Causal Inference and Research Design

María Fernanda Meneses González

June 14 of 2020

1 Github repo and summary

All the information about this assignment is in my repository ”https://github.com/mfmenesesg/RDD” of my Github profile.

Brief summary

The document discusses the effect of stronger punishments on driving under the influence of alcohol in a city in the United States. In this study, the punishments are based on the results of a test that measures alcohol in blood (BAC). According to the document, the results of this test can determine whether there is a drive under the influence situation in an ordinary way (if it exceeds a threshold of 0.08) or a more aggravated case (if it exceeds the threshold of 0.15).

The existence of these predetermined limits allows the use of a discontinuous regression methodology to test the effects of having a BAC given the thresholds in a variable on recidivism. The data used it were the administrative records of the city of Washington between 1995-2011. The main conclusion of the analysis is that recidivism does decrease given the punishments set by BAC levels that exceed the limits. According to the results, it showed that having a BAC above the ordinary threshold reduces recidivism by up to 2 percentage points, and in case of the aggravated threshold is reduced an additional percentage point.

2 Replication

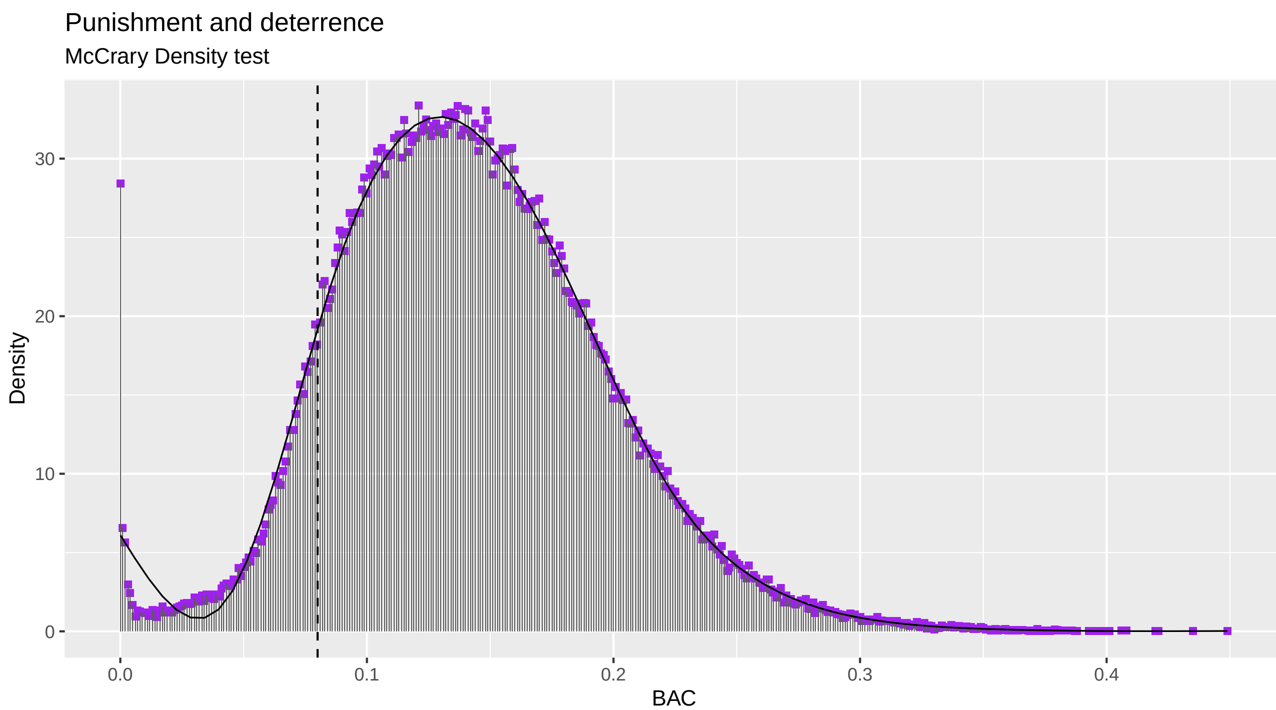
3) To identify which observations are classified as drive under influence (DUI) in an ordinary type, which means that exceeds the threshold of 0.08, a dummy variable was created that takes the value of 1 if the variable bac1 exceeded or equaled this value, and zero in the other cases. Out of a total of 214.558 observations, 191.548 took the value of 1 in the created dummy. In the R script, this variables is called "Dummy".

