Section 3: LATEs and Compliers

Valentine Gilbert

March 3, 2022

Overview

Setup: Angrist et al. (2016)

Angrist et al. want to estimate effects of attendance at Boston's charter high schools on college preparation and enrollment

- Today we're going to use their paper to review IV when treatment effects are heterogeneous.
- With heterogeneous treatment effects, the causal relationship of interest is:

$$Y_i = \alpha + \beta_i D_i + u_i$$

- In potential outcomes notation, the causal effect can also be written as

$$Y_{1i} - Y_{0i} = \beta_i$$

Setup: Identification Strategy

- What's wrong with the naïve OLS regression of various outcomes on charter enrollment?
  - Charter school applicants are selected on various characteristics.
  - What sign do you think the bias will be?
- Angrist et al. use the fact that Boston's oversubscribed charter schools allocate seats via random lottery to estimate causal effects.
- Because take-up was imperfect, they have to use a random encouragement design.
- Specifically, we can estimate causal effects via IV using the following instrument:

$$Z_i = \begin{cases} 1, & i \text{ wins the lottery} \\ 0, & \text{otherwise} \end{cases}$$

- This leads to the following causal estimates:

Table 4
Lottery Estimates of Effects on High School Milestones

MCAS Performance Categories (Math and ELA Combined)		
Category	Noncharter Mean (1)	Effect (2)
Needs Improvement or higher	.975	.014
		(.015)
Proficient or higher	.543	.171**
		(.071)
Advanced or higher	.076	.161***
		(.034)
Meets competency determination	.740	.147**
		(.065)
Eligible for Adams Scholarship using BPS cutoffs	.198	.242***
		(.058)
N	3,608	

The LATE: Setup

- What do these estimates mean when treatment effects are heterogeneous?
- To better understand how to interpret these causal effects, we turn to the Local Average Treatment Effects (LATE) framework.
- Let individual i's treatment status when $Z_i = 1$ and $Z_i = 0$ be denoted by D_{1i} and D_{0i}.
- We can then divide the sample into four groups:
  - Compliers: $D_{1i} = 1$ and $D_{0i} = 0$
  - Always-takers: $D_{1i} = D_{0i} = 1$
  - Never-takers: $D_{1i} = D_{0i} = 0$
  - Defiers: $D_{1i} = 0$ and $D_{0i} = 1$

The LATE: Assumptions

Now assume the following four assumptions hold

- **Relevance:** Instrument affects treatment status
- **Exclusion Restriction:** IV affects outcome only through D_i
- **Independence:** IV as good as randomly assigned
- **Monotonicity:** IV affects everyone in the same direction (i.e. no defiers)

If these assumptions hold, we can simplify the Wald estimator as follows:

$$
\begin{aligned}
\beta_{Wald} &= \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]} \\
&= \frac{E[Y_i(D_{1i}, Z_i = 1) - Y_i(D_{0i}, Z_i = 0)]}{E[D_{1i} - D_{0i}]} \\
&= \frac{Pr(D_{1i} = D_{0i}) \times 0 + Pr(D_{1i} > D_{0i}) \times E[Y_{1i} - Y_{0i}|D_{1i} > D_{0i}]}{Pr(D_{1i} = D_{0i}) \times 0 + Pr(D_{1i} > D_{0i}) \times 1} \\
&= E[Y_{1i} - Y_{0i}|D_{1i} > D_{0i}] = E[\beta_i|D_{1i} > D_{0i}]
\end{aligned}
$$

The LATE: Interpretation

- So, if our assumptions hold, the estimated effect of charter schools is the average effect on compliers – students who attended the charter school *because* they won the lottery.
- For what kinds of policy changes would this be informative? For what kinds of policy changes would this be less informative?
- Do we think that each of the four assumptions holds?
  - Relevance?
  - Exclusion restriction?
  - Independence?
  - Monotonicity?

