Assignment 4: Regression Discontinuity Design

Juan Andrés Rincón†

Causal Inference and Research Design
June 2020

Abstract

This document approaches the regression discontinuity design in econometrics by replicating Hansen's paper *Punishment and deterrence: Evidence from drunk driving* (2015). Based on the general focus of the author, the models and estimations are run using the provided data. Results for the sorting tests are ambiguous as it is highly sensitive to parameters, yet covariate balance tests and overall local regression discontinuity estimations are consistent with the paper's results.

The replication methodology, as well as the empirical procedure can be found at: https://github.com/jarincong98/RDD.

1 Paper Summary

1.1 Research Question

Benjamin Hansen on his paper seeks to evaluate the effects that punishment severity on drivers under the influence has on recidivism (to be caught driving while drunk in a four year follow-up).

1.2 Data

He uses administrative data, counting with 512,964 DUI stops from the state of Washington from 1995 to 2001. The data-base has demographic variables of the individuals, such as gender, race, age and others; the blood alcohol content (BAC) which is used as the running variable in the empiric design that determines the different groups of treatment.

1.3 Research Design

The article uses quasi-experimental evidence to approach the research question. Hansen uses a regression discontinuity design to give consistent estimators, taking advantage on the thresholds determined by the BAC which impose two degrees of punishment on the offenders (0.08 and 0.15 for aggravation). The order of the approach is by testing the possibility of sorting on the running variable, as individuals may be able to manipulate their BACs to fall short of the threshold and avoid the penalties. Then running tests of covariate balance to test the validity of controls for the main test; and finally estimating the effect of the levels of punishment determined by the BAC cutoffs on recidivism.

†Economics student at Universidad de los Andes, Colombia. URosario email: juana.rincon@urosario.edu.co. Uniandes email: ja.rincong@uniandes.edu.co.

1.4 Conclusions

The author finds that crossing the first cutoff of 0.08 BAC has a decrease of 2 percentage points on recidivism, and crossing the aggravated 0.15 cutoff has a reduction on recidivism of an additional percentage point, meaning that additional sanctions are effective in reducing repeat drunk driving.

The mechanisms proposed by the author, based on a Beckerian perspective show that increasing the cost of infraction has a reduction on recidivism, consistent with the rational individual perspective.

2 Replication

2.1 Treatment variable creation

Based on the *blood alcohol content* (BAC) running variable, the treatment dummy DUI (*Driving Under the Influence*) is created as follow:

$$DUI_i = \begin{cases} 1 & \text{if } BAC_i \geq 0.08, \\ 0 & \text{if } BAC_i < 0.08. \end{cases} \tag{2.1}$$

Table 2.1 shows the main descriptive statistics for the newly created variables. It is worthy to mention that 89.3% of the observations are part of the treated.

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
DUI	214,558	0.893	0.309	0	1	1	1

Table 2.1: Treatment variable descriptive statistics.

2.2 McCrary Density Test for manipulation

The literature standard for identifying *sorting* of the data, meaning manipulation of the running variable is the McCrary Density Test (McCrary, 2008). Under the null hypothesis that the density should be continuous at the cutoff point and the alternative that at the cutoff the should be a "jump". The test partitions the assignment variable into bins (as a histogram) and runs a local linear regression taking the partitions as dependent variables.

Firstly, Figure 2.1 shows the density bins and the cutoff point at 0.08 with a polynomial adjustment. Superficially it doesn't seem that there is a jump at the cutoff.

Yet, to be sure, the McCrary test is run in R using the `rdd::DCdensity` and the results are shown in Table 2.2, where Test (1) is estimated with the standard bin-size defined by $bs = 2\sigma_x||x||^{-1/2}$, where σ is the standard deviation of the running variable x and $||x||$ is the number of observations of said variable. And Test (2) is run with the bin size the author used, $bs = 0.001$ (Hansen, 2015). Note that chancing the Bin Size changes dramatically the test results. While maintaining a bin size as in the paper the null hypothesis of no sorting is not rejected, changing the bin size rejects H_0 with very low significance levels.

