**Assignment 4: Due Sunday, June 14th 2020**

**Jorge Alberto Guerra España**

**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.

use https://github.com/scunning1975/causal-inference-class/raw/master/hansen\_dwi, clear

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.

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?

In this article, the authors want to answer the following question: **Do BAC limits as currently administered reduce future drunk driving?** They seek to offer evidence about the limitation of punishments and limitations in reducing recidivism among drunk drivers. For that purpose, they use **administrative data from 512,964 DUI-BAC tests in Washington state from 1999 to 2007**. With these data, the authors analyze the Causal Effect of having BAC threshold of 0.08 (DUI) or 0.15 (aggravated DUI) in recidivism within four years of the original BAC test. The specific limits for DUI and Aggravated DUI allow the use of a **discontinuous regression design as identification strategy**. The authors **concluded that**: having a BAC per threshold of 0.08 or 0.15 is associated with reduced driving in a state of repeated drunkenness in the short and long term. Specifically, they find: 1) having a BAC above the legal limit of 0.08 corresponding to a 2 percent decrease in drunk driving repetition in the next four years; 2) having a BAC above the improved penalty limit of 0.15 is associated with an additional 1 percentage point decrease in drunk driving repetition.

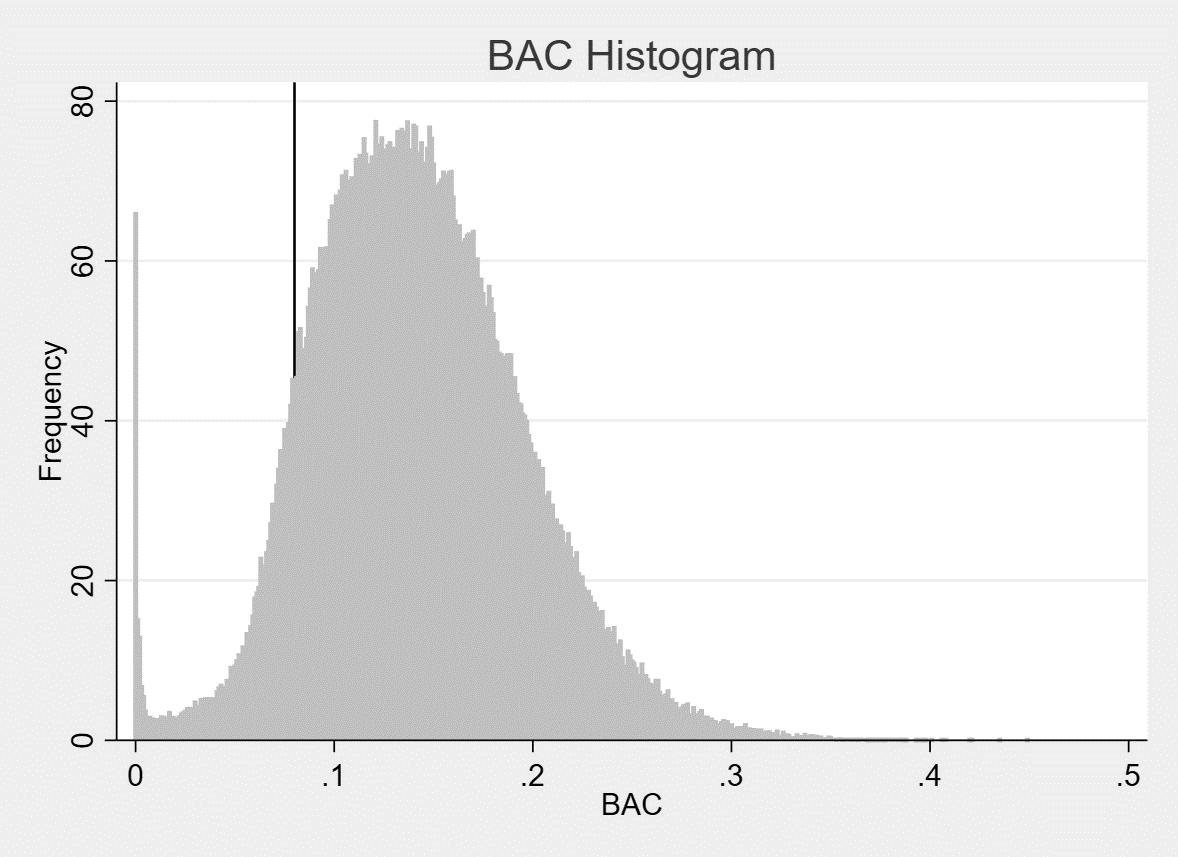
**Replication (worth 6 points)**.[[2]](#footnote-2)

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.
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?

