14.661 Recitation 6: LATE and Discount Rate Bias

Andrea Manera

October 29, 2021

Roadmap

- Review of the LATE theorem (Angrist and Imbens, 1994)
- Application: Discount Rate Bias (Lang, 1993; Card, 1995; Card, 2001)

Setting the Stage

Throughout, we consider a setting with a single instrument, treatment, and outcome:

$$Y_i = \alpha + \beta_i D_i + \eta_i$$
$$D_i = \alpha + \gamma Z_i + \upsilon_i$$

- D_i and Z_i are dummies for treatment and instrument
- β_i a random coefficient that can differ across i.
- Important: outcome can in principle depend on both treatment and instrument:

$$Y_i(d, z)$$

- We want to estimate ATE:

$$\mathbb{E}[\beta_i]$$

First Stage, Independence and Exclusion

For reference, consider first the case $\beta_i = \beta \ \forall i$. Assume:

1. Independence:

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

   - Z_i as good as random, D_{1i}, D_{0i} are potential outcomes for first stage
   - Vietnam lottery example: D_{1i} tells us if individual drawing low number serves in the military, D_{0i} if high.
   - *Causal Interpretation* for the *reduced form*, regression of Y_i on Z_i

2. Exclusion Restriction:

$$Y_i(d,0) = Y_i(d,1) \ \ d = 0, 1,$$

   - Instrument affects outcome only through treatment

3. First Stage: $\mathbb{E}(D_{1i} - D_{0i}) \neq 0$. Instrument affects treatment.

Wald Estimator with Homogeneous β

Super-easy derivation of Wald estimator

- Under assumptions (1)-(3) above, we get:

$$\mathbb{E}\left[Y_i \mid Z_i\right] = \alpha + \beta \mathbb{E}\left[D_i \mid Z_i\right], \; Z_i = 1, 0$$

- Subtract line-by-line the system with $Z_i = 1, Z_i = 0$ and divide through:

$$\beta = \frac{\mathbb{E}\left[Y_i \mid Z_i = 1\right] - \mathbb{E}\left[Y_i \mid Z_i = 0\right]}{\mathbb{E}\left[D_i \mid Z_i = 1\right] - \mathbb{E}\left[D_i \mid Z_i = 0\right]},$$

- Reduced-form causal effect of Z rescaled by the first-stage

LATE, aka when β_i varies

Now consider β_i different across i. Add assumption:

4. Monotonicity:

$$D_{i1} - D_{i0} \geqslant 0,\ \forall i \quad \text{or} \quad D_{i1} - D_{i0} \leqslant 0,\ \forall i$$

- ♢ Instrument shifts treatment only in one direction for everyone
- ♢ We will go with the first: there are only *compliers* or *always takers*

Under (1)-(4) the LATE theorem (Angrist and Imbens, 1994) states:

$$\frac{\mathbb{E}\left[Y_i \mid Z_i = 1\right] - \mathbb{E}\left[Y_i \mid Z_i = 0\right]}{\mathbb{E}\left[D_i \mid Z_i = 1\right] - \mathbb{E}\left[D_i \mid Z_i = 0\right]} = \mathbb{E}\left[Y_{1i} - Y_{0i} \mid D_{1i} > D_{0i}\right] = \mathbb{E}\left[\beta_i \mid D_{1i} > D_{0i}\right]$$

We get an average *of compliers' treatment effects*, differs in general from ATE

LATE proof

$$\begin{aligned}\mathbb{E}\left[Y_i \mid Z_i = 1\right] &= \mathbb{E}\left[Y_{0i} + \left(Y_{1i} - Y_{i0}\right) D_i \mid Z_i = 1\right] \\ &= \mathbb{E}\left[Y_{0i} + \left(Y_{1i} - Y_{i0}\right) D_{1i}\right]\end{aligned}$$

First line by exclusion restriction, second by independence. Similar formula for $Z_i = 0$:

$$\mathbb{E}\left[\left(Y_{1i} - Y_{i0}\right)\left(D_{1i} - D_{0i}\right)\right] = \mathbb{E}\left[\left(Y_{1i} - Y_{i0}\right) \mid D_{1i} > D_{0i}\right] P\left(D_{1i} > D_{0i}\right)$$

Since if $D_{1i} = D_{0i}$, $\left(Y_{1i} - Y_{i0}\right)\left(D_{1i} - D_{0i}\right) = 0$.