Actual Empirical Framework

$$y_i = \sum_j \delta_j d_{ij} + \gamma' X_i + \rho C_i + \varepsilon_i$$

$$C_i = \sum_j \mu_j d_{ij} + \beta' X_i + \pi_1 Z_{1i} + \pi_2 Z_{2i} + \eta_i$$

- C_i : Charter school attendance indicator
- Z_{1i} : Day-of-lottery offer indicator
- Z_{2i} : Waitlist offer indicator
- d_{ij} : Application-set indicators
- X_i : Student characteristics

Compliers 1/5: Characteristics

- We usually can't observe compliers in our data, but we can still learn about them
- Because of independence and monotonicity, we know the probability of being an always taker or a never taker:

$$p_{AT} \equiv Pr(\text{always taker}) = Pr(\text{always taker}|Z_i = 0) = E[D_i|Z_i = 0]$$
$$p_{NT} \equiv Pr(\text{never taker}) = Pr(\text{never taker}|Z_i = 1) = 1 - E[D_i|Z_i = 1]$$

- This allows us to estimate
  - The proportion of compliers in the sample
  - The share of treated who are compliers
  - The relative likelihood of a complier having a given characteristic
  - The average treated and untreated outcomes for compliers
- Let's see how!

Compliers 2/5: Characteristics

- The proportion of compliers in the sample is:

$$\begin{aligned} Pr(D_{1i} > D_{0i}) &= 1 - p_{AT} - p_{NT} \\ &= E[D_i|Z_i = 1] - E[D_i|Z_i = 0] \end{aligned}$$

- The share of treated who are compliers is:

$$\begin{aligned} Pr(D_{1i} > D_{0i}|D_i = 1) &= \frac{Pr(D_i = 1, D_{1i} > D_{0i})}{Pr(D_i = 1)} = \frac{Pr(Z_i = 1, D_{1i} > D_{0i})}{Pr(D_i = 1)} \\ &= \frac{Pr(Z_i = 1)Pr(D_{1i} > D_{0i})}{Pr(D_i = 1)} \end{aligned}$$

- The relative likelihood of a complier having a binary characteristic $x_i = 1$ is:

$$\frac{Pr(x_i = 1|D_{1i} > D_{0i})}{Pr(x_i = 1)} = \frac{E[D_i|Z_i = 1, x_i = 1] - E[D_i|Z_i = 0, x_i = 1]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]}$$

Compliers 3/5: Avg. Treated and Untreated Outcomes

- We can also use the monotonicity assumption to calculate average treated and untreated outcomes for the compliers
- To do so, we exploit the fact that average outcomes for always takers and never takers are easily observable:

$$\bar{Y}_{AT} \equiv E[Y_i|\text{always taker}] = E[Y_i|D_i = 1, Z_i = 0]$$
$$\bar{Y}_{NT} \equiv E[Y_i|\text{never taker}] = E[Y_i|D_i = 0, Z_i = 1]$$

- Now note that the average outcome conditional on $D_i = 0$ and $Z_i = 0$ is a weighted average of $\bar{Y}_{NT}$ and the local average untreated outcome (LAUO) for compliers:

$$E[Y_i|D_i = 0, Z_i = 0] = \frac{Pr(D_{1i} > D_{0i}) \times LAUO + p_{NT} \times \bar{Y}_{NT}}{1 - p_{AT}}$$

- Solving for the $LAUO$, we get:

$$LAUO = \frac{(1 - p_{AT})E(Y_i|D_i = 0, Z_i = 0) - p_{NT}\bar{Y}_{NT}}{Pr(D_{1i} > D_{0i})}$$

Compliers 4/5: Avg. Treated and Untreated Outcomes

- We can similarly derive an expression for the local average treated outcome (LATO) for compliers
- Note that the average outcome conditional on $D_i = 1$ and $Z_i = 1$ is a weighted average of $\bar{Y}_{AT}$ and the LATO:

$$E[Y_i|D_i = 1, Z_i = 1] = \frac{Pr(D_{1i} > D_{0i}) \times LATO + p_{AT} \times \bar{Y}_{AT}}{1 - p_{NT}}$$

