Hansen Replication

Santiago Silva

Feb 19, 2021

This is a replication of some of the figures and tables of Hansen's BAC research using Regression Discontinutiy Design

1)

first we load the data

```
. use https://github.com/scunning1975/causal-inference-class/r
aw/master/hansen_dwi, clear

. eststo clear
```

link to my repo is https://github.com/santiago-silva-1/RDD

we want to know what we are dealing with

```
. quietly estpost sum Alcohol1 Alcohol2 low_score male white r
ecidivism acc aged year bac1 bac2

. esttab, cells("count mean sd min max") compress title(Summar
y Statistics)

Summary Statistics
------------------------------------------------------------
                  (1)

                count       mean         sd        min        max
------------------------------------------------------------
Alcohol1       214558   141.7345   53.90077          0        449
Alcohol2       214558   141.0491    53.8998          0        435
low_score      214558   139.2809   53.36833          0        435
male           214558   .7895115   .4076566          0          1
white          214558   .8615899   .3453307          0          1
recidivism     214558   .1176325   .3221732          0          1
acc            214558   .1472935   .3543991          0          1
aged           214558   34.95732   11.50298         21         80
year           214558   2003.164   2.532329       1999       2007
bac1           214558   .1417345   .0539008          0       .449
bac2           214558   .1410491   .0538998          0       .435
------------------------------------------------------------
N              214558
------------------------------------------------------------
```

also look at the variables in the data editor tab. The outcome variable is "recidivism", which measures whether the person showed back in the data within 4 months.

2)github

The paper tests the hipothesis whether or not a punishemnt for drunk drivers that show back up in the data within 4 months has an effect on drivers showing up back in the data. Hansen gets his data from administrative records on 512,964 DUI stops from the state of Washington. Hansen uses a sharp RDD design and concludes that recivisim punishemnt does have an effect of drunk driver recivisim

3)Create a dummy equaling 1 if bac1>= 0.08 and 0 otherwise

```
. gen D = 0

. replace D = 1 if bac1>=0.08
(191,548 real changes made)
```

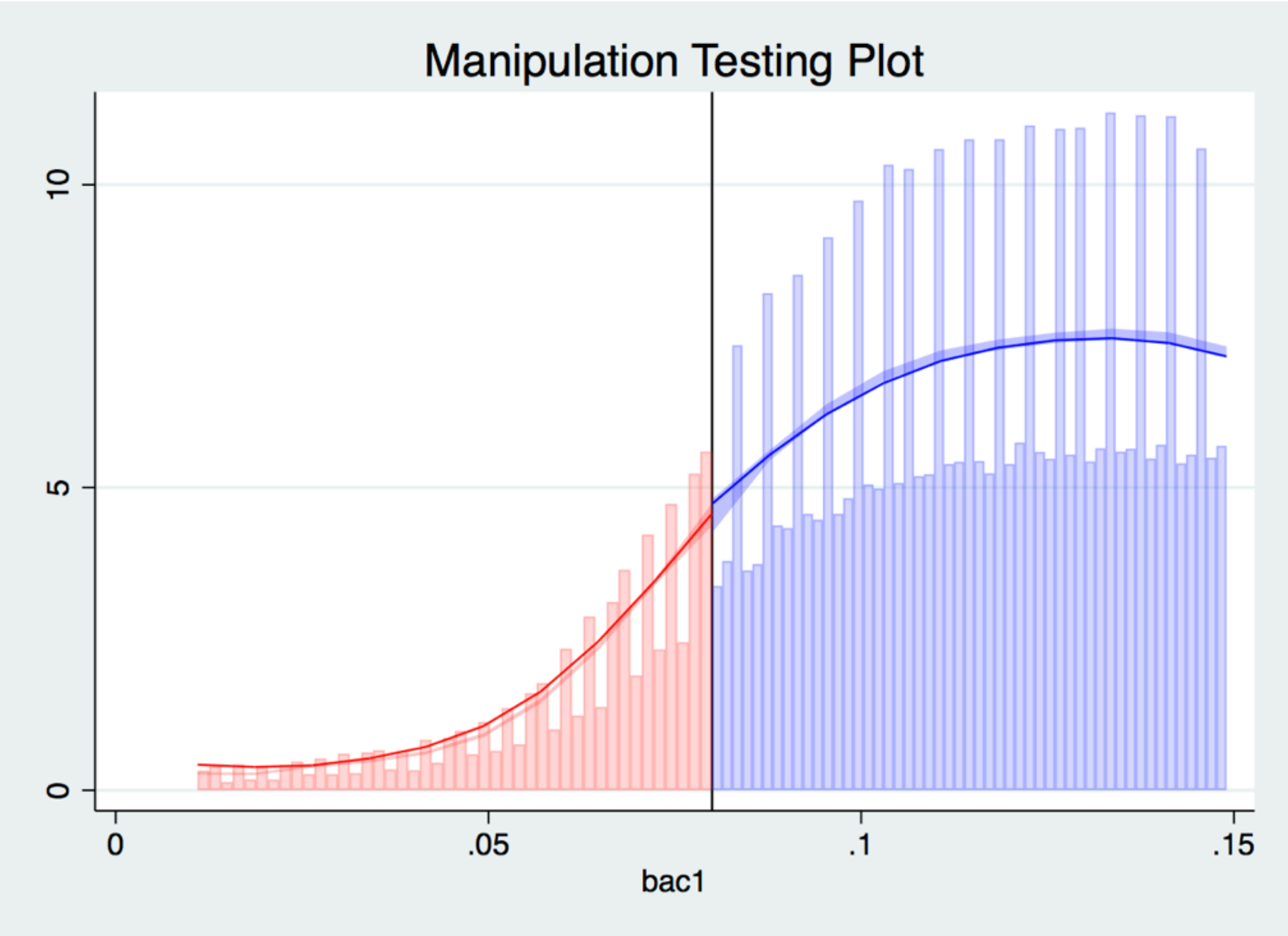
4)

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 manipulatingd 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? Explain your results. Compare what you found to what Hansen found.

Manipulation in the running variable is one of the most common viloations for smothness. We can use the McCrady Desnity test, look at slide 317. Now we run the density test on our running variable, we declare the cutoff at 0.08, and we plot the result

```
. rddensity bac1, c(0.08) plot
Computing data-driven bandwidth selectors.

Point estimates and standard errors have been adjusted for rep
eated observations.
(Use option nomasspoints to suppress this adjustment.)

RD Manipulation test using local polynomial density estimation
.

     c =      0.080 | Left of c  Right of c          Number of
obs =       214558
--------------------+----------------------          Model
    = unrestricted
```

```
    Number of obs |      23010       191548            BW method
    =          comb
Eff. Number of obs |      14727        28946            Kernel
    =    triangular
   Order est. (p) |          2            2            VCE method
    =     jackknife
   Order bias (q) |          3            3
      BW est. (h) |      0.023        0.023

Running variable: bac1.
--------------------------------------------
            Method |        T           P>|T|
-------------------+------------------------
            Robust |   -0.1387        0.8897
--------------------------------------------

P-values of binomial tests. (H0: prob = .5)
-------------------------------------------------------
 Window Length / 2 |        <c         >=c |     P>|T|
-------------------+-----------------------+-----------
             0.000 |       909           0 |    0.0000
             0.000 |       909           0 |    0.0000
             0.000 |       909           0 |    0.0000
             0.000 |       909           0 |    0.0000
             0.000 |       909           0 |    0.0000
             0.000 |       909           0 |    0.0000
             0.000 |       909           0 |    0.0000
             0.000 |       909           0 |    0.0000
             0.000 |       909           0 |    0.0000
             0.000 |       909           0 |    0.0000
-------------------------------------------------------

. graph export Figure11.png , replace
(file /Users/santiagosilva/Downloads/Figure11.png written in P
NG format)
```



.

