Assignment 4: RDD

Juan Martín Londoño

June 14, 2020

1. **Create a new github repo named "RDD".**

   The link for the GitHub repository is the following: https://github.com/JuanMartinL/RDD.

2. **Briefly summarize Hansen paper (2015): "Punishment and Deterrence: Evidence from Drunk Driving"**

   Benjamin Hansen paper has as principal question if there is a "(...) effect of harsher punishments and sanctions on driving under the influence (DUI)." (Hansen, 2015, pg. 1581), specially on recidivism over time. To prove his hypothesis, the author uses a Regression Discontinuity methodology on administrative records on 512.964 DUI stops from the state of Washington, with which he suggests that having a blood alcohol content (BAC) above the DUI threshold and aggravated DUI threshold reduces recidivism (Hansen, 2015, pg. 1581). Specifically, 2 percentage points for the first one and one extra percentage point for the second one.

   Furthermore, his research design, mainly based in the RDD, is heavily tested: checking balanced controls, heterogeneity of the dependent variable, robustness and three primary channels that could affect the causal effect in the process. Also, he uses demographic control variables from the analyzed data. Nevertheless, the regression has some problems in fully passing the tests.

3. **Create a dummy equaling 1 if bac1¿= 0.08 and 0 otherwise in your do file or R file.**

   A dummy is created with the following Stata code:

   ```
   gen dbac = 1 if bac1 >= 0.08
   replace dbac = 0 if bac1 < 0.08
   ```

   Nevertheless, if more details needed, please go to
   https://github.com/JuanMartinL/RDD/Codes/Assignment4.do.

4. **Recreate Figure 1 using the bac1 variable as your measure of blood alcohol content. Do you find evidence for sorting on the running variable?**

   Sorting in RDD is a situation in which the observed units are selecting to get the treatment or not endogenously. The presence of this situation can cause bias in the estimations. For this reason, it is important to evaluate if the observed units are sorting in any way.

   Nevertheless, Hansen recognizes that the decision of drinking more is endogenous.Even so, he defends that the BAC around the threshold is purely random, if seen closely enough.

   To prove this assumption, a McCray Test is made, testing if there is any evidence for sorting. In this way, the Figure 1 shows the results of the tests:

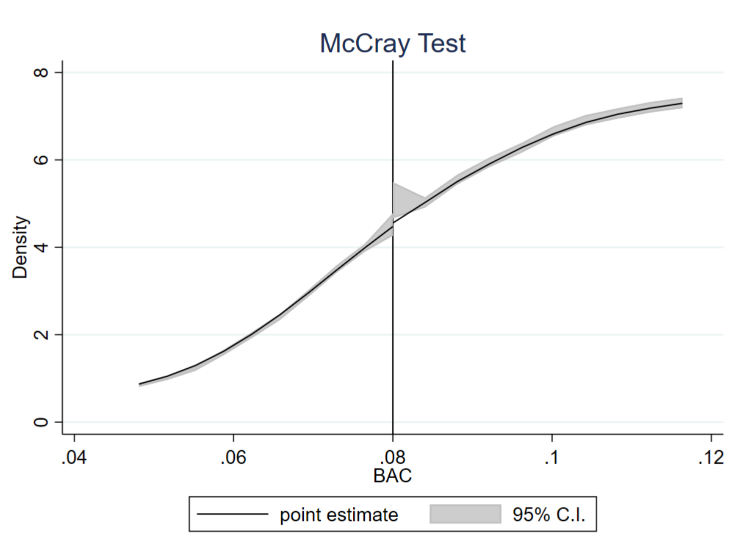

Figure 1: McCray Test for BAC

The plot suggest a piece of evidence for rejecting no sorting in the data. Therefore, the test is made to corroborate this graphical analysis, with the following results:

```
rddensity bac1, c(0.08)

Computing data-driven bandwidth selectors.

RD Manipulation Test using local polynomial density estimation.

    Cutoff c = .08 |Left of c   Right of c          Number of obs =       214558
-------------------+----------------------          Model         = unrestricted
     Number of obs |     23010       191548          BW method     =         comb
Eff. Number of obs |      8895        13730          Kernel        =   triangular
    Order est. (p) |         2            2          VCE method    =    jackknife
    Order bias (q) |         3            3
       BW est. (h) |     0.011        0.012

Running variable: bac1.
-----------------------------------------------
            Method |          T          P>|T|
-------------------+---------------------------
            Robust |      2.2032        0.0276
-----------------------------------------------
```

As seen, there is evidence of sorting around the cutoff taken by the author (0.08).

