Advanced Econometrics Problem Set 1

(Due Monday, October 21 in class)

1. 1. In the potential outcomes framework, suppose that program eligibility is randomly assigned but participation cannot be enforced.To formally describe this situation,For each person i, z_i is the eligibility indicator and x_i is the participation indicator. Randomized eligibility means z_i is independent of (Y_{0i}, Y_{1i}) but x_i might not satisfy the independence assumption.

   i. Explain why the difference in means estimator is generally no longer unbiased.

   ii. In the context of a job training program, what kind of individual behavior(s) would cause bias?

2. In fact, we can think of the policy variable, w, as taking on many different values, and then $y(w)$ denotes the outcome for policy level w. For concreteness, suppose w is the dollar amount of a grant that can be used for purchasing books and electronics in college, $y(w)$ is a measure of college performance, such as grade point average. For example, $y(0)$ is the resulting GPA if the student receives no grant and $y(500)$ is the resulting GPA if the grant amount is \$500.
   For a random draw i, we observe the grant level, $w_i \geq 0$ and $y_i = y(w_i)$. As in the binary program evaluation case, we observe the policy level, w_i, and then only the outcome associated with that level.

   i. Suppose a linear relationship is assumed:

   $$y(w) = \alpha + \beta w + \nu,$$

   where $y(0) = \alpha + \nu$. Further, assume that for all i, w_i is independent of ν_i. Show that for each i, we can write

   $$y_i = \alpha_i + \beta w_i + \nu_i,\ E(\nu_i | w_i) = 0.$$

   ii. In the context of i, how would you estimate β (and α) given a random sample? Justify your answer.

   iii. Now suppose w_i is possibly correlated with ν_i, but for a set of observed variables x_{ij},

   $$E(\nu_i | w_i, x_{i1}, \ldots, x_{ik}) = E(\nu_i | x_{i1}, \ldots, x_{ik}) = \eta + \gamma_1 x_{i1} + \ldots + \gamma_k x_{ik}.$$

The first equality holds if w_i is independent of ν_i conditional on $(x_{i1}, \ldots, x_{ik})$ and the second equality assumes a linear relationship. Show that we can write

$$y_i = \alpha_i + \beta w_i + \gamma_1 x_{i1} + \ldots + \gamma_k x_{ik} + \nu_i,\ E(\nu_i | w_i, x_{i1}, \ldots, x_{ik}) = 0.$$

What is the intercept ϕ?

iv. How would you estimate β (along with ϕ and the γ_j's in part iii)? Explain.

3. 3. The Current Population Survey (CPS) refers to any of the monthly surveys conducted by the US Census Bureau throughout the year, although the March CPS - considered the beginning of the annual survey cycle - is the most significant, and is the data used in this assignment. Broadly, the CPS collects cross-sectional employment data of the participating households, allowing for regression wherein the independent and dependent variables are associated with the same point in time. In this problem, we will explore the relationship between educational attainment on earnings. There are numerous sites you can download the CPS data from. One source among many is http://ceprdata.org/cpsuniform-data-extracts/cps-outgoing-rotation-group/cps-org-data/

   i. Create a figure with hourly wage plotted against educational attainment for men in the US between the ages of 30 and 40 in March of 2019.

   ii. Estimate the CEF using OLS. Why is the CEF linear in this case and show that OLS will generate a consistent estimate of the CEF?

   iii. We wish to estimate the causal effect on earnings of college attendance relative to only completing high school. Please use the framework of the Rubin Causal Model to assess if your comparison from question ii gives you a causal estimate, e.g., what is D_i, what is $E[Y_i(0)|D_i = 1]$, etc.