- Note: Monotonicity $D_{1i} \geqslant D_{0i}$ essential, otherwise there might be *defiers* for which $D_{1i} < D_{0i}$; theorem does not hold!
- If $D_{1i} \leqslant D_{0i}$ $\forall i$, LATE is effect for defiers.

Average Causal Response (Angrist and Imbens, 1995)

- Extension of LATE theorem for treatment with multiple discrete values (intensity levels)
- Notation for return to schooling example:

$$Y_{si} = f_i(s) \quad s \in \{0, 1, \dots, \bar{s}\}$$

- Now there are $\bar{s}$ causal effects for each unit, one for each specific notch:

$$Y_{s,i} - Y_{s-1,i}$$

ACR Theorem

The ACR theorem, under assumptions (1)-(4) with changed notation (see Mostly Harmless):

$$\frac{\mathbb{E}\left[Y_i \mid Z_i = 1\right] - \mathbb{E}\left[Y_i \mid Z_i = 0\right]}{\mathbb{E}\left[D_i \mid Z_i = 1\right] - \mathbb{E}\left[D_i \mid Z_i = 0\right]} = \sum_{j=1}^{\bar{s}} \omega_s \mathbb{E}\left[\left(Y_{j,i} - Y_{j-1,i}\right) \mid s_{1i} > j > s_{0i}\right]$$

$$\omega_s = \frac{P\left[s \mid s_{1i} > s > s_{0i}\right]}{\sum_{j=1}^{\bar{s}} P\left[s \mid s_{1i} > j > s_{0i}\right]}$$

Why ACR is so Cool

$$\frac{\mathbb{E}[Y_i \mid Z_i = 1] - \mathbb{E}[Y_i \mid Z_i = 0]}{\mathbb{E}[s_i \mid Z_i = 1] - \mathbb{E}[s_i \mid Z_i = 0]} = \sum_{j=1}^{\bar{s}} \omega_s \mathbb{E}\left[\left(Y_{j,i} - Y_{j-1,i}\right) \mid s_{1i} > j > s_{0i}\right]$$

$$\omega_s = \frac{P[s \mid s_{1i} > s > s_{0i}]}{\sum_{j=1}^{\bar{s}} P[s \mid s_{1i} > j > s_{0i}]}$$

- IV estimates weighted average of *unit causal responses* at each point s
- ω_s tells us how much *weight* compliers with s have on the estimate
- The rescaled *size of the group of compliers!*
- Estimate weights consistently subtracting CDFs at s for group with $Z_i = 1$ from $Z_i = 0$
- Weights tell us exactly *who* we are learning about (useful for external validity)

Continuous Treatment Extension

Assume the outcome is now a continuous function g with derivative g':

$$Y_i = g_i(s)$$

Note:

$$\begin{aligned}
\mathbb{E}[Y_i \mid Z_i = 1] - \mathbb{E}[Y_i \mid Z_i = 0] &= \mathbb{E}\left[\int_{s_{0i}}^{s_{1i}} g_i'(j)\, dj\right] \\
&= \int \mathbb{E}\left[g_i'(j) \mid s_{1i} > j > s_{0i}\right] P(s_{1i} > j > s_{0i})\, dj
\end{aligned}$$

Where the first line is fundamental thm of calculus, the second uses independence.

