Regression Discontinuity Design (RDD)

Shengqiao Lin (shengqiaolin@utexas.edu (mailto:shengqiaolin@utexas.edu))

2023-03-02

Today's Goal:

- RDD Assumption
- Tutorial
- Replication

Assumption

- Score, Cutoff, Treatment, Outcome
- Continuity Framework: the sole change occurring at the discontinuity point is the shift in the treatment status (de la Cuesta, Brandon; Imai, Kosuke, 2016 (10.1146/annurev-polisci-032015-010115)).
- Local randomization: as-if randomized assignment of treatment
- *Key: No sorting behavior (incentive/ability to change treatment status)*
- Score might correlate with outcomes

Type 1: Winning Margins in Close Elections

- Endogeneity: elected politicians are not randomly assigned
- Assumption: Indifference between candidates who win/lose close elections
- Violation: electoral frauds (self-sorting behavior)
- Treatment: Politician characteristics
- Outcomes: Governance/Policy/Budget
    - David Szakonyi, Private Sector Policy Making: Business Background and Politicians' Behavior in Office (https://www.journals.uchicago.edu/doi/10.1086/709297), JOP 2020.

    - Fowler, Garro, Spenkuch, Quid Pro Quo? Corporate Returns to Campaign Contributions (https://www.journals.uchicago.edu/doi/10.1086/707307), JOP 2020.

    - Potentials: Most focus on US/developed democracies

- Critics: Marshall, Can Close Election Regression Discontinuity Designs Identify Effects of Winning Politician Characteristics? (DOI:%2010.1111/ajps.12741), AJPS 2022.
    - Either that the characteristic of interest does not affect candidate vote shares

    - Or that no compensating differential affects the outcome.

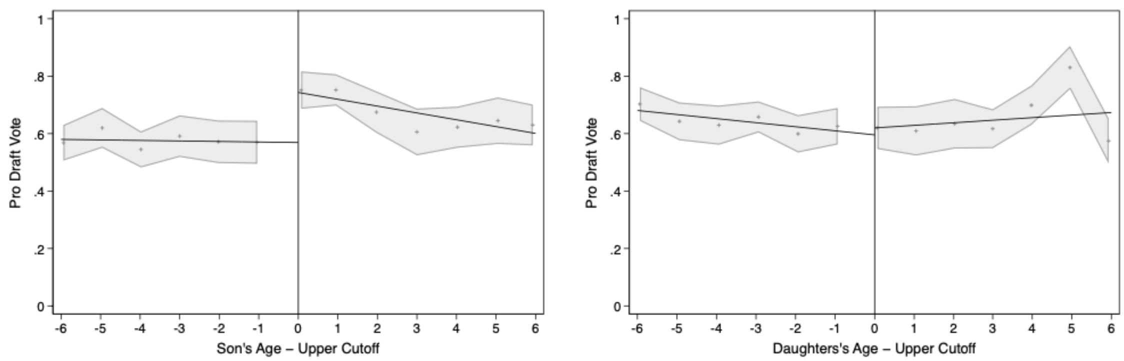
Type 2: Age

- Endogeneity: self-selection into policies
- Assumption: age-based policy design
- Violation: lax enforcement/ noncompliance
- Treatment: policy eligibility/benefits
- Examples:

- Age -> Benefits from education reform -> Political participation. Croke et al., Deliberate Disengagement: How Education Can Decrease Political Participation in Electoral Authoritarian Regimes (doi:10.1017/S0003055416000253), APSR 2016.

- Age -> Public insurance -> Support for public health policies. Lerman and McCabe, Personal Experience and Public Opinion: A Theory and Test of Conditional Policy Feedback (http://dx.doi.org/10.1086/689286), JOP 2017.

- Son's age -> Legislators' pro-conscription voting. McGuirk, Hilger, Miller, No Kin in the Game: Moral Hazard and War in the U.S. Congress (https://www.journals.uchicago.edu/doi/abs/10.1086/724316), JPE 2023.



(a) RD Effect of Son's Age at Upper Cutoff (b) RD Effect of Daughter's Age at Upper Cutoff

Figure 1: **Regression Discontinuity Plots.** These plots correspond to estimates in Table A10. The estimate for ρ in part (a) is 0.1879 ($p<0.05$). The placebo estimate in part (b) is -0.0044 ($p>0.10$).

Type 3: Geography/Boundary (Geographical/Spatial RD)

- Endogeniety: self-selection into policies
- Assumption: geography-based treatment
- Violation: population/resource mobility
- Treatment: policies, colonization, natural disaster etc.
- Examples:
    - North v.s. South Vietnam in history (ruled by Dai Viet before French) -> Economic Growth. Dell, Lane, and Querubin, TheHistorical State, Local Collective Action, and Economic Development in Vietnam (https://doi.org/10.3982/ECTA15122), Econometrica 2018.

    - Japanese colonization -> state building -> township-level governance. Mattingly, Colonial Legacies and State Institutions in China: Evidence From a Natural Experiment (https://doi.org/10.1177/0010414015600465), CPS 2016.

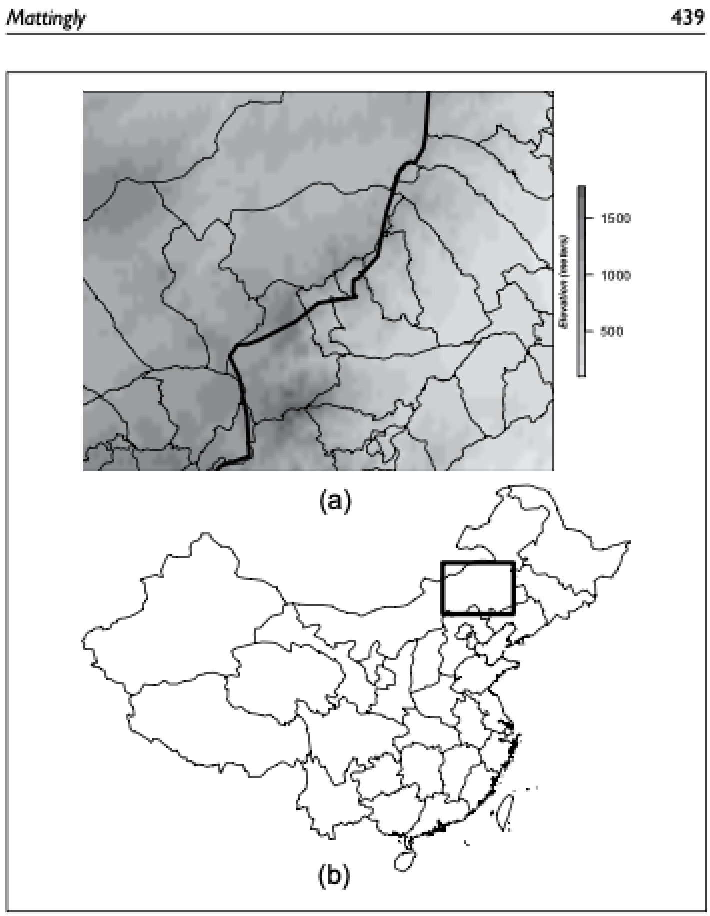


Figure 1. (a) The western border of the Japanese puppet state of Manchukuo, shown with a black line, and (b) The area of detail, which is approximately the size of South Korea.
In Figure 1a, current county borders are shown with lighter black lines, and altitude is shown in the background.

