Section: Module 5

Instrumental Variable

Kentaro Nakamura

GOV 2002

October 10th, 2025

Agenda

- Review of Instrumental Variable
- Estimating Complier for multi-valued treatment
  - Point Estimate / Standard Error (Question 3)
- Weak Instrument

Review of Instrumental Variable

- Setup
  - Z_i: Instrument / Encouragement (randomized)
  - T_i: Treatment (not randomized!)
  - Y_i: Outcome

- Assumptions
  - Randomization of instrument
  - Exclusion restriction (Z_i influences outcome only through T_i)
  - Monotonicity (there is no defiers)

- Check review section's slide for identification

Complier Type (Binary Treatment)

- When treatment is binary, we have four types
  - Compliers: $T_i(Z_i = 1) = 1$ and $T_i(Z_i = 0) = 0$
  - Always-takers: $T_i(Z_i = 1) = T_i(Z_i = 0) = 1$
  - Never-takers: $T_i(Z_i = 1) = T_i(Z_i = 0) = 0$
  - Defiers: $T_i(Z_i = 1) = 0$ and $T_i(Z_i = 0) = 1$

	$Z_i = 1$	$Z_i = 0$
$T_i = 1$	Complier / Always-taker	Defier / Always-taker
$T_i = 0$	Defier / Never-taker	Complier / Never-taker

- We exclude defier by monotonicity
  - As a result, we can identify each principal strata

Estimating Complier for multi-valued treatment (1)

- What if we have multi-valued treatment
  - This is the setting of Question 3

- Consider the case where treatment is three category
  - I.e., $T_i \in \{0, 1, 2\}$
  - We keep instrument binary: $Z_i \in \{0, 1\}$

- How many principal strata do we have?

Estimating Complier for multi-valued treatment (2)

- We have 9 principal strata
  - $\{T_i(Z_i = 0) = 0, T_i(Z_i = 1) = 0\}$
  - $\{T_i(Z_i = 0) = 0, T_i(Z_i = 1) = 1\}$
  - $\{T_i(Z_i = 0) = 0, T_i(Z_i = 1) = 2\}$
  - $\{T_i(Z_i = 0) = 1, T_i(Z_i = 1) = 0\}$
  - $\{T_i(Z_i = 0) = 1, T_i(Z_i = 1) = 1\}$
  - $\{T_i(Z_i = 0) = 1, T_i(Z_i = 1) = 2\}$
  - $\{T_i(Z_i = 0) = 2, T_i(Z_i = 1) = 0\}$
  - $\{T_i(Z_i = 0) = 2, T_i(Z_i = 1) = 1\}$
  - $\{T_i(Z_i = 0) = 2, T_i(Z_i = 1) = 2\}$
- We need to remove some of principal strata to identify the probability

Estimating Complier for multi-valued treatment (2)

- Monotonicity $\rightarrow$ we can remove strata $T_i(Z_i = 0) > T_i(Z_i = 1)$

- We have 6 principal strata (remove 3 strata)
  - $\{T_i(Z_i = 0) = 0, T_i(Z_i = 1) = 0\}$
  - $\{T_i(Z_i = 0) = 0, T_i(Z_i = 1) = 1\}$
  - $\{T_i(Z_i = 0) = 0, T_i(Z_i = 1) = 2\}$
  - ~~$\{T_i(Z_i = 0) = 1, T_i(Z_i = 1) = 0\}$~~
  - $\{T_i(Z_i = 0) = 1, T_i(Z_i = 1) = 1\}$
  - $\{T_i(Z_i = 0) = 1, T_i(Z_i = 1) = 2\}$
  - ~~$\{T_i(Z_i = 0) = 2, T_i(Z_i = 1) = 0\}$~~
  - ~~$\{T_i(Z_i = 0) = 2, T_i(Z_i = 1) = 1\}$~~
  - $\{T_i(Z_i = 0) = 2, T_i(Z_i = 1) = 2\}$

Estimating Complier for multi-valued treatment (3)

- Another Example: $T_i(1) - T_i(0) \in \{-1, 0\}$
  - i.e., difference is at most one

- We have 5 principal strata (remove 4 strata)
  - $\{T_i(Z_i = 0) = 0, T_i(Z_i = 1) = 0\}$
  - ~~$\{T_i(Z_i = 0) = 0, T_i(Z_i = 1) = 1\}$~~
  - ~~$\{T_i(Z_i = 0) = 0, T_i(Z_i = 1) = 2\}$~~
  - $\{T_i(Z_i = 0) = 1, T_i(Z_i = 1) = 0\}$
  - $\{T_i(Z_i = 0) = 1, T_i(Z_i = 1) = 1\}$
  - ~~$\{T_i(Z_i = 0) = 1, T_i(Z_i = 1) = 2\}$~~
  - ~~$\{T_i(Z_i = 0) = 2, T_i(Z_i = 1) = 0\}$~~
  - $\{T_i(Z_i = 0) = 2, T_i(Z_i = 1) = 1\}$
  - $\{T_i(Z_i = 0) = 2, T_i(Z_i = 1) = 2\}$

