**Assignment 4: Due Sunday, June 14th 2020. María Alejandra Martínez- CC 1072706035**

**Directions**: Please turn in your answers on separate paper, typed, and **beautifully written** with **beautiful tables** and **beautiful figures**.**[[1]](#footnote-1)**

**Github repo and summary (worth 2 points)**

1. Download Hansen\_dwi.dta from github at the following address.

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 in your assignment. 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.

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

1. In the writing subdirectory, place your assignment. For the first part of this assignment, read Hansen’s paper in the /articles directory of the main class github entitled “Hansen AER”. **Briefly summarize this paper**. What is his research question? What data does he use? What is his research design, or “identification strategy”? What are his conclusions?

The aim of this document is to estimate the causal effect of punishment for driving under alcohol on recidivism in the short and long term. For this purpose, data was collected from administrative records on BAC tests in Washington state from 1999 to 2007, period where an officer could arrest a driver if the BAC was above 0.08. By understanding the variable measure, a identification strategy based on a local linear regression discontinuity approach were conducted. This was also controlled by gender, race, prior offenses and fixed effects by age, year and county.

The results of this work suggest that those who were above the DUI cut-off point are less likely to repeat drunk driving in 2pp (17%). Likewise, the evidence shows that this effect increases up to 3pp (9%) when analyzing the aggravated DUI threshold. This conclusion is in accordance with economic models, as well as the national model of criminality, in which it is stated that the prominence of punishment contributes to reduce recidivism. To argue this, authors explain this effect can be explained by three possible channels: incapacitation, rehabilitation and deterrence.

**Replication (worth 6 points)**.[[2]](#footnote-2)

1. In the United States, an officer can arrest a driver if after giving them a blood alcohol content (BAC) 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.

gen DUI=0

replace DUI =1 if bac1>=0.08

1. 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?

histogram bac1, freq width(0.001) bcolor(gray) xtitle("BAC") ytitle("Frequency") title("BAC Histogram") xline(0.08, lcolor(black))

**Figure 1:BAC distribution**

Imagen que contiene texto, mapa

Descripción generada automáticamente

rddensity bac1, c(0.08) plot all

-----------------------------------------------

Method | T P>|T|

-------------------+---------------------------

Conventional. | 0.5337 0.5936

Robust | 2.2032 0.0276

-----------------------------------------------

Based on figure 1, there’s no evidence for sorting on the running variable, since at the cutoff point the sample distribution is continuous. This behavior suggests that people are not manipulating the blood alcohol content. This affirmation can also be proved with the Mc Crary density test, under the null hypothesis that density is continuous at the threshold. Given the p-value is 0.59, the hypothesis is not rejected and therefore there is no evidence of manipulation.

1. The second thing we need to do is check for covariate balance. Recreate Table 2 Panel A but only white male, age and accident (acc) as dependent variables. Use your equation 1) for this. Are the covariates balanced at the cutoff? It’s okay if they are not exactly the same as Hansen’s.

gen DUI\_c=DUI-0.08

gen DUI\_bac1=(DUI\_c\*bac1)

global control acc male age white

foreach var of global control{

xi: reg ‘var’ bac1 DUI\_c DUI\_bac1 if inrange(bac1,0.03,0.2), robust

}



Both for gender and race variables, the distribution sample is not affected at the BAC cutoff point, given the lack of significance in both regression coefficients. However, for variables such as age and the presence of an accident of scene, there is no covariates balance at the threshold due to the results show a significance at 95% for both cases.

It is important to mention that, this behavior differs from what was reported at Hansen’s paper due to this estimation didn’t include a bandwidth of 0.05 nor a rectangular kernel for weighting; for contrast it was estimated for a linear regression. Also, it is relevant to point that in this table it was included the 90% of the population, as mentioned in the paper.

1. 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.

