

Causal Inference and Research Design

Assignment 4 -RDD-

María Alejandra Orjuela Pava

Point 1: GitHub repository.

The link to my GitHub RDD Repository is the following:

<https://github.com/MariaAlejandraOrjuelaPava/RDD>

Point 2: Summary of Hansen's paper.

Shortly after the introduction of automobiles, drunk driving emerged as a serious public health issue. At first, identifying the impairment of drivers was difficult as it relied on police's personal experience and the implementation of on field sobriety tests. However, as time went by, breathalyzers were created, providing an accurate method of measuring BAC in a noninvasive manner. Due to their easiness, rapidness and precision, several states began to adopt laws which stipulated strict thresholds for DUI. Within them a BAC of 0.08 was used to establish DUI, whereas a registered value of 0.15 or higher ought to be considered aggravated DUI; both implying a set of punishments which varied on magnitude and degree -fines, jail time, license suspension-. Despite the fact it was believed that these restrictive standards did reduce traffic fatalities, it was well known that multiple criminals returned to committing crimes within a few years of being released from incarceration for their original crime. Taking both factors into consideration, Hansen wrote a research paper in which, by thinking that an appropriate combination of enforcement and punishment was crucial to maximize social welfare, he tried to answer whether these punishments were effective in reducing drunk driving. Precisely, he examined the effects that punishment severity had on the commission of future crime, further investigating the determinants of recidivism by three main channels: incapacitation, rehabilitation, and deterrence.

Within his study, he took advantage of administrative records on 512,964 DUI BAC tests in the state of Washington from 1995 to 2011, noticing that after January 1, 1999, WA applied a 0.08 threshold for determining a DUI, and a 0.15 threshold for an aggravated DUI. Given the BAC thresholds were constant after 1999, he implemented data from 1999–2007 to analyze the causal effect of having a BAC above either the 0.08 or 0.15 threshold on recidivism within four years of the original BAC test. Furthermore, he restricted his attention to those above the legal drinking age given that different cutoffs applied to those under 21.

When regarding the identification strategy which was implemented, it can be affirmed that the local linear regression discontinuity design with a rectangular kernel was chosen, with the slopes allowed to change at the discontinuities. Notice that for it to deliver consistent estimates, several assumptions must be met. Therefore, Hansen proceeded to check upon them. On one hand, he examined the continuity of the underlying conditional regression and distribution functions. Based on the visual and quantitative tests he made, he concluded that both the unobservable and observables remained unchanged across the threshold with only treatment status (or the probability of treatment) changing. On the other hand, he explained why it was plausible to think that it was locally random for a driver to have BAC either just below or just above the BAC thresholds. Settling up the distribution of BAC and founding the p-values of McCrary tests at each cutoff, he concluded that there was none

endogenous sorting to one side of either the thresholds studied. Notice that among all his regression models, clustering at the finest bin at which BAC was measured (0.001) was allowed as this captured potential autocorrelation between individuals which had similar BAC levels. At the same time, the correspondent standard errors were adjusted for heteroskedasticity.

After performing so, the obtained results suggested that harsher punishments and sanctions associated with BAC limits reduced future drunk driving in the short and long term. The estimates suggested having a BAC over the 0.08 legal limit corresponded with a 2 percent point decline in repeat drunk driving over the next four years. Likewise, having a BAC over the 0.15 enhanced punishment limit was associated with an additional 1 percentage point decline in repeat drunk driving -translating the percentage increase in sanctions and the decline in drunk driving into an elasticity, he found evidence that 10 percent increase in sanctions and punishments was linked to a 2.3 percent decline in drunk driving-. Finally, when examining the 3 primary channels related to criminality that could explain the reduction in recidivism, he found that deterrence served as the primary one.

Point 3: Treatment variable creation.