Moreover, Placebo Tests in Table 1 shows no evidence of sorting around different cutoffs, excluding the 0.08 cutoff. These Placebo confirms the data is not sorted in any cutoff, but the specific threshold that decide whether the treatment is applied or not.

Even the histogram of BAC shows a little jump around the cutoff in the data, as shown in the Figure 2, with a bandwidth of 0.001 (suggested by the author). It is not easily perceptible, but the graphic and mathimatical evidence clearly shows a jump in the data.

All this evidence suggests the observed units are capable of deciding whether get treated or not in

Table 1: Placebo Tests

	0.04	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**	**1**
McCray Test P-Value	0.667	0.922	0.995	0.365	0.027	0.879	0.142

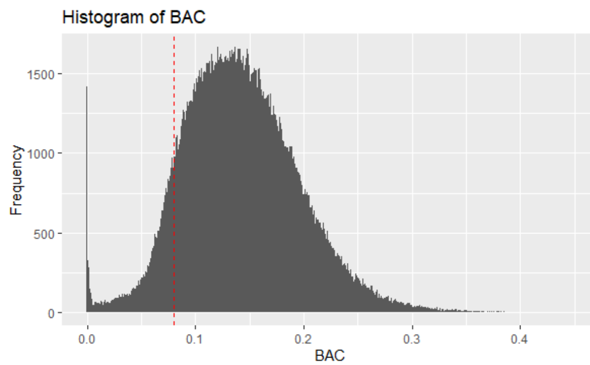

Figure 2: Histogram of BAC

the specific cutoff. This may cause bias in the estimations.

Nevertheless, this probably be explained by the use of a different data set than the author used.

5. **Covariate Balance: Recreate Table 2 but only white male, age and accident as dependent variables**

   The author checked in his paper some demographic factors such as age, race and gender are stable across the DUI punishment thresholds (Hansen, 2015, pg. 1589). In this sense, he regressed the control variables as dependent variable on the DUI punishment, BAC and its interaction. He expected a stability of predetermined characteristics, as they were controls, to avoid bias.

   Given this, the same process is replicated and the results are shown in the Table 2. In this case, the regressions were made with White-Huber Robust Standard Errors to guarantee proper inference.

Table 2: Regression Discontinuity Estimates for the Effect of Exceeding BAC Thresholds on Predetermined Characteristics table

	Driver Demographic Characteristics			
	White (1)	Male (2)	Age (3)	Accident (4)
Panel A. DUI Threshold				
DUI	0.00445 (0.0175)	-0.0184 (0.0198)	-6.224*** (0.586)	-0.154*** (0.0153)
Mean				
Controls	No	No	No	No
Observations	89967	89967	89967	89967

Robust standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Contrary to Hansen results, some estimators are statistically significant. In his paper, Hansen defends his no-significant estimators of the control variables gives credibility to the regression discontinuity design. Nevertheless, there is evidence of statistical significance at least of two control variables. Therefore, the covariates aren't balanced at the cutoff.

However, a disclaimer must be made: his regressions use a Kernel approach, which is not taken into account here. Although a Kernel with lpoly was considered, no similar results were gotten.

6. **Recreate Figure 2 panel A-D. Fit both linear and quadratic with confidence intervals.**

   Following Hansen's paper, recreation of the Figure 2, panel A-D, are calculated for control variables: accident at scene, male, age and white. These are made using linear and quadratic fits. In first place, the linear fit plots are shown in the Figure 2 and the quadratic fit plots are shown in the Figure 3.