ssc install cmogram

global control acc male age white

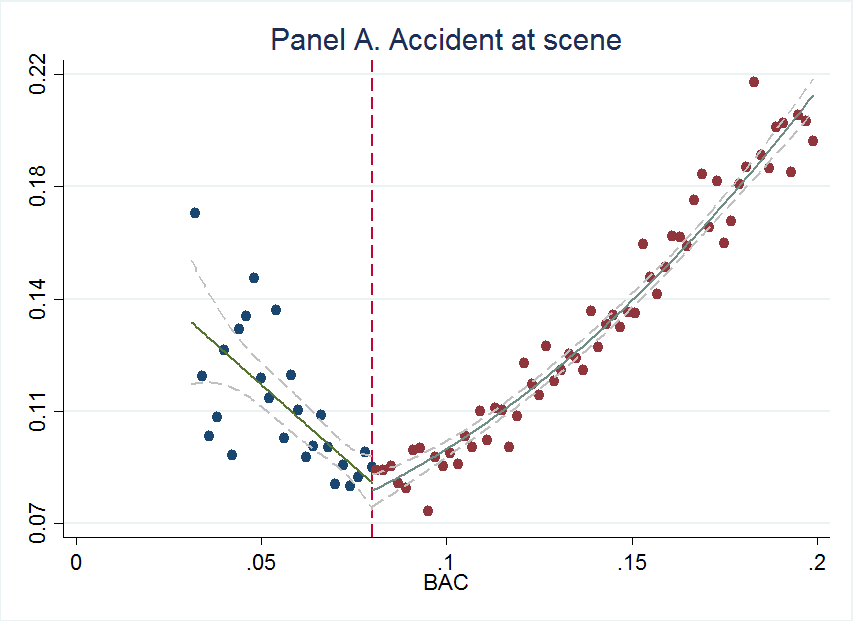
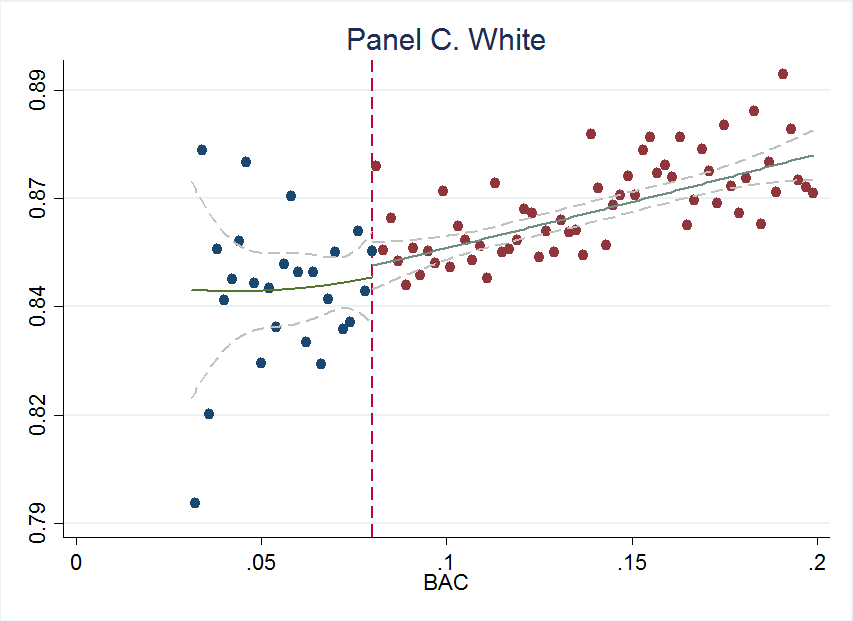
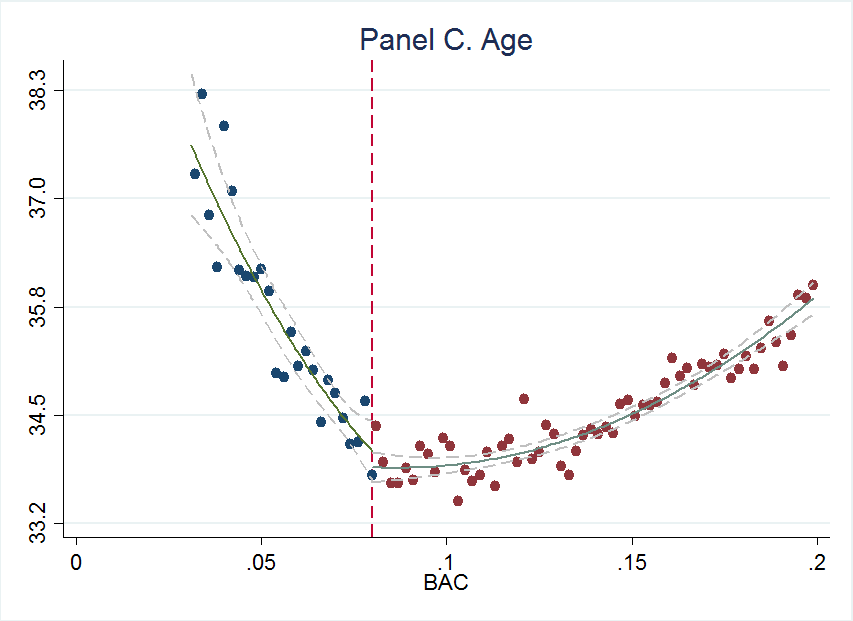
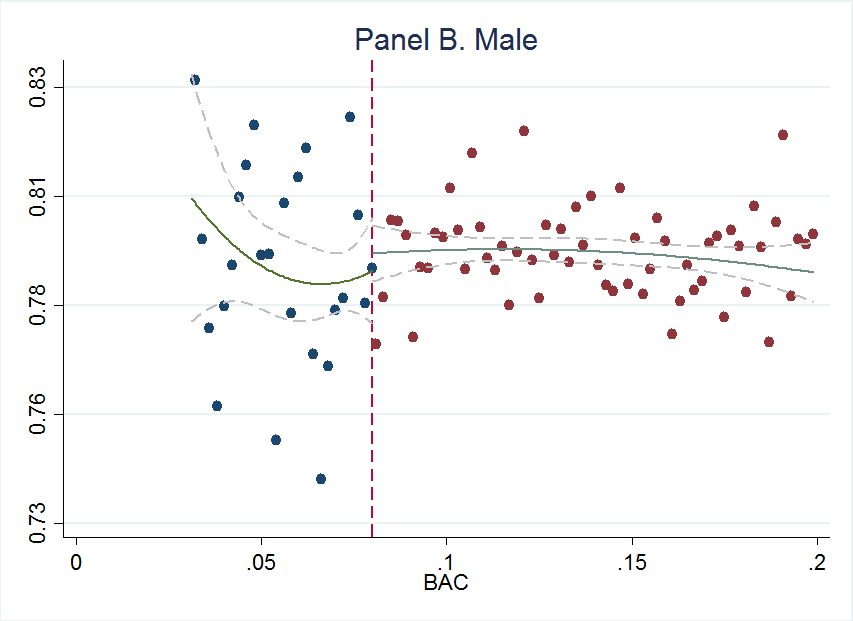
foreach var of global control{