Confidence intervals intersect, no violation of discontiniity. Hansen did not find a violation of density either, this does look like the corresponding part of Hanses's Figure1.

5)Balance test

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.

What we do in a covariate balance test is we run our main equation again, our model, but now instead of using the dependent variable we have been usign-the outcome variable, which is "recid"-, now we are going to run the model on "feature predetermine characteristics of drivers which should remain unchanged at the threshold". In this case we are using: white, male, age, and accident at the scene.

The model is recidivism= beta_0 + b_1 D + b_2 bac1 + b_3 D*bac_1.

If we use command xi and i. we do not need to create an interaction term, ststa will do it for us and will run the regression on the treatment variable (D) , the running variable, and the interaction term.

The research also re-centers the running variable at the cutoff, but this is not in the replication directions.

```
. quietly eststo: xi: reg male i.D*bac1

. quietly eststo: xi: reg white i.D*bac1

. quietly eststo: xi: reg aged i.D*bac1

. quietly eststo: xi: reg acc i.D*bac1

. esttab, title(Table 2 Panel A) compress

Table 2 Panel A
```

	(1) male	(2) white	(3) aged	(4) acc
_ID_1	0.0307***	0.00271	-7.787***	-0.219***
	(4.22)	(0.44)	(-38.12)	(-35.03)
bac1	0.218*	0.154	-56.36***	-1.540***
	(2.01)	(1.68)	(-18.55)	(-16.52)
_IDXbac1_1	-0.311**	0.0170	83.40***	2.656***
	(-2.83)	(0.18)	(26.99)	(28.03)
_cons	0.773***	0.835***	38.57***	0.201***
	(117.99)	(150.44)	(209.84)	(35.61)
N	214558	214558	214558	214558

```
t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

. eststo clear
```

The covariates are balanced at the cutoffs.

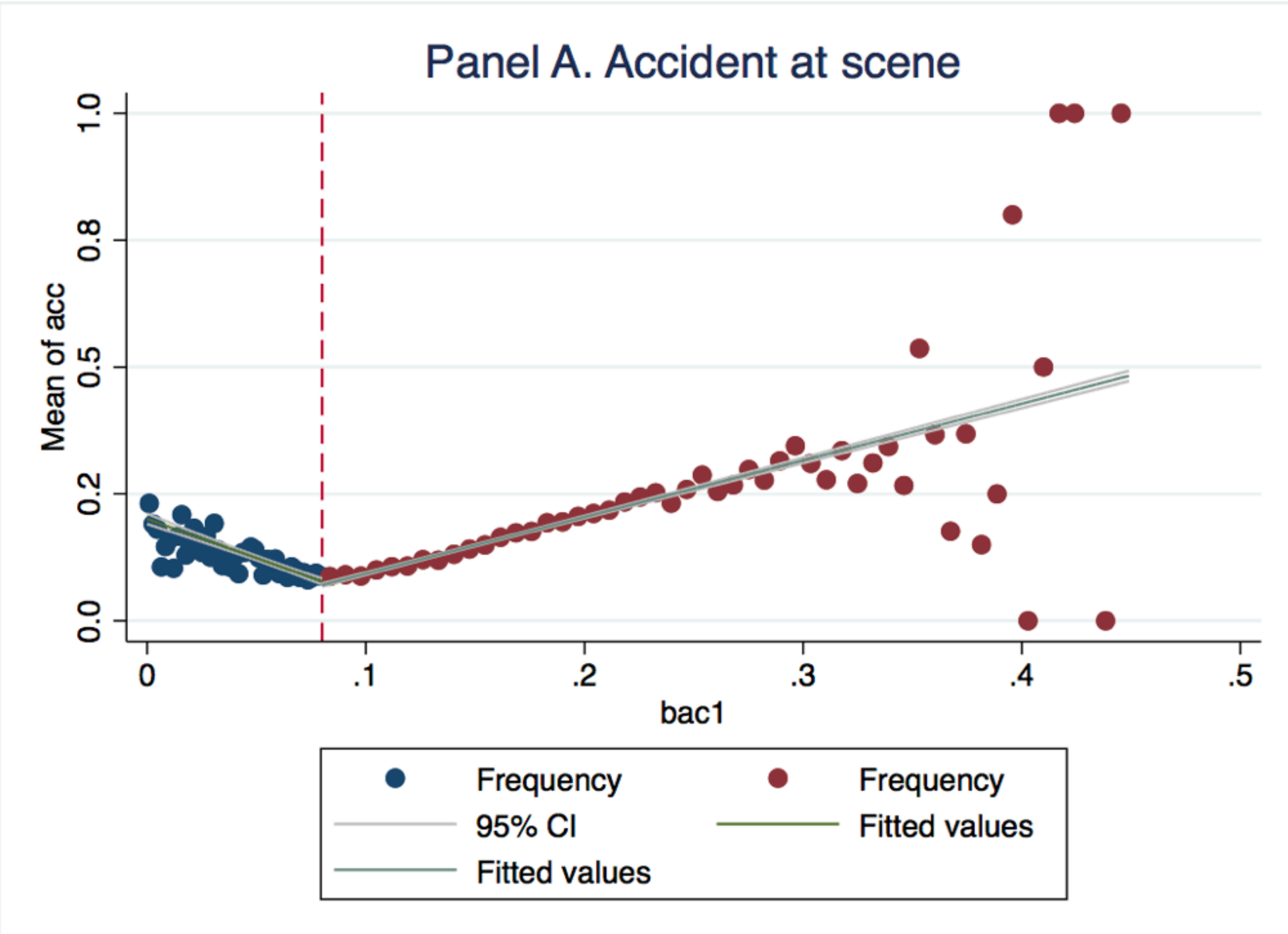
6) Balance Pictures

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.

Now that we checked if the covariates are balanced we also want a picture, the covariates should be balanced across the cutoff. Remember that the running variable is bac1 with a cutoff at 0.08 and the predetermined characteristics we used as the dependent variable for the balance test are: white, male, age, and accident.