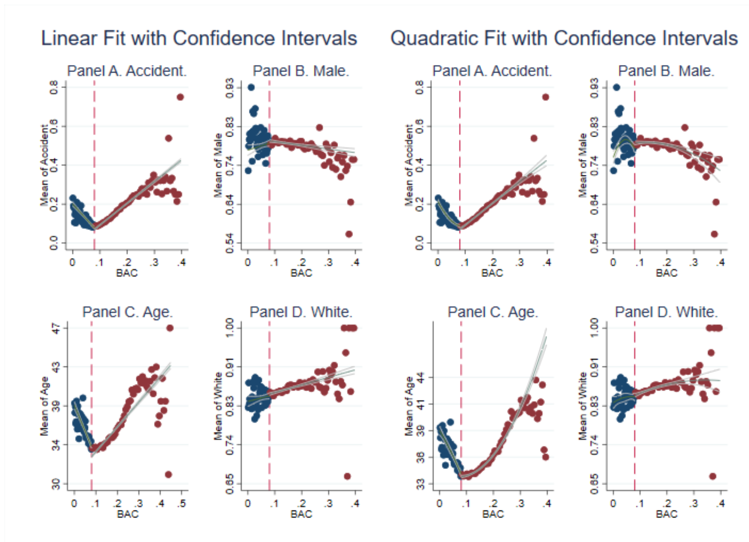

Figure 3: Non parametric fits

In this case, the linear fits are similar to the plots found in Hansen. Nevertheless, Hansen doesn't observe in the paper the quadratic fits, which seems to adjust better for certain variables than the linear one (as seen in age and male). Though this is not considered by Hansen, this observed behaviour of some variables helps the researcher to understand better the analysis that should be done for each control variable.

7. **Estimate equation (1) in Hansen's paper with recidivism (recid) as the outcome. 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 and c. Column 3: interact bac1 with cutoff linearly and as a quadratic.**

   The summary of the results of the regressions are shown in the table 3, differentiating between different bandwidth as the author did. Panel A is calculated with a bandwidth of 0.05 ([0.03-0.13]) and Panel B is calculated with a bandwidth of 0.025 ([0.055, 0.105]). As recommended, Robust standard errors were used.

8. **Recreate the top panel of Figure 3 according to the following rule: a. Fit linear fit using**

Table 3: Regression Discontinuity Estimates for the Effect of Exceeding the 0.08 BAC Thresholds on Recidivism Table

	(1) Recidivism BAC	(2) Recidivism DUI*BAC	(3) Recidivism DUI*BAC^2
Panel A. BAC ϵ [0.03, 0.13]			
DUI	-0.0273***	-0.0591***	0.113
	(0.00403)	(0.0152)	(0.0843)
Mean	-	-	-
Controls	Yes	Yes	Yes
Observations	89967	89967	89967
Panel B. BAC ϵ [0.055, 0.105]			
DUI	-0.0219***	-0.0643	0.371
	(0.00558)	(0.0350)	(0.422)
Mean	-	-	-
Controls	Yes	Yes	Yes
Observations	46957	46957	46957

Robust standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

only observations with less than 0.15 bac on the bac1 and b.Fit quadratic fit using only observations with less than 0.15 bac on the bac1.

Now, following the rules, the plots recreating the top panel of Figure 3 are shown in Figure 4.