The required variable (named as DUI) was created by running the following code in STATA:

```
/*First of all, I define directory. */
clear all
cd "C:\Users\MaríaAlejandra\Desktop\RDD\Data"

/*Secondly, I upload the required dataset. */
use "hansen_dwi"

/*After doing so, I perform a preliminary examination of the data. */
br
describe
count
/*The sample consists of 214,558 observations. */
sort year

*****POINT 3*****
/*In this point, I am required to create a variable which is equivalent to 1 in case bac1>=0.08, and 0
if the contrary situation occurs. In this case, I entitled it as 'DUI'. */

gen DUI=1
replace DUI=0 if bac1<0.08
label DUI "Indicates whether the unit of observation can be arrested based on its bac1 results"
tab DUI

/*When examining the descriptive statistics of this variable, I encounter that 89.28% of the sample
registers a bac1 higher or equal to 0.08. This percentage is equivalent to 191,548 people. On the
opposite side, 10.72% exhibits a value of 0 in this variable, implying that its bac1 is inferior to 0.08.
*/
```

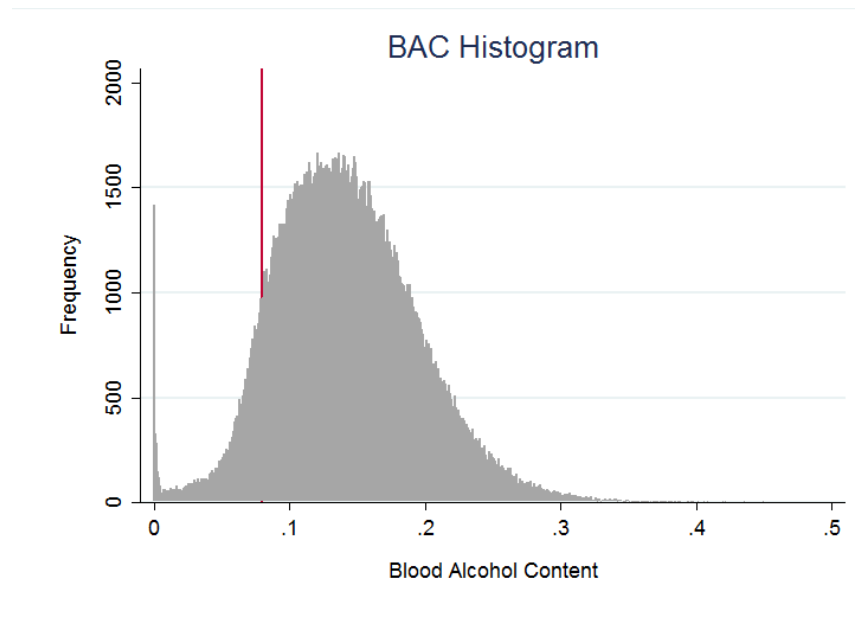
Point 4: Manipulation assumption check.

To begin with, it should be stated that the running variable within this exercise refers to BAC1; the *alcohol blood content* which is registered for each unit of observation among the sample. In this case,

as the author implemented the linear regression discontinuity design, several assumptions should be fulfilled so as to obtain a consistent estimator of the willed causal effect. Precisely, one of those states that there should not be any manipulation of the running variable, implying it is locally random for a driver to have a BAC either just below or just above the cutoff.

In order to check it, I decided to do the histogram -distribution diagram- of BAC1, trying to replicate Hansen's Figure 1 as a first approach.

Figure 1. BAC Histogram



Note: The histogram height on the vertical axis is based on frequency of observations, with BAC on the horizontal axis. The vertical red line represents the established threshold at 0.08. The bin width is 0.0001.

When comparing this diagram to the one displayed within the paper, I can firmly affirm they look alike, showing little evidence of endogenous sorting to one side of the studied threshold. However, it should be stated that Figure 1 just provides a vague idea of the fulfillment of the non-manipulation assumption.

Therefore, I proceeded to check upon it more formally, using the McCrary test. Remember that this process suggests testing the *null hypothesis* of the continuity of the density of the covariate that underlies the assignment at the cutoff, against the *alternative* of a jump in the density function at that same point. Notice that in principle, the design does not require continuity of the density of BAC1 at 0.08, but a discontinuity is suggestive of violations of the no-manipulation assumption.

After performing McCrary test, I found two main results:

Table 1. McCrary Test

Method	T	P> T
Conventional	0.5337	0.5936
Robust	2.2032	0.0276

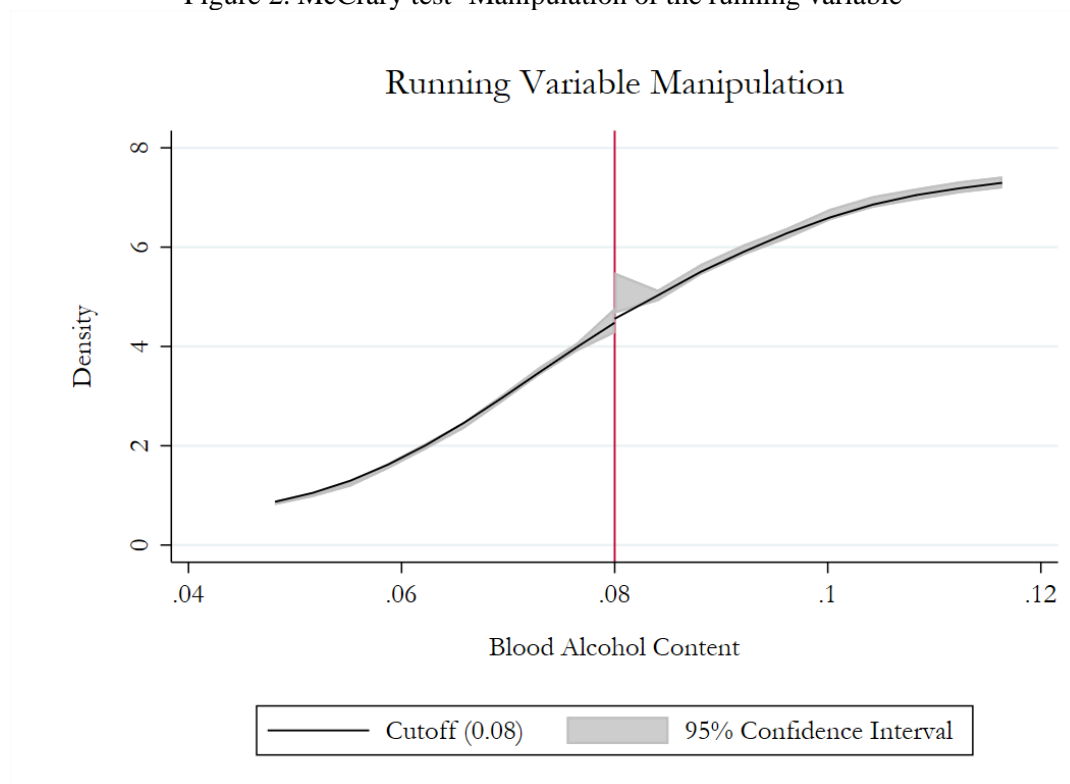
Note: Based on data from the 1999–2007 Washington State Impaired Driver Program
Outcome 1:

The derived p-value, which is equivalent to that one mentioned on Hansen's paper -0.598-, suggests that the null hypothesis cannot be rejected. This implies that there is not sufficient statistic evidence to affirm there was any sort of manipulation of the running variable.

Outcome 2:

The p-value obtained from this test equals to 0.0276. When analyzing its magnitude, it can be said that the null hypothesis is rejected with a 0.05 probability of making a type I error. Therefore, there is evidence of manipulation within the blood alcohol content variable. This can be examined visually if we do a zoom within its density diagram.

Figure 2. McCrary test -Manipulation of the running variable-



Point 5. Covariates balance check.

As a first approach, I ran the following regressions without a specific bandwidth nor a rectangular kernel for weighting. Within them, the dependent variable refers each of the observable controls taking into consideration in the structural model (male, white, age, and accident), whereas the independent ones consist on i.) the treatment -dichotomous variable, which indicates whether the unit of observation is driving under influence-, ii.) the running variable -blood alcohol content- rescaled around the threshold of 0.08 ($BAC1_i^* = BAC1_i - 0.08$), and ii.) the interaction between them.

The parameter of interest for each equation is α_1 . We look forward to its estimator not being statistically significant. If this occur, it is plausible to think there are not systematic differences within

the treated and control groups, satisfying the local continuity assumption required in the regression discontinuity design.

- (1) $Male_i = \alpha_0 + \alpha_1 DUI_i + \alpha_2 BAC1_i^* + \alpha_3 DUI_i * BAC1_i^* + u_i$
- (2) $White_i = \alpha_0 + \alpha_1 DUI_i + \alpha_2 BAC1_i^* + \alpha_3 DUI_i * BAC1_i^* + u_i$
- (3) $Age_i = \alpha_0 + \alpha_1 DUI_i + \alpha_2 BAC1_i^* + \alpha_3 DUI_i * BAC1_i^* + u_i$
- (4) $Accident_i = \alpha_0 + \alpha_1 DUI_i + \alpha_2 BAC1_i^* + \alpha_3 DUI_i * BAC1_i^* + u_i$