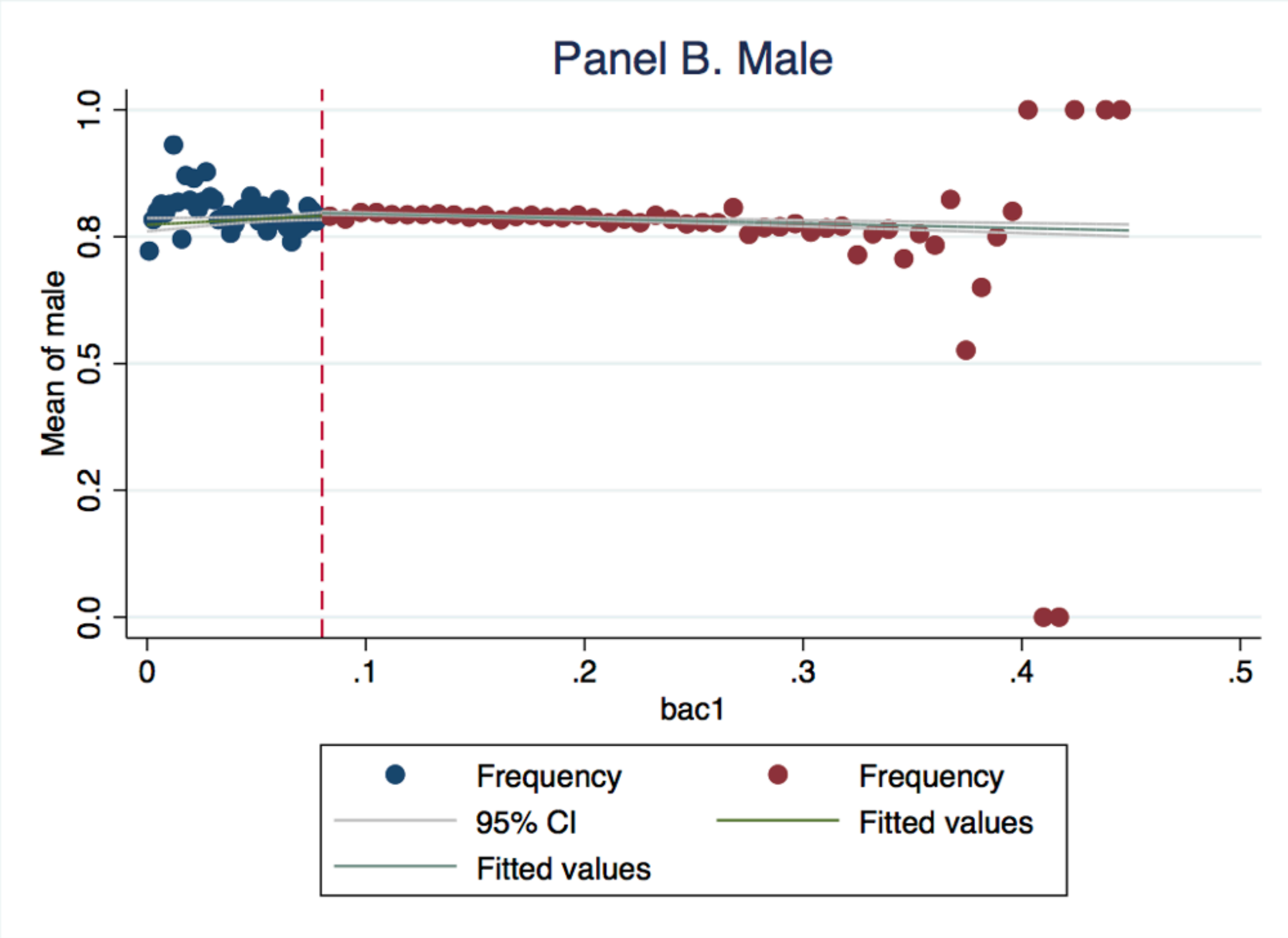
we do linear coefficients first

```
. quietly cmogram acc bac1, cut(0.08) title(Panel A. Accident
at scene) scatter lfitci legend lineat(.08)

. graph export Figure1.png , replace
(file /Users/santiagosilva/Downloads/Figure1.png written in PN
G format)
```



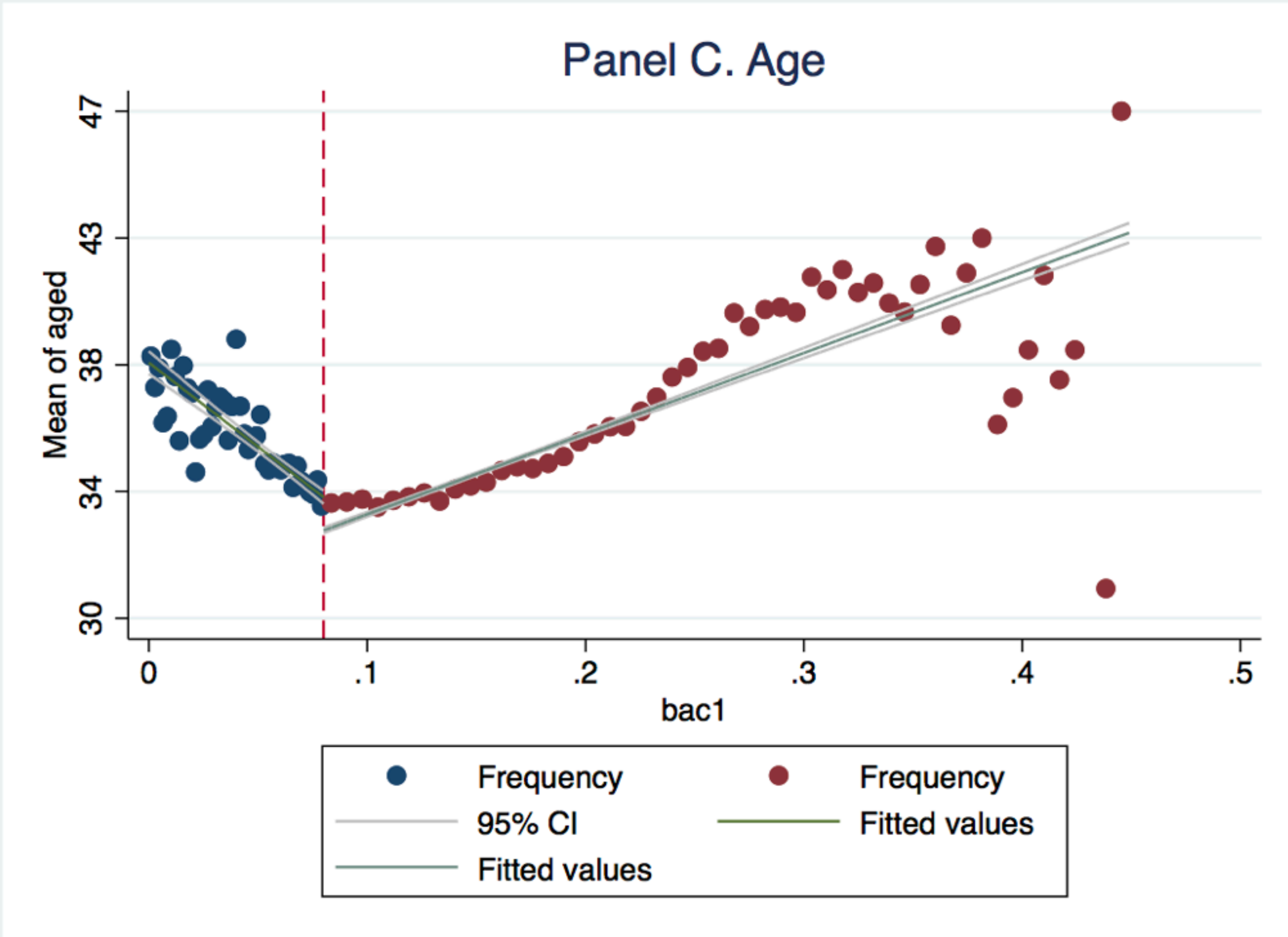
.

```
. quietly cmogram male bac1, cut(0.08) title(Panel B. Male) sc
atter lfitci legend lineat(.08)

. graph export Figure2.png , replace
(file /Users/santiagosilva/Downloads/Figure2.png written in PN
G format)
```