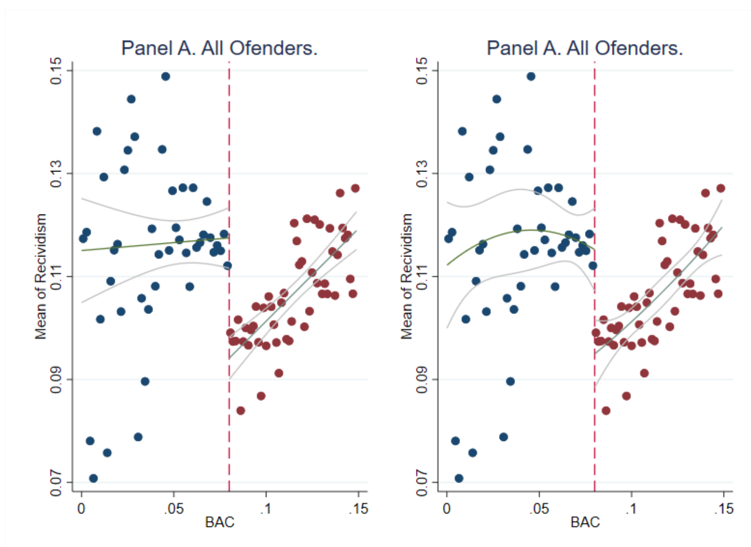

Figure 4: Non parametric fits for BAC and Recidivism

9. **Appendixes**

Table 4: Regression Discontinuity Estimates for the Effect of Exceeding BAC Thresholds on Predetermined Characteristics table

	(1) white	(2) male	(3) aged	(4) acc
bac1	0.0788	-0.210	-69.16***	-1.096***
	(0.214)	(0.240)	(7.220)	(0.186)
dbac	0.00445	-0.0184	-6.224***	-0.154***
	(0.0175)	(0.0198)	(0.586)	(0.0153)
dbac=0 $\times$ bac1	0	0	0	0
	(.)	(.)	(.)	(.)
dbac=1 $\times$ bac1	0.0156	0.307	76.05***	1.888***
	(0.234)	(0.263)	(7.844)	(0.203)
Constant	0.840***	0.801***	39.45***	0.171***
	(0.0143)	(0.0159)	(0.483)	(0.0126)
Observations	89967	89967	89967	89967

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: Regression Discontinuity Estimates for the Effect of Exceeding the 0.08 BAC Thresholds on Recidivism table

	(1) recidivism	(2) recidivism	(3) recidivism
dbac	-0.0273***	-0.0591***	0.113
	(0.00403)	(0.0152)	(0.0843)
bac1	0.321***	-0.0429	2.902
	(0.0748)	(0.187)	(1.637)
aged	-0.000847***	-0.000854***	-0.000854***
	(0.0000849)	(0.0000850)	(0.0000850)
white	0.0162***	0.0162***	0.0162***
	(0.00280)	(0.00280)	(0.00280)
male	0.0332***	0.0332***	0.0332***
	(0.00233)	(0.00233)	(0.00233)
acc	0.00444	0.00421	0.00418
	(0.00345)	(0.00345)	(0.00345)
dbac=0 × bac1		0	0
		(.)	(.)
dbac=1 × bac1		0.438*	-4.210*
		(0.204)	(2.111)
dbac=0 × bac1_sq			-24.72
			(13.74)
dbac=1 × bac1_sq			8.014
			(6.276)
Constant	0.0853***	0.109***	0.0262
	(0.00672)	(0.0131)	(0.0473)
Observations	89967	89967	89967

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6: Regression Discontinuity Estimates for the Effect of Exceeding the 0.08 BAC Thresholds on Recidivism table

	(1) recidivism	(2) recidivism	(3) recidivism
dbac	-0.0219***	-0.0643	0.371
	(0.00558)	(0.0350)	(0.422)
bac1	0.188	-0.196	6.167
	(0.201)	(0.383)	(8.120)
aged	-0.000756***	-0.000758***	-0.000758***
	(0.000115)	(0.000115)	(0.000115)
white	0.0176***	0.0176***	0.0176***
	(0.00381)	(0.00381)	(0.00381)
male	0.0357***	0.0357***	0.0358***
	(0.00317)	(0.00317)	(0.00317)
acc	0.00431	0.00422	0.00423
	(0.00497)	(0.00497)	(0.00497)
dbac=0 × bac1		0	0
		(.)	(.)
dbac=1 × bac1		0.547	-10.52
		(0.449)	(10.61)
dbac=0 × bac1_sq			-46.06
			(58.75)
dbac=1 × bac1_sq			25.20
			(36.58)
Constant	0.0862***	0.113***	-0.104
	(0.0154)	(0.0278)	(0.278)
Observations	46957	46957	46957

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$