Table 2. Covariates Balance

VARIABLES	(1) Male	(2) White	(3) Age	(4) Accident
DUI	0.00581 (0.00421)	0.00407 (0.00368)	-1.115*** (0.121)	-0.00701** (0.00314)
Constant	0.791*** (0.00384)	0.847*** (0.00338)	34.06*** (0.111)	0.0776*** (0.00281)
Observations	214,558	214,558	214,558	214,558
R-squared	0.000	0.001	0.013	0.021

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

Based on the results exhibited in Table 2, it can be affirmed that there are not systematic differences within both groups for characteristics such as gender and race. This implies that, when taking into consideration these two factors, those individuals who are at one side of the cutoff do not differ, on average, to those who are at the opposite one. However, as it can be observed, there are statistical differences in Age and Accident. The derived results suggest that there can be either observable or unobservable characteristics in which both groups differ. Those characteristics do affect the fact of being at one side or the other from the cutoff and they can have huge impacts on the registered outcome (recidivism), generating that the causal effect is biased when estimated. In spite these conclusions, it should be known that they are plausible as I calculated the estimators, alongside with their robust standard errors, using the whole sample (214,558 observations) -be aware that those individuals positioned at the extremes (those who register an extremely high or low BAC, respectively) are likely to be quite heterogeneous when making comparisons among them-.

Due to this, I proceeded to follow Hansen's methodology, implementing a bandwidth of 0.05 and using rectangular kernel for weighting. Before I did it, I expected that the estimations obtained suggested the existence of covariates balance, being able to affirm that the assumption of local continuity was fulfilled in this exercise.

Table 3. Covariates Balance

VARIABLES	(1) Male	(2) White	(3) Age	(4) Accident
DUI	0.00618 (0.00570)	0.00570 (0.00501)	-0.140 (0.164)	-0.00335 (0.00406)

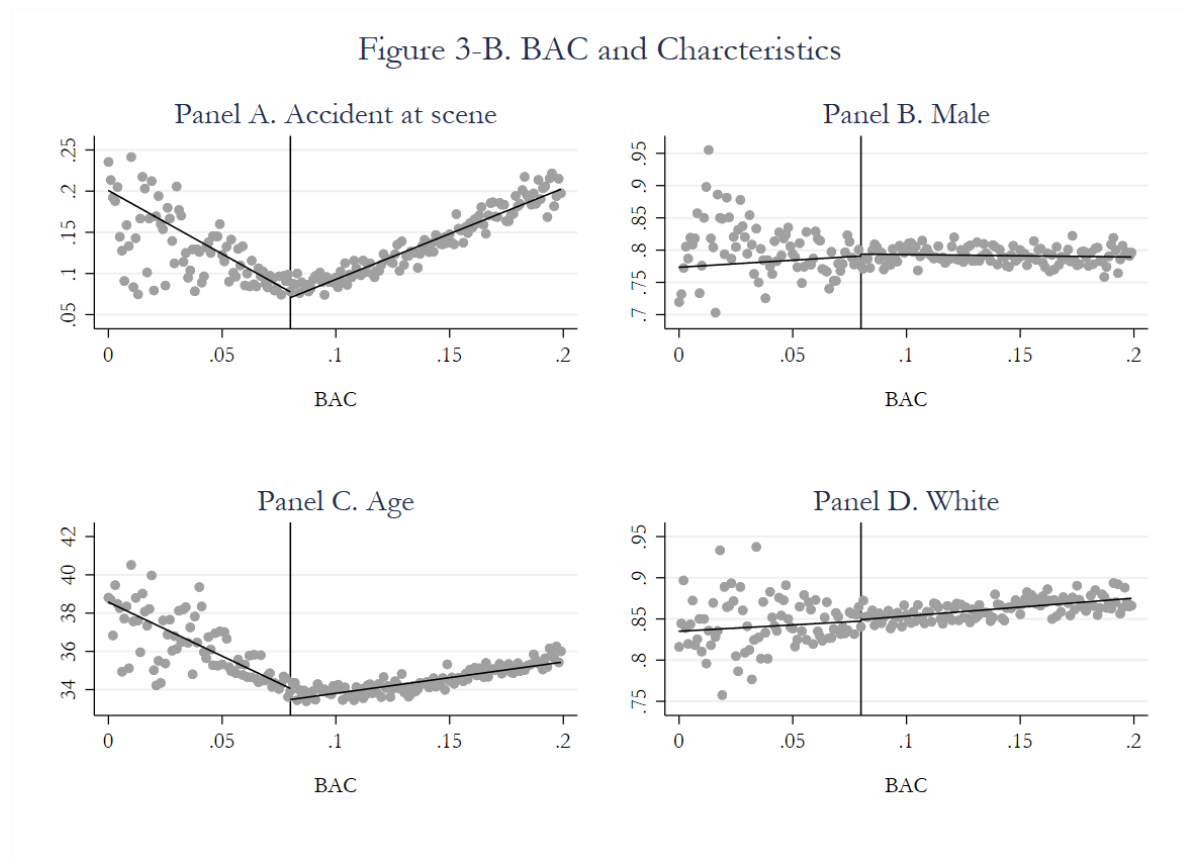
	(0.263)	(0.234)	(7.844)	(0.203)
Constant	0.784***	0.846***	33.92***	0.0834***
	(0.00463)	(0.00409)	(0.135)	(0.00330)
Observations	89,967	89,967	89,967	89,967
R-squared	0.000	0.000	0.002	0.002

