Final - Fall 2024

ECON 5783 — University of Arkansas

1. **(Conditional Expectation Function)** (10pt) Say a researcher wants to model the conditional expectation of wages given gender and college degree. Why is it not 'fully flexible' to include an indicator for gender and an indicator for college degree?
2. **(Selection on Observables)** For the following examples, I will describe an observational study where we observed 'treatment' D_i and outcomes y_i. Let D_i is an indicator variable that equals 1 if a worker went to college; and y_i a variable measuring the worker's health at age 40 Answer the following:
    i. (15pt) Say you use a difference-in-means estimator. What way do you think your ATT would be biased (positively or negatively)? Explain why.
    ii. (10pt) Name a covariate or two that you would think would be really important to control for to make a selection-on-observables argument. Explain why.
3. **(Instrumental Variables)** Say you are in the setting where you have a *randomized* offer to customers for a discounted membership (Z_i). Let D_i denote an indicator for being a member and y_i denote the amount purchased by the customer.
    i. (5pt) Say your discount was very small and it induced very few additional people to sign-up (that wouldn't have signed up anyways). Why might this be a problem for estimating the causal effect of membership on purchase amount?
    ii. (5pt) How might you estimate the proportion of compliers in this setting?
    iii. (5pt) Your boss wants to know *who* the compliers are so that they could better target their discount (not wasting money on people who would sign-up anyways). Describe how you could learn characteristics of the complier group using a 2SLS regression.

4. **(Regression Discontinuity Design)** Say you want to evaluate a policy that guarantees admission to a flagship state school for in-state students who score a certain SAT score or above (you can assume they would go to a lower-quality school if they don't get in). Your outcome is the student's salary at age 30.
    i. (10pt) Describe in words what we must assume in this context for our RDD estimator to identify the causal effect of this policy.
    ii. (15pt) Why might you be concerned that students can take the SAT multiple times? How might you check for this in the data?
5. **(Difference-in-differences)** We consider the roll-out of a COVID-19 lockdown across some states, but not others. For $t \geq 0$, $D_{it} = 1$ for states in the treatment group, and $D_{it} = 0$ for the control states. For $t < 0$, $D_{it} = 0$ for everyone. We're interested in the effect on economic activity Y_i, and will use difference-in-differences.
    i. (5pt) Write out the regression for how you would estimate a single treatment effect in the post-period for the treatment.
    ii. (10pt) Write out what you would need to assume for this estimate to be interpreted causally.
    iii. (10pt) Say you estimate an event-study and plot the estimates. Comment about the pre-trends in relation to the parallel trends assumption.

Figure 1 — Event-study estimates