Assignment #4

Tatiana Mojica Urueña

June 14, 2020

1 GitHub and Summary

1 Download Hansen_dwi.dta from github at the following address.

`https://github.com/scunning1975/causal-inference-class/raw/master/hansen_dwi`

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.

To access the RDD repository you can go to the following link. Folder Data has the database used for this assignment. `https://github.com/tatmoj17/RDD`

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, or "identification strategy"? What are his conclusions?

In the *Punishment and Deterrence: Evidence and Drunk Driving* paper, Benjamin Hansen explores the causal effect of harsher punishments and sanctions on driving under the influence (DUI). The punishments are determined by strict rules on blood alcohol content (BAC) and previous offenses. The author exploits the use of BAC thresholds to determine punishment severity, and the inability of either drivers or police to manipulate BAC.

The database used in this paper is a panel with administrative records on 512,964 DUI stops from the state of Washington. To identify the causal impact of the effectiveness of BAC thresholds in reducing drunk driving, Hansen uses a Regression Discontinuity Design (RDD). Results suggest that having a BAC above the DUI threshold reduces recidivism by up to 2 percentage points (17 percent). Likewise, having a BAC over the aggravated DUI threshold reduces recidivism by an additional percentage point (9 percent). Furthermore, additional sanctions experienced by drunk drivers at BAC thresholds are effective in reducing repeat drunk driving (Hansen, 2015).

2 Replication

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.

In the do-file you can see the variable created with all the specifications. Its called *cutoff.*

2 The first thing to do in any RDD is look at the raw data and see if there's any evidence for manipulation ("sorting on the running variable"). 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? Either recreate Figure 1 using the bac1 variable as your measure of blood alcohol content or use your own density test from software. Do you find evidence for sorting on the running variable?

To determine if people were capable of manipulating their blood alcohol content (BAC), we replicated Figure 1 of the *Punishment and Deterrence: Evidence and Drunk Driving* paper. From it, does not appear to be a problem of manipulation of the running variable around the cutoff value 0.08 - no jumps near this point. However, to be sure, we performed the Cattaneo, Jansson, and Ma manipulation test. Unlike McCrary test, this one is based