Note: Robust standard errors in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. All regressions have a bandwidth of 0.05 and use a rectangular kernel for weighting.

The results illustrated on Table 3 are quite similar to those exhibited in the author's paper. Even though they do differ in magnitude, the conclusions that can be derived from these are the same. On one hand, after locally restricting the sample (using 89,967 observations), it can be observed that demographic factors such as age, race, and gender are stable across the DUI punishment threshold. Likewise, a key source of information that could drive the police to administer a breath test -the presence of an accident at the scene- remains unchanged. This can be affirmed as the correspondent estimators are not statistically significant. On the other hand, it should be stated that this stability of predetermined characteristics gives additional credibility that the regression discontinuity design can deliver unbiased estimates in this scenario, suggesting that neither the impaired driver nor the police officer can manipulate testing on either side of the 0.08 threshold.

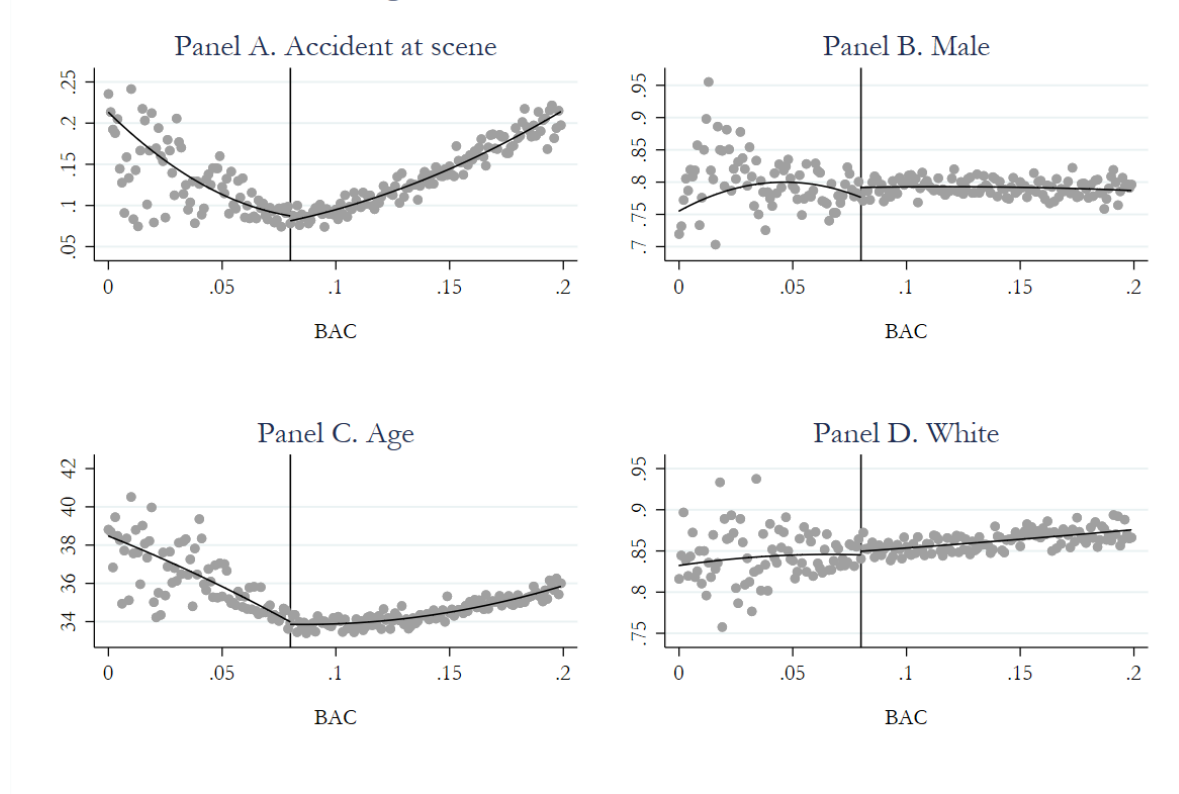
Point 6: Local continuity assumption of the covariates check.