Continuous ACR

$$\frac{\mathbb{E}[Y_i \mid Z_i = 1] - \mathbb{E}[Y_i \mid Z_i = 0]}{\mathbb{E}[s_i \mid Z_i = 1] - \mathbb{E}[s_i \mid Z_i = 0]} = \frac{\int \mathbb{E}\left[g_i'(j) \mid s_{1i} > j > s_{0i}\right] P(s_{1i} > j > s_{0i})\, dj}{\int P(s_{1i} > j > s_{0i})\, dj}.$$

- Similar to before, but derivative turns out to play a very important point in education lit!
- Important special case when:

$$g_i(s) = \alpha_{0i} + \alpha_{1i} s,$$

$$\frac{\mathbb{E}[Y_i \mid Z_i = 1] - \mathbb{E}[Y_i \mid Z_i = 0]}{\mathbb{E}[s_i \mid Z_i = 1] - \mathbb{E}[s_i \mid Z_i = 0]} = \frac{\mathbb{E}[\alpha_{1i}(S_{1i} - S_{0i})]}{\mathbb{E}[S_{1i} - S_{0i}]}.$$

- Weighted average of random coefficients, *weighted by how much schooling shifts.*

Note on Special Case

Thanks to continuity of $g(\cdot)$, we can always use the mean value theorem.

- There exists $\tilde{S}_i \in [S_{i0}, S_{i1}]$ such that:

$$g_i\left(S_{1i}\right) = g_i\left(S_{i0}\right) + g_i'\left(\tilde{S}_i\right)\left(S_{1i} - S_{i0}\right).$$

- Thus we always have:

$$\frac{\mathbb{E}\left[Y_i \mid Z_i = 1\right] - \mathbb{E}\left[Y_i \mid Z_i = 0\right]}{\mathbb{E}\left[s_i \mid Z_i = 1\right] - \mathbb{E}\left[s_i \mid Z_i = 0\right]} = \frac{\mathbb{E}\left[g_i'\left(\tilde{S}_i\right)\left(S_{1i} - S_{0i}\right)\right]}{\mathbb{E}\left[S_{1i} - S_{0i}\right]}.$$

Back to School

Card (2001) assumes HK model with returns to schooling:

$$g(s_i) = \log y_i = \alpha_i + b_i s_i - \frac{1}{2} k_1 s_i^2$$

optimal schooling is then (see Card, 2001 for details):

$$s_i^\star = \frac{(b_i - r_i)}{k},$$

where r_i is the individual's discount rate. Two implications:

1. By concavity, returns to schooling at low s_i are higher: $g''(s_i) < 0$!
2. Individuals with a high discount factor choose less education

Implications of ACR for Returns to Schooling

Given the above, the ACR estimated with an IV that reduces schooling costs is:

$$\frac{\mathbb{E}\left[g_i'\left(\tilde{S}_i\right)\left(S_{1i}-S_{0i}\right)\right]}{\mathbb{E}\left[S_{1i}-S_{0i}\right]}$$

- In the real world, reduction in costs affects more people with *low schooling*, which means:
  - ◇ $S_{1i}-S_{0i}$ is largest for people with lower S_{0i}, get more weight in ACR
  - ◇ $g_i'\left(\tilde{S}_i\right)$ is larger for people with low S
- Therefore:

$$\frac{\mathbb{E}\left[g_i'\left(\tilde{S}_i\right)\left(S_{1i}-S_{0i}\right)\right]}{\mathbb{E}\left[S_{1i}-S_{0i}\right]} > \mathbb{E}\left[g_i'\left(S_i\right)\right]$$

- It is called *discount rate bias* since in the model low S_{0i} happens because of high r_i

Why Does This All Matter?

- The literature on schooling *almost always* finds IV estimates that are larger than OLS
  - e.g., Angrist and Krueger (1991) have OLS 0.07, and IV 0.10
- Ability bias is positive, measurement error is negative
- AK say ability bias is small, so measurement error dominates
- Discount rate bias produces larger estimates regardless!
- Policy relevance:
  - Measurement error has no real-world implications...
  - ... but discount rate has: estimates are high just because of population of compliers