

**ECON 211B**  
**Winter 2017**

**Econometrics**  
**Homework 2**

Please typeset all your work cleanly using LaTeX. All figures and tables should be clearly labeled and have notes that clearly describe their contents.

Please evaluate the effect on annual earnings of the job training provided to participants in the National Supported Work (NSW) Demonstration study. In this study a sample of men were randomly assigned to either receive or not receive job training. Baseline earnings were measured in 1975 which is the year before the they were randomly assigned to either the treatment or control group. Earnings were also measured in 1978 which is several years after the treatment started. More detail on the NSW are available in Robert Lalonde, "Evaluating the Econometric Evaluations of Training Programs," American Economic Review, Vol. 76, pp. 604-620.

The dataset for this homework is called **NSW\_PSID.csv** and includes the following variables: treatment indicator *treat* (1 if treated, 0 if not treated), *age* (age in years), *education* (years of education), *black* (1 if black, 0 otherwise), *hispanic* (1 if Hispanic, 0 otherwise), *married* (1 if married, 0 otherwise), *nodegree* (1 if no high school degree, 0 otherwise), *re75* (earnings in 1975), and *re78* (earnings in 1978). There is one other important variable named *data\_id* which takes on two values. The value "*Lalonde Sample*" means the observation is part of the actual experiment. The value "*PSID*" indicates this is an observation drawn from the Panel Study of Income Dynamics.

1. Provide evidence that the randomization worked by comparing the means of the sample characteristics in the treatment and control group. Please create a clean table that includes columns with the means of each group, the difference between the two groups and the p-value of the difference. The table should be comprehensible on its own. Include a footnote for the table with a description of the dataset. (Hint: Be sure to use the *data\_id* variable to restrict to the right subsample of the dataset). A table

of this form is usually the first one in an empirical study intended to recover a causal estimate. Is the table consistent with the randomization being correctly implemented?

2. Create a table that presents your evaluation of the effect of the NSW experiment on earnings. In the first column present the raw difference in means between the treatment and control group then sequentially add covariates. In the last column include estimates from a difference in differences model. Do the estimates change much as you add covariates or go to a differences in differences estimate. Why or why not? What does this tell you?
3. Do you believe your estimates of the treatment effect are unbiased of the true treatment effect? Why or why not?
4. Estimate and plot the density functions of the 1978 earnings for the treatment and control group.

Now check if you can recover a causal estimate using a non experimental sample. To create the non experimental sample Lalonde took the actual treatment group from the NSW experiment then tried to create a comparison group out of the PSID. To get this sample together take the entire dataset for this homework and restrict to the people from the Lalondes sample as your treatment group and use people from the PSID as your comparison group. To get the correct subsample use ***data.id*** variable in combination with the ***treat*** variable.

5. Create a balance table similar to the one you created to answer question 3. If you have set up your code carefully you can use the same code with the new dataset. Do the treatment group and the comparison group from the PSID look similar prior to treatment? Why or why not? Does this concern you? Why?
6. Create table a table with estimates of the treatment effect from different regression models (similar to what you did in answer to question 2 above). Does the estimated treatment effect you estimates in answer to the prior question match what the experi-

ment tells us the actual effect is? What does the fact that adding variables substantially changes your estimates tell you?

7. Do you believe these estimates have a causal interpretation? Why or why not? If not what assumption do you think failed and why.
8. Estimate the job training effect using the propensity score dropping off the observations with very high and low propensity scores. Be sure to use a flexible implementation of the propensity score. Do not use a canned procedure instead estimate the propensity score using logit and polynomials in the covariates. Plot the histogram of propensity scores for both groups. Then trim the sample and use the binning approach to estimate the treatment effect. Be sure to trim the sample to account for lack of common support. Does this result in estimates closer to the experimental benchmark?
9. Estimate the job training effect using OLS but with a sample restricted using the same trimming as above. What did you find?