

# RDD Replication

Robert Toto

3/5/2021

## One

<https://github.com/rvt245/RDD>

## Two

### Summary of Hansen Paper

#### Research Question

Hansen is asking whether the harsher punishments received for higher blood-alcohol-content (BAC) cutoffs are effective in reducing future drunk driving. In Washington, there are two BAC thresholds for a DUI. If a driver has a BAC above 0.08, he or she receives a certain combination of fines, jail time, and license suspension. However, if the driver has a BAC above 0.15, he or she receives harsher penalties. Hansen wants to know whether the harsher penalties reduce future drunk driving more than the less harsh penalties at the lower BAC cutoff.

#### Data

Hansen begins with data on DUI stops from the state of Washington from 1995 to 2011, including a total of 512,964 DUI BAC tests. He then narrows this to only DUI stops between 1999 and 2007. In 1999, Washington state legislated a new dual-threshold system (0.08 BAC for DUI and 0.15 BAC for “aggravated” DUI). Furthermore, Hansen stops looking at new DUIs in 2007 because he wants to allow for four years (up to 2011) of additional DUI stops to find repeated offenses. This is crucial because Hansen is analyzing “recidivism” (or repeated offenses). In Hansen’s data, recidivism is identified if an offender receives a DUI within four years of his or her initial DUI. The data is also restricted to persons above the drinking age.

#### Research Design

This is a quasi-experiment because there is no direct randomization of the treatment assignment. More specifically, Hansen uses a Regression Discontinuity (RD) design because the BAC levels that legally determine a DUI provide cutoffs that lead to a natural treatment assignment. Since the DUI law generates a deterministic and discontinuous function of BAC, a Sharp RD can be used. Smoothness (or “continuity”) is assumed such that unobservables are expected to remain unchanged across the threshold. This assumption is predicated partly on the assumption that drunk drivers cannot discern their level of impairment prior to driving and thus will not sort themselves out of the running variable by not driving. Identification is established by assuming that it is locally random for a driver to have a BAC either just above or just below the BAC threshold. In other words, some people at the threshold are just lucky, and fall just below the

legal limit. These assumptions are reinforced by the fact that BAC measurements are precise and difficult to manipulate (making sorting at the cutoff very unlikely), yet slight variation between different breathalyzer devices may allow for randomness to occur in the measurement of BAC around the cutoff. The running variable is the minimum BAC measurement at a DUI stop, where two BAC measurements are taken per stop, because the minimum determines legal guilt. For the regression models in the RD design, Hansen uses a local linear regression to estimate the effect of having a BAC above the DUI or aggravated DUI threshold on recidivism. A rectangular kernel is used to weight the regression near the cutoff. The running variable (BAC) is recentered around the relevant threshold (0.08 or 0.15). The outcome variable is recidivism. Pretreatment characteristics related to DUI stops are stable. Bandwidths are varied from 0.05 to 0.025 BAC.

## Conclusions

Hansen generally finds that harsher penalties for BAC cutoffs do in fact reduce future drunk driving. When considering the 0.08 BAC threshold, Hansen finds that this threshold reduces recidivism in the following 4 years by 2 percentage points (and is statistically significant at 0.01 level). This finding is consistent across both 0.05 and 0.025 BAC bandwidths. Individuals with at least one prior BAC test reduce recidivism more than those with no prior tests. When considering both the 0.08 and 0.15 thresholds, Hansen finds a notable drop in recidivism for both, indicating that the increased punishments occurring at each increasing threshold does have an effect of reducing recidivism. Across offender with varying levels of prior tests, having a BAC over the legal cutoffs are associated with lower recidivism. At the 0.15 cutoff, for instance, drivers receiving an aggravated DUI reduce recidivism by an additional 1.1 percentage points. Those with at least one prior test reduce recidivism by an additional 2.2 percentage points, further indicating that harshness of punishment leads to less recidivism, as repeated offenses have harsher punishments. The regression is also carried out on disaggregated definitions of recidivism using a rectangular kernel. In these cases, having a BAC over a legal limit continues to be associated with lowered recidivism, except in the case of refusal to breathalyze. This disaggregation also reveals that having a BAC over either of the thresholds is associated with decreased future automobile accidents. Further analysis is conducted on the effect of exceeding a threshold on the punishments, sanctions, and treatments received in court. Hansen finds that surpassing a threshold is associated with harsher punishments and sanctions. Overall, Hansen finds that exceeding a threshold is associated with reduced drunk driving (recidivism) in the short run and the long run.

