z}', \boldsymbol{Z}, \boldsymbol{D}$$

- By the SUTVA assumption, we can further reduce it to:

$$Y_i(D_i) = Y_i(\boldsymbol{Z}, \boldsymbol{D})$$

- Since other subjects' treatment will not affect this subject

Crucial Assumptions in NON-COMPLIANCE

- Relevance
  - The average causal effect of $\boldsymbol{Z}$ on D, $E[D_i(1) - D_i(0)] \neq 0$
- Monotonicity ("No defiers")
  - $D_i(1) \geq D_i(0)$ for all $i = 1, \ldots, N$

CACE or LATE

- With all these definitions and assumptions, we finally arrived at the most important estimand. The Complier Average Causal Effect or Local Average Treatment Effect as mentioned in Angrist et.al.(1996)

$$\begin{aligned} E[Y_i(1) - Y_i(0)|i : C] &= E[Y_i(1) - Y_i(0)|D_i(1) - D_i(0) = 1] \\ &= \frac{E[Y_i(D_i(1), 1) - Y_i(D_i(0), 0)]}{E[D_i(1) - D_i(0)]} \end{aligned}$$

- Local in the sense that the samples are restricted to compliers only

Proof of the CACE

$$
\begin{aligned}
ITT &= E[Y_i(z=1, d(1))] - E[Y_i(z=0, d(0))] \\
&= E[Y_i(d(1))] - E[Y_i(d(0))]
\end{aligned}
$$

- We can express the ITT as a weighted average of the Never-Takers, Always-Takers, Compliers, and Defiers. We denote them to be $\pi_{NT}, \pi_{AT}, \pi_C, \pi_D$.

$$
\begin{aligned}
ITT = {} & E[Y_i(d(1)) - Y_i(d(0)) | d_i(1) = d_i(0) = 0]\pi_{NT} \\
& + E[Y_i(d(1)) - Y_i(d(0)) | d_i(1) = d_i(0) = 1]\pi_{AT} \\
& + E[Y_i(d(1)) - Y_i(d(0)) | d_i(1) - d_i(0) = 1]\pi_C \\
& + E[Y_i(d(1)) - Y_i(d(0)) | d_i(1) - d_i(0) = -1]\pi_D
\end{aligned}
$$

Proof of the CACE: continue

- By applying the exclusion restriction, we can kick out the always-takers and never-takers. By applying the monotonicity, we rule out the defiers.

$$
\begin{aligned}
ITT_D &= E[d(1) - d(0)|d_i(1) - d_i(0) = 1]\pi_C = \pi_C \\
ITT_Y &= E[Y_i(d(1)) - Y_i(d(0))|d_i(1) - d_i(0) = 1]\pi_C \\
CACE &= \frac{ITT_Y}{ITT_D} = E[Y_i(d(1)) - Y_i(d(0))|d_i(1) - d_i(0) = 1]
\end{aligned}
$$

Mayoral Debate Example

- Contacted 1000 subjects by phone who were successfully interviewed a few days later.
- Treatment: approximately half of the subjects were encouraged to watch the debate. Control: watch a non-political show.
- After the debate,respondents were called back.
- All subjects were asked whether they had watched the debate and whether their views of the candidates had changed

Mayoral Debate Example

- Given the table here, can anyone try to estimate the CACE?

TABLE 6.2

Percent reporting changed views of one or more candidates since watching the New York City mayoral debate

	Treatment group	Control group
% Reporting change (N treated)	59.5 (185)	50.0 (80)
% Reporting change (N untreated)	40.6 (320)	40.2 (415)
% Reporting change (total N)	47.5 (505)	41.8 (495)

Source: Mullainathan, Washington, and Azari 2010.

r

Question

- It's very IMPORTANT, so let's stare at the formula for a while. This is the first time we change the estimand from ATE to something else.

$$CACE = \frac{ITT_Y}{ITT_D} = E[Y_i(d(1)) - Y_i(d(0))|d_i(1) - d_i(0) = 1]$$

- What if the compliers are not representative enough? Can we still use it as a reliable estimator?

Sensitivity of the IV Estimand

Deviation from IV assumptions:

- Suppose i is a non-complier, $D_i(0) = D_i(1)$,
  The causal effect of Z on Y is $H_i = Y_i(1, d) - Y_i(0, d)$,
- $d = 0$ if subject i is a never taker
- $d = 1$ if subject i an always-taker.