Other Smart Designs:

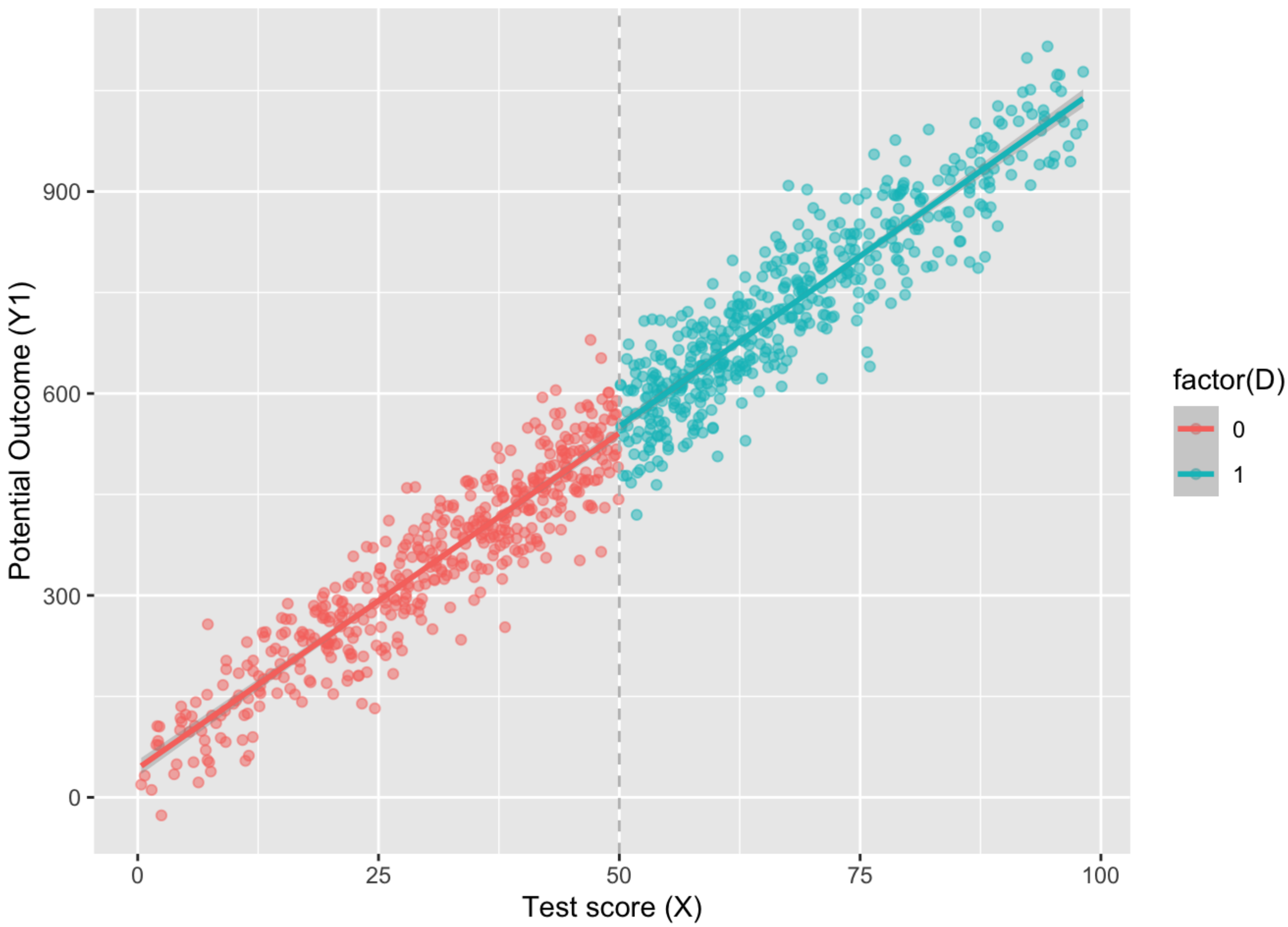
- Timing of unanticipated social events -> Public opinion (weekly survey). Reny and Newman, The Opinion-Mobilizing Effect of Social Protest against Police Violence: Evidence from the 2020 George Floyd Protests (doi:10.1017/S0003055421000460), APSR 2021.
- Exam score -> college education -> political ideology. Apfeld et al., Higher Education and Cultural Liberalism: Regression Discontinuity Evidence from Romania (https://www.journals.uchicago.edu/doi/full/10.1086/720644), JOP 2018.
- Security scores -> military attacks -> economic development. Dell and Querubin, Nation Building Through Foreign Intervention: Evidence from Discontinuities in Military Strategies (https://doi.org/10.1093/qje/qjx037), QJE 2017.

