**Assignment 4: Due Sunday, June 14th 2020**

**Directions**: Please turn in your answers on separate paper, typed, and **beautifully written** with **beautiful tables** and **beautiful figures**.**[[1]](#footnote-1)**

**Github repo and summary (worth 2 points)**

1. Download Hansen\_dwi.dta from github at the following address.

use https://github.com/scunning1975/causal-inference-class/raw/master/hansen\_dwi, clear

Create a new github repo named “RDD”. Inside the RDD directory, put all the subdirectories we’ve discussed in class. Post the link to the repo so I can see it’s done as discussed in your assignment. Save the Hansen\_dwi.dta file into your new /data subdirectory. Note: The outcome variable is “recidivism” or “recid” which is measuring whether the person showed back up in the data within 4 months. (**CHECK)**

1. In the writing subdirectory, place your assignment. For the first part of this assignment, read Hansen’s paper in the /articles directory of the main class github entitled “Hansen AER”. **Briefly summarize this paper**. What is his research question? What data does he use? What is his research design, or “identification strategy”? What are his conclusions?

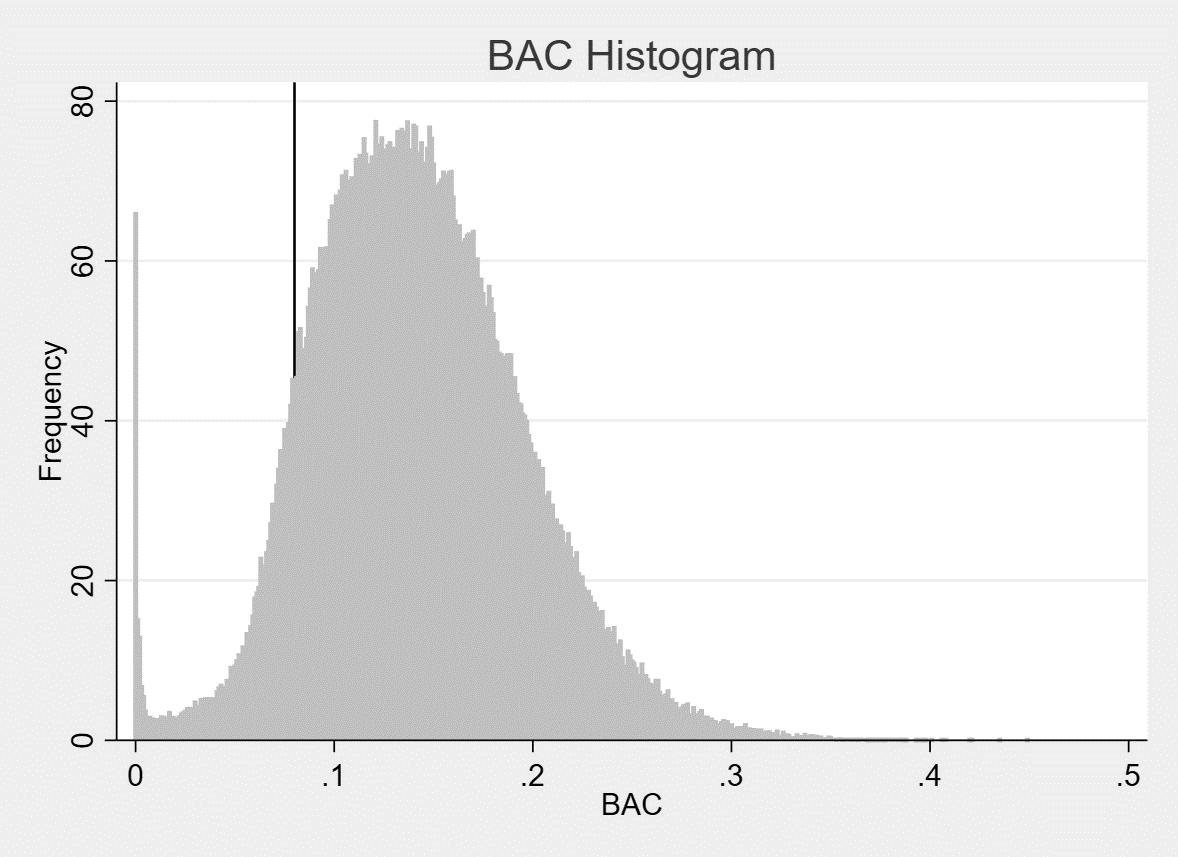
**Replication (worth 6 points)**.[[2]](#footnote-2)

1. In the United States, an officer can arrest a driver if after giving them a blood alcohol content (BAC) test they learn the driver had a BAC of 0.08 or higher. We will only focus on the 0.08 BAC cutoff. We will be ignoring the 0.15 cutoff for all this analysis. Create a dummy equaling 1 if **bac1**>= 0.08 and 0 otherwise in your do file or R file. (**CHECK)**
2. The first thing to do in any RDD is look at the raw data and see if there’s any evidence for manipulation (“sorting on the running variable”). If people were capable of manipulating their blood alcohol content (bac1), describe the test we would use to check for this. Now evaluate whether you see this in these data? Either recreate Figure 1 using the bac1 variable as your measure of blood alcohol content or use your own density test from software. Do you find evidence for sorting on the running variable? (**CHECK)**

