Samuel Snelson

SOC 723 Spring 2023

24 February 2023

**Homework for Chapter 11: Causality with Less Modeling**

1. 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.
   1. A general strategy which sidesteps the need to make numerous statistical adjustments on backdoor paths is identifying front door paths by isolating exogenous sources of variation on the treatment. The challenge here is find such a source of variation that is actually exogenous to the other confounders and the outcome. For instance, we might consider a particular piece of policy or legislation that requires or ceased campus closures.

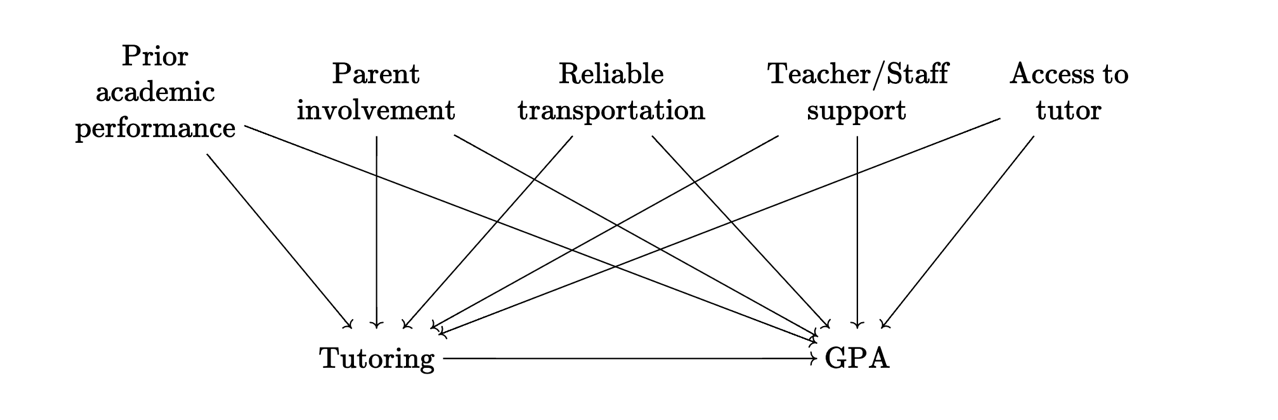
Figure 1: DAG for campus re-opening and COVID-19 cases

Diagram

Description automatically generated

1. Intuitively, why would a method that isolates front doors allow you to ignore back doors related to unmeasured variables?
   1. This method is shown in the general form of the above figure. The logic of isolating front doors follows from the issue of there being so many observed and (more troublesome) unobserved confounds. If the exogenous source of variation is actually independent of the confounders and the outcome, then we can conveniently apply the logic of variation in relationships. A portion of the variation in the treatment is attributable to the exogenous variable (and also to the confounders). If we condition on the treatment for which variation is a function just of the exogenous variable, then we can regard the confounders as an independent causal influence on the outcome (but not the treatment).
2. On robustness tests:
   1. What are robustness tests?
      1. Robustness tests are a general strategy at testing a models’ assumptions and assessing the influence of a models’ assumptions on its estimates. If we assume that a set of variables are or are not related or are related in a specific way (e.g., mediation/fork-confound or collider), we can test whether that is the case (e.g., if we assumed something was a simple collider but association is still transmitted through the path after adjusting for the “collider,” then we would reassess the relationships).
   2. What is the purpose of conducting a robustness test?
      1. As suggested above, the purpose of conducting a robustness test is to test the assumptions we make about a model.
   3. What are placebo tests?
      1. Placebo tests are a particular type of robustness check where we test an assumption that there is no relationship between a set of variables. As the figure in question 1a implies the exogenous source of variation on the treatment is not related to the confounders. If they were demonstrated to be associated, then the exogenous variable wouldn’t actually be exogenous (either directly or indirectly).
3. Suppose you want to study the effect of attending tutoring sessions on grade point averages (GPA). List at least five variables that impact both attendance of tutoring sessions and students’ GPA. Is it feasible to measure and control for all of the variables?

Figure 2: DAG for tutoring and GPA



* 1. Here is a visual of some variables which are plausible related to both attendance of tutoring sessions and GPA. In the context of an educational evaluation project, one might have access to prior academic performance (via self-report or tutor second-hand reports). It seems more unlikely that one would have straightforward access to parental involvement in a child’s education, how reliable a student’s transportation is to tutoring sessions and school generally, the degree of support a student gets from their teachers and other staff, and those students who didn’t get access to the tutor for various reasons.

1. Describe partial identification in your own words.
   1. Partial identification is an approach to accounting for uncertainty around our assumptions about models. Researchers are often more confident about some relationships than others (based on existing knowledge). In order to incorporate this uncertainty into our identification, we could apply the assumptions we are confident about and then consider a number of assumptions for other relationships. Applying different assumptions and then estimating statistics on the relationships provides a wider view of the possible effects in the context of our uncertainty. If we observe a particular treatment effect and think that the adjustment of a backdoor path could change the treatment effect in a more or less predictable way, then we can express the resulting effect in light of the original estimate and uncertainty after adjustment.
2. Pick any causal diagram from the book other than Figure 11.2.
   1. Reproduce that diagram here.

Figure 3: Figure 16.1 in *The Effect*

Diagram

Description automatically generated

* 1. 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).
     1. In this situation, we are interested in the causal effect of, at the country level, being visited by the German chancellor on subsequent trade with Germany. All variables besides the predictor and outcome are fork-confounders and therefore related to the predictor and outcome. Each confounder is independent of one another, so we can go with CurrentPolitics and HistorywithGermany.
  2. What variables would you need to control for that will eliminate any relationship between the two variables (you might not need any).
     1. With the structure of the DAS as given, we would expect no relationship between CurrentPolitics and HistorywithGermany.
  3. 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?
     1. Assuming I correctly estimated the relationship between CurrentPolitics and HistorywithGermany and did find a non-zero relationship, I would conclude that the variables were not actually independent. It is likely that there is a backdoor between them through some common cause (given we’re talking about Germany, politics and historical ties with Germany are likely not independent).

1. 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.
      1. The only modification I would make on this is that I would say the effect “may” be larger (or smaller depending on assumptions).