## Three

```
hansen <- hansen %>%
  mutate(dui = ifelse(bac1 >= 0.08 & bac1 < 0.15, 1, 0))
head(hansen)
```

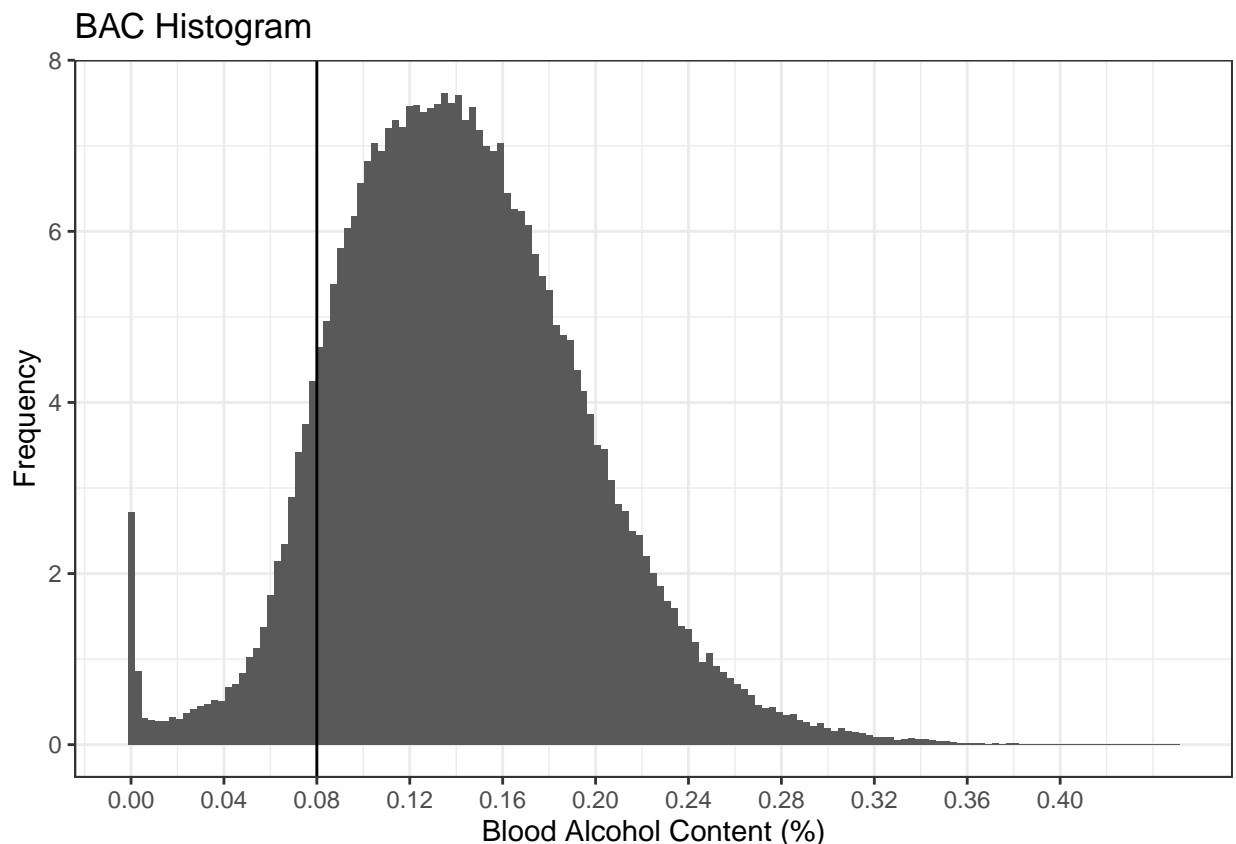
```
##      Date Alcohol1 Alcohol2 low_score male white recidivism acc aged year
## 1 30jul2007      0      0      0      1      1          1    0   48 2007
## 2 20feb2007      0      0      0      1      1          0    0   51 2007
## 3 18mar2003      0      0      0      1      0          0    0   68 2003
## 4 17dec2006      0      0      0      1      1          0    0   51 2006
## 5 07apr1999      0      0      0      1      1          0    0   22 1999
## 6 20mar2003      0      0      0      1      1          0    1   49 2003
##   bac1 bac2 dui
## 1    0    0    0
## 2    0    0    0
## 3    0    0    0
## 4    0    0    0
```

```
## 5    0    0    0
## 6    0    0    0
```

## Four

If people could manipulate their BAC, we would want to use a density test to see whether there is bunching or “sorting” at the 0.08 BAC cutoff. The McCrary Density Test is a popular test to identify such manipulation. The McCrary Density Test gives a null hypothesis that the density in the data is continuous at the cutoff. The alternative hypothesis is that there is greater density (i.e. bunching) around the cutoff. The test separated the data into bins and calculates the frequency of observations in each bin. To reject the null, one would have to see a greater frequency of observations in the bins next to the cutoff, compared to bins further away. In R, we can use “rddensity” to run this test. We can also inspect the histogram of BAC data to look for visual bunching in the running variable near the cutoff.

```
#BAC Histogram
ggplot(data=hansen) +
  geom_histogram(aes(x=bac1, after_stat(density)), binwidth = .003) +
  labs(x = 'Blood Alcohol Content (%)',
       y = 'Frequency',
       title = 'BAC Histogram') +
  theme_bw() +
  scale_x_continuous(breaks=seq(0,0.4,0.04)) +
  geom_vline(xintercept = 0.08)
```



The histogram of BAC observations shows no visual sign of bunching at the 0.08 BAC cutoff for a DUI.

The distribution illustrates a smooth increase in bin frequencies as BAC increases with no noticeable discontinuous jump in the frequency of BAC observations near the cutoff.

```
library(rdd)
library(rddensity)

#Density Test
density <- rddensity(hansen$bac1, c = 0.08, massPoints=FALSE)
#rdplotdensity(density, hansen$bac1)
summary(density)
```

```
##
## Manipulation testing using local polynomial density estimation.
##
## Number of obs =      214558
## Model =             unrestricted
## Kernel =            triangular
## BW method =         estimated
## VCE method =        jackknife
##
## c = 0.08             Left of c           Right of c
## Number of obs       23010                191548
## Eff. Number of obs  14442                28946
## Order est. (p)      2                    2
## Order bias (q)      3                    3
## BW est. (h)         0.023                0.023
##
## Method              T                    P > |T|
## Robust              -0.2299              0.8182
##
##
## P-values of binomial tests (H0: p=0.5).
##
## 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
## 0.000                909     0    0.0000
```

The density test confirms that there is no manipulation of BAC at the cutoff. The test gives a p-value of 0.8182, indicating we should not reject the null (that the density in the data is continuous at the cutoff). While there may be some heaping due to rounding and the choice of bin-widths, the point estimates model of BAC frequency moves smoothly across the cutoff with no discontinuous jump. Overall, this is strong evidence that drivers did not manipulate BAC results. This also aligns with the real-world understanding that breathalyzers are difficult to cheat. This conclusion is also consistent with Hansen’s McCrary density test results which showed “little evidence of endogenous sorting to one side of either of the thresholds studied.” Although, the p-value for the 0.08 cutoff in Hansen’s results was lower at 0.59.

## Five

```
#Check if covariates are balanced at the cutoff
library(stargazer)
library(knitr)
library(kableExtra)
library(data.table)
library(jtools)
library(gridExtra)
library(grid)
library(RDHonest)
library(rdd)

#Recenter the running variable first
hansen_c <- hansen %>%
  mutate(bac1_c = bac1 - 0.08)

#Isolate data to 0.05 Bandwidth
hansen1 <- hansen_c %>%
  filter(bac1_c >=-0.05 & bac1_c <= 0.05)

#Check if covariates are balanced at the cutoff using Regression Discontinuity
#Male
rdmale <- RDestimate(male ~ bac1_c, cutpoint = 0.0, bw = 0.05,
  kernel="rectangular", model = T,
  data=hansen1)
a <- rdmale$est[1]

Hmale <- RDHonest(male ~ bac1_c, cutoff = 0.0, h=0.05,
  kern="uniform", M=0.1, sclass="T", order = 1,
  data=hansen1)
s1 <- Hmale$sd[1]

#White
rdwhite <- RDestimate(white ~ bac1_c, cutpoint = 0.0, bw = 0.05,
  kernel="rectangular", model = T,
  data=hansen1)
b <- rdwhite$est[1]

Hwhite <- RDHonest(white ~ bac1_c, cutoff = 0.0, h=0.05,
  kern="uniform", M=0.1, sclass="T", order = 1,
  data=hansen1)
s2 <- Hwhite$sd[1]

#Age
rdaged <- RDestimate(aged ~ bac1_c, cutpoint = 0.0, bw = 0.05,
  kernel="rectangular", model = T,
  data=hansen1)
c <- rdaged$est[1]

Haged <- RDHonest(aged ~ bac1_c, cutoff = 0.0, h=0.05,
  kern="uniform", M=0.1, sclass="T", order = 1,
  data=hansen1)
s3 <- Haged$sd[1]
```

```

#Accident
rdacc <- RDestimate(acc ~ bac1_c, cutpoint = 0.0, bw = 0.05,
                    kernel="rectangular", model = T,
                    data=hansen1)
d <- rdacc$est[1]

Hacc <- RDHonest(acc ~ bac1_c, cutoff = 0.0, h=0.05,
                 kern="uniform", M=0.1, sclass="T", order = 1,
                 data=hansen1)
s4 <- Hacc$sd[1]

#Build a table of the RDestimate outputs (and RDHonest standard errors)
df1 <- data.frame(Male = a)
df1[2] <- data.frame(White = b)
df1[3] <- data.frame(Age = c)
df1[4] <- data.frame(Accident = d)
df1 <- setattr(df1, "row.names", c("DUI", "Honest SE"))
df1[2,] <- c(s1,s2,s3,s4)
stargazer(df1, type = "text", summary=FALSE,
           title = "Regression Discontinuity for Predetermined Characteristics")

