

# RD Project Summary

Maor Milgrom

2020-03-20

## Introduction

The goal of this exercise is to try to test the robustness of the Regression-Discontinuity analysis to different extreme cases, via simulations, using the 'rdrobust' package.

## This Document: Exercise #1

Replacing observations outside Optimal Bandwidth (OBW) with zeros, or dropping them

### Set Parameters

Here we set the main parameters for the exercise:

```
jump=10          # Size of jump at cutoff
loop=1000
figs.iter.save=5
quadratic=T      # T - quadratic DGP, F - linear
symm_obw="mserd" # mserd - symmetric OBW, msetwo - asymmetric OBW
normal.x=T       # T - normal draws of x around cutoff, F - uniform draws
dgp.sd=10        # sd of normal noise added to DGP
bc=F             # bias-corrected estimates or conventional
treat="zero"     # Type of exercise - "zero" or "bwo" ('bandwidth only')
```

### Constructing dataframes for simulation and for results

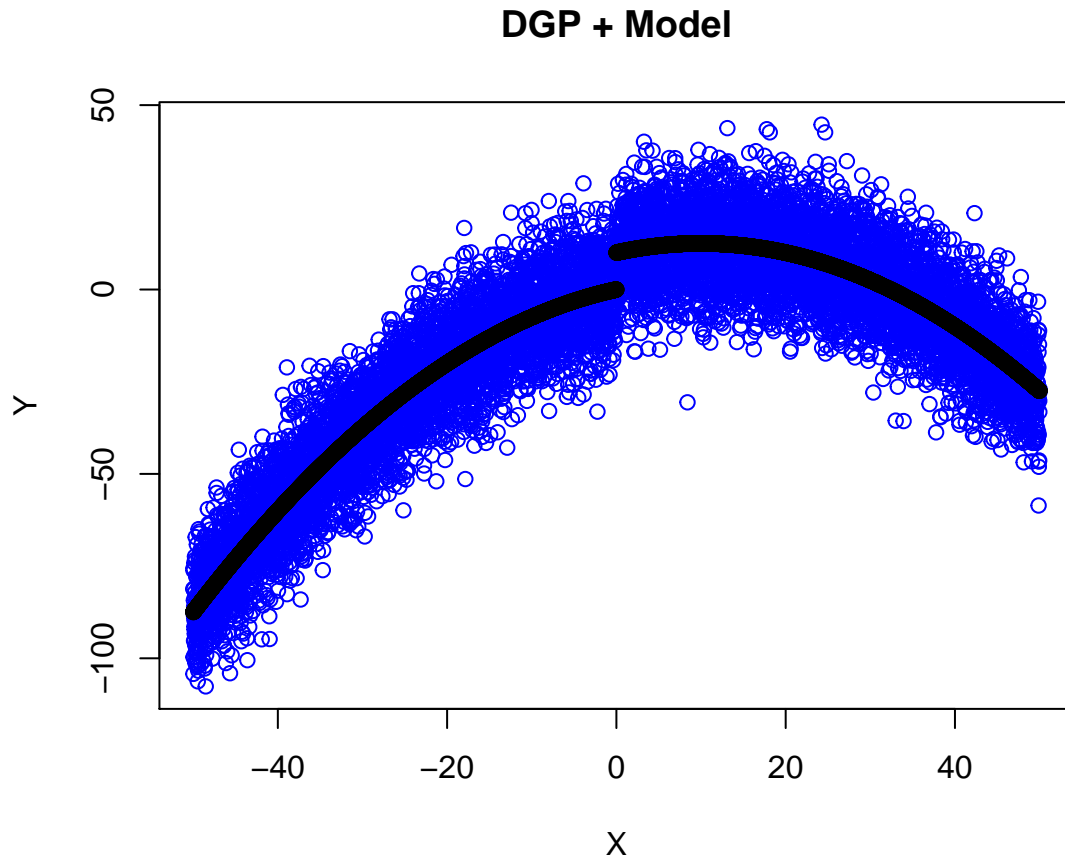
```
df <- as.data.frame(matrix(0, ncol = 0, nrow = length(seq(-100,100,0.01))))
df$x=round(seq(-100,100,0.01), digits=2)
df=subset(df,df$x!=0)
df$treated <- ifelse(df$x>0, 1, 0)
sample.x <- as.data.frame(matrix(0, ncol = 0, nrow = nrow(df)/10))

results<- as.data.frame(matrix(0, ncol = 4, nrow = loop))
colnames(results) <- c("coef","obw","obw_l","obw_r")
results.treat<- as.data.frame(matrix(0, ncol = 6, nrow = loop))
colnames(results.treat) <- c("coef","obw","obw_l","obw_r",
                             "coef_diff","obw_diff")
```