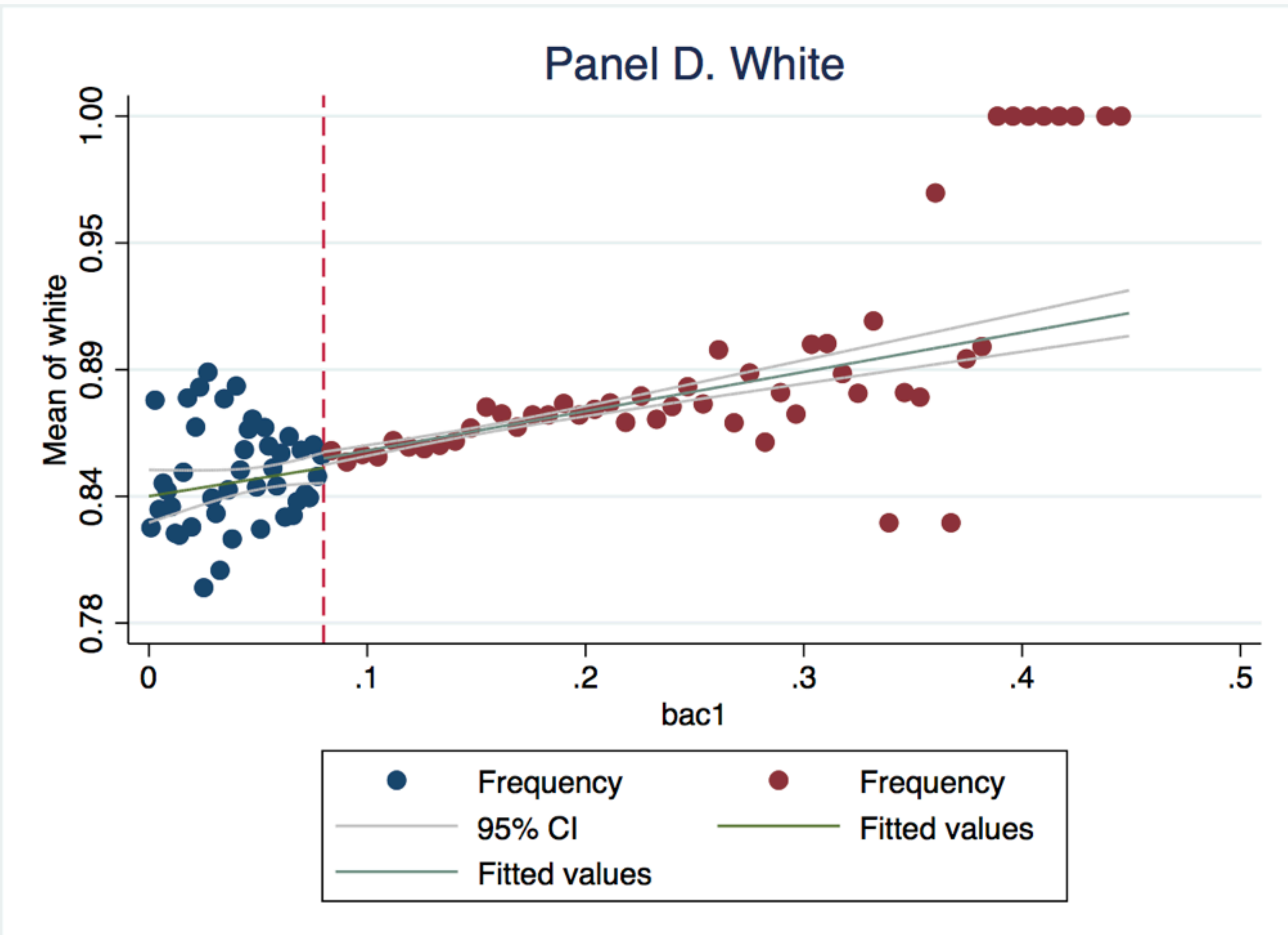
.

```
. quietly cmogram aged bac1, cut(0.08) title(Panel C. Age) sca
tter lfitci legend lineat(.08)

. graph export Figure3.png , replace
(file /Users/santiagosilva/Downloads/Figure3.png written in PN
G format)
```



.