Estimating Complier for multi-valued treatment (4)

- Can we estimate the proportion of each strata with standard errors?
- Because Z_i is randomized, for $k \in \{0, 1, 2\}$

$$\mathbb{P}(T_i(1) = k) = \mathbb{P}(T_i(1) = k \mid Z_i = 1) = \mathbb{P}(T_i = k \mid Z_i = 1)$$
$$\mathbb{P}(T_i(0) = k) = \mathbb{P}(T_i(0) = k \mid Z_i = 0) = \mathbb{P}(T_i = k \mid Z_i = 0)$$

- Moreover, because sum of probability is 1,

$$\begin{aligned} \mathbb{P}(T_i(1) = 2 \mid T_i(0) = 2) + \mathbb{P}(T_i(1) = 1 \mid T_i(0) = 2) &= 1 \\ \mathbb{P}(T_i(1) = 1 \mid T_i(0) = 1) + \mathbb{P}(T_i(1) = 0 \mid T_i(0) = 1) &= 1 \\ \mathbb{P}(T_i(1) = 0 \mid T_i(0) = 0) &= 1 \end{aligned}$$

Estimating Complier for multi-valued treatment (5)

- Now, notice that by the definition of conditional probability,

$$1 = \mathbb{P}(T_i(1) = 0 \mid T_i(0) = 0) = \frac{\mathbb{P}(T_i(1) = 0, T_i(0) = 0)}{\mathbb{P}(T_i(0) = 0)}$$

- This means that

$$\mathbb{P}(T_i(0) = 0) = \mathbb{P}(T_i(1) = 0, T_i(0) = 0)$$

- Hence,

$$\mathbb{P}(T = 0 \mid Z = 0) = \mathbb{P}(T_i(1) = 0, T_i(0) = 0)$$

Estimating Complier for multi-valued treatment (6)

- Now, notice that

$$\mathbb{P}(T_i(1) = 0) = \mathbb{P}(T_i(0) = 1, T_i(1) = 0) + \mathbb{P}(T_i(0) = 0, T_i(1) = 0)$$

- We know $\mathbb{P}(T_i(1) = 0) = \mathbb{P}(T_i = 0 \mid Z_i = 1)$
- We also identify $\mathbb{P}(T_i(0) = 0, T_i(1) = 0) = \mathbb{P}(T = 0 \mid Z = 0)$ from previous step
- Therefore,

$$\mathbb{P}(T_i(0) = 1, T_i(1) = 0) = \mathbb{P}(T_i = 0 \mid Z_i = 1) - \mathbb{P}(T_i = 0 \mid Z_i = 0)$$

Estimating Complier for multi-valued treatment (7)

- What about $\mathbb{P}(T_i(0) = 1, T_i(1) = 1)$?

$$\mathbb{P}(T_i(0) = 1, T_i(1) = 1) = \mathbb{P}(T_i(0) = 1) - \mathbb{P}(T_i(0) = 1, T_i(1) = 0)$$

- As we identify $\mathbb{P}(T_i(0) = 1, T_i(1) = 0)$ already,

$$\begin{aligned}
&\mathbb{P}(T_i(0) = 1, T_i(1) = 1) \\
&= \mathbb{P}(T_i = 1 \mid Z_i = 0) - \big(\mathbb{P}(T_i = 0 \mid Z_i = 1) - \mathbb{P}(T_i = 0 \mid Z_i = 0)\big)
\end{aligned}$$

- We simply repeat the same thing to identify all the strata

Estimating Complier for multi-valued treatment (8)

- Can you estimate the standard error?

$$\widehat{\mathbb{P}(T_i(0) = 1, T_i(1) = 0)} = \widehat{\mathbb{P}(T_i = 0 \mid Z_i = 1)} - \widehat{\mathbb{P}(T_i = 0 \mid Z_i = 0)}$$