Tutorial

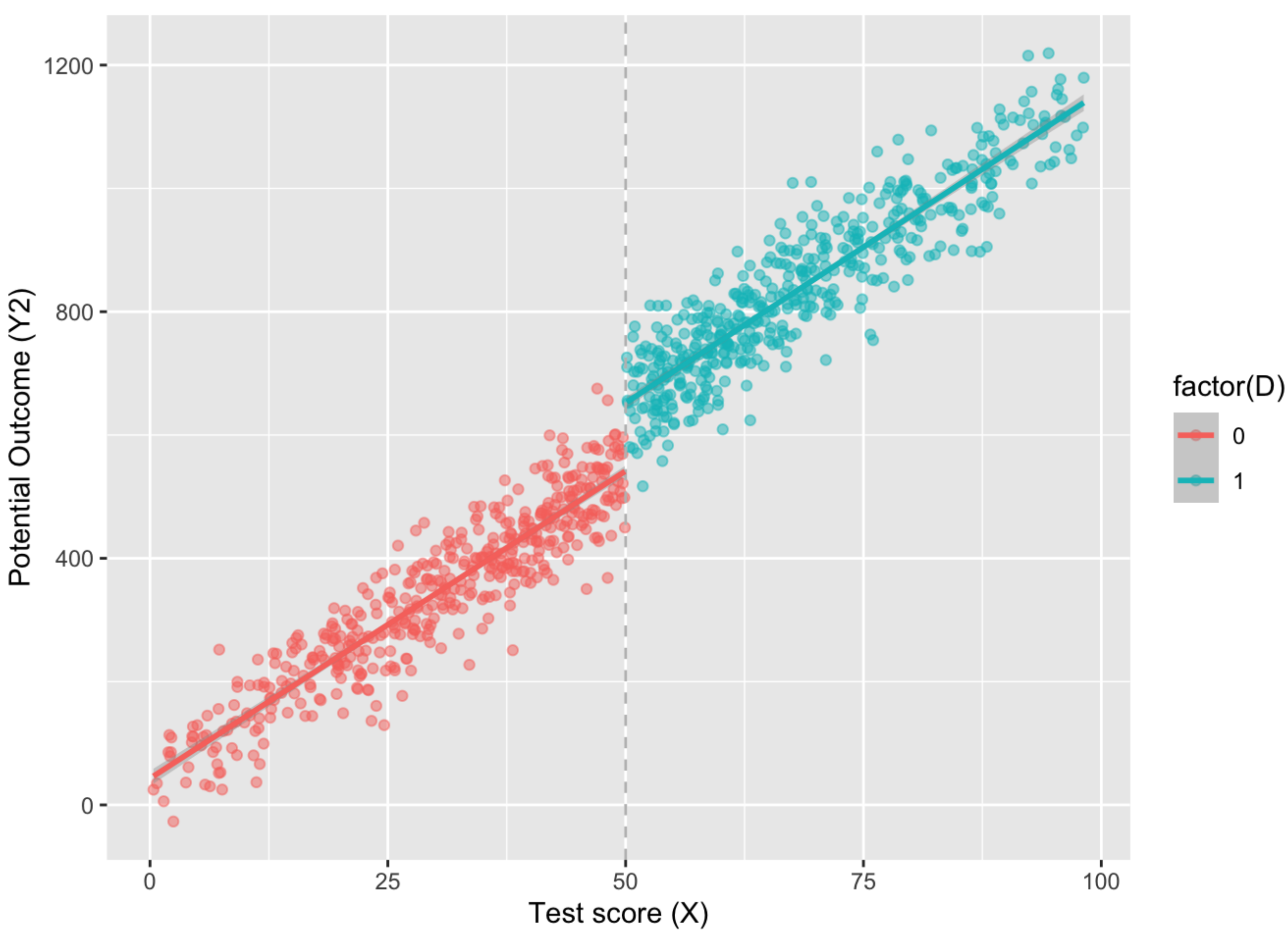
We start with simulating data. C is the confounder, X is the running variable (affected by C), the cutoff is at 25. Y1 is the outcome without a jump at the cutoff and Y2 are affected by the treatment. The real treatment effect is *100*.

```r
library(ggplot2)
library(tidyverse)
set.seed(2023)
dat <- tibble(
  C = rnorm(1000, 10, 5),
  X =  5*C + rnorm(1000, 0, 10),
  D = if_else(X > 50, 1, 0),
  Y1= 0   * D + 30*C + 5 * X + rnorm(1000, 0, 5),
  Y2= 100 * D + 30*C + 5 * X + rnorm(1000, 0, 5))
dat<- subset(dat,X>0&X<100)
```

Here shows the relationship between X and Y1/Y2. Graphically, we can see that Y2 has a jump at the cutoff (X=50) but not Y1.



Potential Outcome (Y1)
0
300
600
900
0
25
50
75
100
Test score (X)
factor(D)
0
1



Potential Outcome (Y2)
0
400
800
1200
0
25
50
75
100
Test score (X)
factor(D)
0
1

Estimation

- We then estimate the effect of D on Y2 using OLS.

```
m0<- lm(Y2~D,dat)
m1<- lm(Y2~D+X,dat)
m2<- lm(Y2~D+X+C,dat)
stargazer::stargazer(m0,m1,m2,type='text',digits = 2,omit.stat = 'f')
```

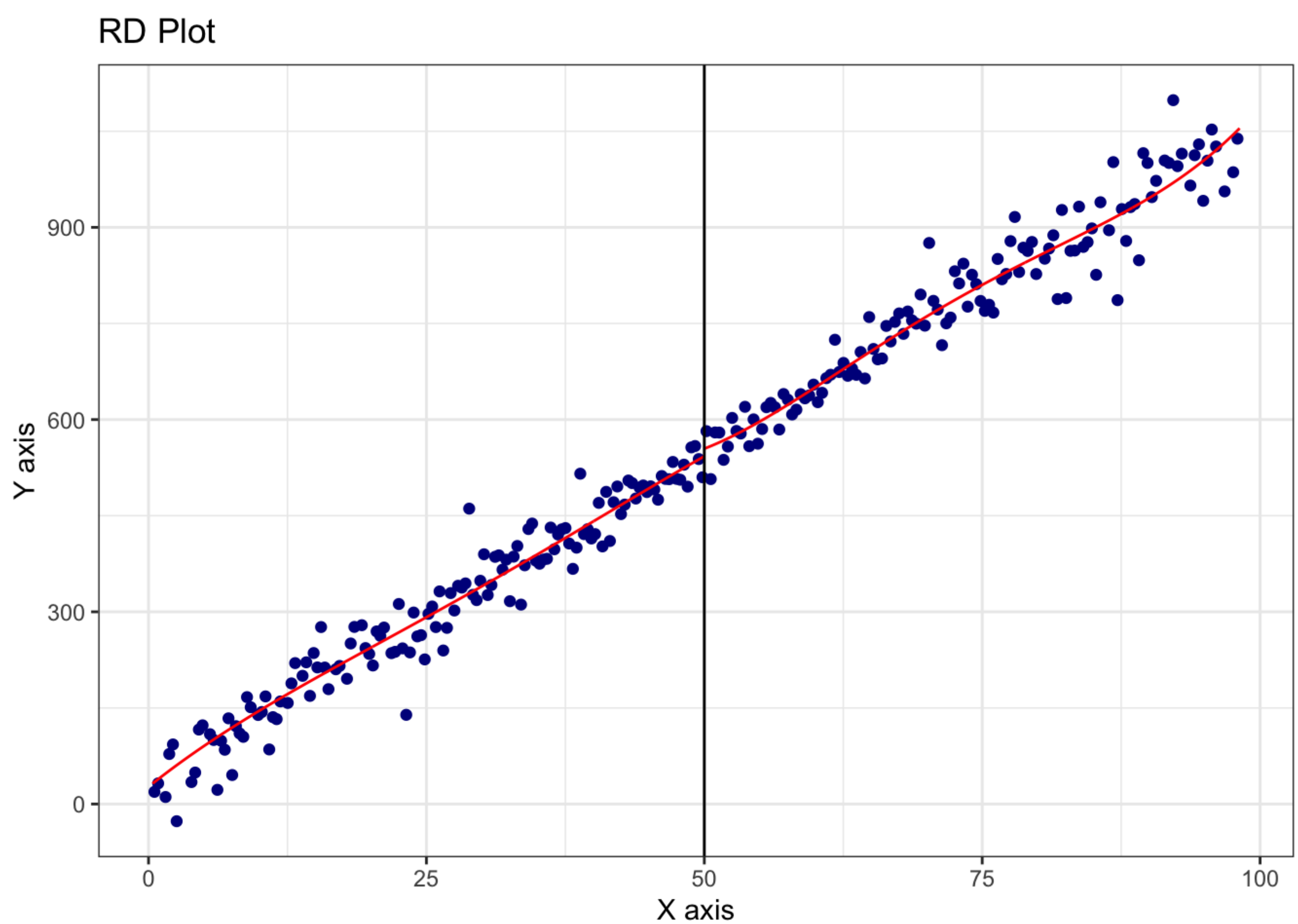
```
##
## ========================================================================
##                                     Dependent variable:
##                        ------------------------------------------------
##                                             Y2
##                              (1)              (2)              (3)
## ------------------------------------------------------------------------
## D                         490.78***        110.21***        100.00***
##                            (9.09)           (6.47)           (0.58)
##
## X                                          10.04***          4.98***
##                                            (0.14)           (0.02)
##
## C                                                           30.15***
##                                                             (0.09)
##
## Constant                  344.42***        40.97***          -0.14
##                            (6.47)           (5.00)           (0.46)
##
## ------------------------------------------------------------------------
## Observations                 939              939              939
## R2                          0.76             0.96             1.00
## Adjusted R2                 0.76             0.96             1.00
## Residual Std. Error 139.30 (df = 937) 55.27 (df = 936) 4.96 (df = 935)
## ========================================================================
## Note:                                       *p<0.1; **p<0.05; ***p<0.01
```

