Week 7: IV with Heterogeneous Effect

JJ Chen

February 27, 2015

Review Plan:

- This week we moved away from traditional IV world and enter the LATE framework
  - It adopts more elaborate notations to describe most of the things we already know
  - Thus it's not about how we should do, but how we should think about IV estimates

Review Plan:

- Today:
  1. The LATE Theorem
  2. Implications of the LATE Theorem in RCTS
     - The Bloom Result: A solution to the compliance problem
     - Two examples: JTPA training and MDVE (since we didn't cover in class)
  3. Counting and characterizing Compliers
- Next week might be further generalizations of LATE: multiple instruments, multi-valued instruments, multi-valued endogenous variable (The ACR Theorem), covariates with heterogeneous effect, Kappa weighting and nonlinear models... Oh, and the midterm

Motivation

- Scenario: You find two valid instruments! But they yields different estimates...
  - Traditional IV says there are something wrong with one of your instruments (Over-id; weak IV)
- An alternative explanation: heterogeneous treatment effect
  - Example: the effect of child caring on female labor supply
    - Two estimates from two (arguably) valid instruments (Table 4.1.4)
- Big picture:
  - Each valid instrument generate its own experiment
  - Internal validity vs. external validity

The LATE Language: New Notations

- Single-indexed potential treatment: D_{1i}, D_{0i}
  - Observed treatment: $D_i = D_{1i} + (D_{1i} - D_{0i})Z_i$
  - What is the average causal impact of Z_i on D_i?
- Double-indexed potential outcome: $Y_i(d, z)$
  - Z_i might have an impact on Y_i... We are going to rule this out, there will just be a IV chain, but no direct impact
  - What is the average causal impact of Z_i on Y_i?

The LATE Assumptions: Independence and Exclusion

- The LATE framework allows us to distinguish between Independence and Exclusion
  - The fact that a instrument is random doesn't guarantee ER

1. Independence: The instrument generates an experiment

$$[\{Y_i(d,z); \forall d, z\}, D_{1i}, D_{0i}] \perp\!\!\!\perp Z_i$$

   - What can we get?

2. Exclusion Restriction: The instruments affects Y_i only via D_i.
   - $Y_i(1,1) = Y_i(1,0) \equiv Y_{1i}$
   - $Y_i(0,1) = Y_i(0,0) \equiv Y_{0i}$
   - ER allows us to collapse the double-indexed into single-indexed
   - How would you write the relationship between the observed outcome and the potential outcome?

The LATE Assumptions: First Stage and Monotonicity

3. First Stage: $\mathrm{E}\left(\mathrm{D}_{1i} - \mathrm{D}_{0i}\right) \neq 0$
4. Monotonicity: $\mathrm{D}_{1i} > \mathrm{D}_{0i}$ for everyone (or vice verse)
    - Monotonicity implies that $\mathrm{E}\left(\mathrm{D}_{1i} - \mathrm{D}_{0i}\right) = \Pr\left(\mathrm{D}_{1i} > \mathrm{D}_{0i}\right)$
    - It gives us a good feature to exploit: the difference in two potential treatment status is the 1st
    - The 1st is the prob that D_{1i} is greater than D_{0i}

The LATE Theorem

- Given the four LATE Assumptions,

$$\frac{E(Y_i | Z_i = 1) - E(Y_i | Z_i = 0)}{E(D_i | Z_i = 1) - E(D_i | Z_i = 0)} = E(Y_{1i} - Y_{0i} | D_{1i} > D_{0i})$$

- The LATE Theorem tells us how to think about the Wald estimator: It's the ATE on the sub-population that has $D_{1i} > D_{0i}$

Name the Sub-population

- The LATE assumptions "partition the world":
  - Compliers: $\mathrm{D}_{1i} > \mathrm{D}_{0i}$
  - Always-takers: $\mathrm{D}_{1i} = \mathrm{D}_{0i} = 1$
  - Never-takers: $\mathrm{D}_{1i} = \mathrm{D}_{0i} = 0$
  - Monotonicity assumes away Defiers
- IV tells us nothing about Always-takers and Never-takers, just like an FE estimate only identify effects for people who have change
  - If an instrument induces an change in your behavior, you're in the conditional set for the LATE

LATE and ATET: Who got treated?

- Note that people who got treated consists of two groups: Always-takes and compliers who are assigned with $Z_i = 1$
- Can be seen from the notation:

$$\begin{aligned} D_i =& D_{0i} + (D_{1i} - D_{0i})Z_i \\ \implies [D_i = 1] =& [D_{1i} = D_{0i} = 1] \cup [(D_{1i} - D_{0i} = 1) \cap (Z_i = 1)] \end{aligned}$$

- The compliers are a proper subset of **the treated**
- Thus we know that ATET is a weighted average of effects on always-takers and compliers

Motivation: LATE and ATET

- Unlike matching, IV can not be used to identify ATET, it only gives LATE
- In the heterogeneous world, when will LATE becomes ATET?
  - When all compliers are the treated (or when there is no always taker).
  - In some scenario, there is no always-takers (or no never-takers).
- The most important scenario is RCTs, in which the control group has no access to the intervention.
  - The group of offered treatment may or may not take the treatment, but the control group has no access to the treatment
  - A one-sided IV scenario, or a Bloom scenario