```

```

##
## Regression Discontinuity for Predetermined Characteristics
## =====
##           Male   White   Age   Accident
## -----
## DUI           0.006 0.006 -0.140  -0.003
## Honest SE     0.006 0.005 0.164   0.004
## -----

```

```

#kable(df1, digits = 4)

```

The BAC was recentered around 0.08 and limited to a bandwidth of 0.05. Regression discontinuity estimates were calculated using a rectangular kernel and the “RDestimate” package in R. These estimates give the local average treatment effect (LATE) of the pretreatment characteristics (i.e. male, white, age, accident) at the BAC cutoff of 0.08. “Honest” standard errors were estimated using the “RDHonest” package. These regression discontinuity estimates in the table show the effect of exceeding the 0.08 BAC threshold on pretreatment characteristics (i.e. male, white, age, accident). These regressions fail to reject the null that the pretreatment characteristics are unrelated to the 0.08 BAC cutoff for a DUI. This indicates that these demographic factors are stable (“balanced”) across the BAC cutoff. These results mirror the findings in Table 2, Panel A of Hansen 2015. However, the regression continuity coefficients values are slightly different from those in Hansen because our analysis only uses only the 0.08 BAC cutoff for a DUI, whereas Hansen uses both cutoffs (0.08 and 0.15). The stability of these demographic factors lends credibility to the regression discontinuity model providing unbiased estimates on recidivism.

Note: I use the “stargazer” package to summarize the LATE estimates of the pretreatment characteristics on DUI, and the respective “honest” standard errors. I first send each coefficient and standard error value to a data frame, which I then send to stargazer for summary.

For fun, I have also created a simple linear model for the predetermined characteristics as well. The results happen to be very similar to the RDD results, and the “lm” model affords a better table presentation. I also show the coefficients visually using the “plot\_summs” function in the “jtools” package to graphically compare the coefficients of BAC, DUI, and the interaction of both by their value and standard deviation. This part is not required by the assignment and is only included for the purposes of practicing with visualizations.

```
library(stargazer)
library(knitr)
library(kableExtra)
library(data.table)
library(jtools)
library(gridExtra)
library(grid)

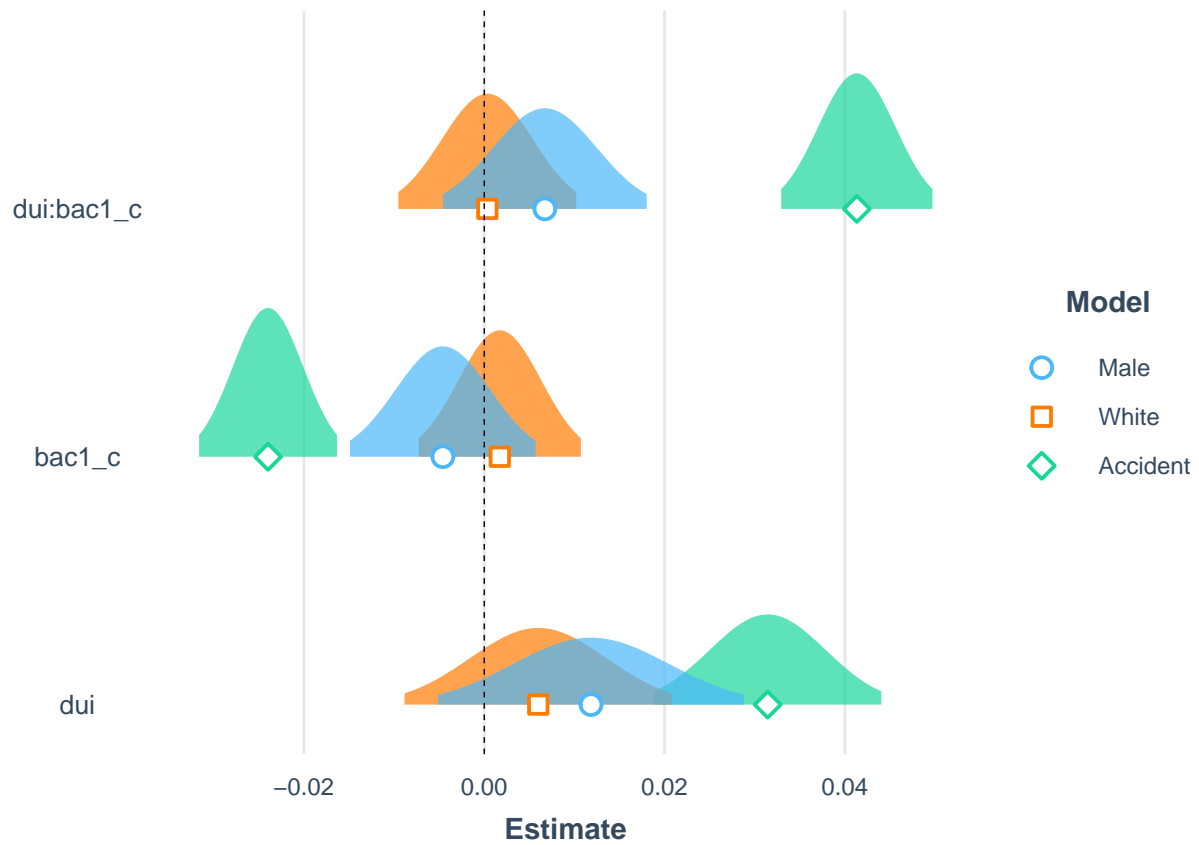
#For fun, also check results using simple linear models
m1 = lm(male ~ dui*bac1_c, data=hansen1)
m2 = lm(white ~ dui*bac1_c, data=hansen1)
m3 = lm(acc ~ dui*bac1_c, data=hansen1)
m4 = lm(aged ~ dui*bac1_c, data=hansen1)

#Table of LM Results
library(jtools)
export_summs(m1, m2, m3, m4, results='asis', ci_level=0.95,
             number_format = "%.3f",
             coefs = c("dui", "bac1_c", "dui:bac1_c"),
             model.names = c("Male", "White", "Accident", "Age"))
```

