Problem Set 9

This problem set is shorter to allow you time to work on your empirical project. Consequently, after discussing the questions of this problem set in class, the rest of the class is left free for you to ask your class teacher for any help you might need with the project.

1. Find an academic article that uses an RDD approach. As before, the easiest way to do this is through google scholar. I recommend searching for: "regression discontinuity" and "causal effect:. You can also use the advanced search options to look in specific economics journals (the same ones I mentioned before). Discuss the causal question they want to answer, is it a sharp or fuzzy design, what is the running variable, what do they find, and any criticisms you have.

2. For this question, we're going to use the Lalonde data again, as we did in Problem Set 8. However, instead of using DiD we are now going to see if matching can recover the causal effect. Recall from the experimental data, we estimated a causal impact of \$886. Here, we use a slightly expanded dataset, you can download this from blackboard, 'Lalonde Observational Big Data'.
   (i) First, run a naive OLS regression. The outcome variable is $re78$, earnings in 1978. Include all other variables as controls.

   We are now going see if matching estimators are different to these OLS results and the experimental results.

   (ii) But first, in order for our matching estimate to obtain a causal effect, what do we need to assume?

   (iii) Why may it not be a good idea to do conventional matching on our entire set of covariates? What do you suggest instead?

   (iv) Estimate and predict the probability of being in the treatment group, as a function of the control variables. Plot the histogram of your predictions for both the treated and the control group. (Hint: hist() will create a histogram of your variable, plot() will plot it) Discuss your findings with regard to the common support assumption.

   (v) Use a matching estimator to obtain the ATET for job training on earnings? (Note that because of the stochastic nature of matches, you can run this estimator several times and get different results. You could try running it a few times and taking an average.) How does your resut compare to the results from the experimental dataset and the OLS estimate?

(vi) Can the result from part (v) be explained by poor matching and therefore unbalanced propensity scores?

(vii) When there is poor overlap (as seen in part (iv)), it can be that some people are used as matches a very large number of times. If this is very extreme, it can lead to our estimates being imprecise. We can check to see how many times people have been used as matches. To calculate the ATET, for each treated person, we find a matching un-treated person. So we want to see if an un-treated person has been used as a match very frequently. Use the following code to create a table which gives you the frequency with which each un-treated person has been used as a match (the first column is the un-treated person's id, the second is the frequency)
as.data.frame(table(ATET$index.control))