

Homework Chapter 11

Pablo Bello

2023-02-26

- Suppose that you are analyzing the effect of universities and colleges opening during a pandemic on increase in the number of positive cases. Name one strategy that you can use to avoid having to collect data on all types of campus characteristic variables that are constant over time that you may have to control for in your analysis.

We can condition our estimates on the university if we have multiple observations for each university with variation in the treatment condition. With this approach, characteristics of the universities that are constant over time won't bias the estimate.

- Intuitively, why would a method that isolates front doors allow you to ignore back doors related to unmeasured variables?

Because, if we have properly identified a front door, the information that the exogenous variable contains on the treatment is independent of all back doors that go into the treatment (precisely because of the exogeneity of our front door).

- On robustness tests:
 1. What are robustness tests?

They are ways to test our assumptions.

2. What is the purpose of conducting a robustness test?

To see the impact of an assumption on our estimates. Either by checking what happens when we relax that assumption or, if possible, empirically checking the assumption.

3. What are placebo tests?

When we run a test for which we can safely assume that the result should be 0 and hope that our model spits out the right answer. It increases our confidence that our model works.

- Suppose you want to study the effect of attending tutoring sessions on grade point averages (GPA). List at least five variables that impact both the attendance of tutoring sessions and students' GPA. Is it feasible to measure and control for all of the variables?

Their interest in school, major, whether the student is also working, family SES and the university they are attending.

- Describe partial identification in your own words.

Instead of trying to make a precise guess of the true estimate we instead make as few assumptions as possible and give a range of values within which we are pretty certain that the true estimate lies.

- Pick any causal diagram from the book other than Figure 11.2.
 1. Reproduce that diagram here.

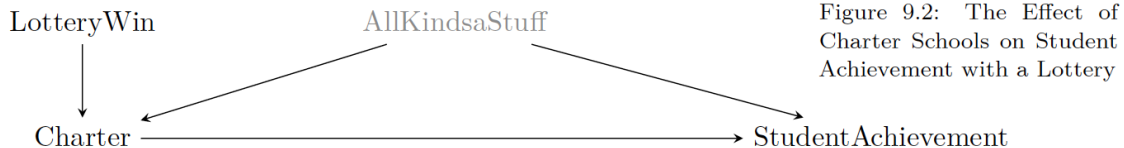


Figure 1: Alt text

2. Select two variables on the diagram without a direct link between them (i.e. no single arrow straight from one of them to the other). **Lottery Win and AllKindsaStuff – AKS – (assuming this was just one variable and observable).**
3. What variables would you need to control for that will eliminate any relationship between the two variables (you might not need any).

There is two paths between LotteryWin AKS but both of them are closed (they have colliders) so there is no relationship between these two variables.

4. If you looked at the relationship between your two variables from part b, while controlling for the variables from part c, and you got a nonzero result, what would you conclude?

In this case, a the zero-order correlation between LotteryWin and AKS would increase our confidence that there is no direct relationship between these two. However, it might still be that there is a direct positive relationship between LotteryWin and AKS and an indirect negative relationship through some other variable not included in this model.

For instance, if AKS is family income, there is an effect of income on winning the lottery (you need money to play the lottery), and an indirect relationship through job status (if you win the lottery you stop working). In that case the zero-order correlation might be nill even though there is a direct relationship between these two. This would imply that the assumption that LotteryWinning is exogenous does not hold, but we wouldn't know.

- What does it mean to say that the effect of financial deregulation on the rate at which firms go bankrupt is “bounded from above” at 2 percentage points?
 1. The effect is 2 percentage points, and it's a positive effect
 2. The effect is 2 percentage points, and it's a negative effect
 3. The effect is at least as large as 2 percentage points
 4. **The effect is no larger than 2 percentage points**
 5. If we're willing to make an additional, stronger set of assumptions, the effect would be larger than 2 percentage points, but without those assumptions it's bounded to be lower.