	Male	White	Accident	Age
dui	0.006 (0.006)	0.006 (0.005)	-0.003 (0.004)	-0.140 (0.162)
bac1_c	-0.210 (0.240)	0.079 (0.210)	-1.096 *** (0.178)	-69.164 *** (6.844)
dui:bac1_c	0.307 (0.263)	0.016 (0.230)	1.888 *** (0.196)	76.049 *** (7.508)
N	89967	89967	89967	89967
R2	0.000	0.000	0.002	0.002

\*\*\* p < 0.001; \*\* p < 0.01; \* p < 0.05.

```
#Plot of LM Results
plot_summs(m1, m2, m3, scale = TRUE, ci_level = 0.95,
           plot.distributions = TRUE,
           model.names = c("Male", "White", "Accident"))
```



```
#plot_summs(m1, m2, m3, scale = TRUE, ci_level = 0.95, inner_ci_level = 0.9,
            #model.names = c("Male", "White", "Accident"))
#plot_coefs(m4, scale = TRUE, ci_level = 0.95, inner_ci_level = 0.9)
```



## Six

```
#Prepare the data for discontinuity modeling of pretreatment characteristics
categories <- hansen$bac1

hansen2 <- hansen %>%
  filter(bac1 >= 0 & bac1 <= 0.2)

means_male <- split(hansen2$male, cut(hansen2$bac1, 96)) %>%
  lapply(mean) %>%
  unlist()
agg_hansen_male <- data.frame(male = means_male, bac1 = seq(0.01, 0.2, by = 0.002))

means_white <- split(hansen2$white, cut(hansen2$bac1, 96)) %>%
  lapply(mean) %>%
  unlist()
agg_hansen_white <- data.frame(white = means_white, bac1 = seq(0.01, 0.2, by = 0.002))

means_acc <- split(hansen2$acc, cut(hansen2$bac1, 96)) %>%
  lapply(mean) %>%
  unlist()
agg_hansen_acc <- data.frame(acc = means_acc, bac1 = seq(0.01, 0.2, by = 0.002))

means_aged <- split(hansen2$aged, cut(hansen2$bac1, 96)) %>%
  lapply(mean) %>%
  unlist()
agg_hansen_aged <- data.frame(aged = means_aged, bac1 = seq(0.01, 0.2, by = 0.002))

hansen2 <- hansen2 %>%
  mutate(gg_group = case_when(bac1 > 0.08 ~ 1, TRUE ~ 0))
```

```

#Model "male" Stability Using Linear Model and Quadratic models
lmale <- ggplot(hansen2, aes(bac1, male)) +
  geom_point(aes(x = bac1, y = male), data = agg_hansen_male) +
  geom_vline(xintercept = 0.08, colour = "grey", linetype = 2) +
  stat_smooth(aes(bac1, male, group = gg_group), method = "lm") +
  labs(x = "BAC")

qmale <- ggplot(hansen2, aes(bac1, male)) +
  geom_point(aes(x = bac1, y = male), data = agg_hansen_male) +
  geom_vline(xintercept = 0.08, colour = "grey", linetype = 2) +