Proposition:
If Exclusion Restriction fails for the non-compliers only, i.e., $H_i \neq 0$ then IV estimand equals LATE plus a bias term given by:

$$
\begin{aligned}
Bias &= \frac{E[Y_i(1, D_i(1)) - Y_i(0, D_i(0))]}{E[D_i(1) - D_i(0)]} \\
&\quad - E[Y_i(1, D_i(1)) - Y_i(0, D_i(0))|i : C] \\
&= E[H_i|i : NC]\frac{P(i : NC)}{P(i : C)}
\end{aligned}
$$

- If there's an effect of assignment on the outcome for noncompliers, it's plausible that there is an effect of assignment on outcome for compliers, therefore:
- Suppose assignment and treatment had additive effects on Y: $Y_i(1,0) - Y_i(0,0) = Y_i(1,1) - Y_i(0,1)$ for all compliers
- Define the causal effect of Z on Y for compliers as $H_i = Y_i(1,d) - Y_i(0,d)$ for $d = 0,1$.
- Define the causal effect of D on Y as $G_i = Y_i(z,1) - Y_i(z,0)$

$$\begin{aligned}
IV &= \frac{E[Y_i(1, D_i(1)) - Y_i(0, D_i(0))]}{E[D_i(1) - D_i(0)]} \\
&= E[G_i | i : C] + \frac{E[H_i]}{P(i : C)} (Bias)(Additivity) \\
&= E[G_i | i : C] + E[H_i | i : C] + E[H_i | i : NC] \frac{P(i : NC)}{P(i : C)}
\end{aligned}$$

Sensitivity of the IV Estimand

- Remember that $H_i = Y_i(1, d) - Y_i(0, d)$
- The second term is due to the direct effect of assignment for those who take the treatment. If compliance perfect, it is 0.
- The third term is directly proportional to the product of the average size of the direct effect of Z for noncompliers and the odds of noncompliance given monotonicity.
- "Stronger" the instrument, the smaller the odds of noncompliance, and less sensitive the IV estimand to the violations of exclusion assumption.

Sensitivity of the IV Estimand

- Violations of the Monotonicity
  Proposition: If Monotonicity fails, then IV estimand equals LATE plus a bias term given by:

$$
\begin{aligned}
Bias = & \frac{E[Y_i(1, D_i(1)) - Y_i(0, D_i(0))]}{E[D_i(1) - D_i(0)]} \\
& - E[Y_i(1, D_i(1)) - Y_i(0, D_i(0))|i : C] \\
& = -\lambda \times E[Y_i(1) - Y_i(0)|i : D] \\
& - E[Y_i(1) - Y_i(0)|i : C] \\
\text{where} \quad \lambda = & \frac{P(i : D)}{P(i : C) - P(i : D)}
\end{aligned}
$$

- λ is related to the proportion of defiers and is 0 under monotonicity.
- Smaller the proportion, smaller the bias. Note that even with a few defiers, the bias can be large

Sensitivity of the IV Estimand

- Second factor is the difference in causal effects of D on Y for the compliers and defiers.
- If the average causal effects of D on Y are identical for defiers and compliers, there will be no bias eventually.
- Without monotonicity, alternatively,

$$IV = (1+\lambda)E[Y_i(1) - Y_i(0)|i : C] - \lambda E[Y_i(1) - Y_i(0)|i : D]$$

- The estimand is still a weighted average of ATE despite the violation of the monotonicity.

Application: Vietnam Era Draft Lottery

- Background Information
  The Effect of Military Service on Civilian Mortality
- Each date of birth in the cohort at risk of being drafted was assigned a random sequence number (RSN) from 1-365
- Men born in 1950,1951,1952 up to RSN 195,125,95 in 1970,1971,1972; the Selective Service called men for induction by RSN up to a ceiling decided by the department, i.e., priority for conscription was allocated to those low RSN, i.e., ≤ 195 in 1970. (based on 1970-1973)
- $Y_i(z, d)$: Indicator variable equal to one if i would have died between 1974 and 1983
- z: Lottery assignment
- d: Military service indicator

Application: Vietnam Era Draft Lottery

- Assessment of Assumptions 1-5
- SUTVA: Veteran status of any man at risk of being drafted in the lottery was not affected by the draft status of others at risk of being drafted. The civilian mortality of any such man was not affected by the draft status of others.
- Random Assignment
- Exclusion Restriction: Civilian mortality risk was not affected by draft status once veteran status is taken into account.
- Relevance: Having a low lottery number increases the average probability of service.
- Monotonicity: No one would have served if given a high lottery number, but not if given a low lottery number.
- Question: Do you think the assumptions are plausible here?

Application: Vietnam Era Draft Lottery

Table 2. Data on Civilian Mortality for White Men Born in 1950 and 1951

Year	Draft eligibility[a]	Number of deaths[b]	Number of suicides[c]	Probability of death[d]	Probability of suicide	Probability of military service[e]
① 1950	Yes $Z_i = 1$	2,601	436	$\boxed{.0204}$	.0034	.3527 actually served
				(.0004)	(.0002)	(.0325)
	No $Z_i = 0$	2,169	352	$\boxed{.0195}$	.0032	.1934
				(.0004)	(.0002)	(.0233)
Difference (Yes minus No)				$\boxed{.0009}$	.0002	$\boxed{.1593}$
			ITT_Y	(.0006)	(.0002)	(.0401) ITT_D
IV estimates[f]				.0056	.0013	
				(.0040)	(.0013)	
1951	Yes	1,494	279	.0170	.0032	.2831
				(.0004)	(.0002)	(.0390)
	No	2,823	480	.0168	.0029	.1468
				(.0003)	(.0001)	(.0180)
Difference (Yes minus No)				.0002	.0003	.1362
				(.0005)	(.0002)	(.0429)
IV estimates				.0015	.0022	
				(.0037)	(.0016)	