### Simulate DGP and plot

```
df$y.model<- 0.5*df$x - 0.025*df$x^2*quadratic + jump*df$treated
df$y=df$y.model+rnorm(length(df$x),0,dgp.sd)
```

```
df %>%
  filter(x > -50 & x < 50) %T>%
  plot(y~x,., ylim = range(c(y,y.model)),
       col="blue", ylab = "Y", xlab = "X") %T>%
  par(new = T) %>%
  plot(y.model~x,., ylim = range(c(y,y.model)),
       axes = FALSE, xlab = "", ylab = "")
  title(main = "DGP + Model")
```



### Iterations

Running 1000 iterations, and saving figures from 5 last iterations to file

```
for(i in 1:loop) {
  if (normal.x==T) {
    sample.x$x <- round(rnorm(nrow(df)/10, 0, 10),digits = 2)
  } else {
    sample.x$x <- round(runif(nrow(df)/10, -20,20),digits = 2)
  }

  sample.x=subset(sample.x, x>-100 & x<100)
  sample=as.data.frame(inner_join(df, sample.x, by="x"))
  results.current=rdrobust(sample$y,sample$x,bwselect = symm_obw)
```

```

results[i,1]=results.current$coef[bc+1]-jump # normalizing to zero
results[i,3:4]=results.current$bws[bc+1,1:2]
results[i,2]=results[i,3]+results[i,4]

if (treat=="zero") {
  sample.treat=sample
  sample.treat$y <- ifelse(sample$x> -results[i,3] &
                           sample$x< results[i,4], sample$y, 0)

} else if (treat=="bwo") {
  sample.treat=subset(sample,x> -results[i,3] & x< results[i,4])
}

results.current=rdrobust(sample.treat$y,sample.treat$x)
results.treat[i,1]=results.current$coef[bc+1]-jump # normalizing to zero
results.treat[i,3:4]=results.current$bws[bc+1,1:2]
results.treat[i,2]=results.treat[i,3]+results.treat[i,4]

results.treat[i,5:6]=results[i,1:2]-results.treat[i,1:2] # computing diff from base

if (i >= begin.figures) {
  coef_base=paste("Base: Coef = ",round(results[i,1],digits = 2)+jump,sep = "")
  coef_treat=paste(treat," : Coef = ",round(results.treat[i,1],digits = 2)+jump,sep = "")

  figure_name=paste(figs.dir,"sample",i,treat,"_",save.ext,"_", "model.png",sep = "")
  png(figure_name)
  temp.treat=sample.treat %>%
    filter(x > -20 & x < 20)
  temp=sample %>%
    filter(x > -20 & x < 20)
  plot(temp$y~temp$x, ylim = range(c(temp.treat$y,temp$y)),
        xlim = range(c(temp.treat$x,temp$x)),
        col="red", ylab = "Y", xlab = "X")
  par(new = T)
  plot(temp.treat$y~temp.treat$x, ylim = range(c(temp.treat$y,temp$y)),
        xlim = range(c(temp.treat$x,temp$x)),
        col="blue", ylab = "Y", xlab = "X")
  par(new = T)
  plot(temp$y.model~temp$x, ylim = range(c(temp.treat$y,temp$y)),
        xlim = range(c(temp.treat$x,temp$x)),
        col="black", ylab = "Y", xlab = "X")
  abline(v = c(-results[i,3], -results.treat[i,3],
               results[i,4], results.treat[i,4]),
         col=c("blue","red", "blue", "red"),
         lty=c(1,2,1,2), lwd=c(1,2,1,2))
  legend("top", legend=c(coef_base, coef_treat),
        col=c("blue", "red"), lty=1:2, cex=0.8)
  dev.off()

  figure_name=paste(figs.dir,"rd_plot",i,"_",save.ext,".png",sep = "")
  png(figure_name)

```

```

rdplot(sample$y,sample$x)
dev.off()

figure_name=paste(figs.dir,"sample",i,"_",save.ext,".png",sep = "")
png(figure_name)
sample %>%
  filter(x > -30 & x < 30) %T>%
  plot(y~x,., ylim = range(c(y,y.model)),
       col="blue", ylab = "Y", xlab = "X") %T>%
  par(new = T) %>%
  plot(y.model~x,., ylim = range(c(y,y.model)),
       axes = FALSE, xlab = "", ylab = "")
dev.off()
}
}
#data_name=paste("base_data","_",treat,"_",save.ext,".RData",sep = "")
#save.image(data_name)