4) To identify if there is any manipulation of the running variable in the data around the limit I would use a McCrary Test. In the context of this test, the manipulation of the variable implies a heaping on the "good side" of the established limit. So what this test means is that under the null hypothesis the density must be continuous in the cutoff. Under the alternative hypothesis, the density increases at that point. Here the idea is to analyze if the density has jumps in the limit. In short, the test performs a regression taking the midpoints of the bins as the dependent or regressing variable, and the frequency as the independent variable or outcome variable. The graphic analysis and the results of the McCrary test are shown below in Figure 1 and Table 1. As it can bee seen in Table 1, the results show rejection of the null hypothesis. Now checking the Figure 1, visually the data around the limit seems to have no manipulation.

Table 1: McCrary Test - Results

Name	Value
Standard error	0.01488215
Z	0.04819338
p	0.9615621
Cutpoint	0.08

Figure 1: Density of the running variable



5) The next step is to identify the balance of the covariates. For this, Table 2 shows the regressions that were used. The main objective now is to use the variables of male, age and accident as dependent variables. According to these results, only white condition has not significance, while the other three are significant. I consider that these results may differ from the paper because no modification was made in relation to the Kernel analysis. Ideally, no variable should come out meaningful. When analyzing the errors, the regression performed with the target and accident variable had the best results.

Table 2: RD effect of exceeding BAC thresholds on predetermined characteristics

	Control variables			
	white	male	aged	acc
	(1)	(2)	(3)	(4)
Dummy	0.003	0.031^{***}	-7.787^{***}	-0.219^{***}
	(0.006)	(0.007)	(0.204)	(0.006)
bac1	0.154^{*}	0.218^{**}	-56.361^{***}	-1.540^{***}
	(0.092)	(0.108)	(3.038)	(0.093)
M	0.017	-0.311^{***}	83.400^{***}	2.656^{***}
	(0.093)	(0.110)	(3.089)	(0.095)
Constant	0.835^{***}	0.773^{***}	38.571^{***}	0.201^{***}
	(0.006)	(0.007)	(0.184)	(0.006)
R^2	0.001	0.0001	0.013	0.021
Adjusted R^2	0.001	0.0001	0.013	0.021
Residual Std. Error (df = 214554)	0.345	0.408	11.431	0.351
F Statistic (df = 3; 214554)	58.401^{***}	9.428^{***}	907.467^{***}	$1,566.761^{***}$

6. In the following figures (Figure 2 - 4) a replica exercise of Figure 2 of the original paper ”BAC and characteristics” is made. Also included are the corresponding confidence intervals for each of the variables related to male, accident, white and age. After that there are the figures with quadratic fit (Figure 6-9).

According to the figures presented, no evidence of non-smoothness is perceived.