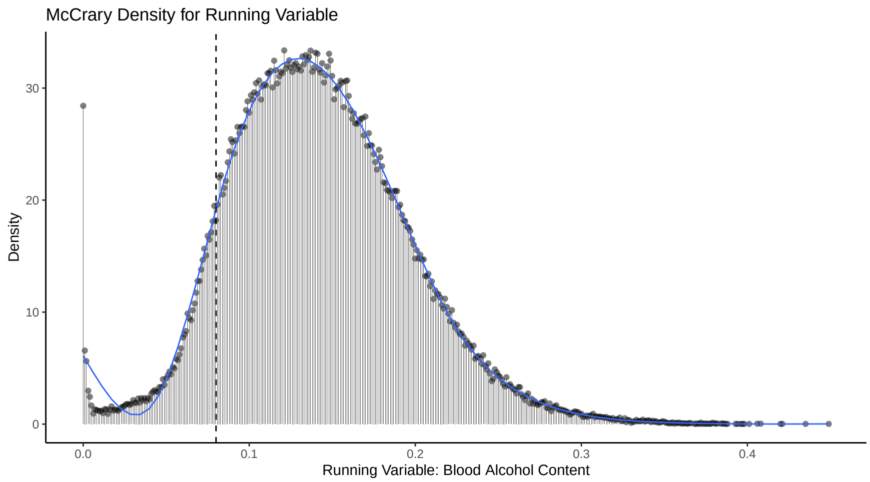

Figure 2.1: Running variable (BAC) density with cutoff.

	Test Statistics						
Test	θ	Standard Error	Z	p-Value	Bin Size	Band Width	Cutpoint
(1)	0.06959	0.0141	4.9436	7.6867e-07	0.0002	0.0501	0.08
(2)	-0.0057	0.0151	-0.3779	0.7055	0.001	0.0432	0.08

Table 2.2: Results for the McCrary Density Test for running variable sorting.

2.3 Covariate Balance

To prove covariate balance, a local regression discontinuity design is estimated using driver demographic characteristics as dependent variables. The model to estimate is:

$$y_i = X_i'\gamma + \alpha_1 DUI_i + \alpha_2 \overline{BAC}_i + \alpha_3(\overline{BAC}_i \times DUI_i) + u_i \ , \tag{2.2}$$

where y_i is the demographic characteristic variable —one of White, Male, Age and Accident—, X_i' is a vector of controls, DUI_i is the treatment dummy and $\overline{BAC}_i$ is the running variable rescaled around the threshold of $c_0 = 0.08$ ($\overline{BAC}_i = BAC_i - c_0$). The regressions are weighted using a rectangular kernel, defined with a threshold of $h = 0.05$ as:

$$K(BAC_i) = \begin{cases} \dfrac{1}{2h} & \text{if } \ |BAC_i - c_0| < h, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, the results of estimation are shown in Table 2.3.

Based on the results shown on Table 2.3, the regression discontinuity coefficient at the cutoff for every demographic characteristic variable is statistically different from zero at the 5% significance level, which, formally means that there is a small jump at the cutoff, they are not balanced.

	Dependent variable:			
	White	Male	Age	Accident
Characteristics	(1)	(2)	(3)	(4)
DUI	0.005^{***}	0.006^{**}	-0.154^{**}	-0.004^{**}
	(0.002)	(0.002)	(0.068)	(0.002)
Mean (at 0.079)	0.843	0.785	35.612	0.118
Observations	214558	214558	214558	214558
Controls	No	No	No	No
Note:		*p<0.1; **p<0.05; ***p<0.01		

Table 2.3: Local Regression Discontinuity Estimates for the effect of exceeding BAC threshold on predetermined characteristics.

2.4 Graphical Covariate Balance

Taking into account the results from the past regressions, a graphical analysis is conducted as well. As Figure 2 Panels A-D in Hansen, 2015, both Figure 2.2 and Figure 2.3 show the discontinuities at the 0.08 cutoff estimated with a local linear and quadratic fitting models. The bin size is set to 0.002 for the grouping of the observations.

Firstly, using the linear fit (Figure 2.2), the small coefficients from Table 2.3 are almost unnoticeable for the variable *accident*, *male* and *white*, while for *age* the confidence intervals are evidently separated. On the other hand, from the quadratic fit perspective (Figure 2.3), all confidence intervals cross at the cutoff, making the apparent unbalance from the linear estimation statistically zero, meaning balance of the covariates.