We then estimate the effect with *rdrobust*.

```
library(rdrobust)
attach(dat)
m1<- rdrobust(y=Y1,#Outcome
           x=X,#Running Variable,
           c=50,#by default#c=0
           covs=C#covariates
           #Kernel=triangular,epanechnikov, and uniform
           #binselect=mserd,
           )
summary(m1)
```

```
## Covariate-adjusted Sharp RD estimates using local polynomial regression.
##
## Number of Obs.                  939
## BW type                       mserd
## Kernel                   Triangular
## VCE method                       NN
##
## Number of Obs.                  463          476
## Eff. Number of Obs.             212          244
## Order est. (p)                    1            1
## Order bias  (q)                   2            2
## BW est. (h)                  16.325       16.325
## BW bias (b)                  26.406       26.406
## rho (h/b)                     0.618        0.618
## Unique Obs.                     463          476
##
## =============================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## =============================================================================
##   Conventional    -0.299     0.942    -0.317     0.751    [-2.145 , 1.548]
##         Robust         -         -    -0.305     0.760    [-2.560 , 1.871]
## =============================================================================
```

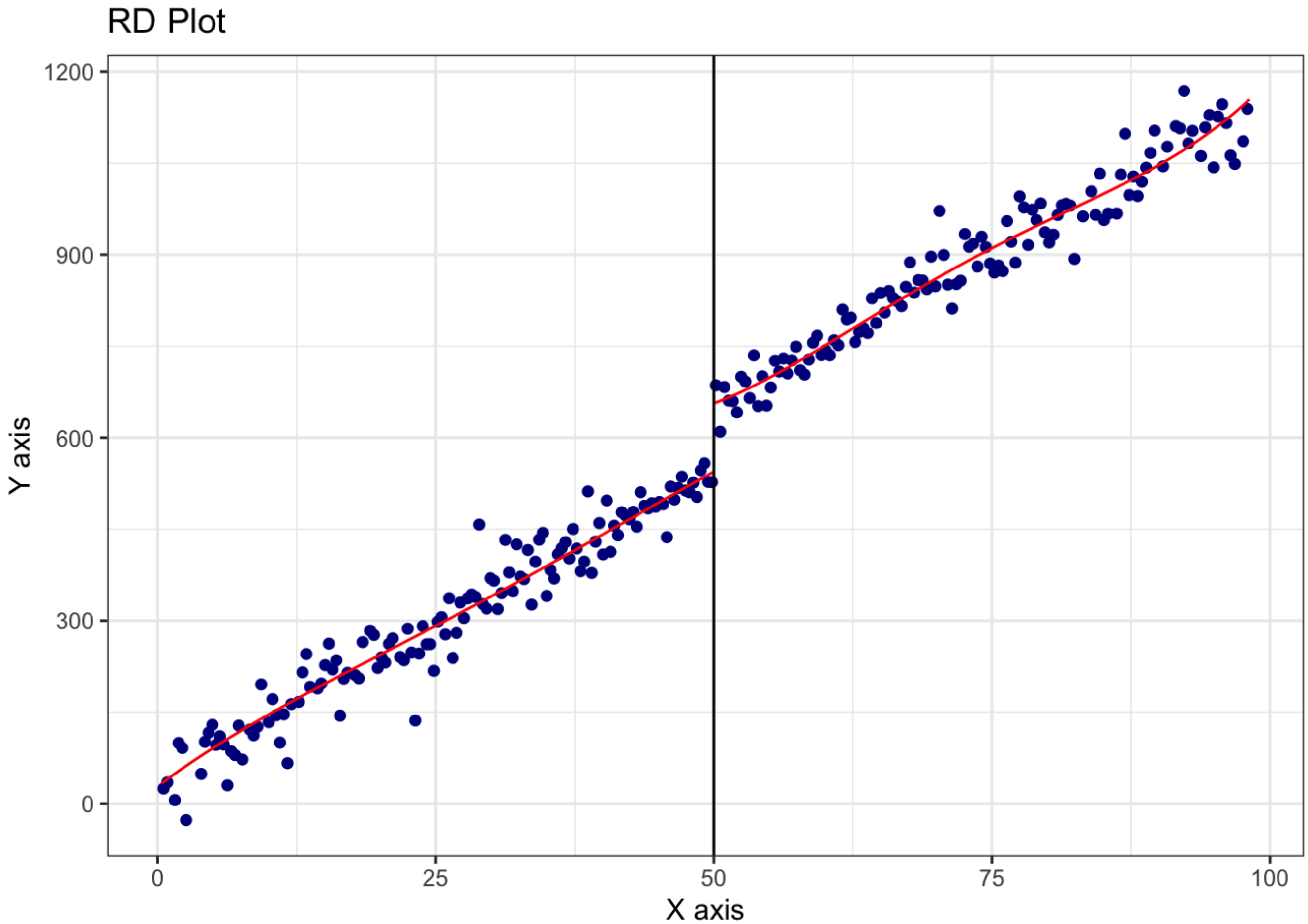
```
rdplot(y=Y1,x=X,c=50)
```



```r
m2<- rdrobust(y=Y2,#Outcome
          x=X,#Running Variable
          c=50,# by default#c=0
          #Kernel=triangular,epanechnikov, and uniform
          #binselect=mserd,
          )
summary(m2)
```

```
## Sharp RD estimates using local polynomial regression.
##
## Number of Obs.                  939
## BW type                       mserd
## Kernel                   Triangular
## VCE method                       NN
##
## Number of Obs.                  463          476
## Eff. Number of Obs.             147          177
## Order est. (p)                    1            1
## Order bias  (q)                   2            2
## BW est. (h)                  11.256       11.256
## BW bias (b)                  18.098       18.098
## rho (h/b)                     0.622        0.622
## Unique Obs.                     463          476
##
## =============================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## =============================================================================
##   Conventional   113.214    15.550     7.281     0.000    [82.736 , 143.691]
##         Robust         -         -     6.404     0.000    [81.668 , 153.711]
## =============================================================================
```