[a] Determined by lottery number cutoff: RSN 195 for men born in 1950, and RSN 125 for men born in 1951.
[b] From California and Pennsylvania administrative records, all deaths 1974–1983. Data sources and methods documented by Hearst et al. (1986). Note: Sample sizes differ from Hearst et al., because non-U.S.-born are included to match SIPP data in the last column.
[c] The mortality figures are tabulated from the data set analyzed by Hearst et al. (1986).
[d] The estimated population at risk is from the author's tabulation of 1970 census data. Estimates by draft-eligibility status are computed assuming a uniform distribution of lottery numbers. Standard errors are given in parentheses.
[e] These figures are taken from Angrist (1990), table 2, and were tabulated using a special version of the SIPP that has been matched to indicators of draft eligibility. Note that probabilities estimated using the SIPP are for the entire country and do not take account of morality. The impact of mortality on differences in the probability of being a veteran by eligibility status is small enough to have only trivial consequences for the estimation.
[f] The standard errors, following econometric practice (e.g. Imbens and Angrist 1994), were calculated based on a normal approximation to the sampling distribution of the ratio of the difference in estimated probability of death/suicide and the difference in estimated probability of serving. We assume independence of numerator and denominator because they were calculated from different data sets. Pooled estimates show a statistically significant increase in risk at conventional significance levels (e.g., Hearst, Newman, Hulley 1986).

Some Statistics

- Prob of service (1950): 35.27% vs. 19.34% served in the military w.r.t $Z_i = 1/0$. Therefore random assignment of draft status had causal effects on the probability of serving by 15.9% on average.
- Prob of death (1950): 2.04% vs. 1.95% w.r.t $Z_i = 1/0$. An estimate of average causal effect of draft on civilian mortality is 0.09%. Then the causal effect of military service on mortality for that 15.9% is $\frac{0.09\%}{15.9\%} = 0.56\%$.

Application: Vietnam Era Draft Lottery

- Sensitivity to the Exclusion Restriction (Civilian mortality risk was not affected by draft status once veteran status is taken into account.)
- Low lottery numbers were more likely to stay in school
- College students were exempt from the draft
- Angrist and Krueger (1992b) showed that men born in 1951 with lottery 1-75 had completed 0.358 more years of schooling than men with numbers above 150

Application: Vietnam Era Draft Lottery

- Evidence: Married white men 25 years old with 1-3 years of college have mortality rates roughly 0.0017 per thousand higher than men with only high school degrees.
- Assume linearity, an additional year of schooling raises mortality by $0.0017 \times (1/3) = 0.00056$
- Estimate of the mortality difference attributable to the effect of draft status on schooling is $0.358 \times 0.00056 = 0.00019$ (1951)
- Bias$=E[H_i]/E[D_i(1) - D_i(0)] = 0.00019/0.1362 = 0.0014$
- 0.1362 comes from difference in the probability being a veteran by draft eligibility status
- Taking account of this potential bias, 0.0015 impact of veteran status on civilian mortality eliminated....

Application: Vietnam Era Draft Lottery

- No evidence of schooling-lottery connection for the 1950 cohort, lottery-based estimates of the effects of service are even larger for men born in 1950 than for the 1951 cohort used in the illustration
- Schooling mortality connection is not well determined
- Thus, a calculation based solely on graduates would indicate no bias

Application: Vietnam Era Draft Lottery

- Sensitivity to the Monotonicity Assumption
- Without monotonicity, ITT_D estimates the difference between the proportions of compliers and defiers. Therefore the table suggests that 15.93% more people are compliers than defiers.
- Suppose 5% of the population are defiers
- Imply that about 21% of the population are compliers
- $\lambda = P(i : D)/(P(i : C) - P(i : D)) = 0.05/0.16 = 0.3125 \approx 0.33$
- Assume the difference between ATE for compliers and defiers is at most 0.0041, then the estimated ATE for compliers could be $0.0056 - 0.33 \times 0.0041 = 0.0042$ or as large as $0.0056 + 0.33 \times 0.0041 = 0.007$
- To reverse the sign of the average causal effect though violations of monotonicity:large group of defiers or large difference between compliers and defiers.

Reference

- Angrist, J. D., Imbens, G., ; Rubin, D. B. (1996). Identification of causal effects using instrumental variables.
- Gerber, A. S., ; Green, D. P. (2012). Field experiments: Design, analysis, and interpretation.
- Angrist, J. D., ; Pischke Jorn-Steffen. (2013). Mostly harmless econometrics: An empiricists companion.