```

### Presenting figures from last iteration for illustration

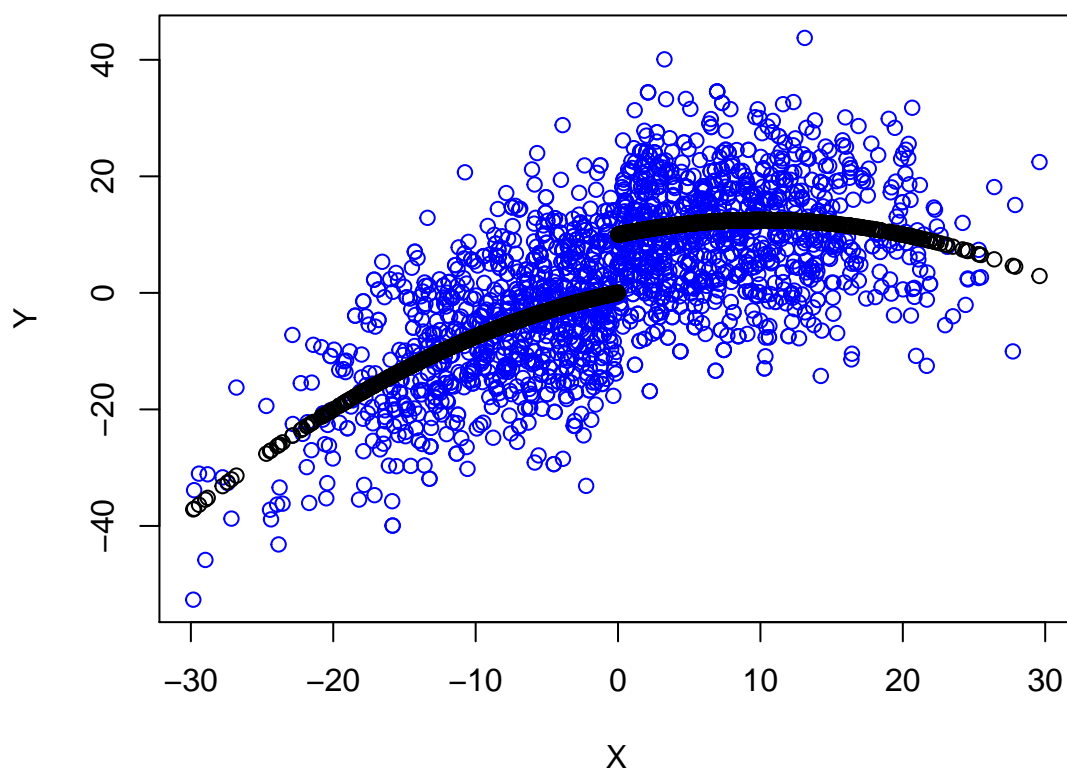
```

i=loop
coef_base=paste("Base: Coef = ",round(results[i,1],digits = 2)+jump,sep = "")
coef_treat=paste(treat,": Coef = ",round(results.treat[i,1],digits = 2)+jump,sep = "")

sample %>%
  filter(x > -30 & x < 30) %T>%
  plot(y~x,., ylim = range(c(y,y.model)),
       col="blue", ylab = "Y", xlab = "X") %T>%
  par(new = T) %>%
  plot(y.model~x,., ylim = range(c(y,y.model)),
       axes = FALSE, xlab = "", ylab = "")
  title(main = "Draw from DGP + Model")

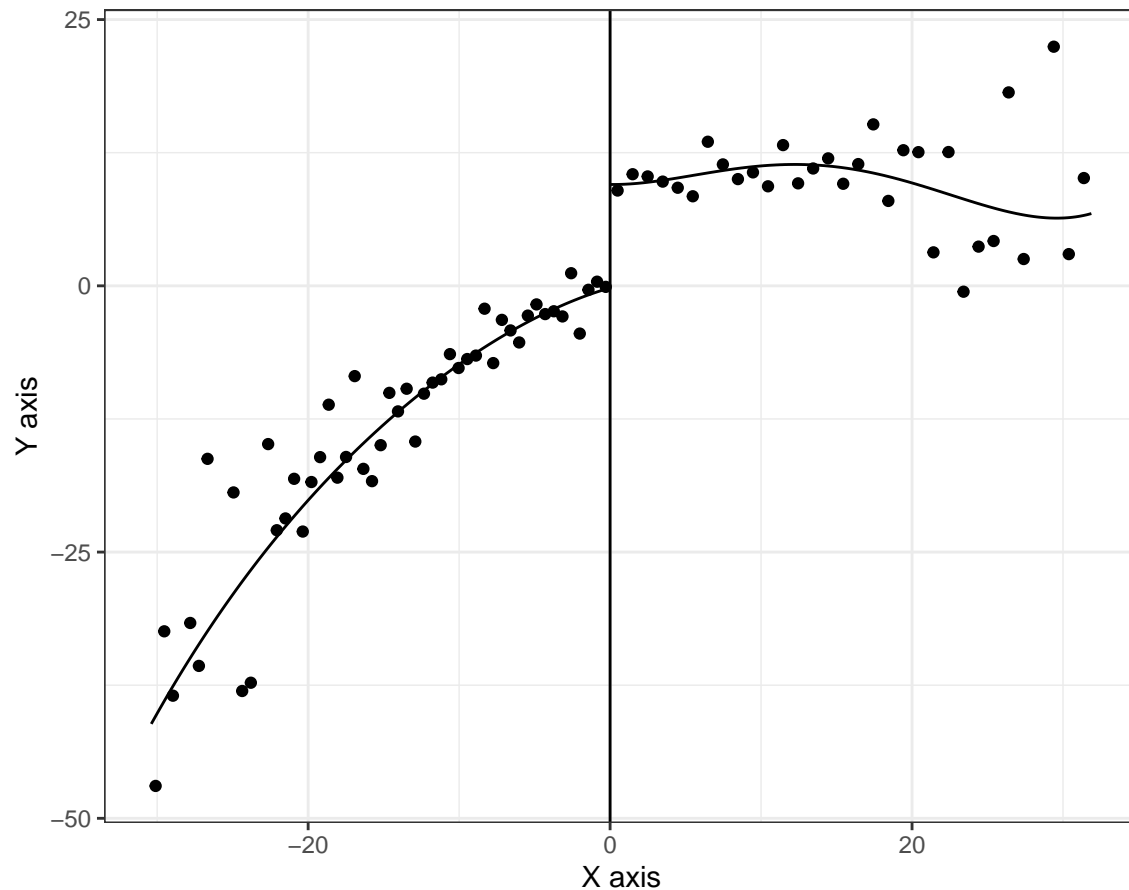
```