The test is the McCrary Density Test. In this test: the null hypothesis contemplates that the density function (in this case of BAC1) must be continuous at the cutoff point (0.08). In contrast, the alternative hypothesis states that the density function has a jump at the cutoff point (0.08). To do the test: 1) Divide the execution variable into bins and calculate the frequency between bins; 2) These frequency counts are treated as a dependent variable in a local linear regression.

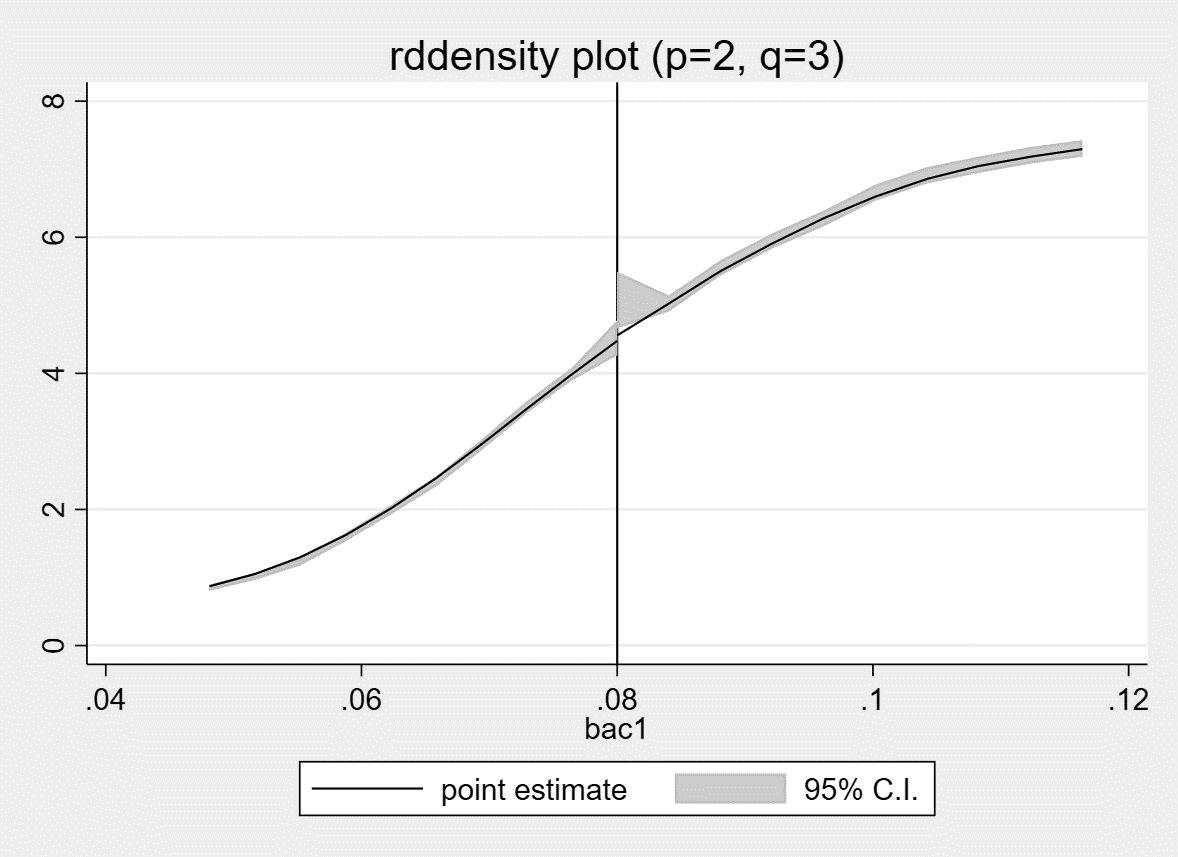
Looking at the data and recreating Figure 1, it appears that the function is continuous at BAC = 0.08.

