**Assignment #3**

**Matching mutual fund managers**

You are studying how mutual fund managers who become hedge fund entrepreneurs change after they transition and are curious about the counter factual. What if they had stayed as mutual fund managers? One possible solution to this problem is to create a matched control sample of non-treated managers who “look just like” the treatment group. The data is organized by dyad-month, where a dyad is unique manager-fund combination. The file is called: mf2hf\_phd2016.dta

1. Before going through a big matching effort first check whether you even need to do this by evaluating whether the treatment group is meaningfully different from the control group?
   1. Do t-tests on the means of all the covariates based on whether they are in the “treated” sample or non-treated/potential control sample. These are: performance (rolling average lag 1), stdv (rolling average standard deviation of performance lag 1), start\_date, log\_age (lag 1), log\_aum (aum=assets under management, lag 1), missing\_aum (lag 1), year\_dum\_\* (a vector of year dummies). Which covariates are statistically different at the 10% level?
   2. Look at the two distributions of propensity score using something like the following code:

*twoway (kdensity pscore if treated==1) || (kdensity pscore if treated==0), legend(label(1 treated) label(2 nontreated))*

Do these distributions look different to you?

1. Let’s say you think the treatment and control group are different enough that matching would be useful. Next you’ll need to generate the propensity score that will be used to facilitate matching.
   1. Run a logit or a probit predicting “treated,” a dummy variable equal to one if a mutual fund manager becomes a hedge fund entrepreneur” based on the covariates. Capture the fitted values from the probit (i.e., the “propensity score”) and store those in a variable called “pscore”.
   2. Study the economic magnitudes of the covariates by looking at the marginal effects (not the raw coefficients). What the baseline probability of being treated. If stdv doubles, how much would that increase the probability of being treated (as a percentage of the baseline probability)?

3. Common support

a. Outliers are hard to match and can skew the common support of the two distributions. Based on the propensity score, trim off outliers below the lowest 1st in the potential control group, and above the 99th percentile for the treatment group. How many observations did you drop?

b. Next, find the common support of the propensity score distributions for the control and treatment group. Eliminate observations that are off the common support. How many observations did you drop?

4. Matching

a. Match treatment group observations to control group observations 1:1 using nearest neighbor matching. One way to do this is with an interactive loop. For example, in STATA you could run a loop starting with:

*forvalues t = 1/`end' {*

and ending with:  *}*. The guts of the loop would involve finding the nearest neighbor for a each treatment. (Try programming this yourself. If you need a hint or some help I’m happy to share my code with you.)

b. Approximately how long (in terms of clock time) did it take to run your matching engine?

c. Now look at the two distributions of propensity score using something like the following code:

*twoway (kdensity pscore if treated==1) || (kdensity pscore if treated==0), legend(label(1 treated) label(2 nontreated))*

Compared to the initial kernel density plot, are the treatment and control groups more similar than they were before matching? Visually do these distributions look similar?

d. Run t-tests on the means of all the covariates in the treatment group and matched sample. Compare these to your initial t-tests. What do you notice?