Figure 3. Replication of Hansen's local continuity assumption of covariates graphs.



Note: Linear fit. Covariates implemented: age, gender, race, and accident at scene. Points represent the averages, with fitted values based on local linear models in black lines. The vertical black line represents the legal threshold at 0.08.

Figure 3-B. BAC and Characteristics



Note: Quadratic fit. Covariates implemented: age, gender, race, and accident at scene. Points represent the averages, with fitted values based on local quadratic models in black lines. The vertical black line represents the legal threshold at 0.08.

The lack of statistical significance in the regression's coefficients exhibited on the previous point, can be supported graphically in Figure 3; this diagram tries to replicate Hansen's one, presenting bins of predetermined characteristics and corresponding fitted linear and quadratic regression lines based on the structural model equation.

On one hand, when regarding panels A-D where linear fit was implemented, it can be observed that for neither the demographic factors taken into consideration nor the presence of an accident at scene, there was a substantial jump at the threshold. This implies that its behavior remains unchanged across the punishment, suggesting that the offenders and police are unable to manipulate the running variable.

On the other hand, if the panels corresponding to the quadratic fit are examined, a small jump is evidenced at the 0.08 cutoff for gender and accident at the scene characteristics. This can suggest that the local continuity assumption might not be entirely fulfilled for these factors, increasing the probability of an existent bias when estimating the causal effect on recidivism. Needless to say, a statistic test ought to be performed for both of them, taking into consideration a 2-degree polynomial so that there is formal evidence of whether this assumption is violated or not.

Point 7: Regression discontinuity design. Effects on recidivism.

To begin with, it should be stated that in order to replicate Hansen's regression discontinuity-based estimates, I decided to implement rectangular kernel for weighting as he did on his own paper. In Panel A the bandwidth was limited to 0.05. Column 1 shows the results when I control for the BAC linearly, Column 2 exhibits them when interacting BAC with the cutoff linearly, and lastly Column 3 is comprised of those ones derived when BAC is interacted with cutoff linearly and as a quadratic. In that sense, Panel B shows off the same columns, but it implements a bandwidth of 0.025.

Table 4. Regression discontinuity-based estimates.

	(1) Recidivism	(2) Recidivism	(3) Recidivism
<i>Panel A. BAC $\in [0.03, 0.13]$</i>			
<i>DUI</i>	-0.0273*** (0.00403)	-0.0240*** (0.00435)	-0.224** (0.0926)
Constant	0.0853*** (0.00672)	0.109*** (0.0131)	0.0262 (0.0473)
Controls	Yes	Yes	Yes
Observations	89,967	89,967	89,967
<i>Panel B. BAC $\in [0.55, 0.105]$</i>			
<i>DUI</i>	-0.0219*** (0.00558)	-0.0206*** (0.00575)	-0.470 (0.437)
Constant	0.0862*** (0.0154)	0.113*** (0.0278)	-0.104 (0.278)
Controls	Yes	Yes	Yes
Observations	46,957	46,957	46,957

Notes: This table contains regression discontinuity-based estimates of the effect of having BAC above the DUI 0.08 threshold on recidivism. Panel A contains estimates with a bandwidth of 0.05 while Panel B has a bandwidth of 0.025, with all regressions utilizing a rectangular kernel for weighting. Controls include indicators for race, gender, and age of the offender as well as accident on the scene. Robust standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

As in Hansen's paper, it was found that having a BAC above 0.08 threshold decreases recidivism by 2 percentage points during a four-year follow-up window. This estimate is statistically significant at the 1 percent level. It should also be mentioned that based on Table 4, it can be affirmed that the effect is consistent across both bandwidths and with the required interactions. Needless to say, the only result which differ from Hansen's is found in Panel B, column 3. This means that when implementing a 2-degree polynomial expression, the effect's magnitude, and significance changes substantially.

