Problem Set #2

Prof. Matt Reimer*

Econ A390: Methods for Public Policy Evaluation

Please answer all the questions below within this Microsoft Word document. For all parts, your answer should consist of a concise explanation, along with any STATA commands and results. See these instructions for importing STATA output and graphs into Microsoft Word documents. You are encouraged to work as a group; however, there may be no more than three students per group. Everyone is responsible for submitting one typed file for grading on Blackboard under Assignments. Please indicate the members of your group at the top of the file. You might find it helpful to read through the assignment, then read your class notes and textbook, and only then try and do the questions. *Attempting this assignment without consulting your notes and the text may be hazardous to your health and your grade!!*

Motivation

In this problem set, you'll carry out a propensity score matching analysis, check for balance, and compare your results to what you would have gotten using regression. You?ll also do regression on your matched sample, which is one way of doing the "doubly-robust" regression.

This problem set is based on a LaLonde (1986) paper on the National Supported Work (NSW) Experiment, which was a randomized experiment carried out in the mid-1970's to examine whether a job training program improved the earnings of low-skilled workers. Since it was a real experiment, there was a control group and a treatment group whose treatments were randomly assigned, which means that it could be used to consistently identify the causal effect of receiving job training.

Lalonde took the data on the treated cases from the experiment (those who participated in the job training program), threw out the real experimental controls, and constructed a new set of controls from observational data, which is what economists and sociologists normally use. More specifically, he took data from the Current Population Survey (CPS) and the Panel Study on Income Dynamics (PSID) to build his new control groups. He then ran a variety of regression models to estimate the effect of job training, and he compared his results to the randomized experiment.

A decade later two more economists, Dehejia and Wahba (1999), came along, armed with their new knowledge of counterfactual causality and matching methods. They wanted to see whether these new methods lived up to the hype. They took the same NSW data as Lalonde and also gathered the same PSID and CPS controls as he did. However, instead of using regression methods, they used matching and then compared their results to the experimental results.

*Telephone: 1-907-786-5430, e-mail: *mreimer2@alaska.edu*.

Following this grand tradition, you will also be analyzing the NSW data with matching and regression methods. LaLonde and Dehejia and Wahba used 6 different observational control groups, but you?ll just two: the experimental control from Lalonde's study and the controls from the PSID, a nationally-representative survey of income and employment. For more detailed information regarding the experiment, read the Dehejia and Wahba paper.

Data

Income in 1978 (one or two years after the job training) is our outcome variable here, and the treatment is whether the individual received job training. The data contain a number of demographic and work-related variables designed to control for other factors that might affect income. The variables included are in Table 1. All the data that you use in this problem have the same variables and all focus on males (i.e. there are no females in the dataset). Note that the experiment was conducted between 1975 and 1977, so that data from 1974 and 1975 are pre-treatment (or baseline), and data from 1978 are post-treatment.

Variable	Description
treatment	1 if in NSW Job Training Program; 0 otherwise
age	Age in years
education	Education (# years)
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
re74	Real earnings in 1974
re75	Real earnings in 1975
re78	Real earnings in 1978 (outcome variable)

Table 1: Variable descriptions for the datasets.

In the following problems, the controls variables will be: *age*, *education*, *black*, *hispanic*, *married*, *nodegree*, *re74* and $re75$.

Problem 1: Experimental Results

Load the nsw_dw.dta dataset into STATA. This is a subset of Lalonde's experimental treatment and control group dataset, used by Dehejia and Wahba (1999).

A. Check for balance between the treated and control groups.

i) If this is a randomized experiment, what do we expect to observe in the results?

ii) Is there a statistically significant difference in any of the characteristics? If so, does this indicate that the experiment was not randomized?

B. Generate the experimental estimate of the effect of job training using $re78$ as the outcome variable, with and without controls.

i) What is the estimated effect of the program? Did adding the controls make a difference in the estimate?

ii) Why would you want to add controls even in the case where you have experimental data?

Problem 2: Observational Controls–Regression

Drop all the observations for the experimental control group (*drop if treatment==0*) from the nsw_dw.dta dataset and append the dataset psid_controls.dta. These are the non-experimental "control group" observations that Lalonde used to compare with his experimental results. Let's see if we can use some of the tools we've discussed in class to identify the causal effect of job training without having the benefit of a control group generated under random assignment.