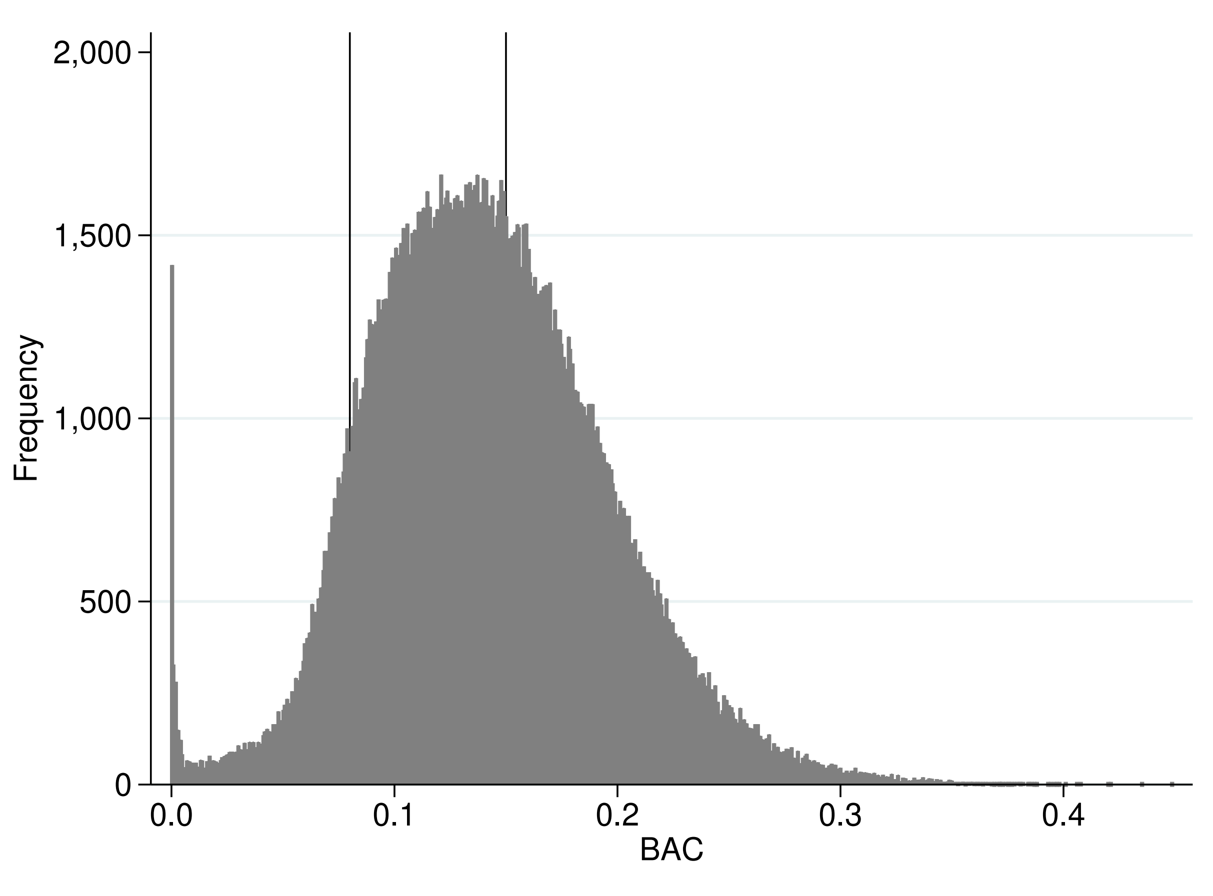

Figure 1: BAC Distribution

on a novel local polynomial density estimation technique that avoids pre-binning of the data and allows for restrictions on other features of the model (Cattaneo et al., 2017). First, we estimated the density of units near the cutoff to conduct a hypothesis test about whether the density is discontinuous. Second, we employed robust bias-correction to obtain valid statistical inference. Under the null-hypothesis, the density of the running variable is continuous near the cutoff, which means there is no evidence of manipulation.

Figure 2 reports the results from the Cattaneo, Jansson, and Ma test. In this test, we used the re-centered variable BAC, so the cutoff point is 0 instead of 0.08. There appears to be a discontinuity in the density near 0. The p-value of the test is 0.027. Therefore, for a significance level of 95%, we reject the null-hypothesis and we conclude there is evidence that people manipulated their blood alcohol content.