**Figure 1. BAC Histogram**



But when doing the McCrary Density Test, the p-value (0.0276) with the robust method returns if there is a discontinuity, that is, people were able to manipulate the BAC.

**Figure 2. McCrary Density Test**



However, the authors find the opposite. When doing when doing the test, they are based on the conventional method that gives a p-value of 0.59 (as it is in the paper) and therefore, there is no evidence to affirm that there is manipulation in the BAC. At this point, placebos effects were estimated for 0.04, 0.04, 0.09 and 0.1 BAC points, but they presented continuity, there was no evidence of jumps.

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.

**Table 2: Regression Discontinuity Estimates for the Effect of Exceeding BAC Thresholds on Predetermined Characteristics**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) |
| VARIABLES | male | white | aged | acc |
|  |  |  |  |  |
| dui | 0.0307\*\*\* | 0.00271 | -7.787\*\*\* | -0.219\*\*\* |
|  | (0.00729) | (0.00617) | (0.204) | (0.00627) |
| bac1 | 0.218\*\* | 0.154\* | -56.36\*\*\* | -1.540\*\*\* |
|  | (0.108) | (0.0917) | (3.038) | (0.0932) |
| duibac1 | -0.311\*\*\* | 0.0170 | 83.40\*\*\* | 2.656\*\*\* |
|  | (0.110) | (0.0933) | (3.089) | (0.0948) |
| Constant | 0.773\*\*\* | 0.835\*\*\* | 38.57\*\*\* | 0.201\*\*\* |
|  | (0.00655) | (0.00555) | (0.184) | (0.00564) |
|  |  |  |  |  |
| Observations | 214,558 | 214,558 | 214,558 | 214,558 |
| Mean | 0.789 | 0.861 | 34.957 | 0.354 |
| Controls | NO | NO | NO | NO |

Standard errors in parentheses

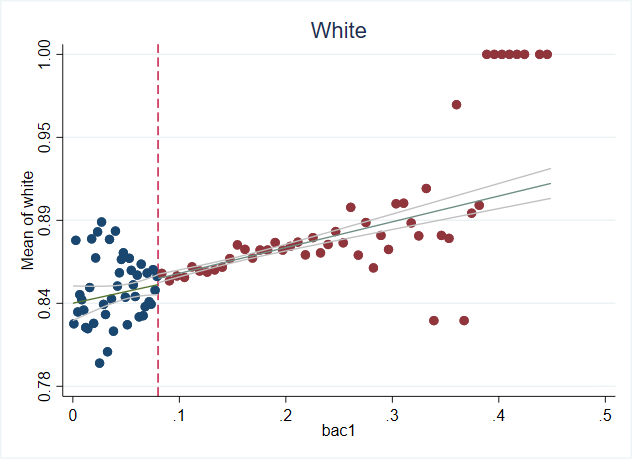
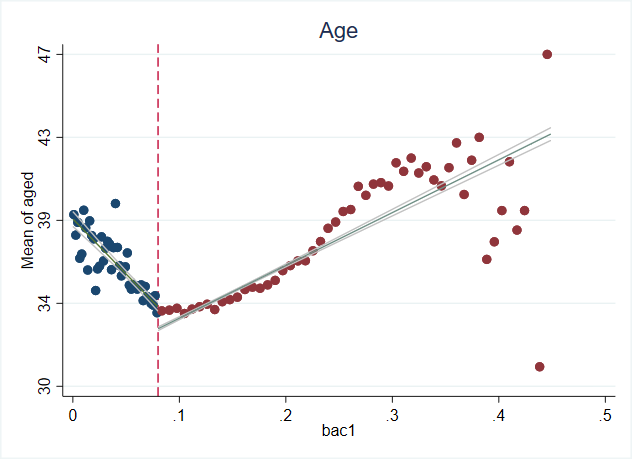
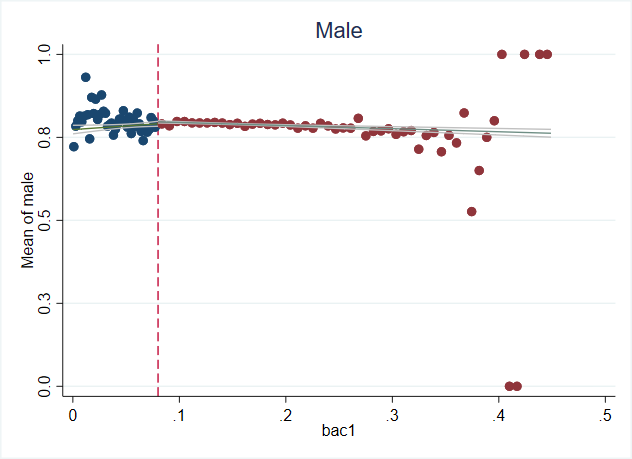
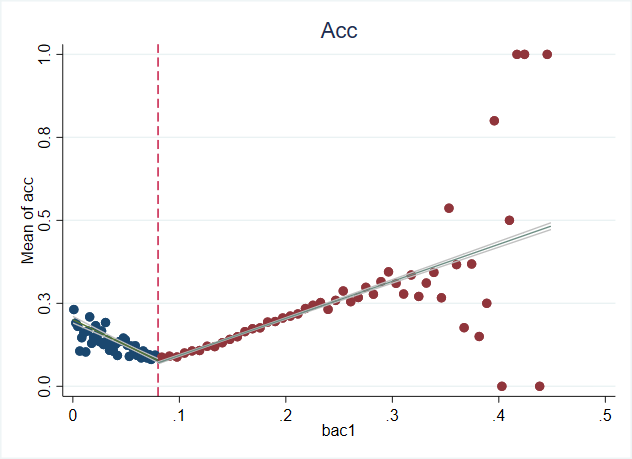
\*\*\* p<0.01, \*\* p<0.05, \* p<0.