  stat_smooth(aes(bac1, male, group = gg_group), method = "lm", formula = y ~ x + I(x^2)) +
  labs(x = "BAC")

#Model "white" Stability Using Linear Model and Quadratic models
lwhite <- ggplot(hansen2, aes(bac1, white)) +
  geom_point(aes(x = bac1, y = white), data = agg_hansen_white) +
  geom_vline(xintercept = 0.08, colour = "grey", linetype = 2) +
  stat_smooth(aes(bac1, white, group = gg_group), method = "lm") +
  labs(x = "BAC")

qwhite <- ggplot(hansen2, aes(bac1, white)) +
  geom_point(aes(x = bac1, y = white), data = agg_hansen_white) +
  geom_vline(xintercept = 0.08, colour = "grey", linetype = 2) +
  stat_smooth(aes(bac1, white, group = gg_group), method = "lm", formula = y ~ x + I(x^2)) +
  labs(x = "BAC")

#Model "accident" Stability Using Linear Model and Quadratic models
lacc <- ggplot(hansen2, aes(bac1, acc)) +
  geom_point(aes(x = bac1, y = acc), data = agg_hansen_acc) +
  geom_vline(xintercept = 0.08, colour = "grey", linetype = 2) +
  stat_smooth(aes(bac1, acc, group = gg_group), method = "lm") +
  labs(x = "BAC")

qacc <- ggplot(hansen2, aes(bac1, acc)) +
  geom_point(aes(x = bac1, y = acc), data = agg_hansen_acc) +
  geom_vline(xintercept = 0.08, colour = "grey", linetype = 2) +
  stat_smooth(aes(bac1, acc, group = gg_group), method = "lm", formula = y ~ x + I(x^2)) +
  labs(x = "BAC")

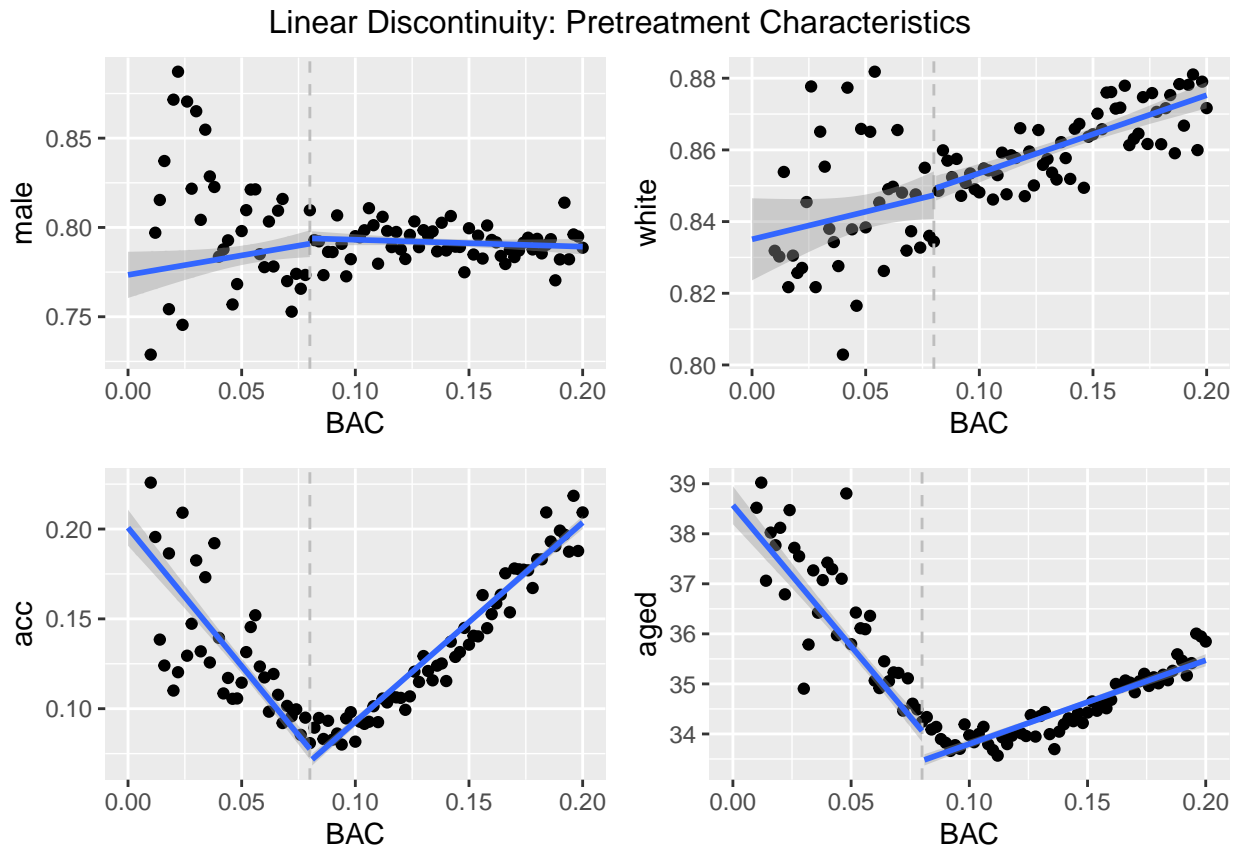
#Model "age" Stability Using Linear Model and Quadratic models
lage <- ggplot(hansen2, aes(bac1, aged)) +
  geom_point(aes(x = bac1, y = aged), data = agg_hansen_aged) +
  geom_vline(xintercept = 0.08, colour = "grey", linetype = 2) +
  stat_smooth(aes(bac1, aged, group = gg_group), method = "lm") +
  labs(x = "BAC")

qage <- ggplot(hansen2, aes(bac1, aged)) +
  geom_point(aes(x = bac1, y = aged), data = agg_hansen_aged) +
  geom_vline(xintercept = 0.08, colour = "grey", linetype = 2) +
  stat_smooth(aes(bac1, aged, group = gg_group), method = "lm", formula = y ~ x + I(x^2)) +
  labs(x = "BAC")

library(gridExtra)

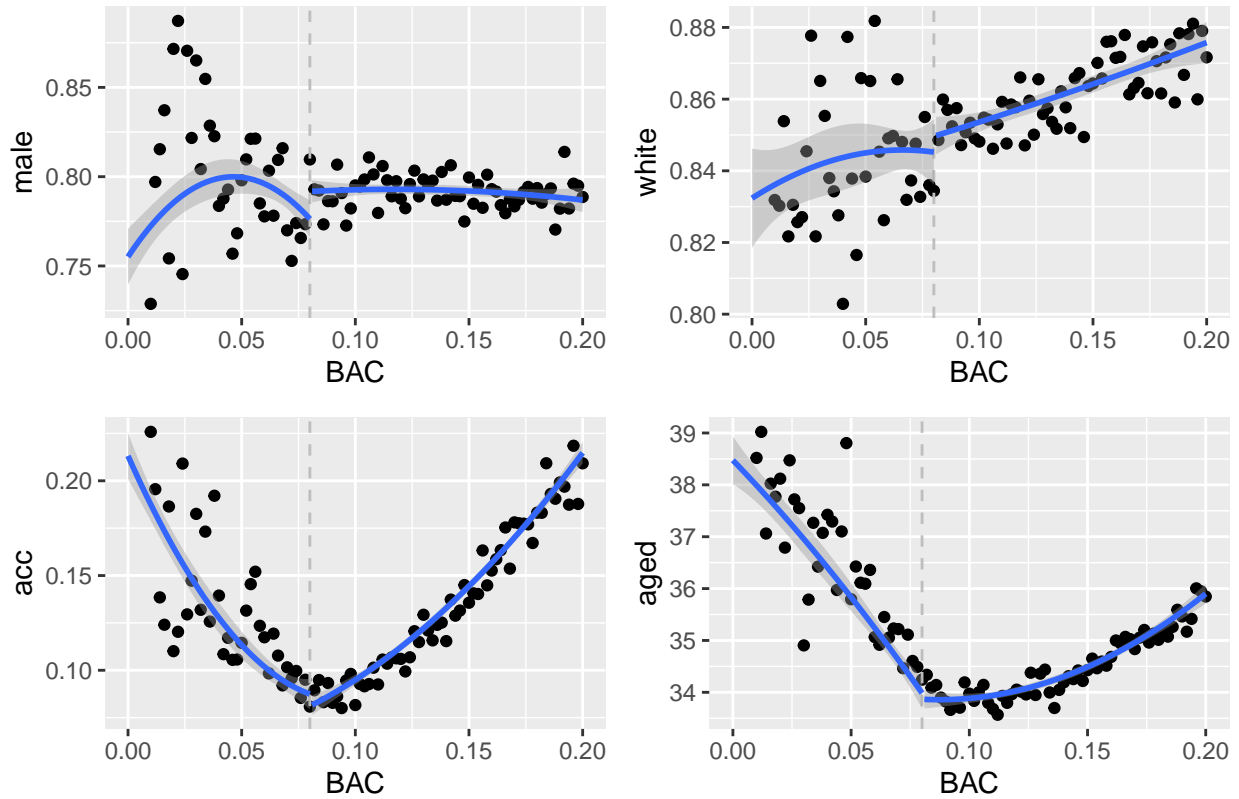
```

```
grid.arrange(lmale, lwhite, lacc, lage, nrow=2, top="Linear Discontinuity: Pretreatment Characteristics")
```



```
grid.arrange(qmale, qwhite, qacc, qage, nrow=2, top="Quadratic Discontinuity: Pretreatment Characteristics")
```