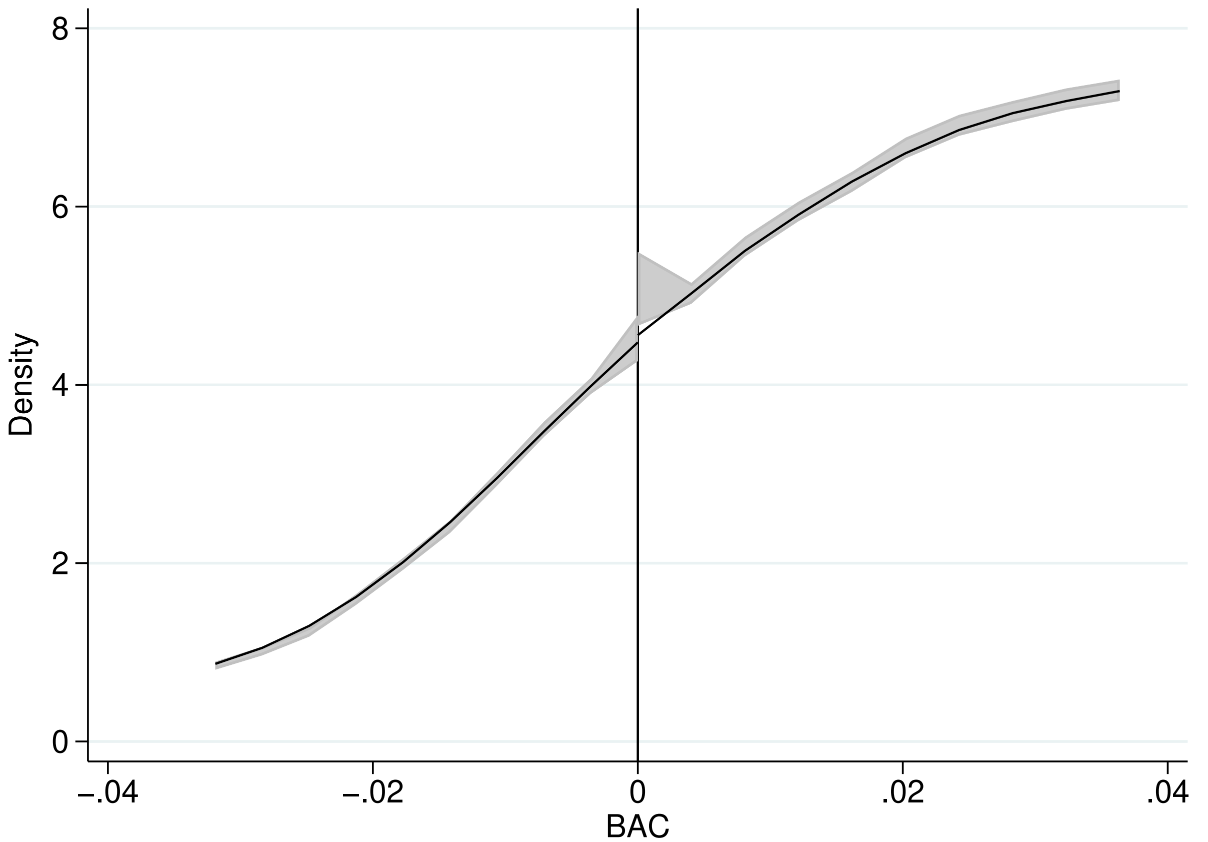

Figure 2: Manipulation Test

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.

Table 1 recreates Panel A of *Punishment and Deterrence: Evidence and Drunk Driving* paper Table 2. Through this table, the authors test if there was sorting by examining regression models that feature predetermined characteristics of the drivers - gender, race, and age - which should remain unchanged at the threshold. Also, they include the presence of an accident at the scene to determine whether police endogenously test people around the threshold. For a significance level of 95%, we fail to reject the null-hypothesis that the predetermined characteristics are unrelated to the BAC cutoff for DUI. Therefore, the

Table 1: Regression Discontinuity Estimates on Characteristics

	Driver Characteristics			
	Male	White	Age	Accident
Characteristics	(1)	(2)	(3)	(4)
Panel A. DUI threshold				
DUI	0.006	0.006	-0.140	-0.003
	[0.006]	[0.005]	[0.164]	[0.004]
Mean at (0.079)	0.792	0.853	34.17	0.101
Controls	No	No	No	No
Observations	89,967	89,967	89,967	89,967

Heteroskedastic robust standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1

coefficient are statistically equal to zero, and we conclude that the covariates are balanced at the cutoff.

Finally, the confidence intervals - displayed in the following figures - indicate that the coefficient estimations are quite precise. So, we are confident that there are no covariate imbalances in observable characteristics that could lead to biased treatment effect estimates.

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.

Figures 3 and 4 present bins of predetermined characteristics, corresponding fitted regression lines and confidence intervals based on equation (1), which should remain unchanged across the thresholds if drivers or police are unable to manipulate the running variable. The first figure is fitted with a linear regression model, while the second is fitted with a quadratic regression model. The quadratic model appears to be a better fit for the results, even though the linear model is closer to the one in Hansen's paper. According to him, the higher order polynomials offer little improvement in model fit, and there is relatively little empirical value to expanding to larger bandwidths.