Figure 2: Male

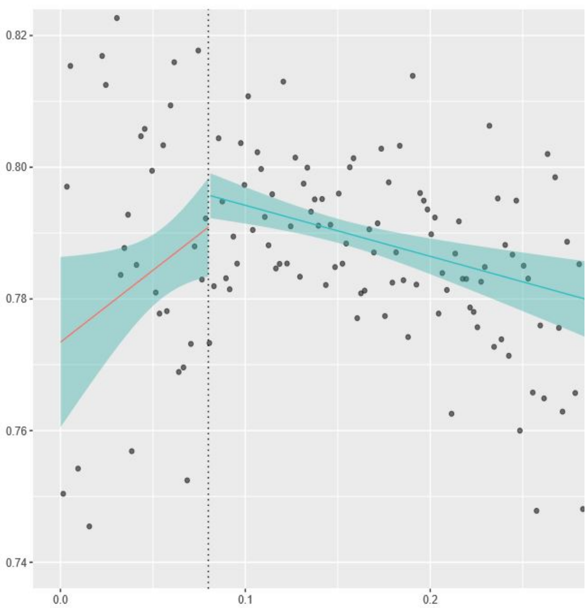


Figure 3: White

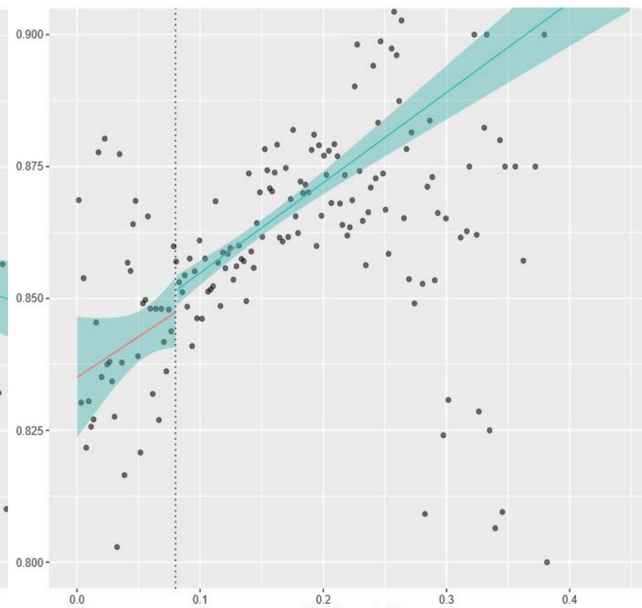


Figure 4: Age

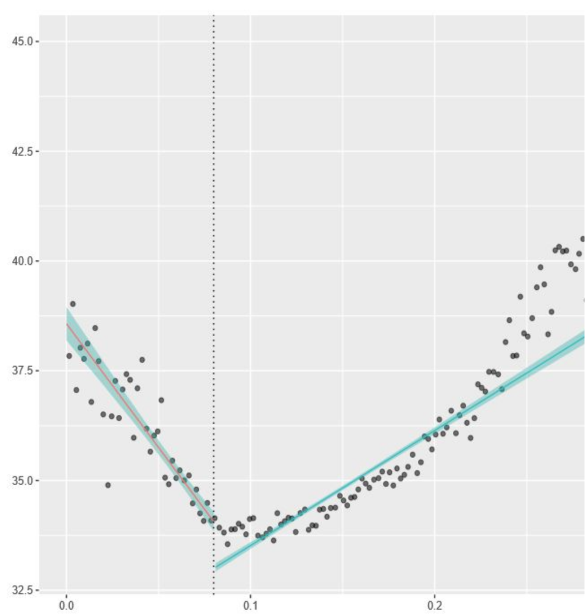


Figure 5: Acc

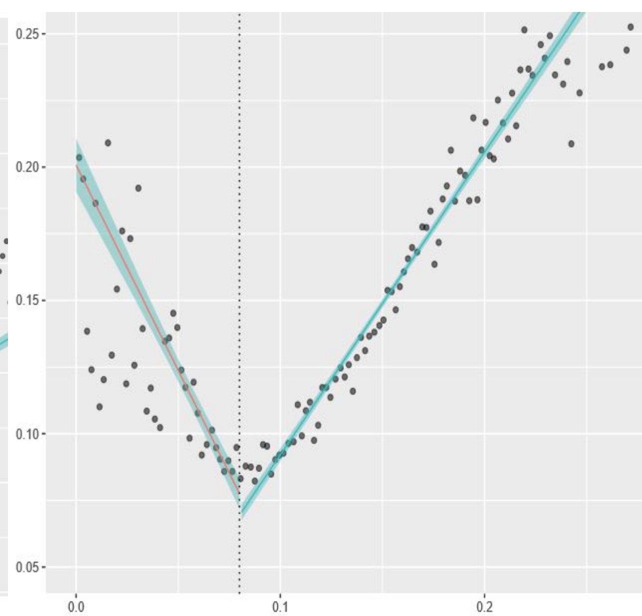