## Draw from DGP + Model



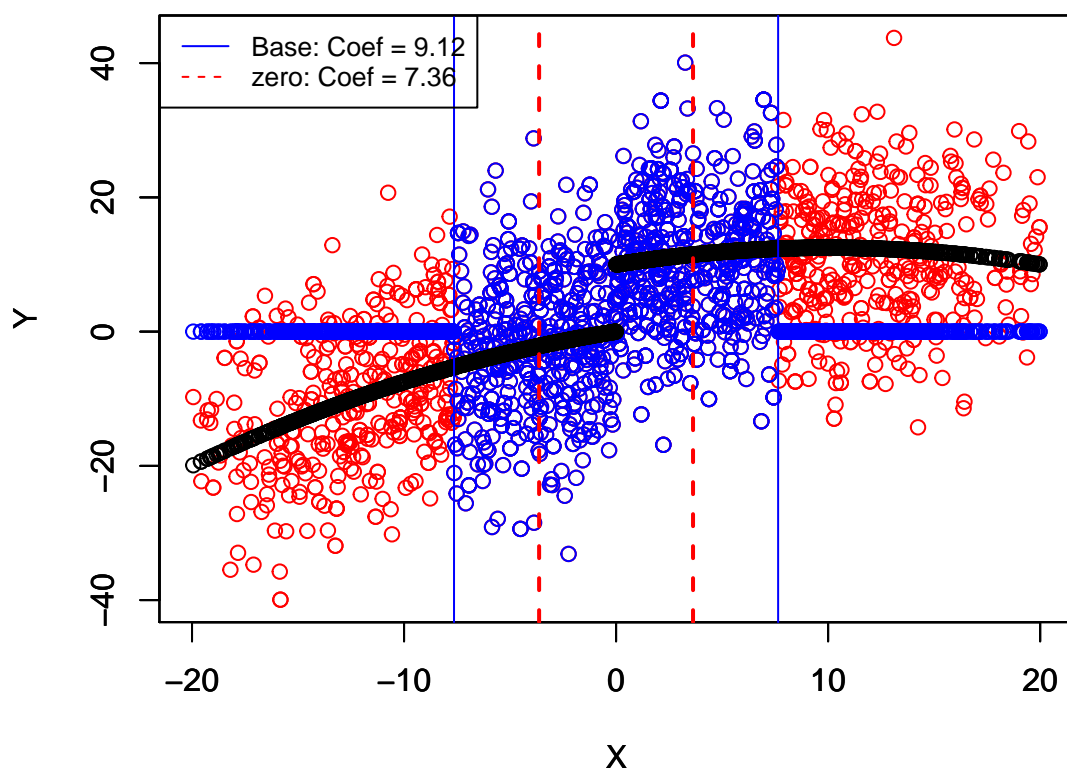
```
rdplot(sample$y,sample$x, title = "RD Plot for baseline")
```