- Recall that $\mathbb{V}[X - Y] = \mathbb{V}[X] + \mathbb{V}[Y] - 2\mathrm{Cov}(X, Y)$
- Hence,

$$\begin{aligned}
&\mathbb{V}\Big(\widehat{\mathbb{P}(T_i(0) = 1,\ T_i(1) = 0)}\Big) \\
&= \mathbb{V}\Big(\widehat{\mathbb{P}(T_i = 0 \mid Z_i = 1)} - \widehat{\mathbb{P}(T_i = 0 \mid Z_i = 0)}\Big) \\
&= \mathbb{V}\Big(\widehat{\mathbb{P}(T_i = 0 \mid Z_i = 1)}\Big) + \mathbb{V}\Big(\widehat{\mathbb{P}(T_i = 0 \mid Z_i = 0)}\Big) \\
&\quad - 2\ \underbrace{\mathrm{Cov}\Big(\widehat{\mathbb{P}(T_i = 0 \mid Z_i = 1)}, \widehat{\mathbb{P}(T_i = 0 \mid Z_i = 0)}\Big)}_{=0(\because \text{Independence})}
\end{aligned}$$

Estimating Complier for multi-valued treatment (9)

- What about $\mathbb{P}(T_i(0) = 1, T_i(1) = 1)$?

$$\widehat{\mathbb{P}(T_i(0) = 1, T_i(1) = 1)}$$
$$= \widehat{\mathbb{P}(T_i = 1 \mid Z_i = 0)} - \left(\widehat{\mathbb{P}(T_i = 0 \mid Z_i = 1)} - \widehat{\mathbb{P}(T_i = 0 \mid Z_i = 0)}\right)$$

- Hence,

$$\mathbb{V}\Big(\widehat{\mathbb{P}(T_i(0) = 1, T_i(1) = 1)}\Big)$$
$$= \mathbb{V}\left(\widehat{\mathbb{P}(T_i = 1 \mid Z_i = 0)} - \left(\widehat{\mathbb{P}(T_i = 0 \mid Z_i = 1)} - \widehat{\mathbb{P}(T_i = 0 \mid Z_i = 0)}\right)\right)$$

- **Caution**: Notice that $\widehat{\mathbb{P}(T_i = 1 \mid Z_i = 0)}$ and $\widehat{\mathbb{P}(T_i = 0 \mid Z_i = 0)}$ are not independent!
  - They are on the same sample ($Z_i = 0$)
  - Thus, you need to take into account **covariance**

Estimating Complier for multi-valued treatment (9)

$$
\begin{aligned}
&\mathbb{V}\Bigg(\widehat{\mathbb{P}(T_i = 1 \mid Z_i = 0)} - \left(\widehat{\mathbb{P}(T_i = 0 \mid Z_i = 1)} - \widehat{\mathbb{P}(T_i = 0 \mid Z_i = 0)}\right)\Bigg) \\
&= \mathbb{V}(\widehat{\mathbb{P}(T_i = 1 \mid Z_i = 0)}) \\
&\quad + \mathbb{V}(\widehat{\mathbb{P}(T_i = 0 \mid Z_i = 1)}) \\
&\quad + \mathbb{V}(\widehat{\mathbb{P}(T_i = 0 \mid Z_i = 0)})) \\
&\quad - 2\underbrace{\mathrm{Cov}(\widehat{\mathbb{P}(T_i = 1 \mid Z_i = 0)}, \widehat{\mathbb{P}(T_i = 0 \mid Z_i = 1)})}_{=0} \\
&\quad + 2\underbrace{\mathrm{Cov}(\widehat{\mathbb{P}(T_i = 1 \mid Z_i = 0)}, \widehat{\mathbb{P}(T_i = 0 \mid Z_i = 0)})}_{\text{Non-Zero!}} \\
&\quad - 2\underbrace{\mathrm{Cov}(\widehat{\mathbb{P}(T_i = 0 \mid Z_i = 1)}, \widehat{\mathbb{P}(T_i = 0 \mid Z_i = 0)})}_{=0}
\end{aligned}
$$

Estimating Complier for multi-valued treatment (10)

- How can we calculate covariance term?

$$\mathrm{Cov}(\widehat{\mathbb{P}(T_i = 1 \mid Z_i = 0)}, \widehat{\mathbb{P}(T_i = 0 \mid Z_i = 0)})$$

- Why is it non-zero? $\rightarrow$ If $\widehat{\mathbb{P}(T_i = 1 \mid Z_i = 0)}$ becomes larger, $\widehat{\mathbb{P}(T_i = 0 \mid Z_i = 0)}$ should be smaller

- From theory of multinomial distribution:

$$\mathrm{Cov}(\widehat{\mathbb{P}(T_i = 1 \mid Z_i = 0)}, \widehat{\mathbb{P}(T_i = 0 \mid Z_i = 0)}) = -\frac{\mathbb{P}(T_i = 1 \mid Z_i = 0)\mathbb{P}(T_i = 0 \mid Z_i = 0)}{\text{Number of } Z_i = 0}$$

- Same for other covariance terms

Added: What is multinominal distribution?

