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]

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.

Repository:
https://github.com/danielcabarcasv/RDD

2. 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?

Hansen uses a type of research design known as regression discontinuity in order to test the effects of sanctions and punishments on driving under the influence of alcoholic substances. In the United States, the blood alcohol content (BAC) has some cutoffs which determine whether the individual is driving under the influence or not, called the DUIs. Hansen intents to determine if having a BAC above both the 0.08 and 0.15 DUI thresholds reduces recidivism so that offenders are discouraged to repeat drunk driving. To this end, Hansen uses administrative records on 512'964 DUC BAC tests in the state of Washington from 1995 to 2011, and uses other demographic information that helps discerning on the smoothness assumption. Finally, he concludes that having a BAC over the 0.08 threshold reduces and recidivism by up to 2 percentage points, and having a BAC over 0.15 reduces recidivism by an additional percentage point.

Replication (worth 6 points).[2]

---

[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.
[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.

3. 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.

See Do file.
gen D=0
replace D=1 if bac1>=0.08

4. 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?
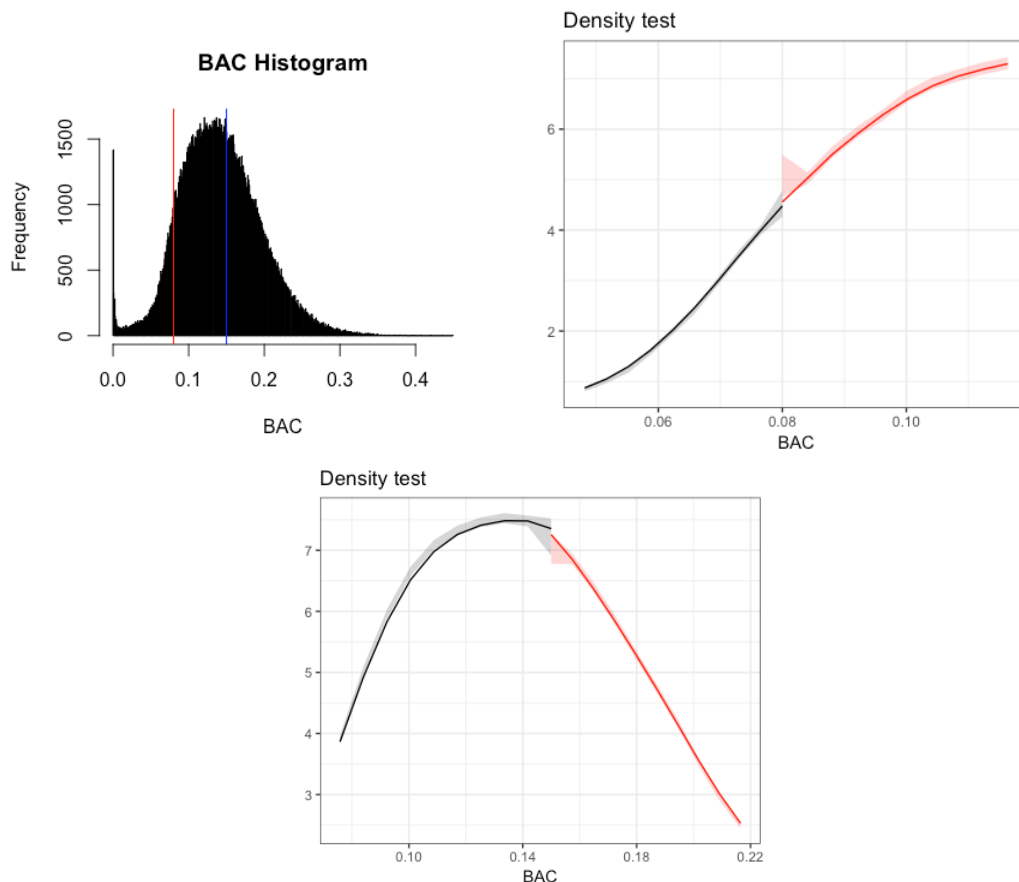
The histogram suggests there is no manipulation of the BAC around either the 0.08 or 0.15 DUI limits. However, a density test based on McCrary and used by Cattaneo, Jansson and Ma (2019) suggests there may be reasons to believe there is sorting on the running variable at the 0.08 cutoff as a 0.027 p-value was obtained. The test shows no evidence of sorting at the 0.15 cutoff.

5. 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.

| VARIABLES	(1) Male	(2) White	(3) Age	(4) Accident
DUI threshold	0.00733**	0.0129***	-1.412***	-0.00535**
	(0.00297)	(0.00258)	(0.0845)	(0.00230)
Controls	No	No	No	No
Observations	124,642	124,642	124,642	124,642

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Results are similar to those obtained by Hansen and for each of the demographics we fail to reject the null that the predetermined characteristics are unrelated to the BAC cutoffs for DUI.

6. 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.