In the paper the authors conclude that the covariates are balanced at the cutoff, but with the sample to this assignment the results changed. Only the covariate "white" is balanced on the cutoff. On the contrary, the variables male, age and accidents (acc) are unbalanced in the cutoff, according to the significance of the coefficients that accompany the DUI variable and the interaction with the continuous variable BAC (duibac1).

Speculating from the results, it could be said that it is relatively easier to change the age, the existence of an accident than the race. However, the reason why violators want to change gender is not so intuitive.

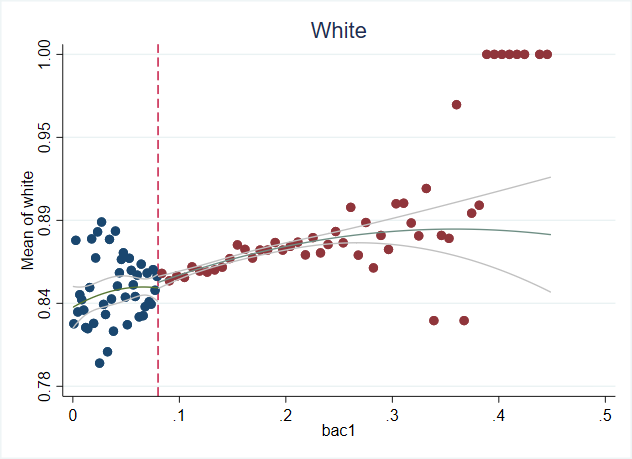
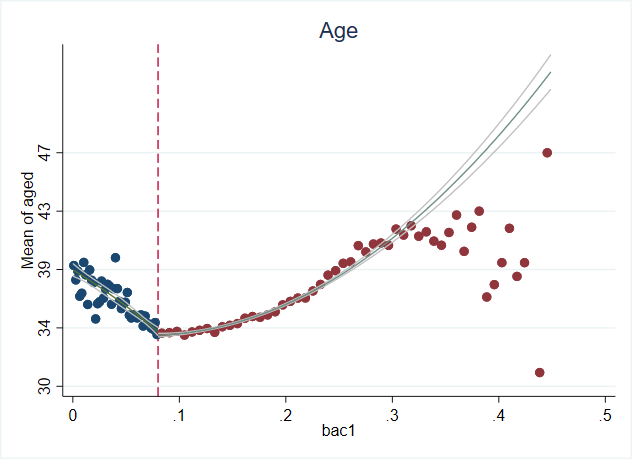
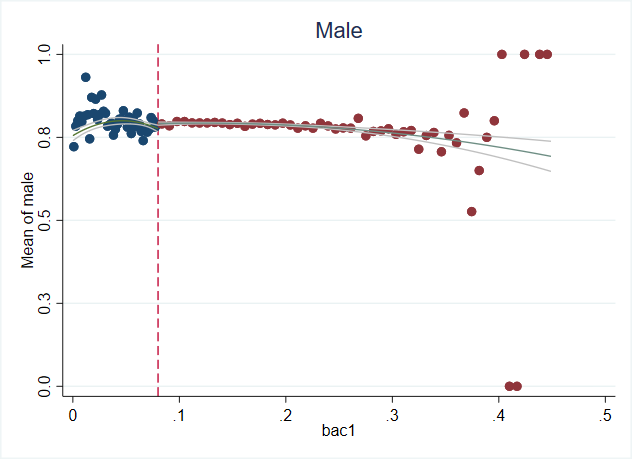
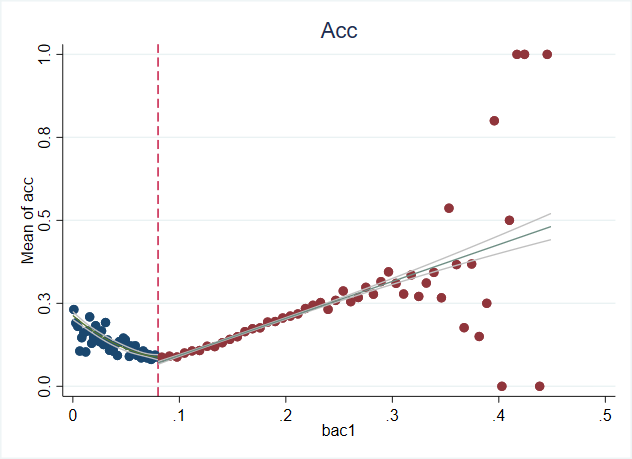
1. Recreate Figure 2 panel A-D. You can use the -cmogram- command in Stata to do this. Fit both linear and quadratic with confidence intervals (CI). Discuss what you find and compare it with Hansen’s paper.

**Figure 3: BAC and Characteristics with linear CI**



Through figure 3 and 4, it can be seen that the presence of accidents, gender, age and race (White) show stability in the DUI cutoff (0.08), with quadratic and linear confidence intervals. The