LUIS OMAR HERRERA-PRADA

(l.herreraprada@stud.uni-goettingen.de)

Assignment. 4.

Github repo and summary (worth 2 points)

1. *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. 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.*

   https://github.com/luomar/RDD.git

2. *In the writing subdirectory, place your assignment. For the first part, read Hansen's paper in the articles directory of the main class github entitled "Hansen AER". Briefly summarize this paper. What is his question? What data does he use? What is his research design? What does he find.*

   This paper analize how driviers who have received punishment for being driving under the influence (DUI) are less likely to reincid in driving drunk; as the higher their blood alcohol control (BAC) score was at the moment of the penalty less probability of drink and drive. Hansen (2015) uses administrative records from the Washington State Impaired Driver Testing Program, 1999–2007. By using a regression discontinuity approach, Hansen (2015) analizes the two BAC triggers cut-offs 0.08 and 0.015 into the reincidence of drunk drivers. Hansen (2015) finds evidence that suggest the effectiveness of punishments and sanctions in reducing recidivism among drunk drivers. The paper also shows that having a BAC over the 0.15 (the high cut-off) reduces in 1 percentage point the probability of drink and drive again; and if the punishment increases 10 percent then, the probability of being drunk while driving fall 2.3 percent.

Replication (worth 6 points).

1. *In the United States, an officer can arrest a driver if after giving them a blood alcohol content 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.*

   gen select=bac1>= 0.08

2. *The first thing to do in any RDD is look at the raw data and see if there's any evidence for manipulation. 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? Recreate Figure 1 using the bac1 variable as your measure of blood alcohol content. Do you find evidence for sorting on the running variable? .*

Figure 1: RDD in GitHub

Notes: The figure shows the new folders inside RDD

Basically we have to look for a bumps before or after the cut line that show manipulation. In the replication the value to replicate the chart is width=.0001 no width=0.001 as showed in the paper. There is not an evident manipulation or bump before the cut-off, after the cut off is easy to see an small bump. This is consistent with the value of the

[Table 1 about here.]

3. *The second thing we need to do is check for covariate balance. Recreate Table 2 but only white male, age and accident (acc) as dependent variables. Use your equation (1) for this. Are the covariate balanced at the cutoff? It's okay if they are not exactly the same as Hansen's.*

   With the results from eq (1) and (2) for males and withe, I found not relationship with the treatment variable. However, for age and accident I found a significant value for the treatment variable what makes me think there is not independence using this covariates.

[Table 2 about here.]

4. *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.*

   We are only analizing the 0.08 cut-off. My results with the covariantas are similar to the ones of Hansen (2015). There is not evidence of a jump caused because of the cut-off. This allows us to continue with the RDD analizys.

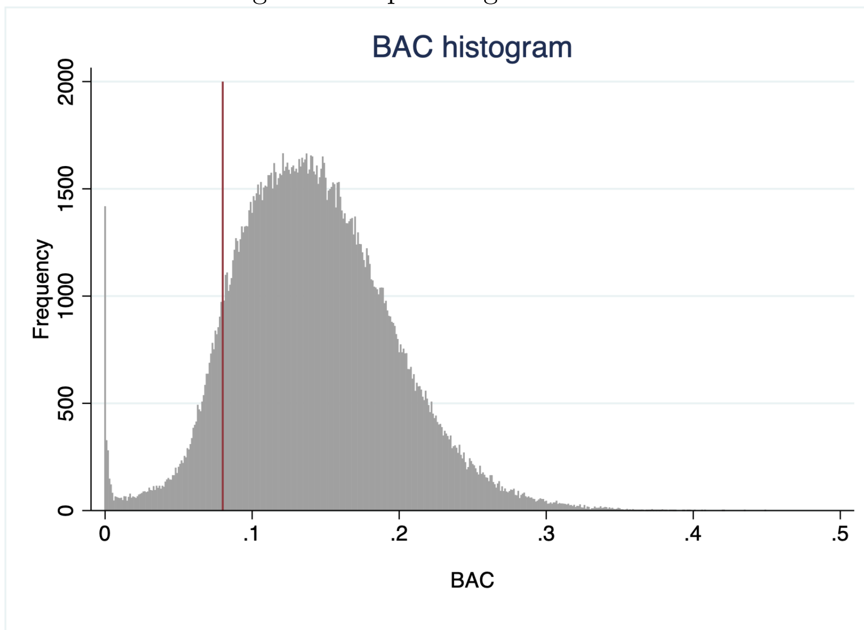
5. *Estimate equation (1) with recidivism (recid) as the outcome. This corresponds to Table 3 column 1, but since I am missing some of his variables, your sample size will be the entire dataset of 214,558. Nevertheless, replicate Table 3, column 1, Panels A and B. Note that these are local linear regressions 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.:*

[Table 3 about here.]

[Table 4 about here.]