RD Plot for baseline



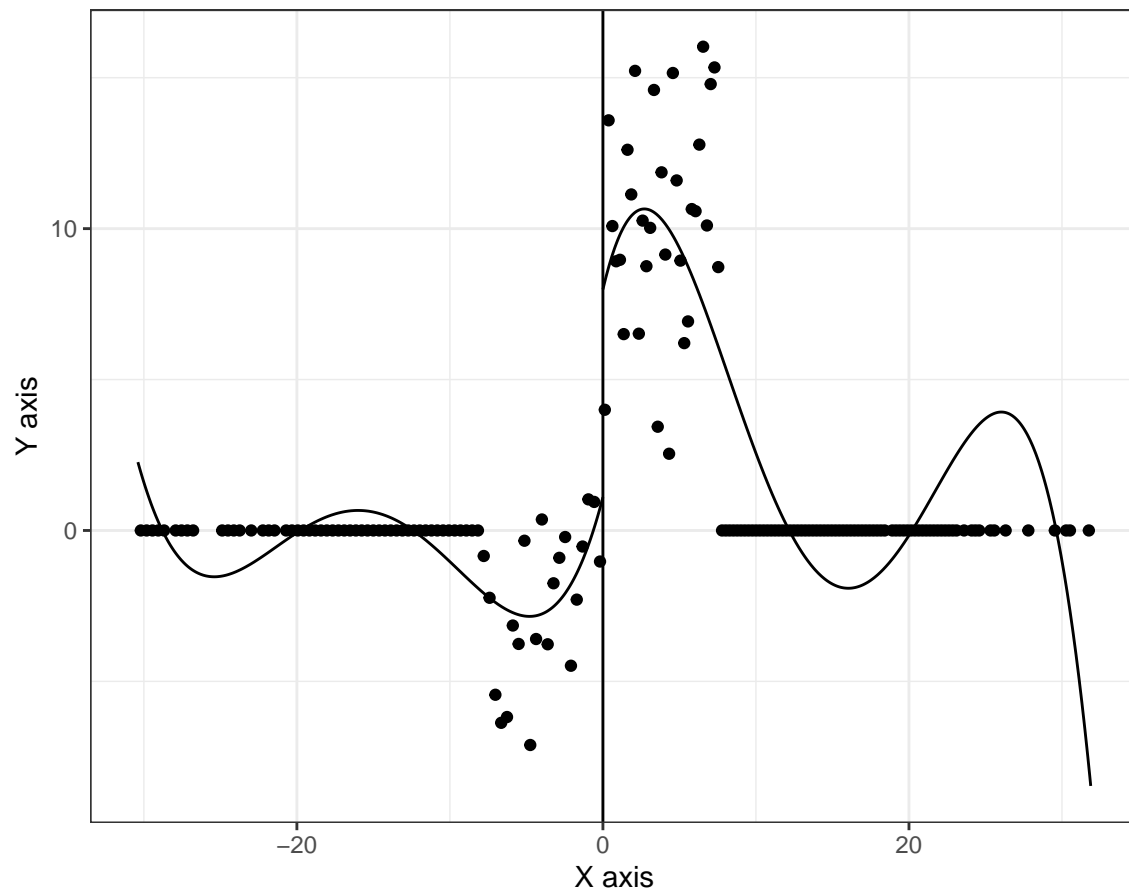
```
temp.treat=sample.treat %>%
  filter(x > -20 & x < 20)
temp=sample %>%
  filter(x > -20 & x < 20)
plot(temp$y~temp$x, ylim = range(c(temp.treat$y,temp$y)),
      xlim = range(c(temp.treat$x,temp$x)),
      col="red", ylab = "Y", xlab = "X")
par(new = T)
plot(temp.treat$y~temp.treat$x, ylim = range(c(temp.treat$y,temp$y)),
      xlim = range(c(temp.treat$x,temp$x)),
      col="blue", ylab = "Y", xlab = "X")
par(new = T)
plot(temp$y.model~temp$x, ylim = range(c(temp.treat$y,temp$y)),
      xlim = range(c(temp.treat$x,temp$x)),
      col="black", ylab = "Y", xlab = "X")
abline(v = c(-results[i,3], -results.treat[i,3],
              results[i,4], results.treat[i,4]),
       col=c("blue","red", "blue", "red"),
       lty=c(1,2,1,2), lwd=c(1,2,1,2))
legend("topleft", legend=c(coef_base, coef_treat),
       col=c("blue", "red"), lty=1:2, cex=0.8)
title(main = paste0("Comparing baseline to ",treat))
```

