The University of Texas at Austin
ECO 348K (Advanced Econometrics)
Prof. Jason Abrevaya
Fall 2016

PROBLEM SET #2 (due Thursday, September 29th, 8am)

1. Consider the following question:

    What is the causal effect of studying Economics rather than History on the salaries of UT-Austin graduates?

    For simplicity, you may assume that all UT-Austin students are either Economics or History majors.

    (a) What is the outcome variable Y_i and the treatment D_i?

    (b) Define the counterfactual outcomes Y_{0i} and Y_{1i}.

    (c) What plausible causal channel runs directly from the treatment to the outcome?

    (d) If you did a simple comparison of average outcomes by treatment status, what would be a possible source of selection bias? Which way would you expect the selection bias to go, and why?

2. Repeat the parts of Question #1 for the following question:

    What is the causal effect of divorce on life expectancy (i.e., age at death)?

    Think of the population as ever-married individuals. For simplicity, assume that there are no second marriages; once you are divorced, you are divorced forever.

3. Back to Question #1, consider the following initiative created by the outstanding Economics department chair: For incoming UT-Austin students, 20% are randomly selected to receive a $5,000 payment if they major in Economics (but no such payment if they major in History). You remain interested in the causal effect of studying Economics on graduates' salaries.

    (a) What are the outcome Y_i, treatment D_i, and instrument Z_i here?

    (b) Do you think the monotonicity assumption is likely to hold here? Explain why or why not.

    (c) For this subsidy program, explain why you might be concerned with the assumption that the effect of Z_i on Y_i is only acting through D_i.

    (d) What four quantities would you need to know in order to estimate the local average treatment effect (LATE) of majoring in Economics upon salary?

    (e) If a UT-Austin administrator cares only about the salaries of UT-Austin graduates, explain why the intent-to-treat (ITT) effect may be more relevant from a cost-benefit perspective.

4. For this question, use the dataset **bertrandmull.dta** from the Canvas site. These data are from a randomized experiment described in the paper "Are Emily And Greg More Employable Than Lakisha And Jamal? A Field Experiment On Labor Market Discrimination" (*American Economic Review*, 2004) by Marianne Bertrand and Sendhil Mullainathan. In the study, the authors sent 4,870 fake resumes for advertised job openings in Boston and Chicago. The resumes differed on various dimensions, including the names of applicants. Some applicant names were distinctly "black sounding" (like Lakisha and Jamal) and some were not (like Emily and Greg). The question of interest was whether the resumes with black-sounding names would receive fewer interview requests than those with non-black-sounding names. Black-sounding names are denoted by a value of 1 for the **black** indicator variable in the data.

   (a) Check for balance of the randomized experiment. The other explanatory variables of interest are whether the applicant is female (**female**), whether the applicant has computer skills (**computerskills**), the number of previous jobs (**ofjobs**), years of previous experience (**yearsexp**), and a categorical variable for education level (**education**). For all the variables but **education**, use regression analysis in order to check for balance. What are the relevant p-values for determining balance with respect to each variable?

   (b) Running a regression of **education** on **black** is not really appropriate since **education** is a categorical variable. How could you check for balance for this variable?

   (c) The outcome of interest is the indicator variable **call** (1 if there is an interview callback). Using a simple linear regression, what is the estimated average treatment effect of **black** on the probability of an interview callback? How does the ATE compare to the baseline (control) callback probability?

   (d) Now, add all of the other explanatory variables into the regression, including indicators for the educational categories. How does the estimated (treatment) effect of **black** compare to the previous regression? Are you surprised by what you find in the comparison?

   (e) From the regression in part (d), what other effects of interest do you find? Does there seem to be a treatment effect of gender upon callback probability?