Point 8: Hansen's Figure 3 replication.

Figure 4. Effect of BAC's threshold on recidivism.

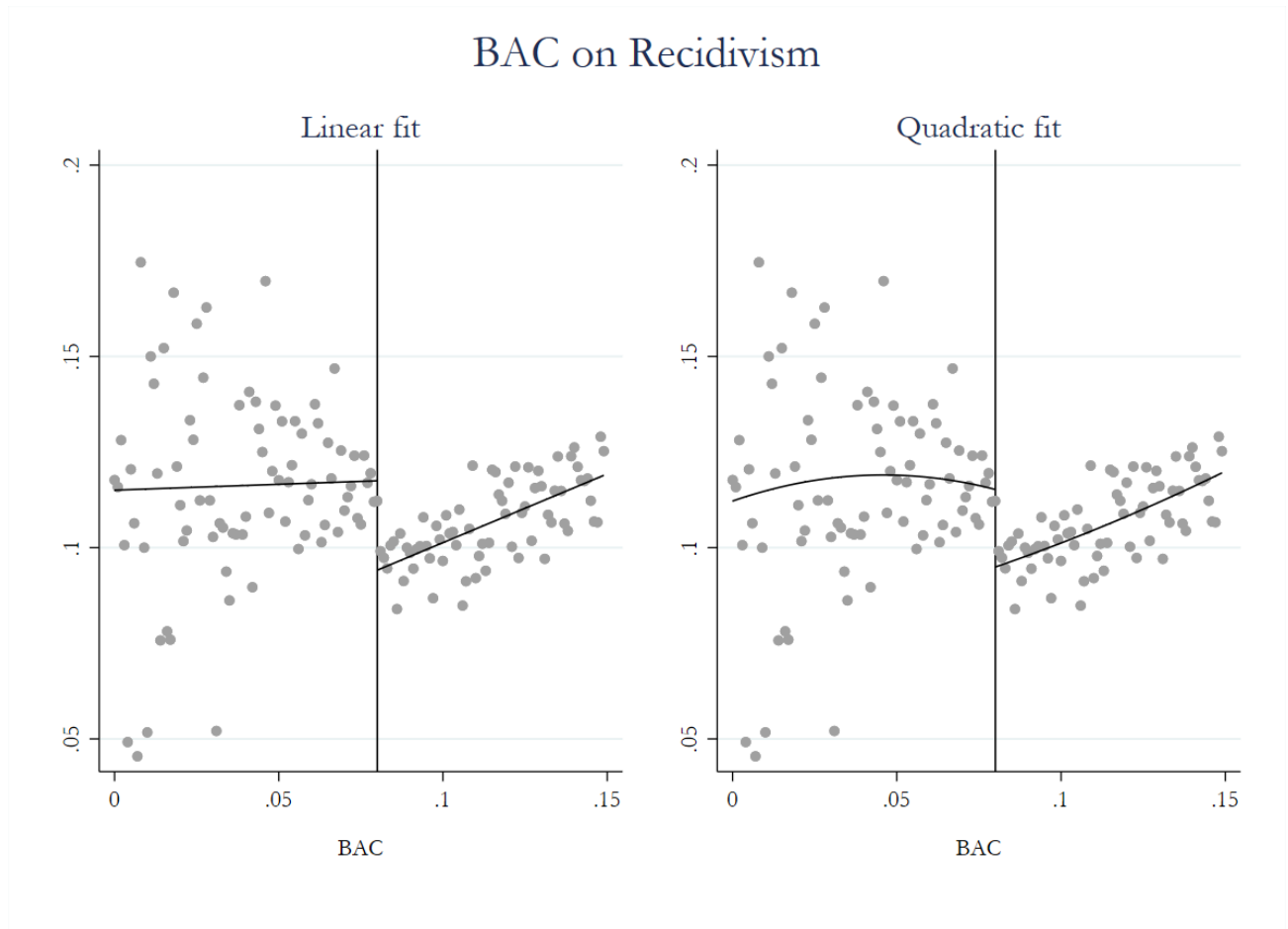


Figure 4 shows a statistically significant discontinuity at 0.08 threshold, supporting the results exhibited in Table 4. For both linear and quadratic fits, a reduction of nearly 2 percentage points on recidivism is evidenced at the cutoff.