A. Check for balance between the treated and control groups.

i) Since these groups were not created by a randomized experiment, what do we expect to observe in the results?

ii) Is there a statistically significant difference in any of the characteristics?

B. Estimate the causal effect of job training via regression, using $re78$ as the outcome variable, with and without controls.

i) What assumptions must be made in order for regression to identify the causal effect? Why do we think adding control variables to the regression will help us?

ii) What is the estimated effect of the job training program? Did adding the controls make a difference in the estimate? How so?

iii) How does your estimated effect of the job training program compare to the results from the experimental dataset? Do you have confidence in the assumptions you identified in part (i)?

iv) Based on the table of means for the treatment and control group that you generated above, is this regression result surprising? What do you think is happening?

Problem 3: Observational Controls–Matching

To answer the following questions, you will need to download and install an add-on package in Stata. At the Stata prompt, type: `ssc install psmatch2, replace`. For details on how to use the `psmatch2` command, type `help psmatch2` at the Stata prompt. Continue to use the same data as you did for Problem 2.

A. Perhaps matching methods will prove to be a more useful tool than regression for estimating the causal effect of job training.

i) What assumptions must be made in order for matching to identify the causal effect?

ii) Why is exactly matching treatment and control observations using our control variables problematic? How does matching on a propensity score address this problem?

iii) Estimate and predict the probability of being in the treatment group, as a function of the control variables. Plot your predictions for both the treated and the control group. Discuss your findings with regard to common support. *Hint: Use the following commands in Stata*:

```
logit treat age education black hispanic married nodegree re74 re75
predict phat
histogram phat, by(treat)
```

B. Estimate the causal effect of job training by matching treatment and control observations on the propensity scores you estimated in Part A. Use nearest neighbor matching with replacement.

i) What is the estimated causal effect of job training on earnings? How does this compare to the results from the experimental dataset? *Hint: Use the following command in Stata*:

```
psmatch2 treatment, outcome(re78)pscore(phat)
```

ii) Check the balance between the treatment group and the control group formed by nearest neighbor matching. How does the balance after matching compare the balance before matching? Are there any variables in particular that look unbalanced even after matching? *Hint: Use the following command in Stata after the* `psmatch2` *command*: `pstest age education black hispanic nodegree re74 re75, both`

iii) How do the propensity scores for the treated observations compare to the propensity scores for their matched counterparts in the control group? Do all treated observations have a "good" match? *Hint: Use the following command in Stata*: `histogram _pdif`

iv) Because we are "matching with replacement", some units in the control group may serve as a match for multiple units in the treatment group. Is this true for the analysis above? If so, how many units from the control group are used as matches? How often are they used? Do the frequently used control units have high or low propensity scores? Why would this be the case? *Hint: Use the following commands in Stata*:

```
tab _weight if treatment==0
scatter _weight _pscore if _weight>1 & _weight!=.
```

v) Do your answers in parts (ii)-(iv) raise or lower your confidence in the matching results in part (i)? Why?

C. Some people may or may not be happy with the results from Part B. Explore different specifications for your matching results by relaxing/changing some of the following implicit assumptions made in Part B:

1. Explore different combinations for the control variables used to estimate the propensity scores.
2. Explore different matching techniques: e.g. caliper or kernel matching; with or without replacement; with or without common support.

Only report the results from your final specification.

i) Explain why/how you chose your final specification. Do you have more/less/the same confidence in the matching results from your new specification relative to Part B? Why?

D. Discuss the validity (focus on internal and statistical validity) of your analyses in Problems 2 and 3. Did the regression and matching analyses provide similar/consistent results? Why or why not. Were the regression and matching analyses successful in estimating the causal effect of job training that we found in Problem 1? Why or why not?

References

Rajeev H Dehejia and Sadek Wahba. Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs. *Journal of the American Statistical Association*, 94(448):1053–1062, 1999.

R J LaLonde. Evaluating the Econometric Evaluations of Training Programs with Experimental Data. *The American Economic Review*, pages 604–620, 1986.