Figure 6: Male

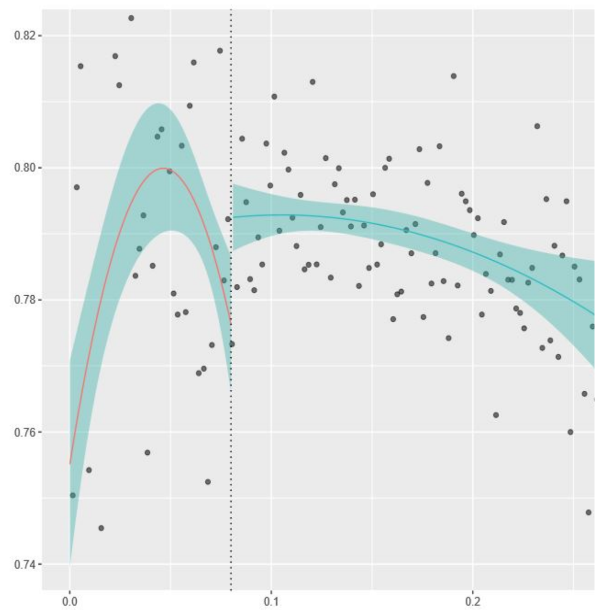


Figure 7: White

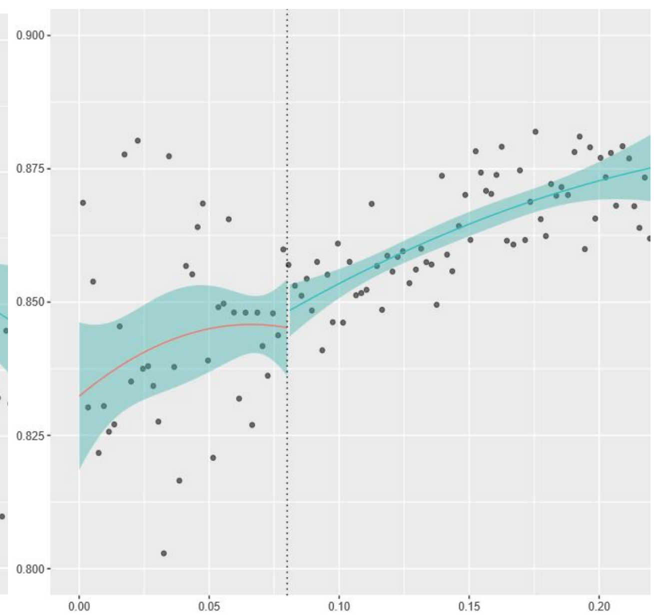


Figure 8: Age

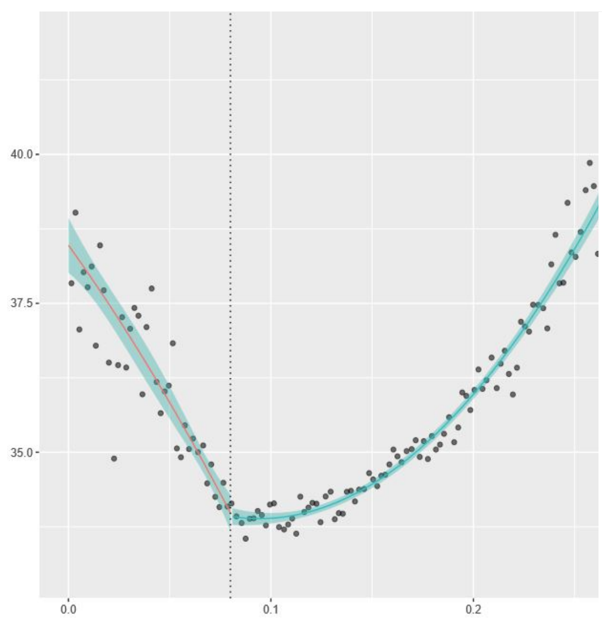
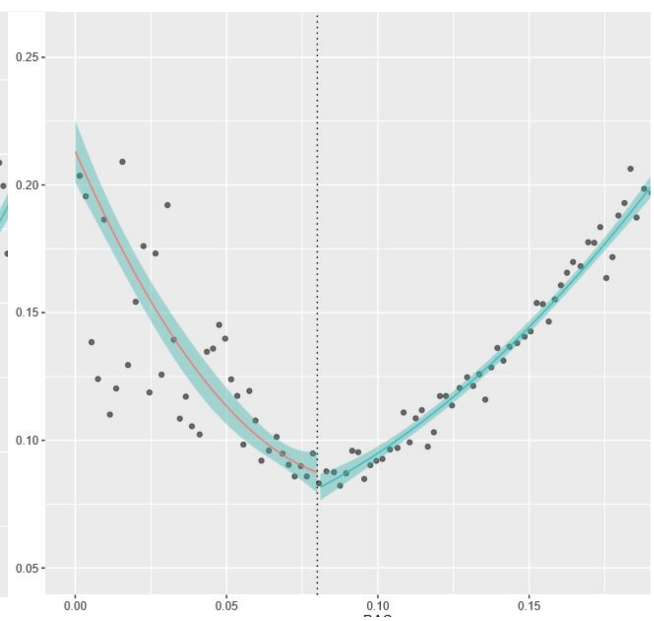