- Suppose there are K categories
- Think about how many times the trial where each T_i falls into one of K categories with probability

$$\mathbb{P}(T_i = k) = p_k \quad \text{with} \quad \sum_{k=1}^{K} p_k = 1$$

- We care about the counts of each category: i.e., $X_k = \sum_{i=1}^{n} \mathbb{1}\{T_i = k\}$
- The joint distribution of counts $(X_1, \cdots, X_K)$ follows multinominal distribution with $\sum_{k=1}^{K} X_k = n$
  - $\mathbb{E}[X_k] = np_k$
  - $\mathbb{V}[X_k] = np_k(1 - p_k)$
  - $\mathrm{Cov}(X_k, X_l) = -np_k p_l$
- As a result, the covariance of each probability estimates is

$$\mathrm{Cov}(\hat{p}_k, \hat{p}_l) = \mathrm{Cov}\left(\frac{X_k}{n}, \frac{X_l}{n}\right) = \frac{1}{n^2}\mathrm{Cov}(X_k, X_l) = -\frac{p_k p_l}{n}$$

Two Stage Least Squares

- Consider the following models:

$$Y_i = \alpha + \beta T_i + \epsilon_i$$
$$T_i = \gamma Z_i + \eta_i$$

where $\mathbb{E}[\epsilon_i \mid Z_i] = \mathbb{E}[\eta_i \mid Z_i] = 0$

- **Wald Estimator**:

$$\hat{\beta}_{\text{IV}} := \frac{\widehat{\text{Cov}(Y_i, Z_i)}}{\widehat{\text{Cov}(T_i, Z_i)}} = \frac{\text{Effect of Z on Y}}{\text{Effect of Z on T}}$$

Two Stage Least Squares: Why it works?

$$\begin{aligned}
\beta_{\text{IV}} &= \frac{\text{Cov}(Y_i, Z_i)}{\text{Cov}(T_i, Z_i)} \\
&= \frac{\text{Cov}(\alpha + \beta T_i + \epsilon, Z_i)}{\text{Cov}(\gamma Z_i + \eta_i, Z_i)} \\
&= \frac{\text{Cov}(\beta T_i, Z_i)}{\text{Cov}(\gamma Z_i, Z_i)} \quad (\because \text{Exogeneity}) \\
&= \frac{\text{Cov}(\beta(\gamma Z_i + \eta_i), Z_i)}{\text{Cov}(\gamma Z_i, Z_i)} \\
&= \frac{\beta\gamma\mathbb{V}[Z_i]}{\gamma\mathbb{V}[Z_i]} = \beta
\end{aligned}$$

Weak Instrument (1)

- IV is unstable when instrument weakly affects treatment $\gamma \approx 0$
  - Let's see how bias appears
  - For the sake of simplicity, assume $\bar{Z} = 0$
- Then, Wald estimator is written as

$$\hat{\beta}_{\mathrm{IV}} := \frac{\widehat{\mathrm{Cov}(Y_i, Z_i)}}{\widehat{\mathrm{Cov}(T_i, Z_i)}} = \frac{\frac{1}{n}\sum_{i=1}^{n} Y_i Z_i}{\frac{1}{n}\sum_{i=1}^{n} T_i Z_i}$$

Weak Instrument (2)

$$
\begin{aligned}
\hat{\beta}_{\text{IV}} - \beta &= \frac{\frac{1}{n}\sum_{i=1}^{n} Y_i Z_i}{\frac{1}{n}\sum_{i=1}^{n} T_i Z_i} - \beta \\
&= \frac{\frac{1}{n}\sum_{i=1}^{n}(\alpha + \beta T_i + \epsilon_i) Z_i}{\frac{1}{n}\sum_{i=1}^{n} T_i Z_i} - \beta \\
&= \frac{\frac{1}{n}\sum_{i=1}^{n}(\beta T_i Z_i + \epsilon_i Z_i)}{\frac{1}{n}\sum_{i=1}^{n} T_i Z_i} - \beta \\
&= \beta + \frac{\frac{1}{n}\sum_{i=1}^{n} \epsilon_i Z_i}{\frac{1}{n}\sum_{i=1}^{n} T_i Z_i} - \beta = \frac{\frac{1}{n}\sum_{i=1}^{n} \epsilon_i Z_i}{\frac{1}{n}\sum_{i=1}^{n} T_i Z_i}
\end{aligned}
$$

- If $\gamma = 0$, then $T_i = \gamma Z_i + \eta_i = \eta_i$. So,

$$
\hat{\beta}_{\text{IV}} - \beta = \frac{\frac{1}{n}\sum_{i=1}^{n} \epsilon_i Z_i}{\frac{1}{n}\sum_{i=1}^{n} \eta_i Z_i}
$$