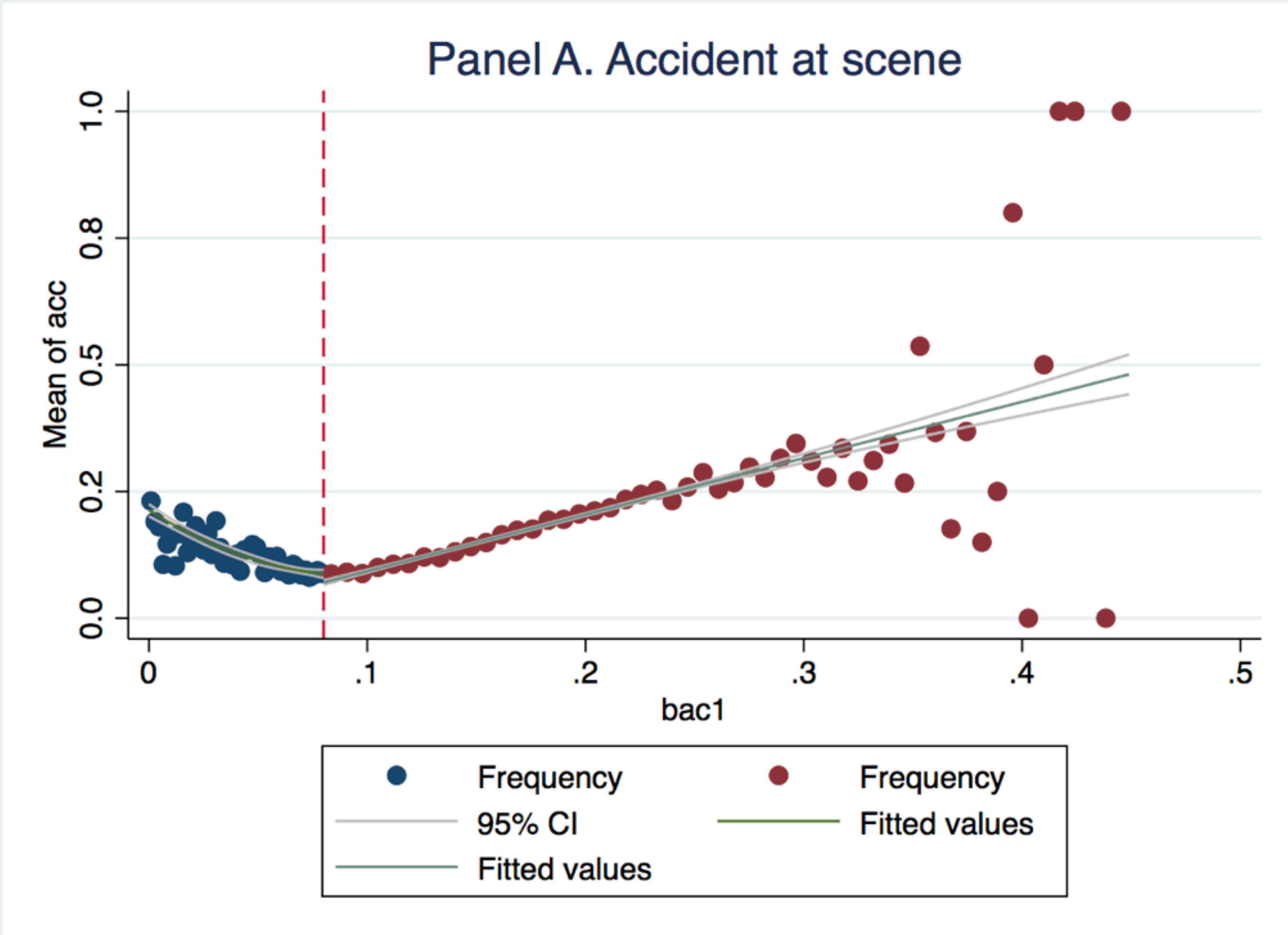
```
. quietly cmogram white bac1, cut(0.08) title(Panel D. White)
scatter lfitci legend lineat(.08)

. graph export Figure4.png , replace
(file /Users/santiagosilva/Downloads/Figure4.png written in PN
G format)
```



.

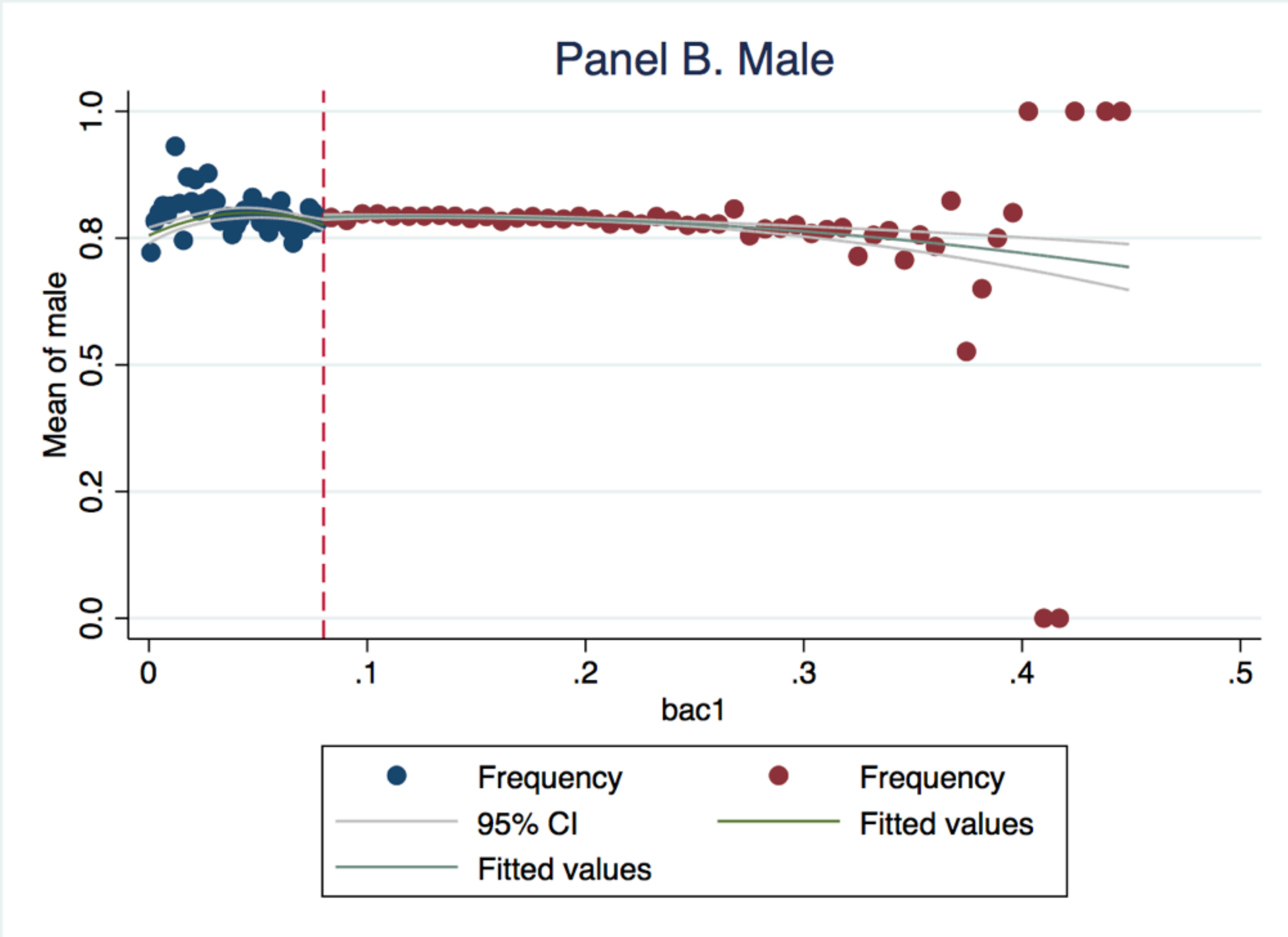
Now repeat for quadratic fit

```
. quietly cmogram acc bac1, cut(0.08) title(Panel A. Accident
at scene) scatter qfitci legend lineat(.08)

. graph export Figure5.png , replace
(file /Users/santiagosilva/Downloads/Figure5.png written in PN
G format)
```



