**Homework # 3**

Please email me your answers (in Word or pdf) and attach the log file and do file. Name all three files as “HW4\_Yourlastname”. Please do not copy and paste from the stata output window. Make sure that your do file runs from the beginning to the end; de-bug the program until it runs smoothly.

**Matching**

This problem makes use of data on National Supported Work Demonstration job-training program conducted by the Manpower Demonstration Research Corporation in the mid 1970s. The impact of training was analyzed by LaLonde (1986). LaLonde used data on NSW to show non-experimental estimates generated through typical econometric techniques are quite different from the results obtained from experimental data. Dehejia and Wahba (1999) showed that matching techniques can be used to obtain results that are similar to experimental estimates. The data have been also used by Heckman and Hotz (1989) and Smith and Todd (2005). The user-written Stata program psmatch2.ado can help you in this exercise.

**Data set:**

File name: nsw.dta. The dataset contains the following variables:

sample –

1 for the experimental sample (the union of the treatment and control groups)

2 for the comparison group from the Current Population Survey (CPS)

3 for the comparison group from the Panel Study of Income Dynamics (PSID)

treated –

1 for the experimental treatment group

0 for the experimental control group

Missing for the comparison groups generated from the PSID and CPS

age – age in years

educ – years of schooling

black – 1/0 black

hisp – 1/0 Hispanic

married – 1/0 married

nodegree – 1/0 no high school degree

re74 – real earnings in 1974

re75 – real earnings in 1975

re78 – real earnings in 1978

dwincl – 1/0 included in the Dehejia and Wahba sample.

early\_ra – 1/0 included in the early random assignment sample in Smith and Todd (2005)

1. Focus first on the experimental sample (treated=0,1). Compare variable means for the treatment and control group. Test whether the means are statistically significant different from one another at the 5% level for a two-sided test (show your results in a table). If this is a randomized experiment, what do you expect? Now generate an experimental estimate by running a regression of earnings in 1978 on the “treated” variable along with age, age squared, education, black, Hispanic, married, no degree and earnings in 1974 and 1975. Explain why you want to include covariates in this regression even with experimental data.

2. Now we want to analyze the quality of the PSID comparison group. Create a variable called *d* that equals 1 for all experimental observations and 0 for all PSID. Compare the means of the experimental sample and the non-experimental PSID sample. What do you think about simply using the PSID sample as the comparison group? What assumptions do we need to make for matching based on a propensity score to give us unbiased estimates of the effect of program participation?

3. Estimate two sets of propensity scores for the PSID comparison group using a probit model (*d* is your depvar). The first set should include the variables age, age squared, education, black, Hispanic, married, and no degree (score set I). The second set should contain the variables in the first set plus earnings in 1974 and 1975 (score set II). Explain what is going on with the 135 observations that are “completely determined” (see note at the end of Stata program)?

4. Examine the distributions of estimated propensity scores for the experimental and comparison group samples (sum, d). What do the descriptive statistics suggest about the common support condition in these data? What do they suggest about the comparability of the PSID comparison group?

5. Construct histograms of the estimated propensity scores for the combined experimental treatment and control groups and for the PSID comparison group. Using a command such as: histogram phat, start(0.0) width(0.05) by(d, col(1)) will make it easy to compare the histograms, where phat is the estimated propensity score and where d = 1 for the experimental sample and d = 0 for the comparison group sample. What do the histograms suggest about the common support condition in these data? What do they suggest about the comparability of the PSID comparison group? Why is it important to have overlap in the distribution across groups?

6. Drop the experimental treatment group. Note that if there is no selection into the control group, the matching estimator should be close to zero in this case. Construct non-experimental biased estimates for both sets of estimated propensity scores using single nearest neighbor matching *without* replacement (use the option “common” to impose common support condition). What can you say about the difference between actual control group and the comparison group from PSID? Does the score set II perform better (i.e., result in lower bias estimates) than the score set I?

7. Repeat Problem 6 but using single nearest neighbor matching *with* replacement. How much does allowing comparison group observations to be reused in the matching change the estimates in this context? Explain. Perform a balancing test by quartiles of the propensity score distribution (pstest). Compare results before and after matching from the balancing test.

8. Generate bootstrap standard errors for the estimates obtained using the score set II in Problem 7. First use 10 replications then use 100 replications. Discuss your findings.

9. Estimate a weighted regression that includes a control group dummy variable and the same conditioning variables as in the score set II. The weights equal one for the observations from the experimental control group. The weights equal the number of times each observation gets matched for the comparison sample. Helpfully, psmatch2 keeps exactly these weights around in the variable “\_weight” for you to use in estimating this regression. You should tell Stata to interpret these as frequency weights. How does the regression-adjusted matching estimate differ, if at all, from the corresponding estimate obtained in Problem 7?

10. Create propensity score matching estimates using the score set II and kernel matching with a Gaussian (normal) kernel and bandwidths of 0.02, 0.2 and 2. Describe the resulting impact estimates. How do the estimates change as the bandwidth increases? How do the estimates differ from the single nearest neighbor matching with replacement estimates obtained in Problem 7?

11. Repeat Problem 10 but using local linear matching rather than kernel matching. How do the estimates change as the bandwidth increases? How do the estimates differ from the single nearest neighbor matching with replacement estimates obtained in Problem 7 and the kernel matching estimates obtained in Problem 10?