## Comparing baseline to zero



```
rdplot(sample.treat$y,sample.treat$x, title = paste0("RD Plot for ",treat))
```

RD Plot for zero



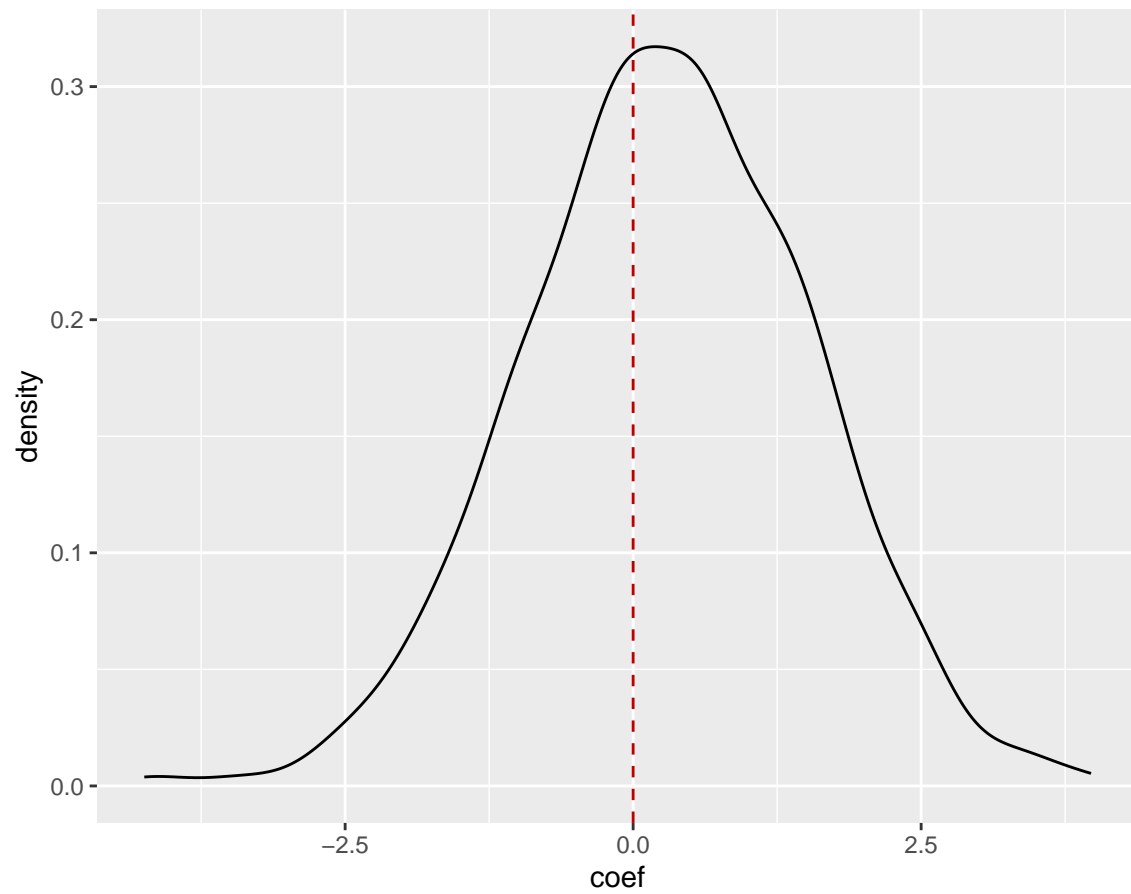
## Results

Figures summarizing iterations

```
## PDF'S OF LEVELS
ggplot(results, aes(coef)) + stat_density(geom="line") +
  geom_vline(aes(xintercept=0), colour="#BB0000", linetype="dashed") +
  ggtitle("Density of RD coefficients: baseline")
```