cmogram ‘var’ bac1, cut(0.08) scatter line(.08) qfitci histopts(width(.002))ciopts(95

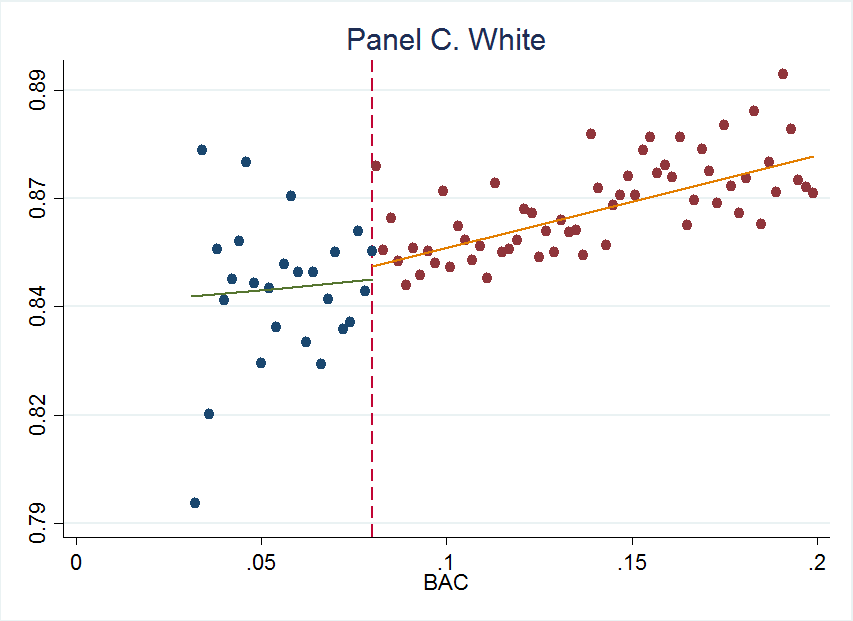
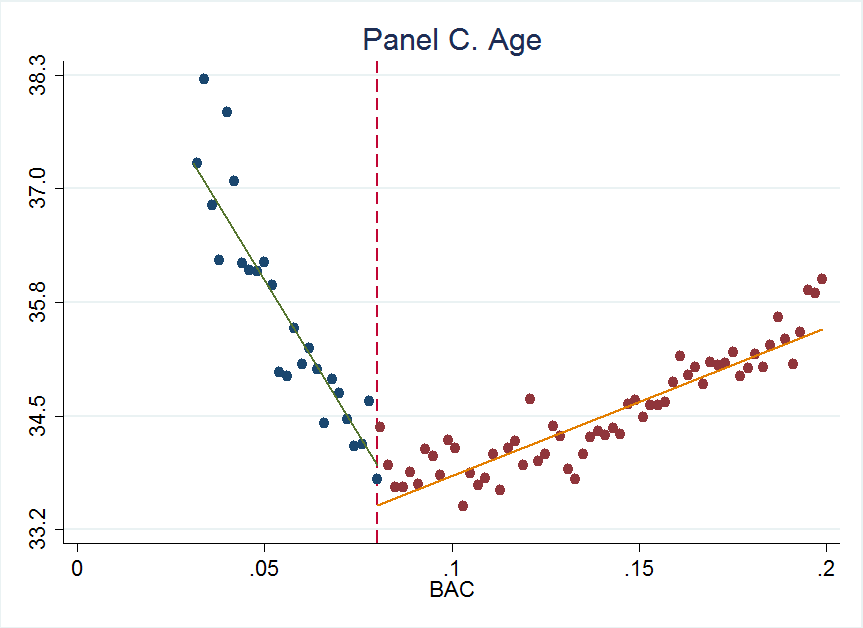