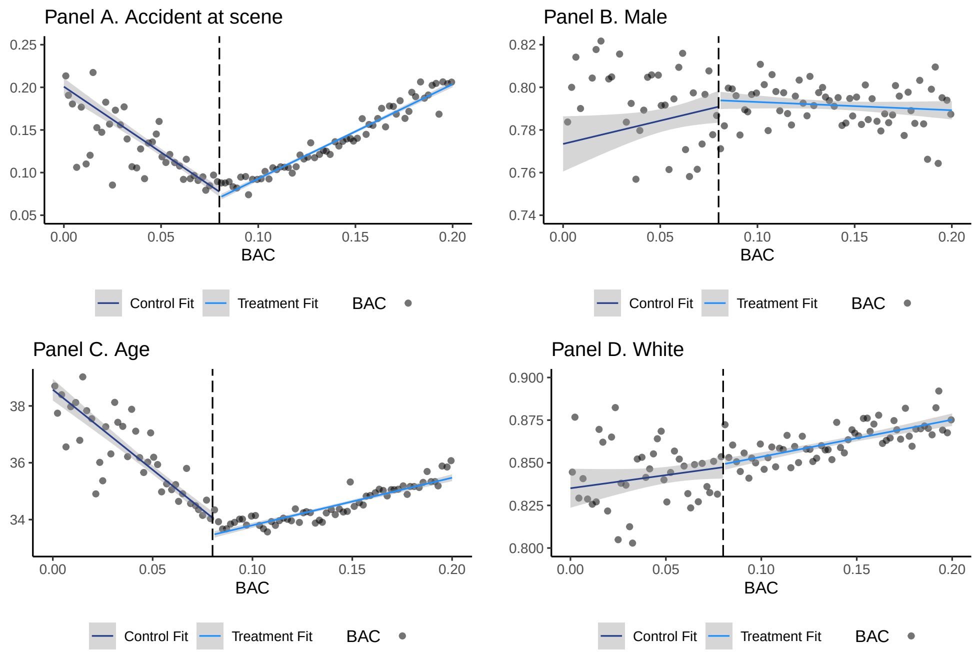

Figure 2.2: Covariate Balance graphical analysis of BAC on demographic variables with linear fit.

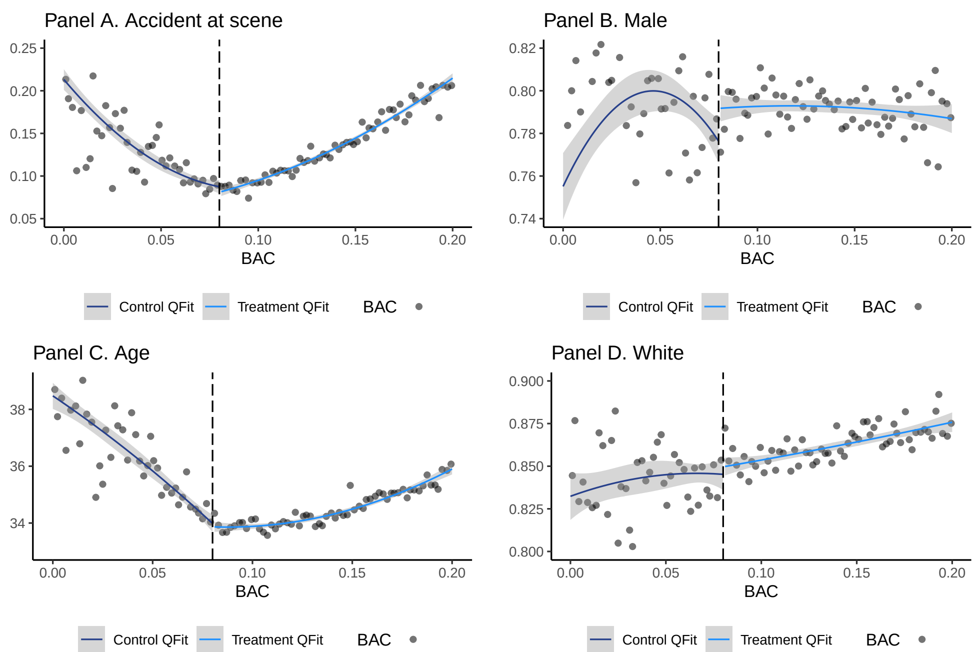

Figure 2.3: Covariate Balance graphical analysis of BAC on demographic variables with quadratic fit.