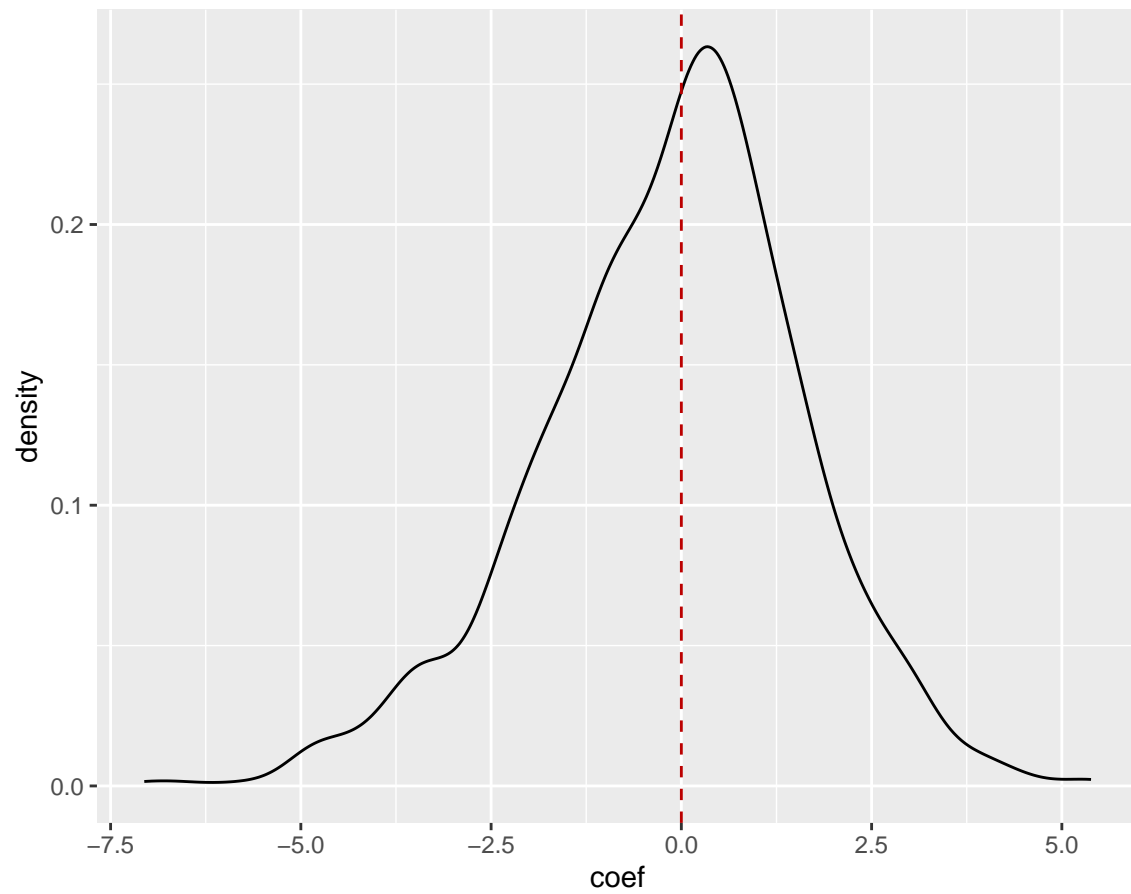


Density of RD coefficients: baseline



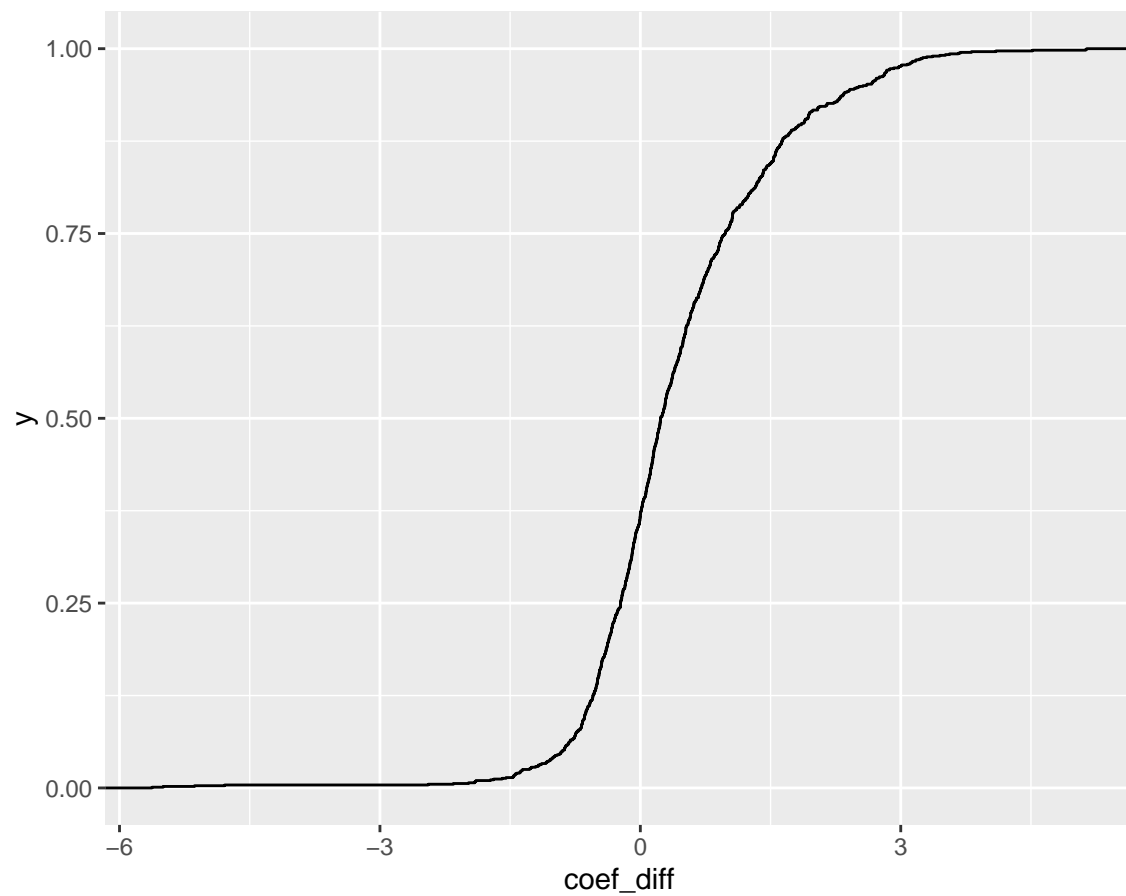
```
ggplot(results.treat, aes(coef)) + stat_density(geom="line") +  
  geom_vline(aes(xintercept=0), colour="#BB0000", linetype="dashed") +  
  ggtitle(paste0("Density of RD coefficients: ",treat))
```