Covariate: White

Covariate: Age

Covariate: Accident

Covariate: Male
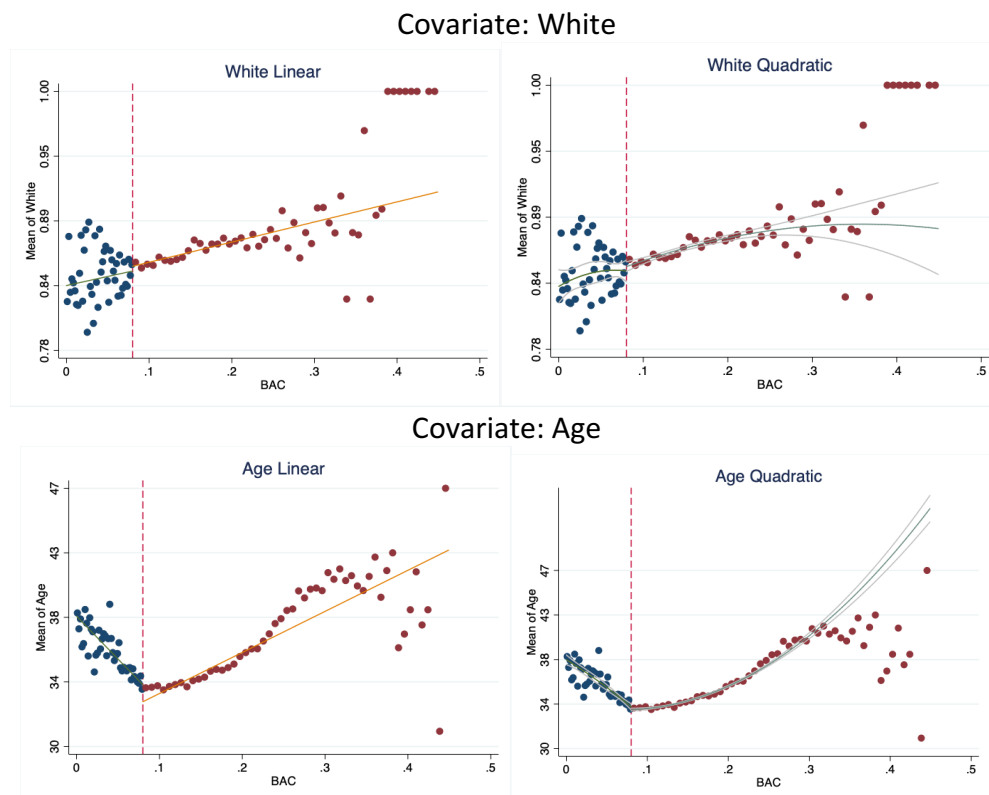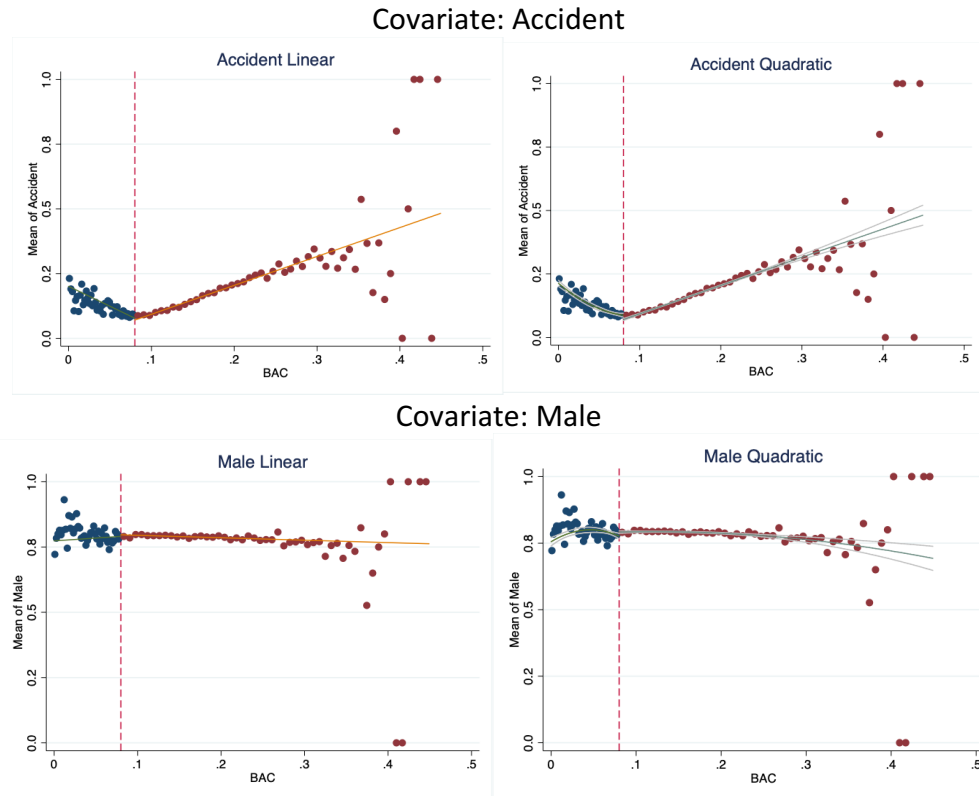
Findings are similar to those on Hansen's paper, and changes in the behavior of the demographic variables are clearly noticeable for previous accidents and age.

7. 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.:
   a. Column 1: control for the bac1 linearly
   b. Column 2: interact bac1 with cutoff linearly
   c. Column 3: interact bac1 with cutoff linearly and as a quadratic
   d. For all analysis, use heteroskedastic robust standard errors.

VARIABLES	(1) Linear BAC	(2) BAC*DUI	(3) BAC*DUI2
PANEL A **[0.03, 0.13]**			
DUI threshold	-0.0273***	-0.0591***	-0.0367***
	(0.00403)	(0.0151)	(0.00569)
Controls	Yes	Yes	Yes

Observations	89,967	89,967	89,967
PANEL B [0.055, 0.105]			
DUI threshold	-0.0219*** (0.00558)	-0.0643* (0.0345)	-0.0395*** (0.0147)
Controls	Yes	Yes	Yes
Observations	46,957	46,957	46,957

8. Recreate the top panel of Figure 3 according to the following rule:
   a. Fit linear fit using only observations with less than 0.15 bac on the bac1
   b. Fit quadratic fit using only observations with less than 0.15 bac on the bac1