Problem Set 4 - Matching and Propensity Score Matching

Refer to the Dehejia-Wahba paper for this problem set.

The sample you have in this paper is rows 3, 4, and 5 of Table 1 in the paper, namely the "RE74 subset" of the experimental data (i.e., experimental treatment and controls units) and the PSID-1 comparison group (which is a non-experimental comparison group).

Part I. Direct Matching

1. Tabulate the covariates by treatment and control group for the RE74 sample, indicating also the differences in means and standard errors, and do the same for the combination of RE74 (experimental) treated and PSID (non-experimental comparison units). How do the non-experimental controls compare to the experimental ones to the treatment group?
2. Direct matching: begin by using nnmatch to match the RE74 treated group to the PSID comparison group, using RE78 as the outcome and RE74 and RE75 as the covariates, using 1:1 matching. What is the estimated treatment effect? How does this compare to the experimental treatment effect (using the RE74 treatment and control groups). Plot RE74 for the treatment observation against the average of RE74 for matching comparison units. Do the same for RE75. How well are these being matched? (Hint: This is tricky, but here's a start on how to do it. When you run nnmatch, use the "keep" option to save the matching info. Load this. Type "list id index treat" and you'll see that many treatment observations are matching to multiple comparison units. Why? Because even though you've asked for 1:1 matching, there are many ties with these variables, so when there's a tie all comparison units are used. Keep just the treat==1 observations, and next collapse the data by id. When you're done you should have one row for each treatment observation with information about the matches.)
3. For the previous matching procedure, how well is education matched if you match on the two income variables? Give an average difference, and plot some relevant figures.
4. Redo the matching using as covariates RE74 RE75, educ, black, hispanic, married, and quadratic terms of the two income variables. Assess the quality of the matches for each covariate, and compare the estimated treatment effect to the treatment effect you would estimate from the experimental treatment and control groups.

Part II. Propensity Score Matching

5. Begin by estimating the propensity score using the covariates specified in Question 4 above. After estimating the propensity score check whether the covariates are balanced across treatment and comparison groups and plot a figure that depicts the balance of the estimated propensity score across treatment and comparison groups. (Hint: check out pstest and psgraph.
6. What is the estimated treatment effect and how does it compare to the experimental treatment effect?
7. Based on covariates that were not balanced in step 5 try to improve your propensity score specification and redo steps 5 and 6.

Part III. Propensity Score Weighting

8. Now use your specifications in 5 and 7 above to estimate the treatment effect via inverse propensity score weighting. Bonus points: bootstrap the standard errors, remembering that this is a two-step procedure.