**GitHub Link:** <https://github.com/juanmolina77/RDD>

**Replication (worth 6 points)**.[[1]](#footnote-1)

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.

It is in the R file.

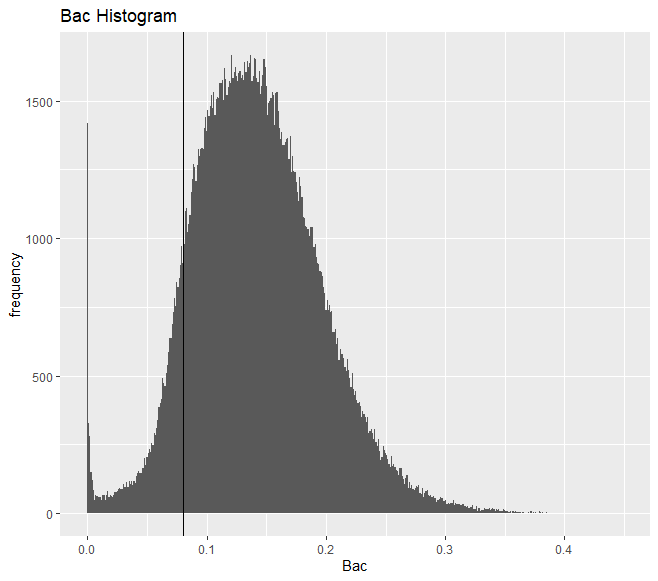
1. The first thing to do in any RDD is look at the raw data and see if there’s any evidence for manipulation (“sorting on the running variable”). If people were capable of manipulating their blood alcohol content (bac1), describe the test we would use to check for this. Now evaluate whether you see this 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?

The test we would use to see if there is manipulation of the running variable is the **McCrary Density Test**, in this case this is plausible that happens because people knows approximately how much they can drink without going over the first Bac limit that is 0.08, or even there could be people that bet not to exceed the limit of 0.15.

In the case of RDD the manipulation implies heaping on the “good side” of c0 (the cutoff)

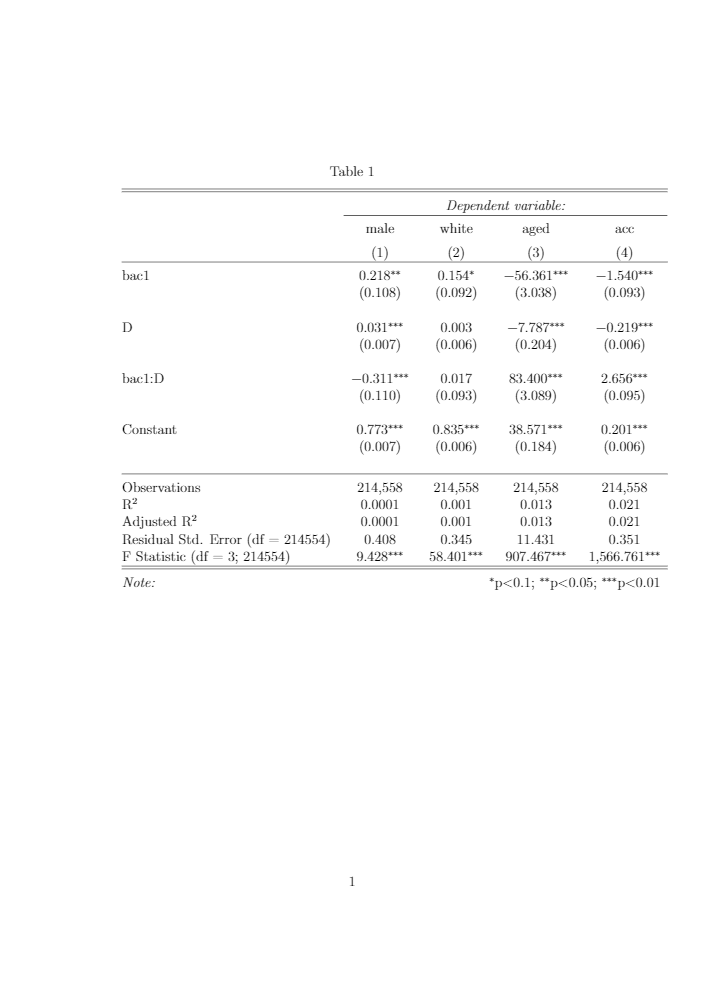
McCrary Density Test: under the null the density should be continuous at the cutoff point and under alternative hypothesis, the density should increase at the kink. Seeing the histogram, I think that there is not clear evidence of manipulation since we do not observe big jumps, like the one we observe in the study of the Marathon where there was a big jump in the cutoff of 4 hours.

**When running McCrary Density test in R**, we reject the null hypothesis, which means there could be manipulation, however if we run placebos at different cutpoints (0.06,0.07,0.09,0.1) it also rejects the null at that numbers, that leads us to think that there is no evidence of sorting on the running variable.



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.