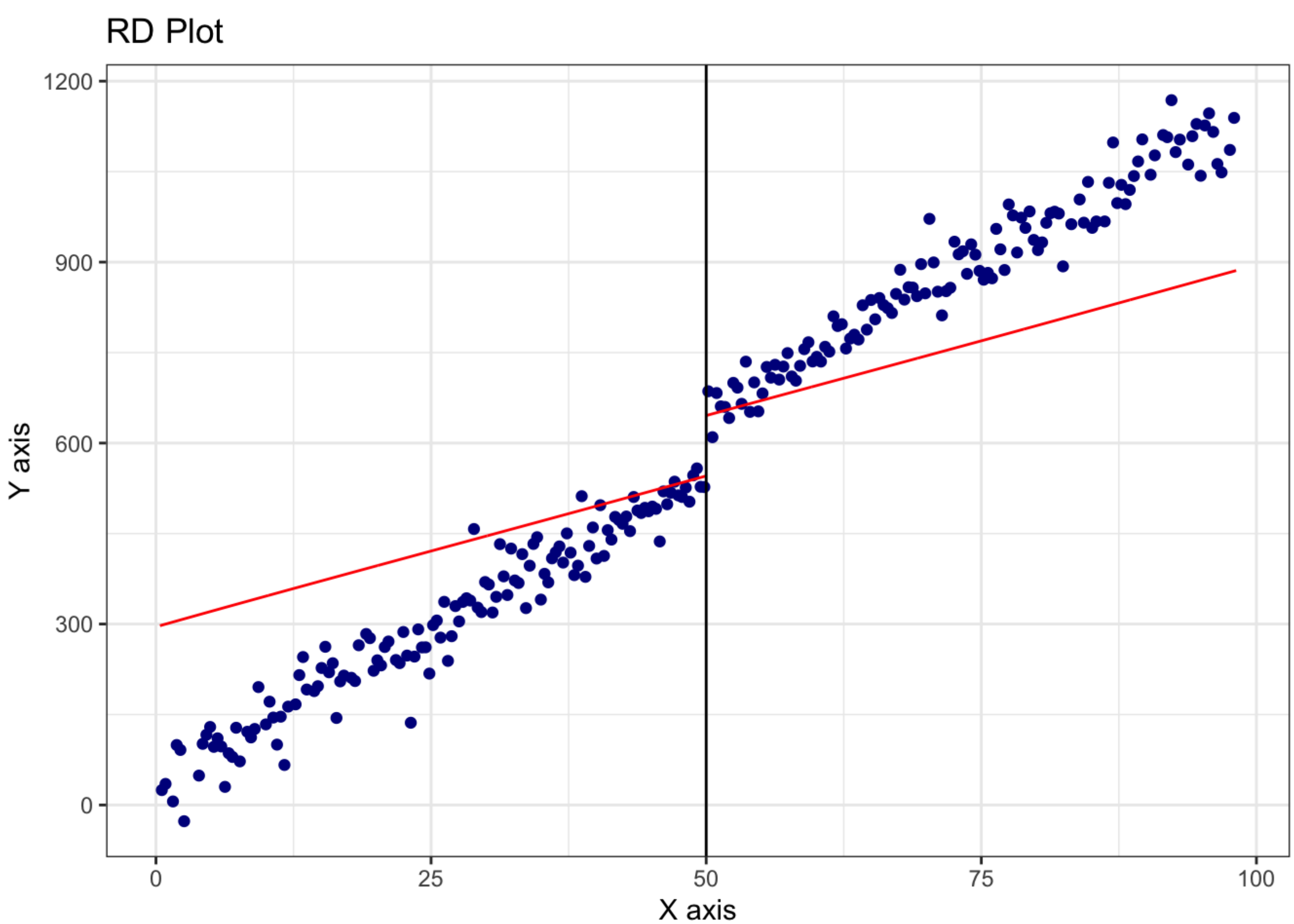
```r
rdplot(y=Y2,x=X,c=50)
```




```
m3<- rdrobust(y=Y2,#Outcome
          x=X,#Running Variable
          c=50,# by default#c=0
          covs=C#covariates
          #Kernel=triangular,epanechnikov, and uniform
          #binselect=mserd,
          )
summary(m3)
```

```
## Covariate-adjusted Sharp RD estimates using local polynomial regression.
##
## Number of Obs.                  939
## BW type                       mserd
## Kernel                   Triangular
## VCE method                       NN
##
## Number of Obs.                  463          476
## Eff. Number of Obs.             173          205
## Order est. (p)                    1            1
## Order bias  (q)                   2            2
## BW est. (h)                  13.060       13.060
## BW bias (b)                  21.431       21.431
## rho (h/b)                     0.609        0.609
## Unique Obs.                     463          476
##
## ================================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## ================================================================================
##   Conventional   100.674     1.182    85.167     0.000    [98.357 , 102.991]
##         Robust         -         -    72.320     0.000    [98.260 , 103.735]
## ================================================================================
```

```
rdplot(y=Y2,x=X,c=50,covs = C)
```