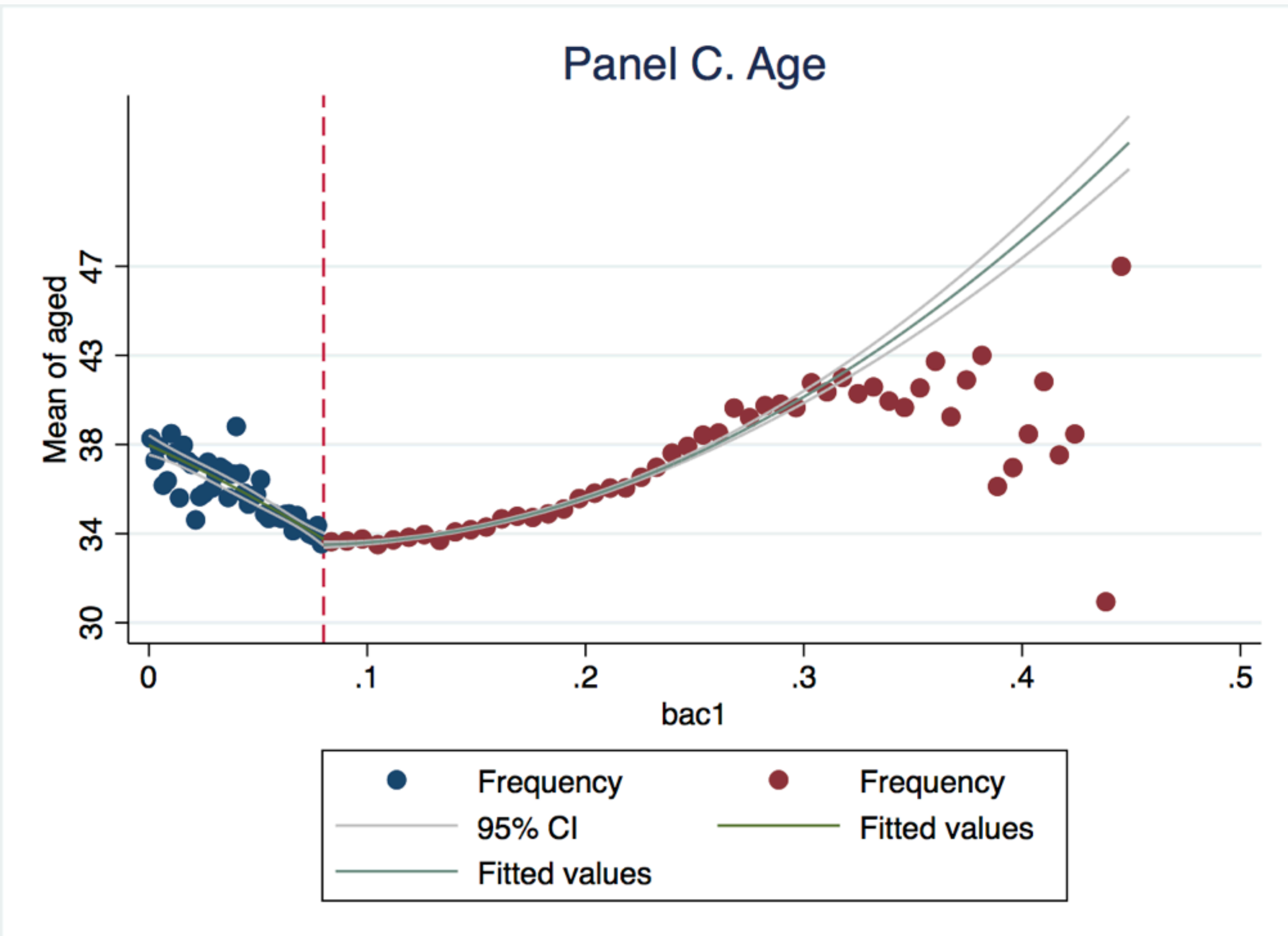
.

```
. quietly cmogram male bac1, cut(0.08) title(Panel B. Male) sc
atter qfitci legend lineat(.08)

.     graph export Figure6.png , replace
(file /Users/santiagosilva/Downloads/Figure6.png written in PN
G format)
```



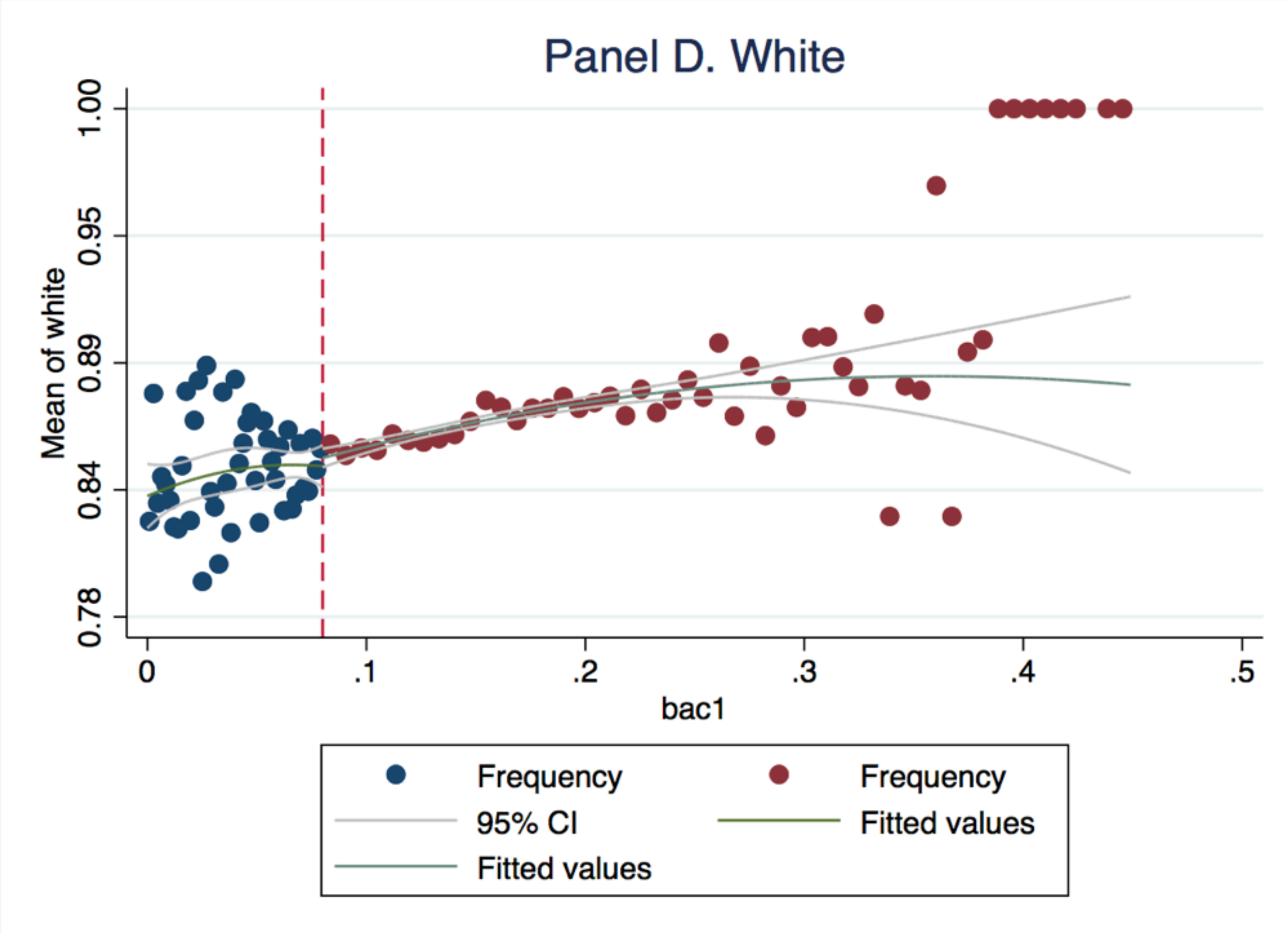
.

```
. quietly cmogram aged bac1, cut(0.08) title(Panel C. Age) sca
tter qfitci legend lineat(.08)

.     graph export Figure7.png , replace
(file /Users/santiagosilva/Downloads/Figure7.png written in PN
G format)
```