### Quadratic Discontinuity: Pretreatment Characteristics



There is not an observable discontinuous change in the average values of the four pretreatment characteristic covariates (male, white, accidents, and age) around the 0.08 BAC cutoff. The use of linear and quadratic models helps show this. While there appears to be a small discontinuity for the male covariate at the 0.08 cutoff under the quadratic model, this gap disappears under the linear model. Similarly, the small gap in the aged variable at the cutoff using the linear model disappears when a quadratic model is applied. Generally, regardless of linear or quadratic models, any gaps at the cutoff across these four covariates is extremely small, indicating that they remain unchanged across the threshold. This provides indirect evidence to support the continuity assumption of the regression discontinuity design on recidivism. In other words, these covariates appear to be *invariant* to changes in the treatment assignment. Because these predetermined characteristics appear to be *stable* across the BAC threshold, we are more confident in the regression discontinuity design's ability to provide unbiased estimates on changes in recidivism caused by the cutoff, as these predetermined characteristics do not appear to be interacting with the cutoff, implying that the recidivism effect is not biased by covariates when data points cross the cutoff.

Just as in Hansen, I used a bin-width of 0.002 to summarize the data-points into point-averages in the chart space. The linear models results in this analysis are nearly identical to the Hansen's results in Figure 2 for these four predetermined characteristics (i.e. male, white, age, accident) at the 0.08 BAC cutoff. The linear fits for white, age, and accident are exact replications, but there appears to be a difference in the male figure. While the distribution of bins is visually the same, the pre-cutoff linear model in Hansen is decreasing, while mine is increasing. This may be due to differences in the bin-averaging process between R and Stata. The process is much more hands-on in R, whereas cmogram does everything automatically in Stata.

Note: I was able to create these in Stat using cmogram as well. I used R instead because I felt it was a challenge. While it took significantly more work, it was great fun to work through the process of tailoring the code in R to produce these results, which look great!

## Seven

```
library(RDHonest)
library(jtools)
library(stargazer)

##--BINWIDTH 0.05--##
#Col1: Binwidth 0.05: control for bac1 linearly
rd_recid_05_1 <- RDestimate(recidivism ~ bac1, cutpoint = 0.08, bw = 0.05,
                           kernel="rectangular", model = T,
                           data=hansen)
r1_05 <- rd_recid_05_1$est[1]

H_recid_05_1 <- RDHonest(recidivism ~ bac1, cutoff = 0.08, h=0.05,
                        kern="uniform", M=0.1, sclass="T", order = 1,
                        data=hansen)
H1_05 <- H_recid_05_1$sd[1]

#Col2: Binwidth 0.05: interact bac1 with cutoff linearly
rd_recid_05_2 <- RDestimate(recidivism ~ bac1 | bac1*dui, cutpoint = 0.08, bw = 0.05,
                           kernel="rectangular", model = T,
                           data=hansen)
r2_05 <- rd_recid_05_2$est[1]

H_recid_05_2 <- RDHonest(recidivism ~ bac1 + bac1*dui, cutoff = 0.08, h=0.05,
                        kern="uniform", M=0.1, sclass="T", order = 1,
                        data=hansen)
H2_05 <- H_recid_05_2$sd[1]

#Col3: Binwidth 0.05: interact bac1 with cutoff linearly and as a quadratic
rd_recid_05_3 <- RDestimate(recidivism ~ bac1 | bac1*dui + (bac1^2)*dui, cutpoint = 0.08,
                           bw = 0.05, kernel="rectangular", model = T,
                           data=hansen)
r3_05 <- rd_recid_05_3$est[1]

H_recid_05_3 <- RDHonest(recidivism ~ bac1 + bac1*dui + (bac1^2)*dui, cutoff = 0.08,
                        h=0.05, kern="uniform", M=0.1, sclass="T", order = 1,
                        data=hansen)
H3_05 <- H_recid_05_3$sd[1]

#Prepare Summary Table for Panel A (0.05 Bandwidth_
df_recid <- data.frame(BAC = r1_05)
df_recid[2]<- data.frame(Linear_Interact. = r2_05)
df_recid[3]<- data.frame(Linear_Quadratic_Interact. = r3_05)
df_recid <- setattr(df_recid, "row.names", c("DUI", "Honest_SE"))
df_recid[2,] <- c(H1_05,H2_05,H3_05)

##--BINWIDTH 0.025--##
#Col1: Binwidth 0.025: control for bac1 linearly
rd_recid_025_1 <- RDestimate(recidivism ~ bac1, cutpoint = 0.08, bw = 0.025,
                           kernel="rectangular", model = T,
```

```

                                data=hansen)
r1_025 <- rd_recid_025_1$est[1]

H_recid_025_1 <- RDHonest(recidivism ~ bac1, cutoff = 0.08, h=0.025,
                          kern="uniform", M=0.1, sclass="T", order = 1,
                          data=hansen)
H1_025 <- H_recid_025_1$sd[1]

#Col2: Binwidth 0.025: interact bac1 with cutoff linearly
rd_recid_025_2 <- RDestimate(recidivism ~ bac1 | bac1*dui, cutpoint = 0.08,
                             bw = 0.025, kernel="rectangular", model = T,
                             data=hansen)
r2_025 <- rd_recid_025_2$est[1]

H_recid_025_2 <- RDHonest(recidivism ~ bac1 + bac1*dui, cutoff = 0.08, h=0.025,
                          kern="uniform", M=0.1, sclass="T", order = 1,
                          data=hansen)
H2_025 <- H_recid_025_2$sd[1]

#Col3: Binwidth 0.025: interact bac1 with cutoff linearly and as a quadratic
rd_recid_025_3 <- RDestimate(recidivism ~ bac1 | bac1*dui + (bac1^2)*dui,
                             cutpoint = 0.08, bw = 0.025, kernel="rectangular",
                             model = T, data=hansen)
r3_025 <- rd_recid_025_3$est[1]

H_recid_025_3 <- RDHonest(recidivism ~ bac1 + bac1*dui + (bac1^2)*dui,
                          cutoff = 0.08, h=0.025, kern="uniform", M=0.1,
                          sclass="T", order = 1,
                          data=hansen)
H3_025 <- H_recid_025_3$sd[1]

#Prepare Summary Table for Panel B (0.025 Bandwidth)
df_recid2 <- data.frame(BAC = r1_025)
df_recid2[2]<- data.frame(Linear_Interact. = r2_025)
df_recid2[3]<- data.frame(Linear_Quadratic_Interact. = r3_025)
df_recid2 <- setattr(df_recid2, "row.names", c("DUI","Honest_SE"))
df_recid2[2,] <- c(H1_025,H2_025,H3_025)