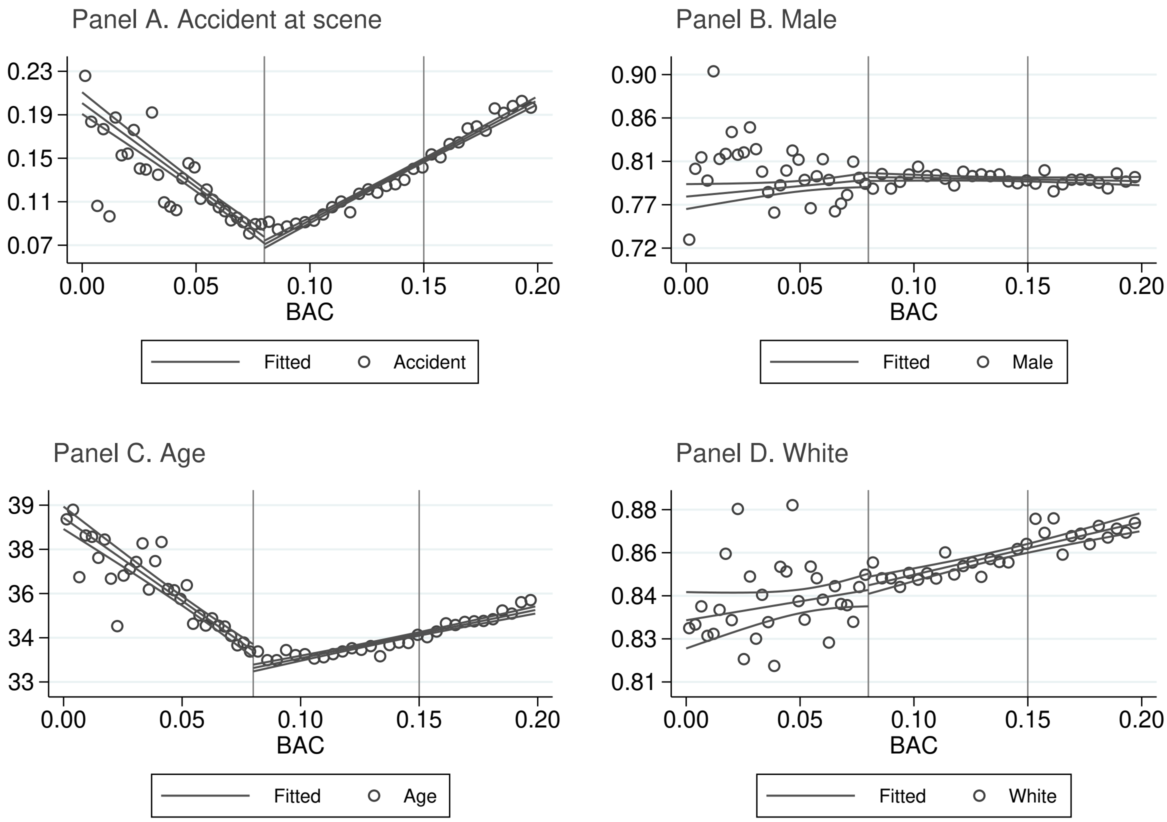

Figure 3: BAC and Characteristics - Linear

In both figures, we observe that demographic characteristics - gender, race, and age - are stable across the DUI thresholds. The presence of an accident at the scene is also unchanged. These findings go in line with the ones in Hansen's paper. The stability of the covariables gives credibility that the regression discontinuity design can deliver unbiased estimates in this scenario.

An advantage of the regression discontinuity design is that it only requires predetermined characteristics to be unaffected by BAC thresholds and not necessarily by the whole distribution. Therefore, we allow for unbalance in covariables at extreme points of the BAC distribution. Weaker assumptions are easier to fulfilled and facilitate the causal estimation of the treatment on the outcome.