Density of RD coefficients: zero



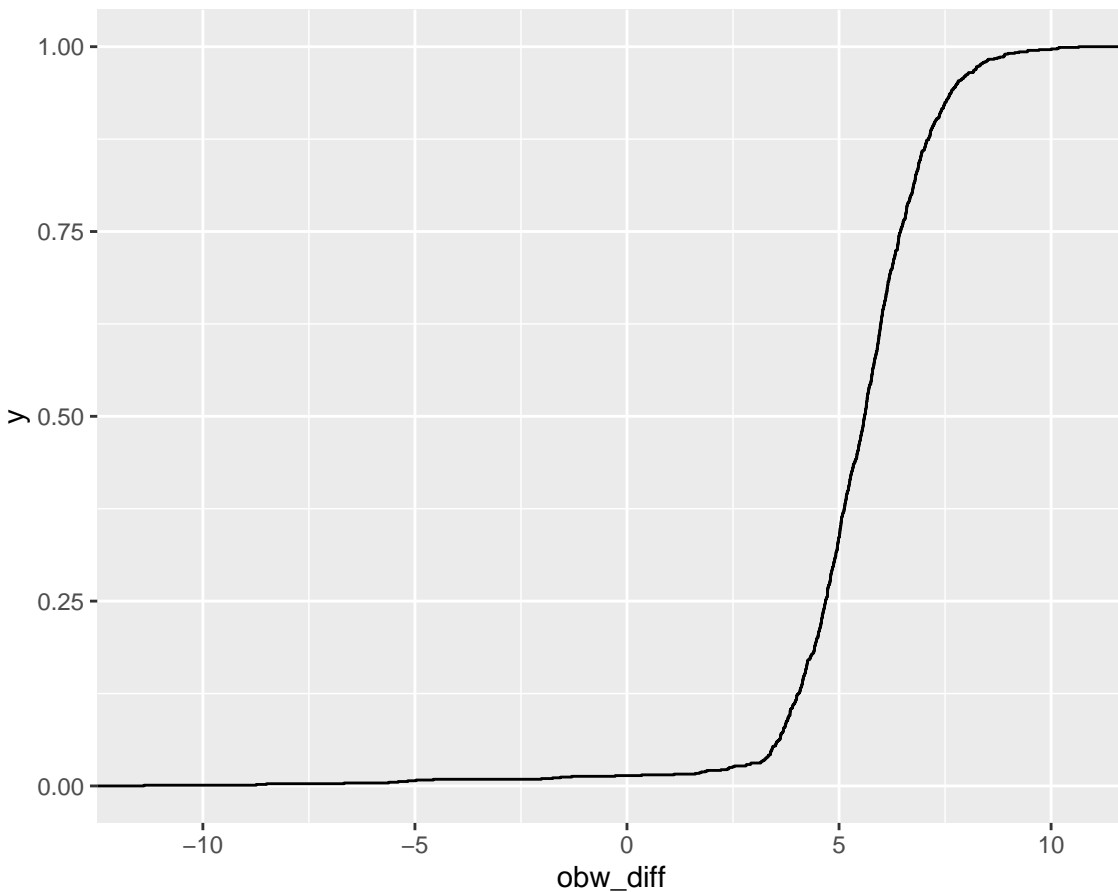
```
### treat_base_diff
ggplot(results.treat, aes(coef_diff)) + stat_ecdf(geom = "step") +
  ggtitle(paste0("CDF of difference b/w coefficients of baseline and ", treat))
```

CDF of difference b/w coefficients of baseline and zero



```
ggplot(results.treat, aes(obw_diff)) + stat_ecdf(geom = "step") +  
  ggtitle(paste0("CDF of difference b/w OBW of baseline and ", treat))
```

CDF of difference b/w OBW of baseline and zero



Correlation between baseline OBW and difference in OBW

```
cor(results$obw,results.treat$obw_diff)
```

```
## [1] 0.6837858
```

Correlation between baseline coefficients and difference in coefficients

```
cor(results$coef,results.treat$coef_diff)
```

```
## [1] -0.1076403
```

Summary results - TABLE

```
results_table=as.data.frame(matrix(0, ncol = 3, nrow = 0))
results_table[1,1]=round(mean(results$coef),digits = 4)
results_table[1,2]=round(mean(results.treat$coef),digits = 4)
results_table[1,3]=round(mean(results.treat$coef_diff),digits = 4)
results_table[2,1]=round(mean(results$obw),digits = 4)
results_table[2,2]=round(mean(results.treat$obw),digits = 4)
results_table[2,3]=round(mean(results.treat$obw_diff),digits = 4)
colnames(results_table) = c("base",treat,"diff_base")
```

```
rownames(results_table) = c("coef", "obw")
kable(results_table, caption = "Summary Table")
```

Table 1: Summary Table

	base	zero	diff_base
coef	0.2886	-0.1448	0.4333
obw	13.1101	7.6357	5.4744