```

```
#Panel A Table
stargazer(df_recid, type = "text", summary=FALSE,
          title = "Panel A: RDD Recidivism at 0.08 BAC (0.03 to 0.13)")
```

```
##
## Panel A: RDD Recidivism at 0.08 BAC (0.03 to 0.13)
## =====
##          BAC   Linear_Interact. Linear_Quadratic_Interact.
## -----
## DUI          -0.024        -0.024                -0.024
## Honest_SE 0.004          0.004                0.004
## -----
```

In Panel A, with the wider binwidth, the results are very similar to Table 3, Panel A in Hansen. My results show that having a BAC above the 0.08 BAC cutoff (getting a DUI) decreases recidivism by 2.4 percentage points, and Hansen finds a decrease of 2.1 percentage points. Furthermore, my results are significant, with “honest” standard errors of just 0.004, which is exactly the same as Hansen’s standard error when controlling just for BAC. When I add the linear and quadratic interactions (columns 2 and 3), the causal effects are the same as when only controlling for BAC. This implies that the regression coefficient on the interaction terms is essentially zero in both cases, meaning these interaction terms do not contain predictive information for recidivism discontinuity.

```
#Panel B Table
stargazer(df_recid2, type = "text", summary=FALSE,
          title = "Panel B: RDD Recidivism at 0.08 BAC (0.055 to 0.105)")
```

```
##
## Panel B: RDD Recidivism at 0.08 BAC (0.055 to 0.105)
## =====
##           BAC   Linear_Interact. Linear_Quadratic_Interact.
## -----
## DUI         -0.020       -0.020             -0.020
## Honest_SE 0.006         0.006             0.006
## -----
```

In Panel B, with the tighter binwidth, the results are again very similar to Table 3, Panel B in Hansen. My results show that having a BAC above the 0.08 BAC cutoff (getting a DUI) decreases recidivism by 2.0 percentage points, and Hansen finds a decrease of 1.9 percentage points. Furthermore, my results are significant, with “honest” standard errors of just 0.006, which is very close to Hansen’s (0.005) when controlling for BAC. Again, When I add the linear and quadratic interactions (columns 2 and 3), the causal effects are the same as when only controlling for BAC. This implies that the regression coefficient on the interaction terms is essentially zero in both cases, meaning these interaction terms do not contain predictive information for recidivism discontinuity.



## Eight

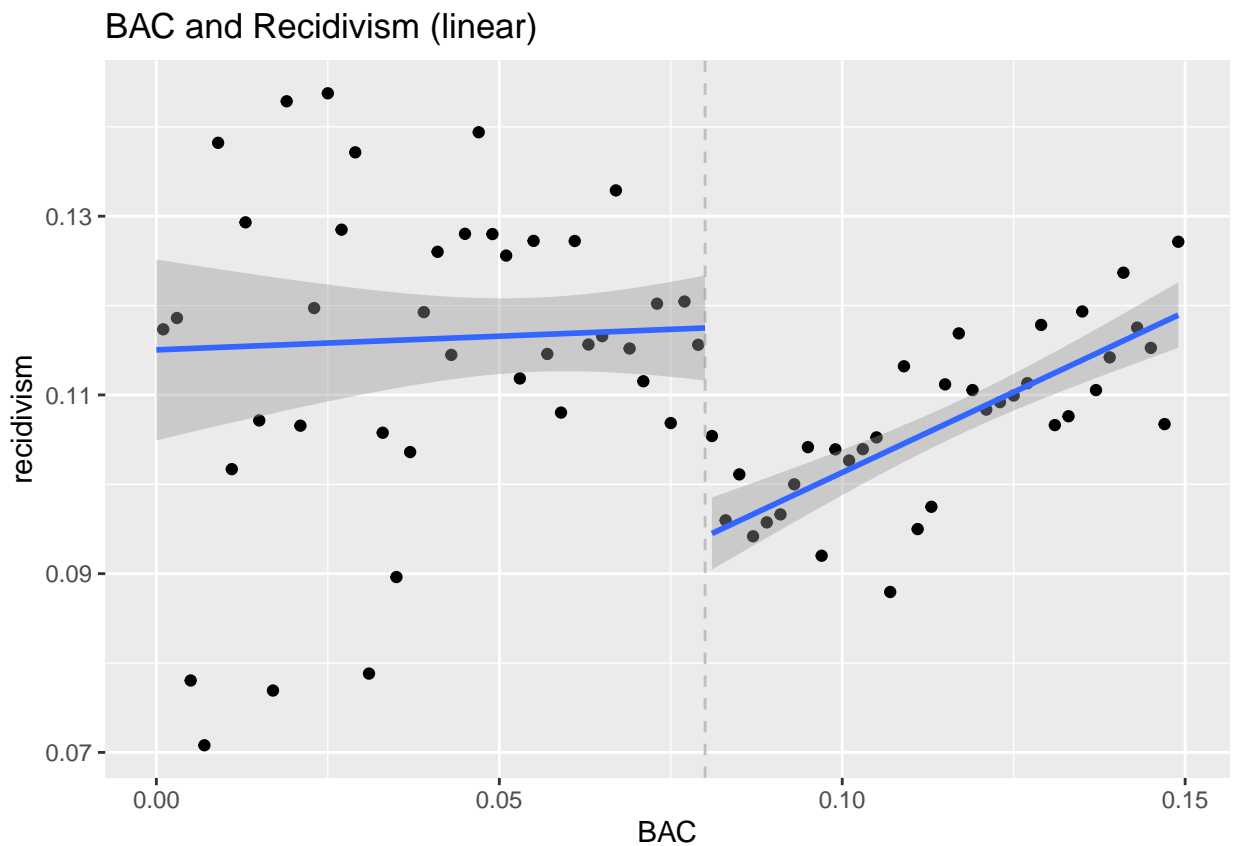
```
categories <- hansen$bac1

hansen3 <- hansen %>%
  filter(bac1 >= 0 & bac1 <= 0.15)

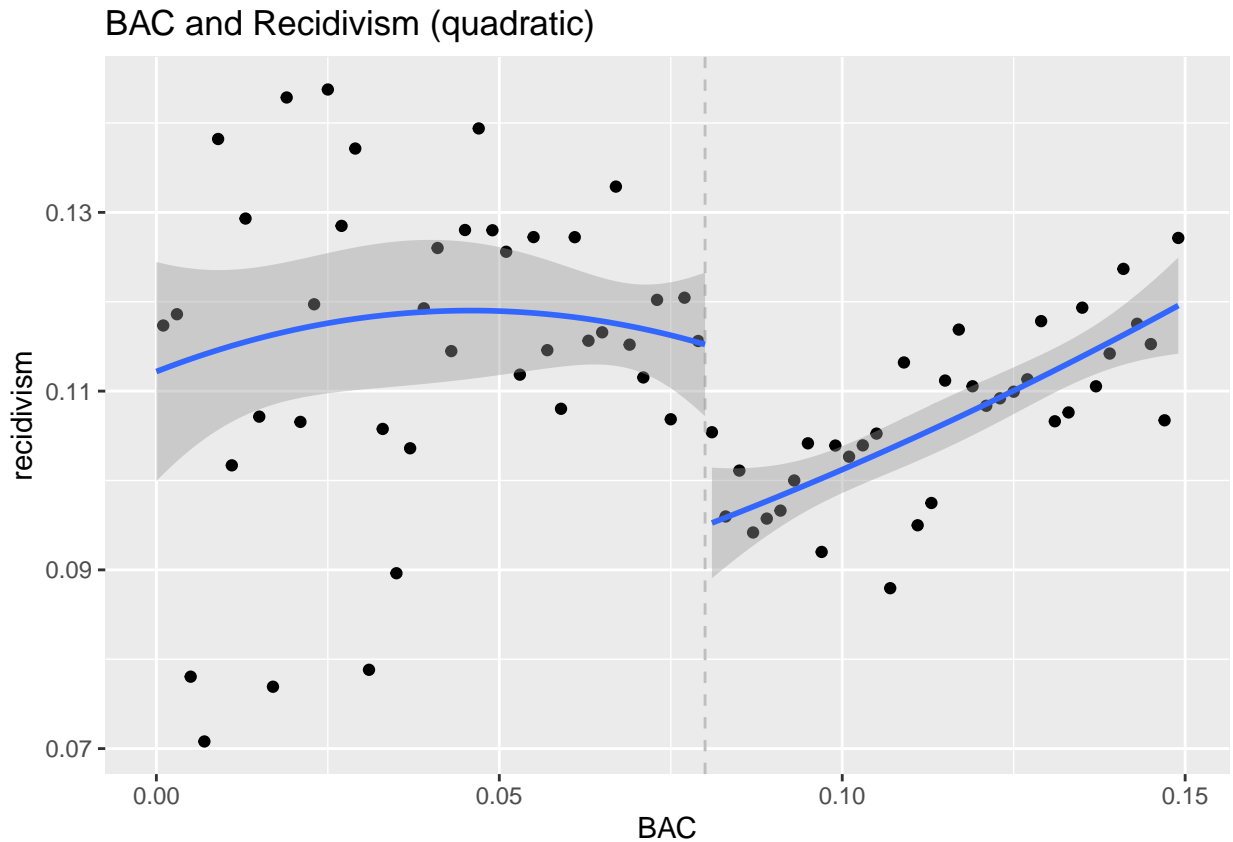
means_recid <- split(hansen3$recidivism, cut(hansen3$bac1, 75)) %>%
  lapply(mean) %>%
  unlist()
agg_hansen_recid <- data.frame(recidivism = means_recid, bac1 = seq(0.001, 0.15, by = 0.002))

hansen3 <- hansen3 %>%
  mutate(gg_group2 = case_when(bac1 > 0.08 ~ 1, TRUE ~ 0))

#Model "recidivism" RDD Using Linear Model and Quadratic models
lrecid <- ggplot(hansen3, aes(bac1, recidivism)) +
  geom_point(aes(x = bac1, y = recidivism), data = agg_hansen_recid) +
  geom_vline(xintercept = 0.08, colour = "grey", linetype = 2) +
  stat_smooth(aes(bac1, recidivism, group = gg_group2), method = "lm") +
  labs(x = "BAC", title = "BAC and Recidivism (linear)")
lrecid
```



```
qrecid<- ggplot(hansen3, aes(bac1, recidivism)) +
  geom_point(aes(x = bac1, y = recidivism), data = agg_hansen_recid) +
  geom_vline(xintercept = 0.08, colour = "grey", linetype = 2) +
  stat_smooth(aes(bac1, recidivism, group = gg_group2), method = "lm", formula = y ~ x + I(x^2)) +
  labs(x = "BAC", title = "BAC and Recidivism (quadratic)")
qrecid
```