.

```
. quietly cmogram white bac1, cut(0.08) title(Panel D. White)
scatter qfitci legend lineat(.08)

.     graph export Figure8.png , replace
(file /Users/santiagosilva/Downloads/Figure8.png written in PN
G format)
```



.

The graphs are very similar to Hansen's, we do not have jumping where there should not be jumping, so we pass the balance test.

7)Table 3

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.:

The model is recidivism= beta_0 + b_1 D + b_2 bac1 + b_3 D*bac_1

Column 1 of OUR table will be a simpler version of the model, it will not have the interaction term, and will only control for the bac1 linearly by including a dummy variable for the treatment, D, this will be 0 if bac<0.08 and 0 otherwise

```
. eststo clear

. quietly eststo: reg recidivism D bac1 if bac1>0.055 & bac1<0
.105, robust

. quietly eststo: reg recidivism D bac1 if bac1>0.03 & bac1<0.
13, robust
```

Column 2 is the model Hansen uses. We run the regression on the interaction term, the tunning variable, and the interaction term. Reacall that to run a regression with an interaction term we can use the **xi** command

```
. quietly eststo: xi: reg recidivism i.D*bac1 if bac1>0.055 &
bac1<0.105, robust

. quietly eststo: xi: reg recidivism i.D*bac1 if bac1>0.03 & b
ac1<0.13, robust
```

Column 3 adds a quadratic interaction term to Hansen's model. We create the quadratic interaction term, we square the running variable, bac1, not the treatment variable obviously.The dependent variable, our outcome is still recidivism. Now the independent variables are: D , bac1, bac1_sq , interaction term D*bac1 , and second interaction term D*bac1_sq.

```
. gen bac1_sq = bac1^2

. quietly eststo: xi: reg recidivism D##c.(bac1 bac1_sq) if ba
c1>0.055 & bac1<0.105, robust

. quietly eststo: xi: reg recidivism D##c.(bac1 bac1_sq) if ba
c1>0.03 & bac1<0.13, robust

. esttab, title(Table 3 Column 1)

Table 3 Column 1
--------------------------------------------------------------
----------------------------------------------
                     (1)             (2)             (3)
        (4)             (5)             (6)
              recidivism      recidivism      recidivism
 recidivism      recidivism      recidivism
--------------------------------------------------------------
----------------------------------------------
D                -0.0218***      -0.0266***

                 (-3.89)         (-6.58)


bac1               0.216           0.331***        -0.150
    0.00588           5.116           2.939
                  (1.07)          (4.42)         (-0.39)
     (0.03)          (0.63)          (1.79)

_ID_1                                             -0.0623
    -0.0549***
                                                  (-1.78)
    (-3.61)

_IDXbac1_1                                          0.523
      0.392
                                                   (1.16)
```

```
     (1.92)

0.D
                          0                0

                        (.)              (.)

1.D
                      0.337            0.108

                     (0.80)           (1.28)

bac1_sq
                     -38.12           -24.61

                    (-0.65)          (-1.79)

0.D#c.bac1
                          0                0

                        (.)              (.)

1.D#c.bac1
                     -9.465           -4.083

                    (-0.89)          (-1.93)

0.D#c.bac1~q
                          0                0

                        (.)              (.)

1.D#c.bac1~q
                      63.42            31.87*

                     (0.92)           (2.11)

_cons               0.101***        0.0953***         0.127***
      0.117***     -0.0528           0.0338
                   (7.02)           (17.59)           (4.68)
     (9.38)        (-0.19)           (0.71)
----------------------------------------------------------------
----------------------------------------------
N                   46957            89967            46957
      89967         46957            89967
----------------------------------------------------------------
----------------------------------------------
t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001
```

Used robust standard errors like the directions asked.