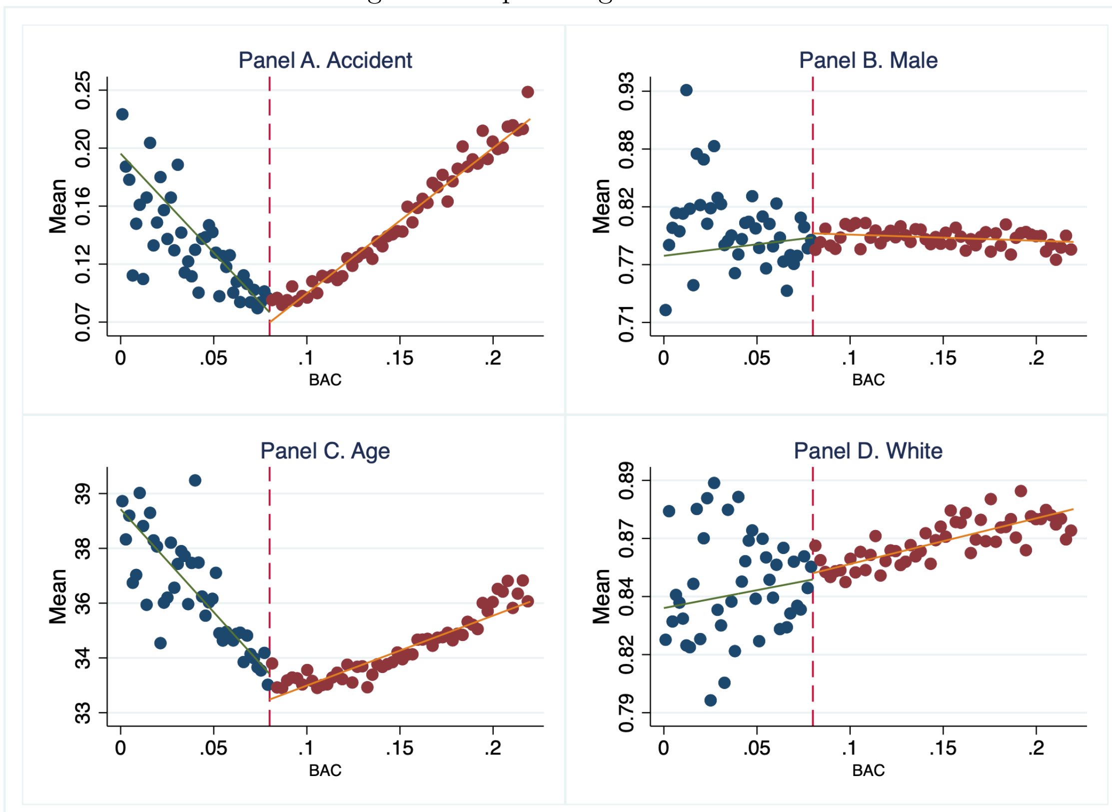
6. *Recreate the top panel of Figure 3 according to the following rule:*

Figure 2: Replica Figure 1Hansen



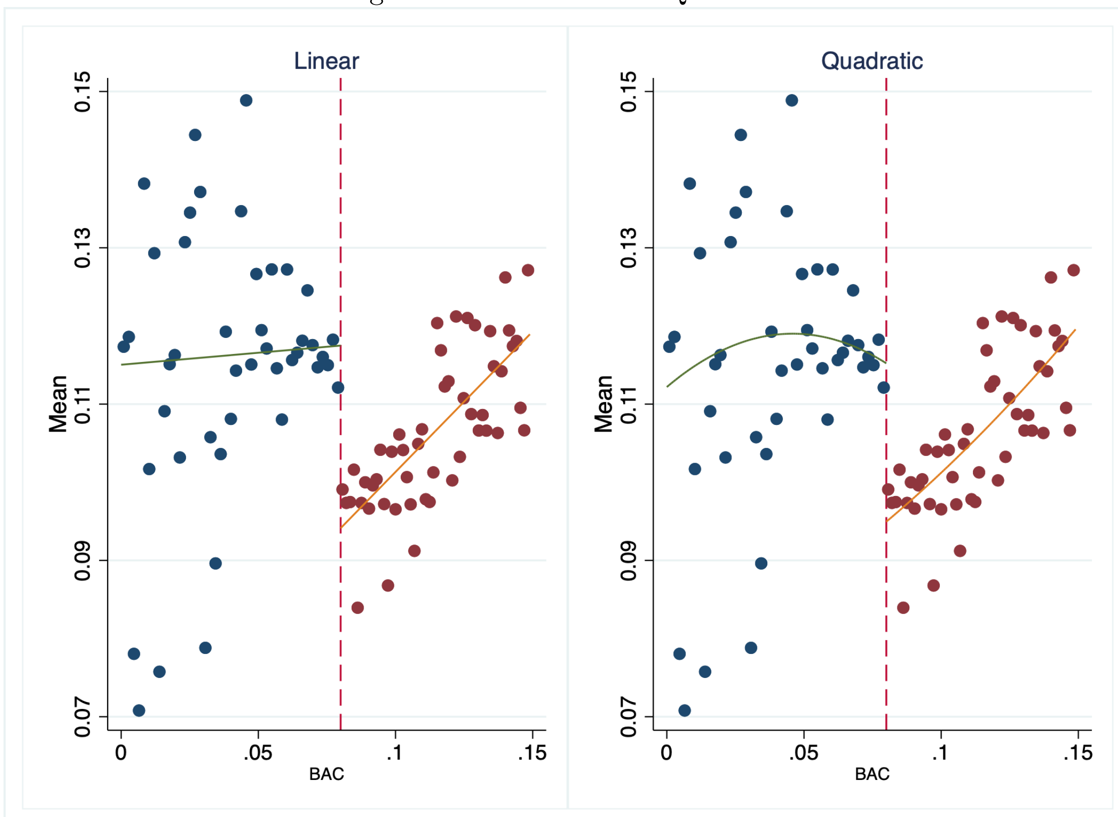
Replica from Hansen (2015). Author chart is "Based on administrative records from the Washington State Impaired Driver Testing Program, 19992007. The histogram height on the vertical axis is based on frequency of observations, with BAC on the horizontal axis. The vertical black lines represent the two legal thresholds at 0.08 and 0.15. The bin width is 0.001, the original precision used on the breathalyzers." Hansen (2015). However, there is a typo in the note from the chart. Width needed to replicate is 0.0001 instead the 0.001 mentioned.