## Nine

The graphical results in part eight reflect the findings from Panels A and B in part seven, which estimated that recidivism decreases by roughly 2 percentage points when a person receives a DUI (BAC is past the 0.08 cutoff). In both the linear and quadratic fits in part eight, the visual discontinuity is roughly a 2 percentage point drop in recidivism at the cutoff, from about 11.7% recidivism just before the cutoff to about 9.5% recidivism just after the cutoff. These results are very similar to the results in Hansen, visually. This is expected since the regression discontinuity estimation was already very similar to Hansen in part seven. Given that the RD causal effect estimation in part seven was statistically significant, and given that the predetermined characteristics (i.e. male, white, age, accident) were stable at the cutoff, we can be confident that getting a DUI does in fact reduce recidivism by about 2 percentage points. This reflects Hansen's findings at the 0.08 BAC cutoff, confirming those results.

As implied above, the hypothesis being tested was the following: "Getting a DUI reduces the likelihood of drunk driving in the future." Using regression discontinuity design, I found support for this hypothesis by reconstructing the analysis from scratch in R and confirming many of Hansen's findings.

I found that getting a regular DUI (BAC beyond 0.08) reduces recidivism by about 2 percentage points. Moreover, the effect is slightly larger when a larger binwidth is used (more data points further from the cutoff). With a binwidth of 0.03 to 0.13, the decrease in recidivism was 2.4 percentage points. With a smaller binwidth of 0.055 to 0.105, the decrease in recidivism was 2.0 percentage points. Thus, the closer the data points are to the cutoff, the smaller the effect of a DUI on reducing recidivism, but only slightly. And both results were statistically significant under “honest” standard errors of the RD coefficients.

Based on this analysis, I am confident in Hansen’s findings for three reasons. First, the density test showed that the running variable (BAC) is continuous at the cutoff. Second, the pretreatment characteristics (i.e. male, white, age, accident) were stable at the cutoff. Third, the regression discontinuity estimates for recidivism were statistically significant. The first two points are crucial to gain indirect evidence that the stability and continuity (smoothness) assumptions are met for RDD. With these in place, we can be confident that the statistically-significant decrease in recidivism at the cutoff is based on unbiased RDD estimation and not suffering from confounding effects in the pretreatment variables. As a result, I am confident in Hansen’s conclusion.

Personally, this exercise was quite challenging but extremely valuable. Using R instead of Stata was *much* more time-consuming, but it paid off, and I am glad I used R. Along the way, I ran a parallel analysis in Stata to verify my R results. So, in the end, I got great practice with both tools.