The power of RCTs – Teacher Performance Pay in India

Econ 140, Section 6

Jonathan Old

Roadmap

1. OVB, Bad Controls
2. Omitted Variable Bias
3. Recap: RCTs
4. Exam review

Any comments about the midterm?

OVB, Bad Controls

Omitted Variable Bias

Recap: OVB (Very important!)

We can summarize everything of OVB in three equations. Let Y_i be the outcome variable, X_i our regressor of interest, and Z_i the "omitted" variable.

$$\begin{aligned}
\text{[Long regression]} \quad & Y_i = c_1 + \beta_L X_i + \lambda Z_i + \gamma_L C_i + e_i \\
\text{[Short regression]} \quad & Y_i = c_2 + \beta_S X_i \qquad\quad + \gamma_S C_i + u_i \\
\text{[Auxiliary regression]} \quad & Z_i = c_3 + \pi X_i \qquad\quad + \gamma_A C_i + v_i
\end{aligned}$$

Then, the **Omitted variable bias formula** states that:

$$\underbrace{\beta_S}_{\text{Short}} = \underbrace{\beta_L}_{\text{Long}} + \underbrace{\lambda}_{\text{Omitted}} \cdot \underbrace{\pi}_{\text{Included}}$$

We call $\lambda\pi$ the **omitted variable bias**. We can appraise the direction of the bias by multiplying our guesses for the signs of λ and π. **If either $\lambda = 0$ or $\pi = 0$, then OVB is zero!**

Bad controls

- Some controls are called ”bad controls”. These are:
  1. Variables that are themselves outcomes of a treatment: What happens if you control for the change in English test scores in the regression below?
  2. Variables that moderate the treatment effect: If gender affects earnings because gender affects occupational choice, then controlling for occupation **shuts down** this channel
- Rule of Thumb: Good controls are either pre-determined or immutable characteristics.
- Example: Education, income, ZIP code at birth, current ZIP code

Recap: RCTs

RCTs

- In RCTs, the treatment is randomly assigned: So it is, by definition, independent of any other characteristics **at baseline**
- Can we test whether this is true in the data?
- Therefore, any ”good control” is not necessary to remove OVB
- But: We still need to be careful about bad controls, because these happen after the treatment

Balance table

TABLE 2
SAMPLE BALANCE ACROSS TREATMENTS

	Control (1)	Group Incentive (2)	Individual Incentive (3)	*p*-Value (Equality of All Groups) (4)
	A. Means of Baseline Variables			
School-level variables:				
1. Total enrollment (baseline: grades 1–5)	113.2	111.3	112.6	.82
2. Total test takers (baseline: grades 2–5)	64.9	62.0	66.5	.89
3. Number of teachers	3.07	3.12	3.14	.58
4. Pupil-teacher ratio	39.5	40.6	37.5	.66
5. Infrastructure index (0–6)	3.19	3.14	3.26	.84
6. Proximity to facilities index (8–24)	14.65	14.66	14.72	.98
Baseline test performance:				
7. Math (raw %)	18.5	18.0	17.5	.69
8. Math (normalized; in SD)	.032	.001	−.032	.70
9. Telugu (raw %)	35.1	34.9	33.5	.52
10. Telugu (normalized; in SD)	.026	.021	−.046	.53

- If we test 100 baseline characteristics: How often would we expect p-values below 0.05 if our randomization worked?
- Do you think randomization worked in this case?

Exam review

How to approach an exam question

1. Think: About the question, about the real world
2. Start with the numbers you see
    - One-sentence summary
    - Direction (positive or negative?)
    - Statistical significance (significant or insignificant, level?)
    - (Economic) magnitude (big or small?)
3. Then: Establish whether estimated relationship is causal or not
    - What do the results mean? Correlation (interesting) or causality (policy-relevant)
    - Is X-variable randomized? Do we have valid counterfactuals?
    - If not: Do you expect bias? Of which sort (OVB, reverse causality, bad controls, . . .)?
    - Find a plausible story for bias (using the OVB formula)

There are no traps!