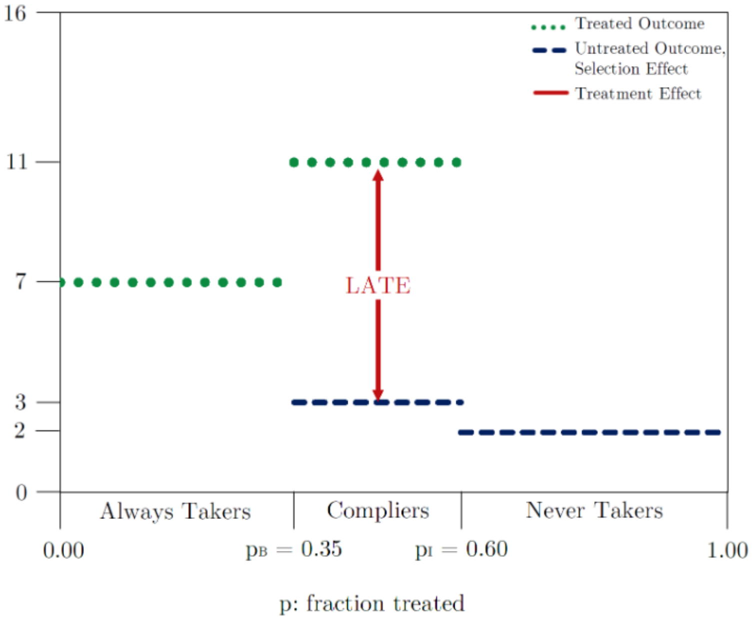
- Solving for the *LATO* gives us:

$$LATO = \frac{(1 - p_{NT})E[Y_i|D_i = 1, Z_i = 1] - p_{AT}\bar{Y}_{AT}}{Pr(D_{1i} > D_{0i})}$$

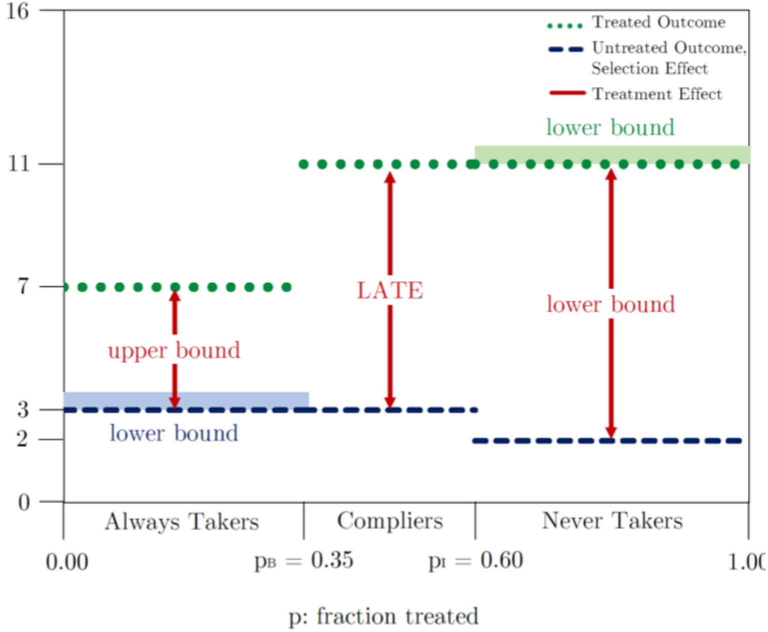
- You can also make use of the identity $LATE = LATO - LAUO$

Compliers 5/5: Heterogeneity

- If we make additional assumptions, we can use our estimates of $\bar{Y}_{AT}$, $\bar{Y}_{NT}$, the LAUO, and the LATO to learn about treatment effect heterogeneity
- Imagine ranking individuals according to how likely they are to take up treatment
- If we assume untreated outcomes are monotonic with respect to this latent propensity score, we can place bounds on the average untreated outcome of always takers
- Assuming treated outcomes are monotonic with respect to the latent propensity score, we can place bounds on the average treated outcomes of never takers
- These bounds can potentially tell us about treatment heterogeneity
- Note: These monotonicity assumptions help with external validity and are very different from the monotonicity assumption needed for internal validity of the LATE!



16
11
7
3
2
0
Treated Outcome
Untreated Outcome, Selection Effect
Treatment Effect
LATE
Always Takers
Compliers
Never Takers
0.00
pB = 0.35
pI = 0.60
1.00
p: fraction treated



16
11
7
3
2
0
Treated Outcome
Untreated Outcome,
Selection Effect
Treatment Effect
lower bound
LATE
lower bound
upper bound
lower bound
Always Takers
Compliers
Never Takers
0.00
pB = 0.35
pI = 0.60
1.00
p: fraction treated

Random Encouragement: Kling, Liebman, and Katz (2007)

- What is the causal effect of neighborhoods on the outcomes of residents? Kling, Liebman, and Katz study this question by evaluating the Moving to Opportunity program.
- Residents of high-poverty public housing projects in Baltimore, Boston, Chicago, LA, and New York were randomized to three treatment arms:
    - Control: Status quo
    - Section 8: Received traditional Section 8 vouchers
    - Experimental: Received mobility counseling and Section 8 vouchers restricted to low poverty neighborhoods
- Take-up of the program was imperfect: 47% of the experimental group and 60% of the Section 8 group used their vouchers.