2.5 RDD Estimation

Based on equation (2.2), with covariate vector X_i' composed of white, male, age and accident variables, three local regression discontinuity models are estimated: Controlling for BAC linearly (2.3), controlling for the interaction (2.4) and controlling for the interaction squared

(2.5).

$$y_i = X_i'\gamma + \alpha_1 DUI_i + \alpha_2 \overline{BAC}_i + u_i \tag{2.3}$$
$$y_i = X_i'\gamma + \alpha_1 DUI_i + \alpha_2 \overline{BAC}_i + \alpha_3(\overline{BAC}_i \times DUI_i) + u_i \tag{2.4}$$
$$y_i = X_i'\gamma + \alpha_1 DUI_i + \alpha_2 \overline{BAC}_i + \alpha_3(\overline{BAC}_i \times DUI_i) + \alpha_4(\overline{BAC}_i \times DUI_i)^2 + u_i \tag{2.5}$$

Regression results for the treatment coefficient of models 2.3 to 2.5 are shown on Table 2.4. Panel A runs the models with a bandwidth of $h = 0.05$ and Panel B with $h = 0.025$, the rectangular kernel adjusts for the weights of the observations due to the reduced area of estimation and thus the observations drop. For the case of a wider bandwidth, all three models expose significant treatment estimators at the $\alpha = 0.01$ level. This means that having a BAC above the 0.08 threshold decreases recidivism by 2 percentage points during a four year follow-up window. Even, for the smaller bandwidth, the results are the same.

	Dependent variable: Recidivism		
	Linear Control	Interaction	Squared Interaction
	(1)	(2)	(3)
Panel A $\in [0.03, 0.13]$			
DUI	-0.027^{***}	-0.024^{***}	-0.020^{***}
	(0.002)	(0.002)	(0.002)
Mean	0.111	0.106	0.106
Controls	Yes	Yes	Yes
Observations	88373	88373	88373
Panel B $\in [0.055, 0.105]$			
DUI	-0.022^{***}	-0.021^{***}	-0.021^{***}
	(0.001)	(0.001)	(0.001)
Mean	0.101	0.098	0.98
Controls	Yes	Yes	Yes
Observations	46957	46957	46957
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 2.4: Regression Discontinuity results for bandwidths of 0.05 (Panel A) and 0.025 (Panel B) weighted with rectangular kernel and Robust Standard Errors.

2.6 Graphical Regression Discontinuity Design

Consistently with the results from Table 2.4, Figure 2.4 shows the statistically significant discontinuity at the 0.08 threshold of the Blood Alcohol Content running variable on the Recidivism variable. For both linear and quadratic fits, the confidence intervals do not cross and the effect is a reduction or 2 percentage points on recidivism.

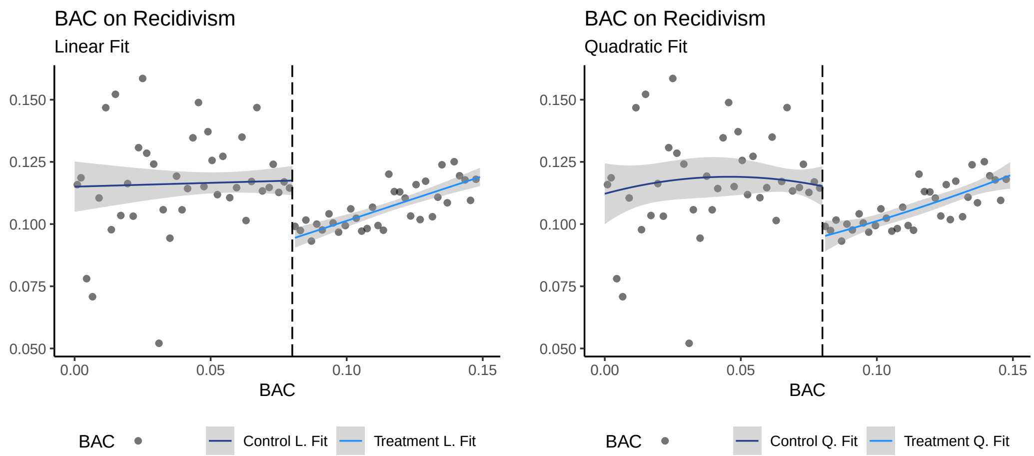

Figure 2.4: BAC and Recidivism discontinuity. Linear and Quadratic fits.

References

Hansen, B. (2015). Punishment and deterrence: Evidence from drunk driving. *American Economic Review*, 105(4): 1581–1617. https://doi.org/10.1257/aer.20130189

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142(2): 698–714. https://doi.org/j.jeconom.2007.05.005