results are the same as the authors find in the paper. As the authors say, the stability of these variables supports the design of the discontinuous regression and the credibility of its effects.

**Figure 4: BAC and Characteristics with quadratic CI**



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 (**1’)**
   2. Column 2: interact bac1 with cutoff linearly **(2’)**
   3. Column 3: interact bac1 with cutoff linearly and as a quadratic **(3’)**
   4. For all analysis, use heteroskedastic robust standard errors.

**Table 3: Regression Discontinuity Estimates for the Effect of Exceeding the 0.08 BAC Threshold on Recidivism**

|  |  |  |  |
| --- | --- | --- | --- |
|  | (1) | (2) | (3) |
| VARIABLES | PANEL A (1’) | PANEL A (2’) | PANEL A (3’) |
|  |  |  |  |
| dui | -0.0273\*\*\* | -0.0591\*\*\* | 0.0298 |
|  | (0.00403) | (0.0152) | (0.0709) |
| bac1 | 0.321\*\*\* | -0.0429 | -0.0429 |
|  | (0.0748) | (0.187) | (0.187) |
| duibac1 |  | 0.438\*\* | -1.265 |
|  |  | (0.204) | (1.346) |
| duibac1q |  |  | 8.014 |
|  |  |  | (6.276) |
| Constant | 0.0853\*\*\* | 0.109\*\*\* | 0.109\*\*\* |
|  | (0.00672) | (0.0131) | (0.0131) |
|  |  |  |  |
| Observations | 89,967 | 89,967 | 89,967 |
| Mean | 0.0984 | 0.0984 | 0.0984 |
| Controls | YES | YES | YES |

