Problem Set 3 Treatment Effects and Matching

Read in the databank "matching.tlb" from the L-drive into TSP. The data are from the National Supported Work Demonstration (NSW), a job training experiment which took place in the U.S. in the mid-1970s (see LaLonde (1986) and Dehejia and Wahba (1999, 2002)). As matching methods require comparison with another sample, the experimental sample (N = 445) is complemented by non-experimental controls from the Current Population Survey. The data set contains the following variables in the given order:

Variable name	Variable description
treat	Dummy variable for treatment (1 if assigned to job training)
age	Age in years
educ	Years of education
black	Dummy for skin color (1 if black)
hisp	Dummy for Hispanic origin (1 if hispanic)
married	Dummy for marriage (1 if married)
nodegree	Dummy for no schooling degree (1 if no schooling)
re74	Real income 1974
re75	Real income 1975
re78	Real income 1978
nsw	Sample dummy (1 if experimental data)
nexp	Sample dummy (1 if non-experimental data)

 Do some descriptive statistics. For example, compare the experimental and the non-experimental control groups with respect to observed pre-treatment variables. Then compare the non-treated individuals in the two control groups with the treated individuals. What do you observe? Note: You do not need to check for significant sample differences. Estimate the ATE as a difference-in-means estimator for the experimental and for the non-experimental sample. The outcome variable is real earnings in 1978 (*re78*). Under which assumption this estimator will give you the "true" treatment effect?

In the following we will estimate the ATE and the ATT using only non-experimental data.

- 3. Propensity score matching
- (a) Estimate the propensity score for training doing a probit regression of the treatment dummy on a *constant, age, age2, educ, nodegree, married, black, hisp, re74, re75*. Analyze the results. Then predict the propensity score for both treated and non treated individuals. What is the average propensity score in the treatment group? And in the comparison group?
 Hint: The variable *age2* means the second order polynomial in age and has to be generated.
- (b) Before matching we want to check the assumption of common support between treatment and control group. To prove this, analyze graphically the density distribution of the propensity score in both groups. What do you conclude? Hint: Use the TSP command *hist*.
- (c) Estimate the ATT using nearest neighbor matching with one nearest neighbor and with replacement. Interpret the results.
- (d) In the following we want to check whether the matching procedure from above was successful. To do this, determine the standardized bias before and after matching. Analyze the results. Which other approaches do you know to assess the matching quality?
- 4. Regression based on estimated propensity score

Estimate the ATE from a simple OLS regression of *re78* on a constant, the treatment dummy, and the estimated propensity score. What is the idea behind this regression?