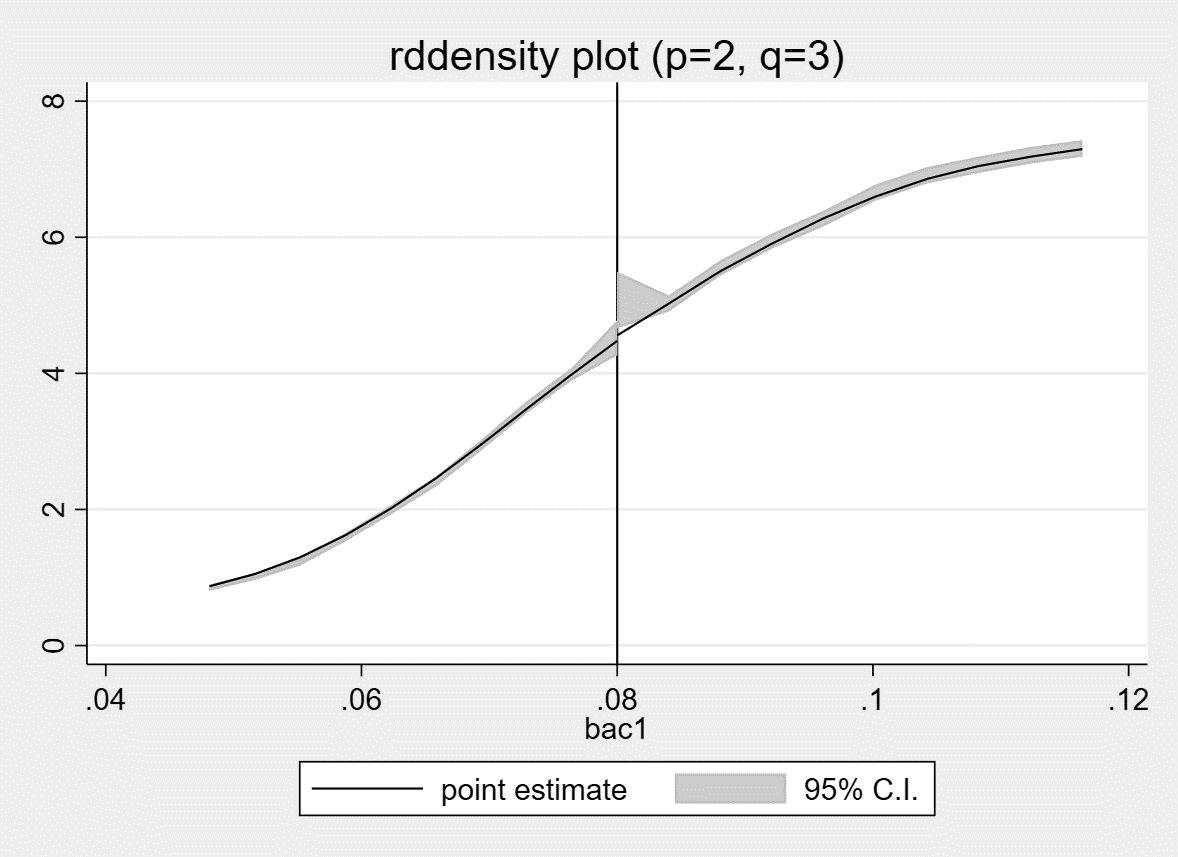
The test is the McCrary Density Test. In this test: the null hypothesis contemplates that the density function (in this case of BAC1) must be continuous at the cutoff point (0.08). In contrast, the alternative hypothesis states that the density function has a jump at the cutoff point (0.08). To do the test: 1) Divide the execution variable into bins and calculate the frequency between bins; 2) These frequency counts are treated as a dependent variable in a local linear regression.

Looking at the data and recreating Figure 1, it appears that the function is continuous at BAC = 0.08.

Figure 1. BAC Histogram



But when doing the McCrary Density Test, the p-value (0.0276) with the robust method returns if there is a discontinuity, that is, people were able to manipulate the BAC.



However, the authors find the opposite. When doing when doing the test, they are based on the conventional method that gives a p-value of 0.59 (as it is in the paper) and therefore, there is no evidence to affirm that there is manipulation in the BAC.

1. The second thing we need to do is check for covariate balance. Recreate Table 2 Panel A but only white male, age and accident (acc) as dependent variables. Use your equation 1) for this. Are the covariates balanced at the cutoff? It’s okay if they are not exactly the same as Hansen’s.
2. Recreate Figure 2 panel A-D. You can use the -cmogram- command in Stata to do this. Fit both linear and quadratic with confidence intervals. Discuss what you find and compare it with Hansen’s paper. (**CHECK**)
3. Estimate equation (1) with recidivism (recid) as the outcome. This corresponds to Table 3 column 1, but since I am missing some of his variables, your sample size will be the entire dataset of 214,558. Nevertheless, replicate Table 3, column 1, Panels A and B. Note that these are local linear regressions and Panel A uses as its bandwidth 0.03 to 0.13. But Panel B has a narrower bandwidth of 0.055 to 0.105. Your table should have three columns and two A and B panels associated with the different bandwidths.:
   1. Column 1: control for the bac1 linearly
   2. Column 2: interact bac1 with cutoff linearly
   3. Column 3: interact bac1 with cutoff linearly and as a quadratic
   4. For all analysis, use heteroskedastic robust standard errors.
4. Recreate the top panel of Figure 3 according to the following rule:
   1. Fit linear fit using only observations with less than 0.15 bac on the bac1
   2. Fit quadratic fit using only observations with less than 0.15 bac on the bac1

1. Again, my preference is that you attempt to create automated tables and automated figures as much as you can. I’ve placed a simple estout program called ols.do in the estout subdirectory. You just need to edit. [↑](#footnote-ref-1)
2. Much of this advice applies to Stata commands, but you can check the R files for lmb.r to see ways of doing the same in R. [↑](#footnote-ref-2)