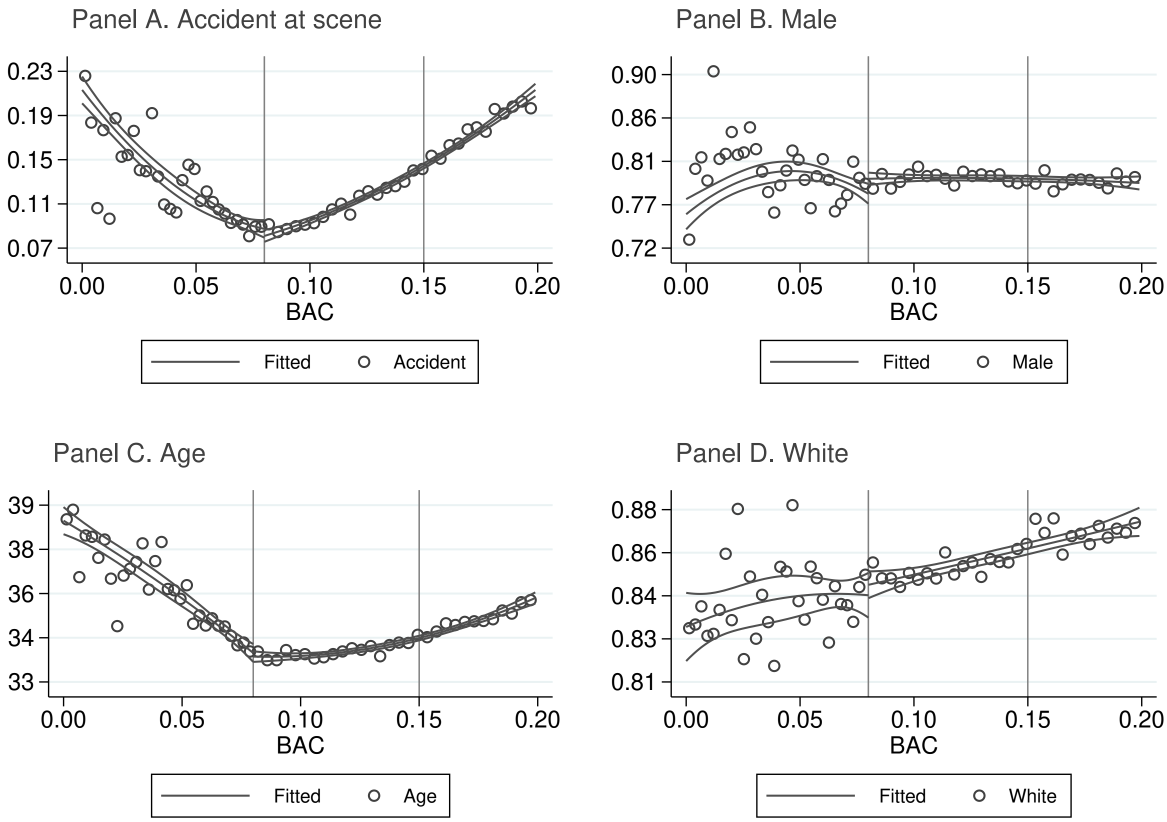

Figure 4: BAC and Characteristics - Quadratic

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. 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.:

(a) Column 1: control for the bac1 linearly

(b) Column 2: interact bac1 with cutoff linearly

(c) Column 3: interact bac1 with cutoff linearly and as a quadratic

(d) For all analysis, use heteroskedastic robust standard errors.

For Table 2, we used driver's demographic characteristics - age, gender and race - and the presence of an accident at the scene as **control variables.**

Table 2: Regression Discontinuity Estimates on Recidivism

	Linear (1)	Linear Int. (2)	Quadratic Int. (3)
Panel A. BAC $\in$ *[0.03, 0.13]*			
DUI	-0.027***	-0.024***	-0.014**
	[0.004]	[0.004]	[0.006]
Mean	0.107	0.107	0.107
Controls	Yes	Yes	Yes
Observations	89,967	89,967	89,967
Panel B. BAC $\in$ *[0.055, 0.105]*			
DUI	-0.022***	-0.021***	-0.014*
	[0.006]	[0.006]	[0.008]
Mean	0.105	0.105	0.105
Controls	Yes	Yes	Yes
Observations	46,957	46,957	46,957

Heteroskedastic robust standard errors in brackets

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

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

(a) Fit linear fit using only observations with less than 0.15 bac on the bac1.