Replication

David Szakonyi, Private Sector Policy Making: Business Background and Politicians' Behavior in Office (https://www.journals.uchicago.edu/doi/10.1086/709297), JOP 2020

- Research Question: Do businessperson politicians actually govern differently?
- Empirical evidence: mayor elections and outcomes in Russia
- Argument: Pro-business policies, government efficiency
- Research Design: RDD on margin of victory

Replication

- We start with a simple model without any setting

```
library(rdrobust)
library(stargazer)
load("vrn_b.RData")
m_test<-with(vrn_b, rdrobust(y=competitive_construction,
                         x=businesswinmargin))
summary(m_test)
```

```
## Sharp RD estimates using local polynomial regression.
##
## Number of Obs.                 1662
## BW type                       mserd
## Kernel                   Triangular
## VCE method                       NN
##
## Number of Obs.                 1299          363
## Eff. Number of Obs.             478          188
## Order est. (p)                    1            1
## Order bias  (q)                   2            2
## BW est. (h)                   0.284        0.284
## BW bias (b)                   0.460        0.460
## rho (h/b)                     0.616        0.616
## Unique Obs.                    1295          363
##
## =============================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## =============================================================================
##   Conventional    -0.185     0.063    -2.941     0.003    [-0.309 , -0.062]
##         Robust         -         -    -2.714     0.007    [-0.341 , -0.055]
## =============================================================================
```

- We then replicate the results and see what were set in the specification
    - BW=.05
    - Kernel='uni'
    - p=1

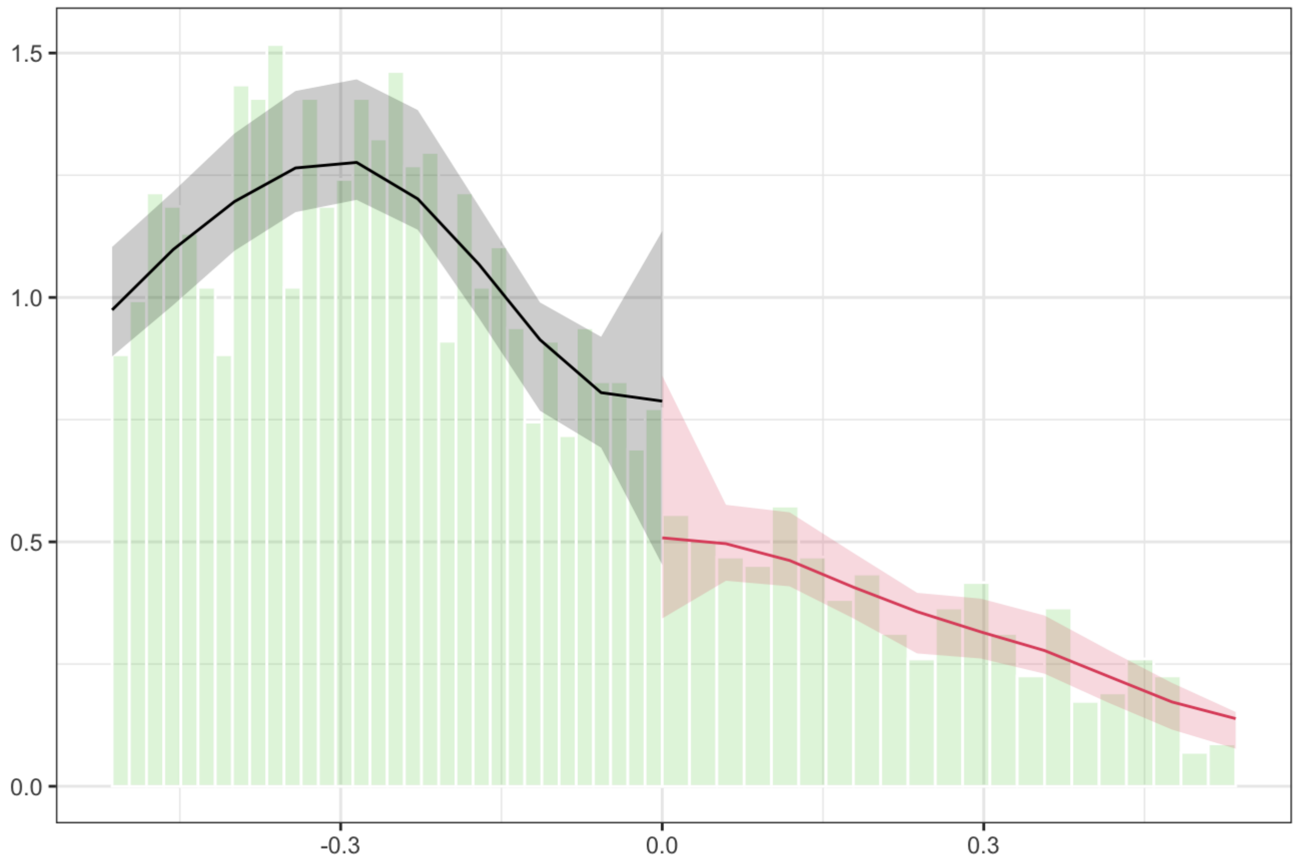
```
m_published<-with(vrn_b, rdrobust(y=competitive_construction,
                         x=businesswinmargin,
                         bwselect="mserd",
                         covs=cbind(factor(unit_type),totalexpend_log_year0),
                         all=TRUE, kernel='uni',p=1,cluster=region,h=.05))
summary(m_published)
```

```
## Covariate-adjusted Sharp RD estimates using local polynomial regression.
##
## Number of Obs.                 1662
## BW type                      Manual
## Kernel                      Uniform
## VCE method                       NN
##
## Number of Obs.                 1299          363
## Eff. Number of Obs.              58           43
## Order est. (p)                    1            1
## Order bias  (q)                   2            2
## BW est. (h)                   0.050        0.050
## BW bias (b)                   0.050        0.050
## rho (h/b)                     1.000        1.000
## Unique Obs.                    1299          363
##
## =============================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## =============================================================================
##   Conventional    -0.323     0.104    -3.096     0.002    [-0.528 , -0.119]
## Bias-Corrected    -0.356     0.104    -3.408     0.001    [-0.561 , -0.151]
##         Robust    -0.356     0.197    -1.808     0.071    [-0.742 , 0.030]
## =============================================================================
```

Test of Manipulation

- Is there any self-sorting behavior?

```
#Density Test
library(rddensity)
rdd<- rddensity(X = vrn_b$businesswinmargin)
rddplot<- rdplotdensity(rdd,X = vrn_b$businesswinmargin)
```