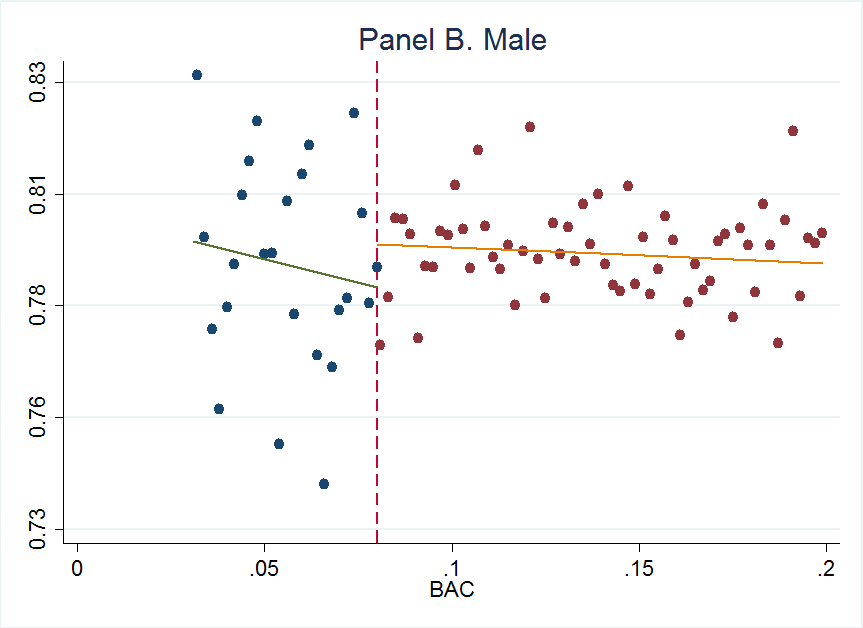
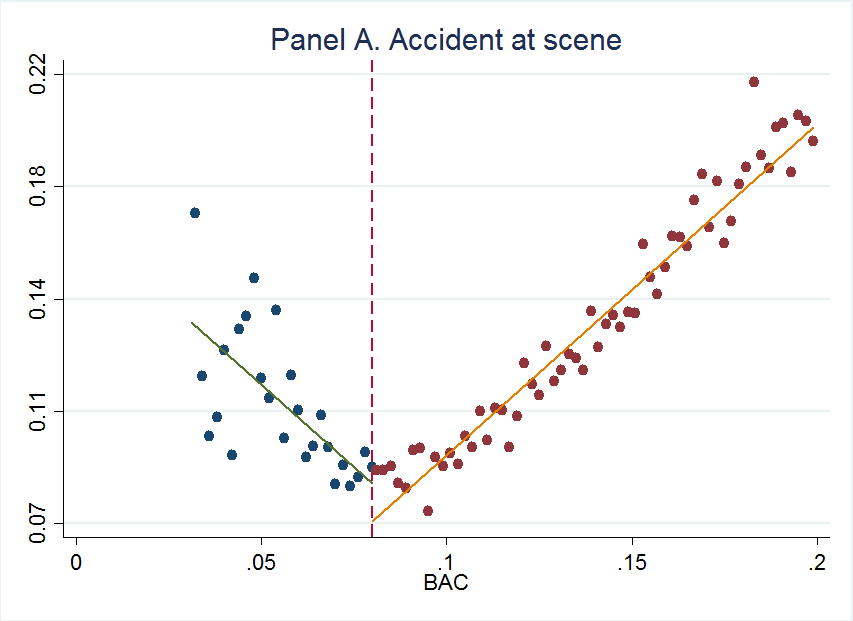
cmogram ‘var’ bac1, cut(0.08) scatter line(.08)lfit histopts(width(.002))ciopts(95)