Figure 3: Replica Figure 2- Hansen



Replica from Hansen (2015) Figure 2.

Figure 4: Fit Linear and Quadratic



Replica from Hansen (2015). Database provided in class.

Table 1: Manipulation test

RD Manipulation Test using local polynomial density estimation.		
	(1)	(2)
	Left of c	Right of c
Obs	23010	191548
Eff. Obs	8895	13730
Order est.	2	2
Order bias	3	3
BW est	0.011	0.012
T	2.2	
Prob>T=	0.028	

Manipulation test over 0.08. We find evidence of manipulation at the cut-off.

Table 2: Results

VARIABLES	(1) Panel B. Male	(2) Panel D. White	(3) Panel C. Age	(4) Panel A. Accident
select = 1	0.005	0.010	-5.343***	-0.184***
	(0.025)	(0.012)	(0.389)	(0.017)
bac1	0.218	0.154	-56.361***	-1.540***
	(0.321)	(0.140)	(3.165)	(0.207)
0b.select#co.bac1	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
1.select#c.bac1	-0.080	-0.069	63.255***	2.334***
	(0.342)	(0.160)	(4.465)	(0.227)
Constant	0.773***	0.835***	38.571***	0.201***
	(0.021)	(0.009)	(0.201)	(0.014)
Observations	95,661	95,661	95,661	95,661
R-squared	0.000	0.000	0.007	0.005

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Regressions were estimated using database provided by Professor. However, the sample to replicate table 2 from Hansen (2015) was not possible to replicate. Results shown are form the closest sample using only population with less than 0.13 in the BAC score (0.08 + 0.05, according with the text from original author). *** p<0.01, ** p<0.05, * p<0.1

Table 3: Results

VARIABLES	(1) recidivism	(2) recidivism	(3) recidivism
bac1	-0.055	0.006	2.939*
	(0.048)	(0.187)	(1.641)
select = 1		-0.055***	0.108
		(0.015)	(0.084)
0b.select#co.bac1		0.000	0.000
		(0.000)	(0.000)
1.select#c.bac1		0.392*	-4.083*
		(0.204)	(2.116)
bacsq			-24.614*
			(13.771)
0b.select#co.bacsq			0.000
			(0.000)
1.select#c.bacsq			31.871**
			(15.138)
Constant	0.112***	0.117***	0.034
	(0.005)	(0.012)	(0.047)
Observations	89,967	89,967	89,967
Type	Linear	Interaction	Quadratic
mean	0.107	0.107	0.107
N	89967	89967	89967

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Regressions were estimated using database provided by Professor.

Table 4: Results

	(1)	(2)	(3)
VARIABLES	recidivism	recidivism	recidivism
bac1	-0.442***	-0.150	5.116
	(0.112)	(0.383)	(8.133)
select = 1		-0.062*	0.337
		(0.035)	(0.423)
0b.select#co.bac1		0.000	0.000
		(0.000)	(0.000)
1.select#c.bac1		0.523	-9.465
		(0.450)	(10.623)
bacsq			-38.124
			(58.838)
0b.select#co.bacsq			0.000
			(0.000)
1.select#c.bacsq			63.419
			(69.302)
Constant	0.143***	0.127***	-0.053
	(0.010)	(0.027)	(0.279)
Observations	46,957	46,957	46,957
Type	Linear	Interaction	Quadratic
mean	0.105	0.105	0.105
N	46957	46957	46957

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Regressions were estimated using database provided by Professor.