Random Encouragement: Kling, Liebman, and Katz (2007)

- What's wrong with the naïve OLS regression of outcomes on neighborhood poverty rates?
  - Individuals endogenously sort into neighborhoods! So OLS estimates suffer from selection bias.
  - What sign do you think the bias will be?
- What's wrong with the reduced form OLS regression of outcomes on treatment group indicators?
- How can we use the MTO experiment to address these concerns?
  - Use randomly assigned treatment arm (interacted with MTO site) as an instrument for exposure to neighborhood poverty
- What would be the interpretation of our IV estimates?

Random Encouragement: Kling, Liebman, and Katz (2007)

Does our instrument satisfy the necessary assumptions?

- Relevance?
- Exclusion restriction?
- Independence?
- Monotonicity?

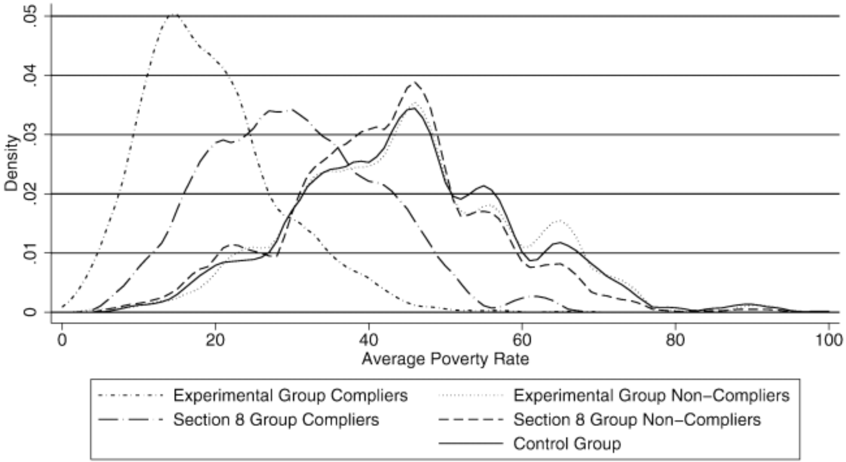


FIGURE 1.—Densities of average poverty rate, by group. Average poverty rate is a duration-weighted average of tract locations from random assignment through 12/31/2001. Poverty rate is based on linear interpolation of 1990 and 2000 Censuses. Density estimates used an Epanechnikov kernel with a half-width of 2.

B. Female Youth Mental Health

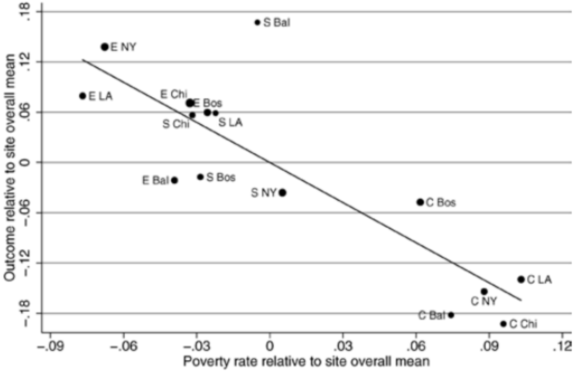


FIGURE 2.—Partial regression leverage plots. The index on the horizontal axis is expressed in standard deviation units relative to the control group overall standard deviation for each variable. The components of the overall and mental health indices are described in the notes to Table II. The poverty rate is an average across tracts since random assignment, weighted by residential duration, using linear interpolation between the 1990 and 2000 Censuses. The line passes through the origin with the slope from 2SLS estimation of Equation (3) of the outcome on poverty rate and site indicators, using group-by-site interactions as instrumental variables. The points are from a partial regression leverage plot of the group outcome means on the group poverty rate means, conditional on site main effects, as described in the text. The size of each point is proportional to the sample size of that group and, correspondingly, to the weight each point receives in the 2SLS regression.