}

**Figure 2. BAC and Characteristics - Quadratic fitted**

**Figure 3. BAC and Characteristics - Linear fitted**



Given the above graphics, it can be seen that the sample distribution of demographic variables, such as race, gender and age, remains unchanged across the punishment cutoff points. Also, for information like the presence of an accident at the scene. This behavior is observed both for linear and quadratic fit, implying that there is no evidence that the police nor the driver were able to manipulate the running variable.

However, by recreating these graphics from Hansen’s paper, it can be observed that, since the authors include in the analysis an additional aggravated threshold (0.15), the trend of the fitted regression lines can change. In effect, while in the paper it is reported a positive trend for the variable gender both for the ranges 0.08-0.15 and 0.15-0.2; in this case the accumulative effect in the total range is negative. Furthermore, the authors only establish a linear fit, but it seems that for factors like accidents at the scene and age, a quadratic fit represents the behavior of the distribution in a better way.

1. 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. Nevertheless, replicate Table 3, column 1, Panels A and B. 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.:
   1. Column 1: control for the bac1 linearly
   2. Column 2: interact bac1 with cutoff linearly
   3. Column 3: interact bac1 with cutoff linearly and as a quadratic
   4. For all analysis, use heteroskedastic robust standard errors.

gen DUI\_c=DUI-0.08

gen DUI\_bac1\_c= (DUI\_c\*bac1)

gen DUI\_bac1quad\_c= (DUI\_c\*(bac1^2))

\*\*Panel A: bandwidth (0.03-0.13)

reg recid DUI\_c bac1 acc male age white if inrange(bac1,0.03,0.13), robust

estimates store resid1

reg recid DUI\_c bac1 DUI\_bac1\_c acc male age white if inrange(bac1,0.03,0.13), robust

estimates store resid2

reg recid DUI\_c bac1 DUI\_bac1\_c DUI\_bac1quad\_c acc male age white if inrange(bac1,0.03,0.13), robust

estimates store resid3

outreg2 [resid1] using Table7\_1.doc, replace

outreg2 [resid2] using Table7\_1.doc, append

outreg2 [resid3] using Table7\_1.doc, append

\*\*Panel B: bandwidth (0.055-0.13)

reg recid DUI\_c bac1 acc male age white if inrange(bac1,0.055,0.105), robust

estimates store resid4

reg recid DUI\_c bac1 DUI\_bac1\_c acc male age white if inrange(bac1,0.055,0.105), robust

estimates store resid5

reg recid DUI\_c bac1 DUI\_bac1\_c DUI\_bac1quad\_c acc male age white if inrange(bac1,0.055,0.105), robust

estimates store resid6

outreg2 [resid4] using Table7\_2.doc, replace

outreg2 [resid5] using Table7\_2.doc, append

outreg2 [resid6] using Table7\_2.doc, append



1. Recreate the top panel of Figure 3 according to the following rule:
   1. Fit linear fit using only observations with less than 0.15 bac on the bac1

Keep if bac1<0.15

cmogram rec bac1, cut(0.08) scatter line(0.08) lfit histopts(width(.002)) ciopts(95)

Imagen que contiene texto, mapa, foto, computadora

Descripción generada automáticamente

* 1. Fit quadratic fit using only observations with less than 0.15 bac on the bac1

Keep if bac1<0.15

cmogram rec bac1, cut(0.08) scatter line(0.08) qfitci histopts(width(.002))ciopts(95)

Imagen que contiene texto, mapa

Descripción generada automáticamente

1. Again, my preference is that you attempt to create automated tables and automated figures as much as you can. I’ve placed a simple estout program called ols.do in the estout subdirectory. You just need to edit. [↑](#footnote-ref-1)
2. Much of this advice applies to Stata commands, but you can check the R files for lmb.r to see ways of doing the same in R. [↑](#footnote-ref-2)