|  |  |  |  |
| --- | --- | --- | --- |
|  | (1) | (2) | (3) |
|  | PANEL B (1’) | PANEL B (2’) | PANEL B (3’) |
|  |  |  |  |
| dui | -0.0219\*\*\* | -0.0643\* | 0.154 |
|  | (0.00558) | (0.0350) | (0.318) |
| bac1 | 0.188 | -0.196 | -0.196 |
|  | (0.201) | (0.383) | (0.383) |
| duibac1 |  | 0.547 | -4.158 |
|  |  | (0.449) | (6.838) |
| duibac1q |  |  | 25.20 |
|  |  |  | (36.57) |
| Constant | 0.0862\*\*\* | 0.113\*\*\* | 0.113\*\*\* |
|  | (0.0154) | (0.0278) | (0.0278) |
|  |  |  |  |
| Observations | 46,957 | 46,957 | 46,957 |
| Mean | 0.0863 | 0.0863 | 0.0863 |
| Controls | YES | YES | YES |

Robust standard errors in parentheses

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

In panel A (bandwidth from 0.03 to 0.13) and with the specification (1 '), the causal effect of having a BAC above the legal limit of 0.08, corresponds to a 2.73% decrease in the possibility of drunk driving again in the next 4 years (recidivism), this result is significant at 1%. With specification (2'), the effect is 5.9%, with 1% significance. However, the causal effect in the specification (3’) is not significant.

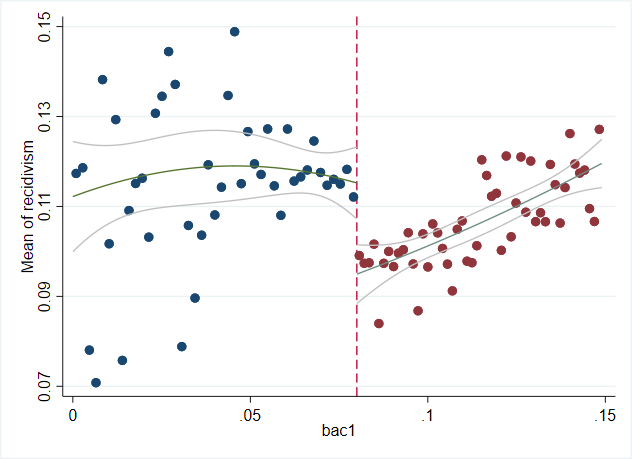
The authors estimate that the effect, with the specification (2 ''), is 2%, which differs by 4 p.p. with the effect found in this exercise. The authors 'results are more like the specification (1').

In panel B (bandwidth from 0.055 to 0.105) and with the specification (1 '), the causal effect of having a BAC above the legal limit of 0.08, corresponds to a 2.19% decrease in the possibility of drunk driving again in the next 4 years (recidivism), this result is significant at 1%. With specification (2 '), the effect is 6.43%, whit 10% significance. However, the causal effect in the specification (3’) is not significant.

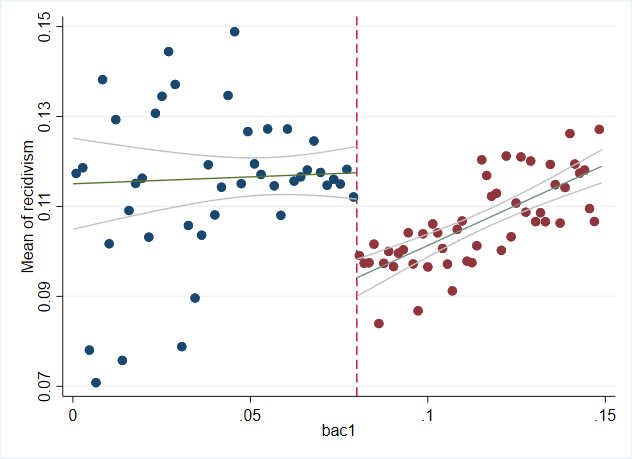
The authors estimate that the effect, with the specification (2 ''), is 2%, which differs by 4.43 p.p. with the effect found in this exercise. The authors 'results are more like the specification (1').

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

**Figure 5: BAC and Recidivism with linear CI**



**Figure 5: BAC and Recidivism with quadratic CI**



In the graphs it can be seen that, regardless of the type of adjustment (quadratic or linear), there is an evident jump in BAC1, that under the assumptions of smoothness, no accumulation under the running variable and continuity in the covariates, the causal effect can to be found.

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)