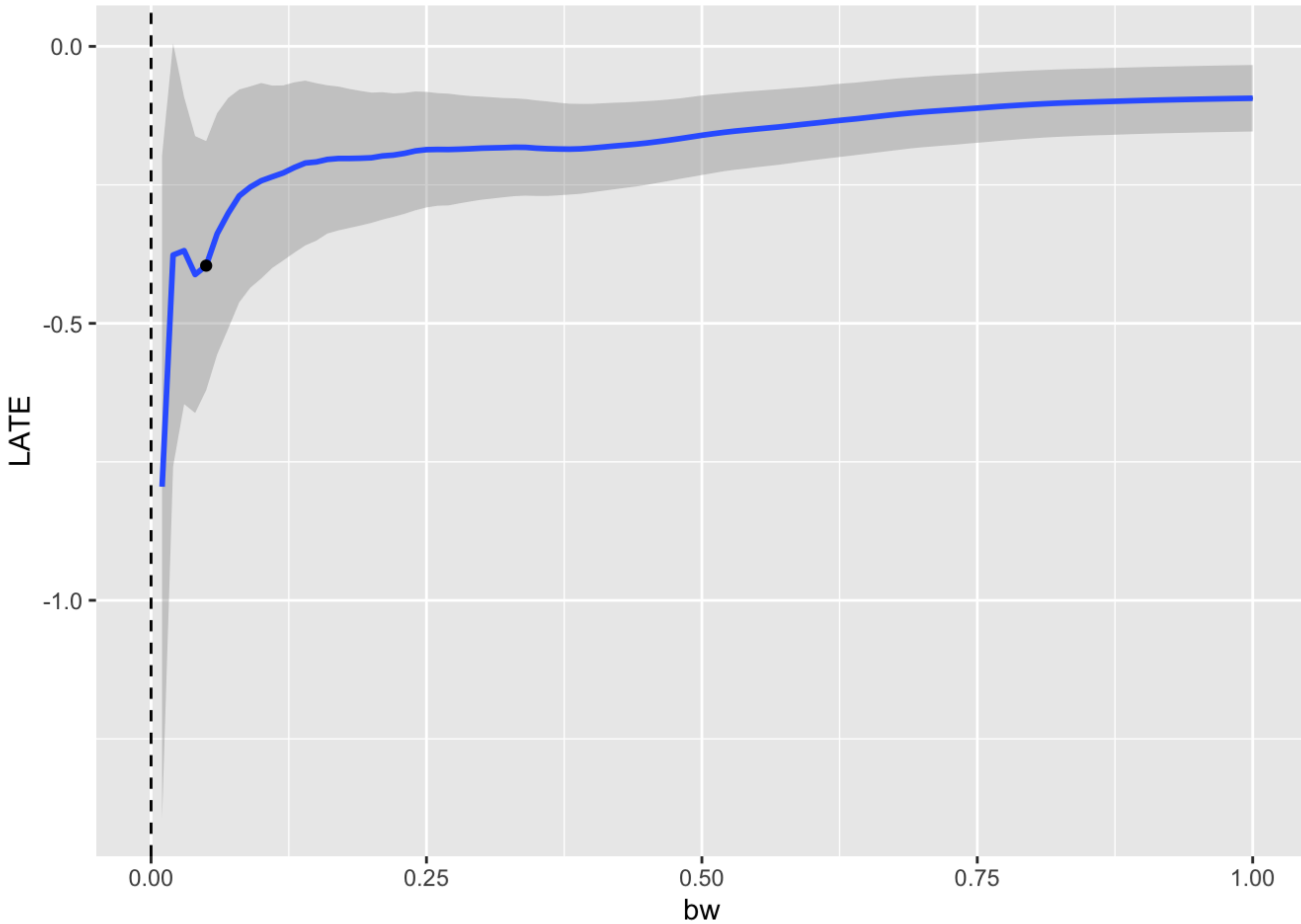

```
summary(rdd)
```

```
##
## Manipulation testing using local polynomial density estimation.
##
## Number of obs =       2260
## Model =               unrestricted
## Kernel =              triangular
## BW method =           estimated
## VCE method =          jackknife
##
## c = 0                 Left of c           Right of c
## Number of obs         1751                509
## Eff. Number of obs    330                 198
## Order est. (p)        2                   2
## Order bias (q)        3                   3
## BW est. (h)           0.171               0.178
##
## Method                T                   P > |T|
## Robust                -0.9534             0.3404
##
##
## P-values of binomial tests (H0: p=0.5).
##
## Window Length / 2          <c     >=c    P>|T|
## 0.014                      22      20    0.8776
## 0.028                      46      41    0.6683
## 0.042                      68      56    0.3232
## 0.056                      97      69    0.0358
## 0.070                     124      83    0.0053
## 0.084                     154      97    0.0004
## 0.098                     178     112    0.0001
## 0.112                     205     130    0.0000
## 0.126                     228     147    0.0000
## 0.141                     257     166    0.0000
```

Test of Sensitivity (BW Select)

- Is the effect unique to the bandwidth choice (h=.05)?