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

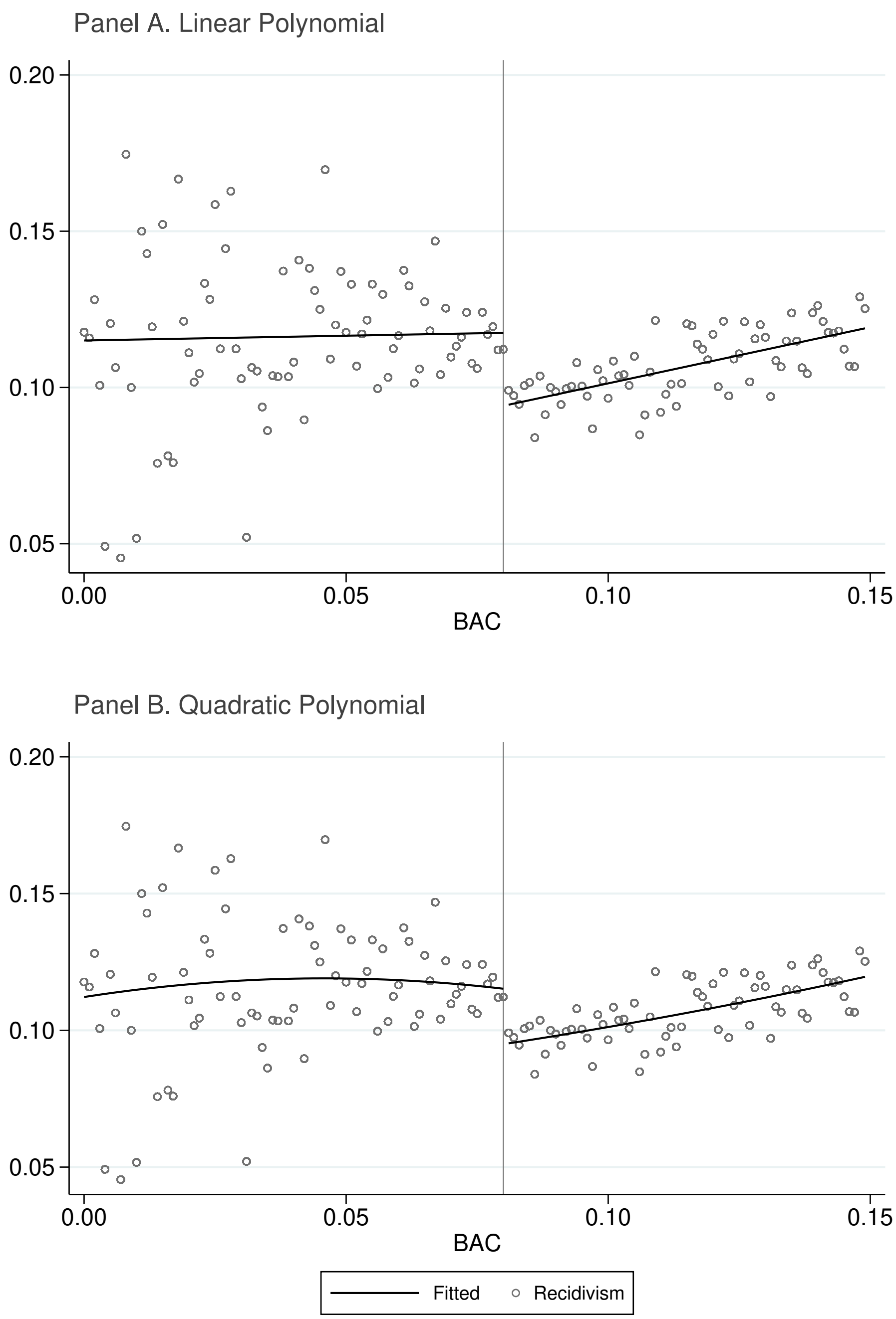

Figure 5: BAC and Recidivism

References

[1] Hansen, B. (2015). "Punishment and Deterrence: Evidence from Drunk Driving." *American Economic Review*. 105(4): 1581–1617

[2] Cattaneo, M. D., Jansson, M., and Ma, X. (2017) "rddensity: Manipulation Testing Based on Density Discontinuity." *The Stata Journal*. ii: 1-24.