The Bloom Result: IV Solves the Compliance Problem

- A one-sided IV scenario implies that there is no always-takers
  - No controls are treated (Assuming the controls don't have access), $\mathrm{E}(\mathrm{D}_i \mid \mathrm{Z}_i = 0) = 0$
- Recall that *{Treated}* = *{Alwayer-takers}* + *{Compliers}*
  - So "treated" IS "compliers" in RCTs
- LATE = ATET
- The LATE Theorem implies The Bloom Result:

$$\frac{\mathrm{E}(\mathrm{Y}_i \mid \mathrm{Z}_i = 1) - \mathrm{E}(\mathrm{Y}_i \mid \mathrm{Z}_i = 0)}{\mathrm{E}(\mathrm{D}_i \mid \mathrm{Z}_i = 1)} = \frac{\text{ITT}}{\text{Compliance Rate}}$$
$$= \mathrm{E}(\mathrm{Y}_{1i} - \mathrm{Y}_{0i} \mid \mathrm{D}_i = 1)$$

Examples: JTPA and MDVE

1. JTPA
   - Table 4.4.1
   - The 1st is approximately the compliance rate. How come?
   - OLS is bias (why?); ITT is diluted (Why?)
2. MDVE
   - Research Question: What is the best response to domestic violence?
   - Table 1 (Summary Stat) and Table 2 (IV Estimates) from Angrist (2006)
   - LATE helps a lot of experiments: changing subjects' behavior might be problematic (polices should decide what to do for a given case), but LATE says only changing the likelihood of doing something is enough

Motivation

- Each instrument generates its own experiment for it's compliers
- Knowing more about the compliers helps to reconcile different IV estimates
  - Also homogeneous vs. heterogeneous effect, interval validity vs. external validity
- How to use the information from counting and characterizing?
  1. Similar characteristics, different IV estimates: Implications?
  2. Different characteristics, similar IV estimates: Implications?
  3. Extrapolation and external validity for other sample

Counting

- Table 4.4.2
- Given monotonicity, 1st are

$$\Pr\left(\mathrm{D}_{1i} > \mathrm{D}_{0i}\right)$$

- Among the treated, we have

$$\begin{aligned}
&\Pr\left(\mathrm{D}_{1i} > \mathrm{D}_{0i} \mid \mathrm{D}_i = 1\right) \\
&= \frac{\Pr\left(\mathrm{D}_i = 1 \mid \mathrm{D}_{1i} > \mathrm{D}_{0i}\right)\Pr\left(\mathrm{D}_{1i} > \mathrm{D}_{0i}\right)}{\Pr\left(\mathrm{D}_i = 1\right)} \\
&= \frac{\Pr\left(\mathrm{D}_i = 1 \mid \mathrm{D}_{1i} > \mathrm{D}_{0i}\right)\left[\mathrm{E}\left(\mathrm{D}_i \mid \mathrm{Z}_i = 1\right) - \mathrm{E}\left(\mathrm{D}_i \mid \mathrm{Z}_i = 0\right)\right]}{\Pr\left(\mathrm{D}_i = 1\right)} \\
&= \frac{\Pr\left(\mathrm{Z}_i = 1\right)\left[\mathrm{E}\left(\mathrm{D}_i \mid \mathrm{Z}_i = 1\right) - \mathrm{E}\left(\mathrm{D}_i \mid \mathrm{Z}_i = 0\right)\right]}{\Pr\left(\mathrm{D}_i = 1\right)},
\end{aligned}$$

where the first equality use the definition of conditional

Characterizing

- Table 4.4.3
- Are the compliers more likely to be high school graduates?

$$\frac{\Pr\left(\mathrm{X}_{1i}=1\left|\mathrm{D}_{1i}>\mathrm{D}_{0i}\right.\right)}{\Pr\left(\mathrm{X}_{1i}=1\right)}$$
$$=\frac{\Pr\left(\mathrm{X}_{1i}=1\cap\left(\mathrm{D}_{1i}>\mathrm{D}_{0i}\right)\right)}{\Pr\left(\mathrm{D}_{1i}>\mathrm{D}_{0i}\right)\Pr\left(\mathrm{X}_{1i}=1\right)}$$
$$=\frac{\Pr\left(\mathrm{D}_{1i}-\mathrm{D}_{0i}\left|\mathrm{X}_{1i}=1\right.\right)}{\Pr\left(\mathrm{D}_{1i}>\mathrm{D}_{0i}\right)}$$
$$=\frac{\mathrm{E}\left(\mathrm{D}_{i}\left|\mathrm{Z}_{i}=1,\mathrm{X}_{1i}=1\right.\right)-\mathrm{E}\left(\mathrm{D}_{i}\left|\mathrm{Z}_{i}=0,\mathrm{X}_{0i}=1\right.\right)}{\mathrm{E}\left(\mathrm{D}_{i}\left|\mathrm{Z}_{i}=1\right.\right)-\mathrm{E}\left(\mathrm{D}_{i}\left|\mathrm{Z}_{i}=0\right.\right)}$$

References I

Angrist, Joshua D. 2006. "Instrumental Variables Methods in Experimental Criminological Research: What, Why and How." *Journal of Experimental Criminology* 2 (1). Springer: 23–44.