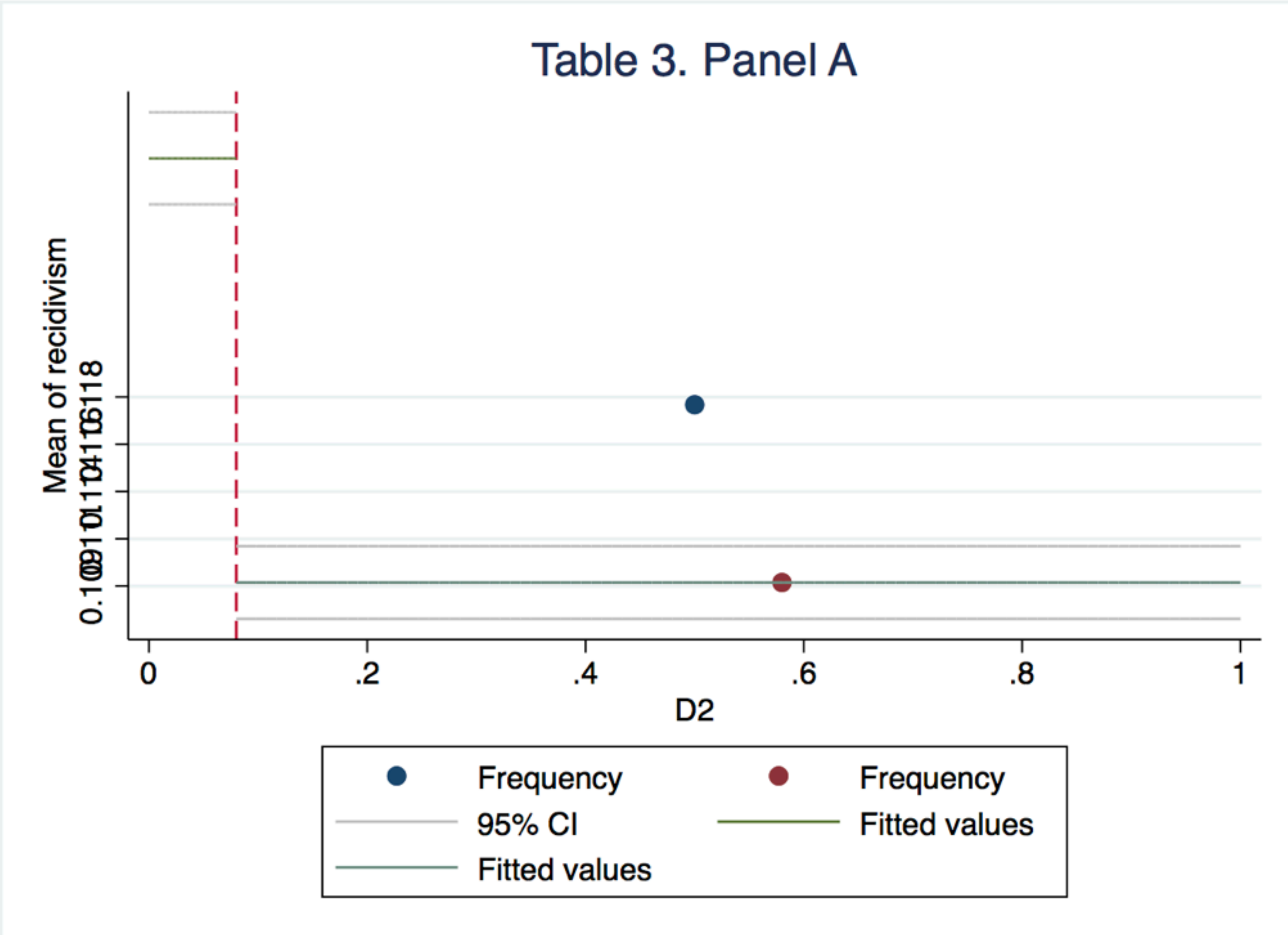
8)Figure 3, top panel

"Firue 3 plots means of recidivism rates based on simple regression models for all offedners and highlights the stark changes which occur at the 0.08 threshold"

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

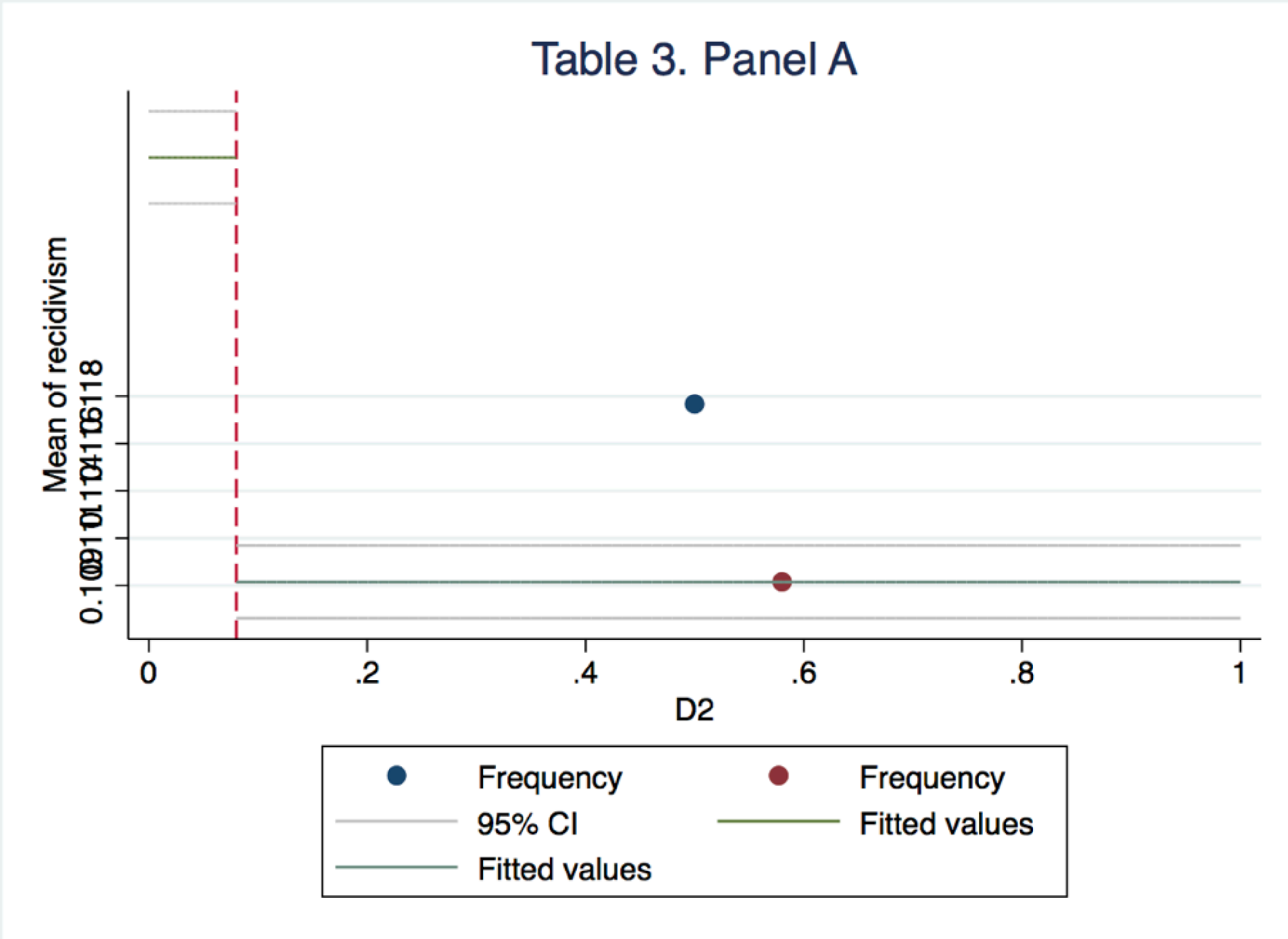
to do this we can create a new variable wich includes only the values for which bac1<0.15

```
. generate D2= 0

. replace D2 = 1 if bac1<0.15
(124,642 real changes made)

. quietly cmogram recidivism D2, cut(0.08) title(Table 3. Pane
l A) scatter lfitci legend lineat(.08)

.     graph export Figure9.png , replace
(file /Users/santiagosilva/Downloads/Figure9.png written in PN
G format)
```



. we do see a jump as expected, the causal effect.

```
. quietly cmogram recidivism D2, cut(0.08) title(Table 3. Pane
l A) scatter qfitci legend lineat(.08)

.     graph export Figure10.png , replace
(file /Users/santiagosilva/Downloads/Figure10.png written in P
NG format)
```



. we do see a jump as expected, the causal effect

9)Discussion

Hansen’s research tests the hipothesis whether or not a punishemnt for drunk drivers that show back up in the data within 4 months has an effect on drivers showing up back in the data. I learned how to code an RDD design, and how it is supposed to work. I am confident in Hansen’s because the did a proper RDD, which is therorically correct.