Figure 9: Acc



7. The regressions performed will take recidivism as a dependent variable.The Tables 3-4 contain three columns indicating: i) the regression controlling linearly bac1, ii) includes the interaction bac1 with the linear limit and iii) the previous example but now added a fit control between bac1 and the limit in quadratic form. In the first set of regressions (Table 3) only the coefficients of the first two models were significant. According to the errors, the first model is better. In the other set of regressions (Table 4), the results again show significance only for the two first models.

Table 3: 0.03 - 0.13

	Dependent variable:Recidivism		
	Linear	Interaction - Linear	Interaction - Quadratic
	(1)	(2)	(3)
Dummy	-0.027^{***}	-0.059^{***}	0.030
	(0.004)	(0.015)	(0.071)

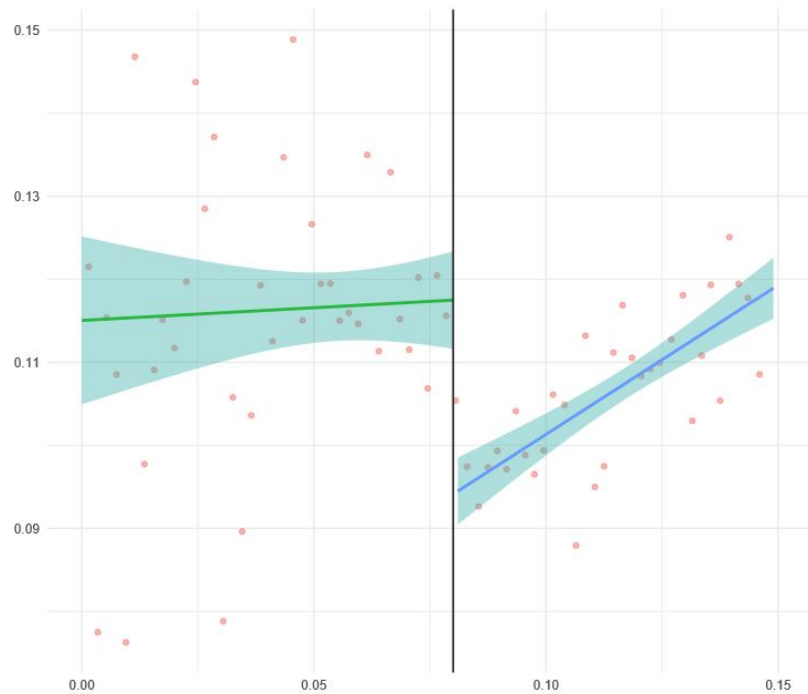
Table 4: 0.055 - 0.105

	Dependent variable:Recidivism		
	Linear	Interaction - Linear	Interaction - Quadratic
	(1)	(2)	(3)
Dummy	-0.022^{***}	-0.064^{*}	0.154
	(0.006)	(0.035)	(0.318)
Note:			*p<0.1; **p<0.05; ***p<0.01

8. Here I recreated the top panel of the Figure 3 according to a fit linear and quadratic. In both figures the data set is reduced only to those observations of the bac1 variable that are less than 0.15.

Figure 10: Fit using only observations with less than 0.15 in the running variable

(a) Linear



(b) Quadratic