Values are different than in Hansen´s paper since we are not using bandwidth neither kernel specifications. Other factor could be that this data is different, we have more observations. However, the signs are okay, so the interpretation is quite the same. Having the “controls” as dependent variables is need se the balance if covariates by a running variable. It´s easier to see it if we graph because there if there is smoothness through the cutoff, that implies balance, but if non-collider covariates jump at the cut off that implies no balance



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.

I set the Y axis not exactly as Hansen did it, I wide them a little bit.

I recreated both, linear and quadratic forms for each panel from A to D. However, the quadratic figures I do not paste them here because Hansen only did the linear ones, if the reader wants to see the quadratic recreations please access to my GitHub account: <https://github.com/juanmolina77/RDD/tree/master/Figures>

Imagen que contiene texto, mapa

Descripción generada automáticamenteImagen que contiene texto, mapa

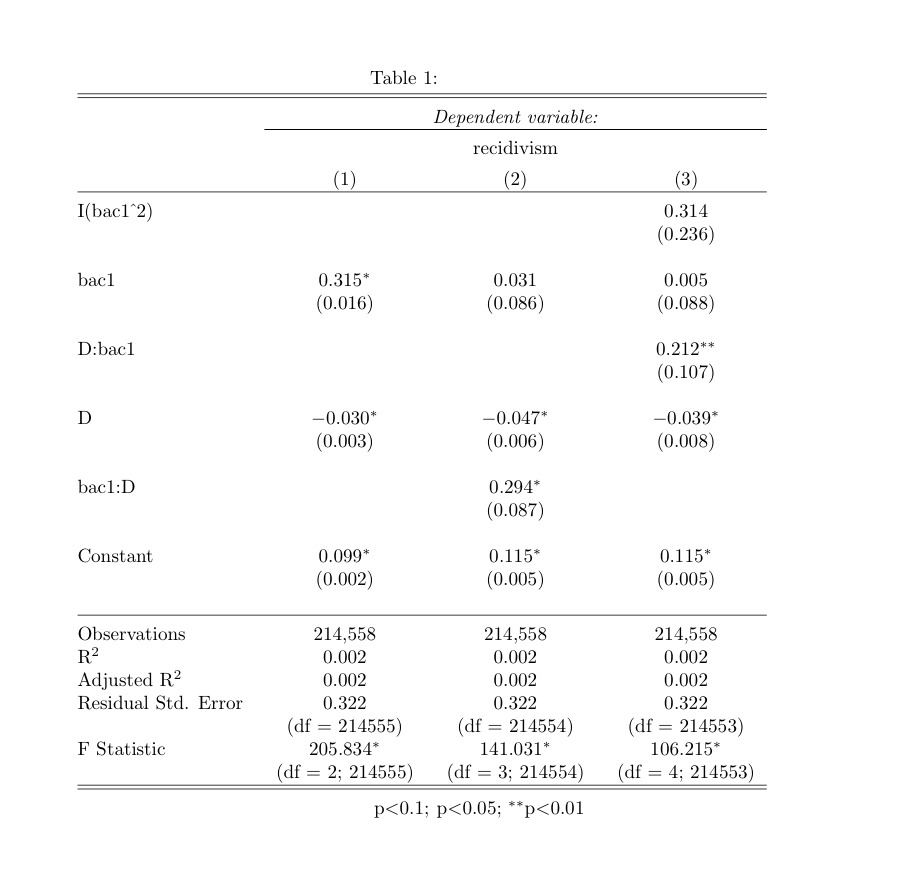
Descripción generada automáticamente

Imagen que contiene texto, mapa

Descripción generada automáticamenteImagen que contiene texto, mapa

Descripción generada automáticamente

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.

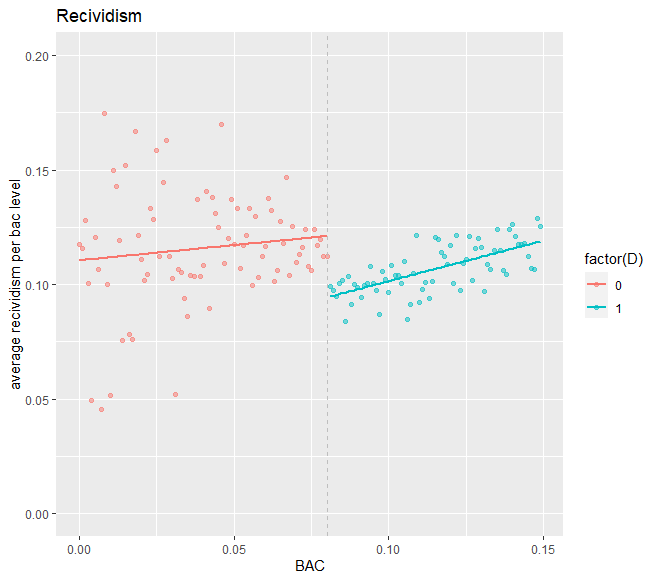


D:bac1 and bac1:D are the same

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

Recreation of Figure 3 panel A that is the one of all-ofenders.

In comparison with Hansen, the first part, before bac 0.08 is different, mine is upward sloping and Hansen´s is downward sloping. However, the second part is slightly the same

Imagen que contiene texto, mapa

Descripción generada automáticamente

1. 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-1)