**Replication 2: Abadie (2005)**

This is a simple replication, but I want you to do this manually, not using a packaged command. In the matching chapter of the mixtape, I discuss the LaLone (1986) paper in the AER studying a job trainings program. The chapter is here:

<https://mixtape.scunning.com/matching-and-subclassification.html>

In this 1986 article, LaLonde took a randomized controlled trial in which volunteers were randomly assigned to a job trainings program or nothing. The treatment was the program. But to illustrate the problem with econometrics at the time, LaLonde then dropped the experimental control group and replaced it with the PSID and the CPS in separate analysis. Both were a random sample of the entire country. The experimental results showed that the treatment effect was positive, but when he used econometric modeling with a *non-experimental* control group, the effect was consistently negative. A couple of papers by economists Dehejia and Wahba in the late 1990s and early 2000s introduced a new method called “propensity score based analysis”. Sometimes it’s called matching and sometimes weighting. That’s because there are numerous ways to use the propensity score. One the propensity score is trimmed, this analysis regularly finds estimates that are nearly the same as the experimental results.

For this project, I want you to implement Abadie’s 2005 Restud on semiparametric Diff-in-diff. Use the data contained in nsw\_pscore.do at the mixtape chapter, which appends the CPS data to the original job trainings program data. Then answer these questions.

1. Calculate a propensity score using the same covariates as used in the mixtape only use a series of polynomials for each one. You will do the following analysis twice: once using a logit, once using OLS, to fit the propensity score.
   1. You will fit a linear probability model (OLS) for one of the following and you will fit a logit for the second.
   2. Fit one propensity score using up to a quadratic for each variable for one set of analysis, and a cubic for a separate set of analysis.
   3. Create a histogram showing the distribution of the propensity score for the treatment and control group. What is the max and min values of the propensity score for the treatment group? What is it for the control group?
   4. Drop all units whose propensity scores are less than 0.1 and more than 0.9 then repeat 1.c.
2. Calculate a before and after first difference for each unit.
3. Construct a weighted difference-in-differences using the first equation at the following substack entry:

<https://causalinf.substack.com/p/callaway-and-santanna-dd-estimator>

Ignore inference issues. We are only going to calculate the point estimate. Compare your answers to that I found in the ipw.do program at section 5.3.5 of the Mixtape. I found an ATT of $1806 or $2006 depending on which weighting scheme I used. Compare your answers to those.