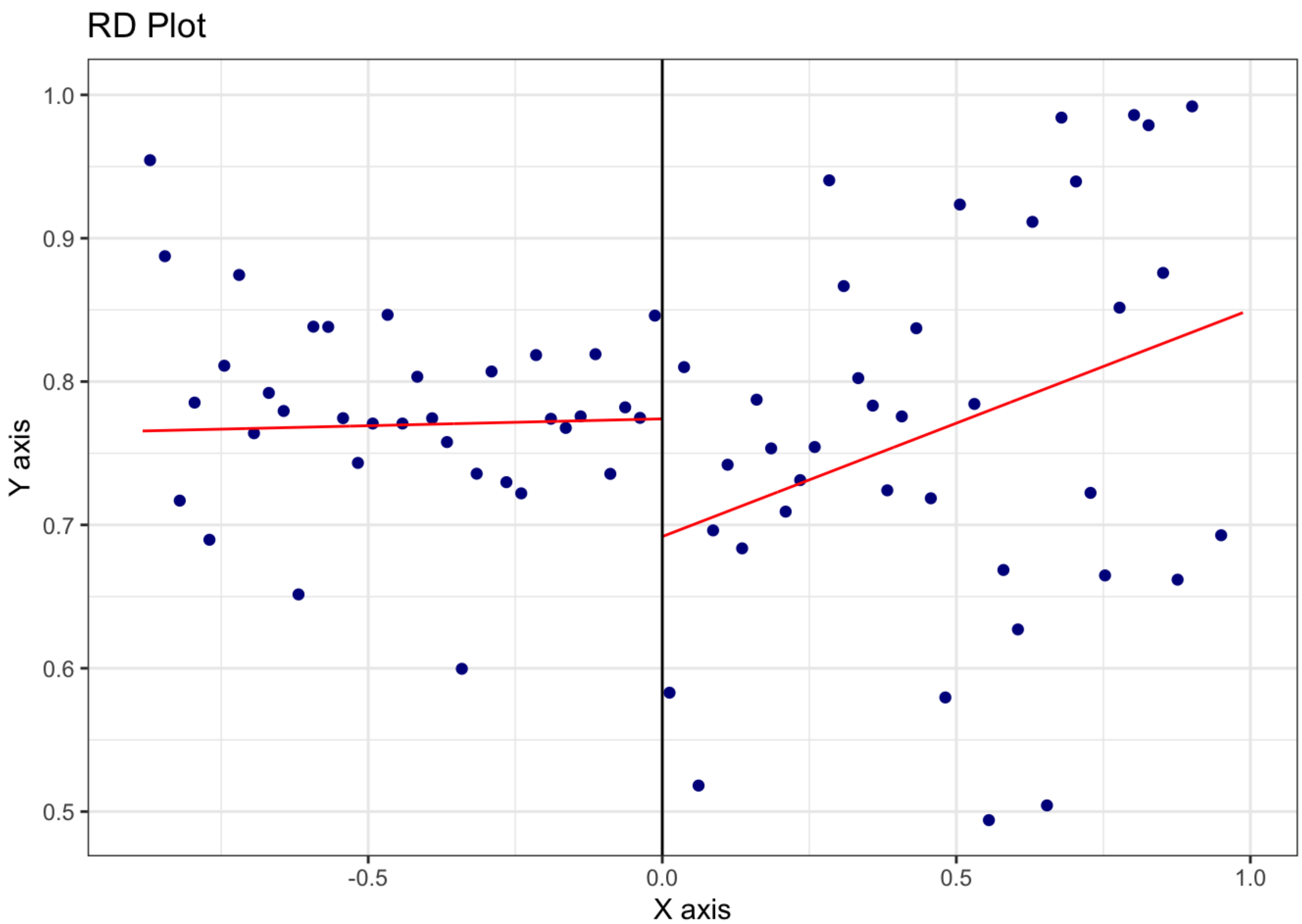
```
library(rddtools)
dat2 <- rdd_data(y = vrn_b$competitive_construction,x = vrn_b$businesswinmargin,cutpoint
= 0)
m1<- rdd_reg_np(rdd_object=dat2,bw=0.05)
plotSensi(m1, from = 0.01, to = 1, by = 0.01)
```



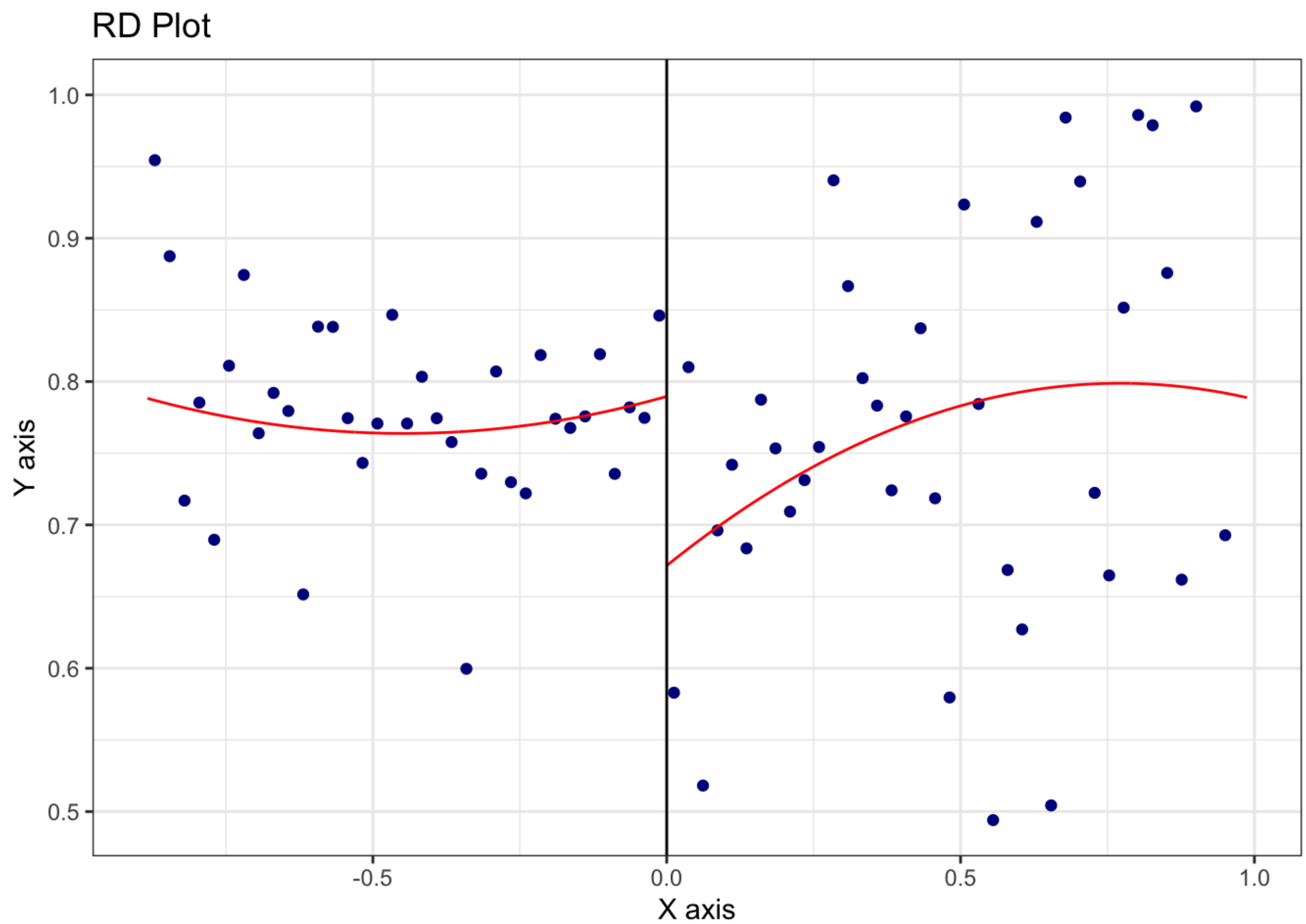
Test of Sensitivity (Order)

- Does the setting of order (p1) matter

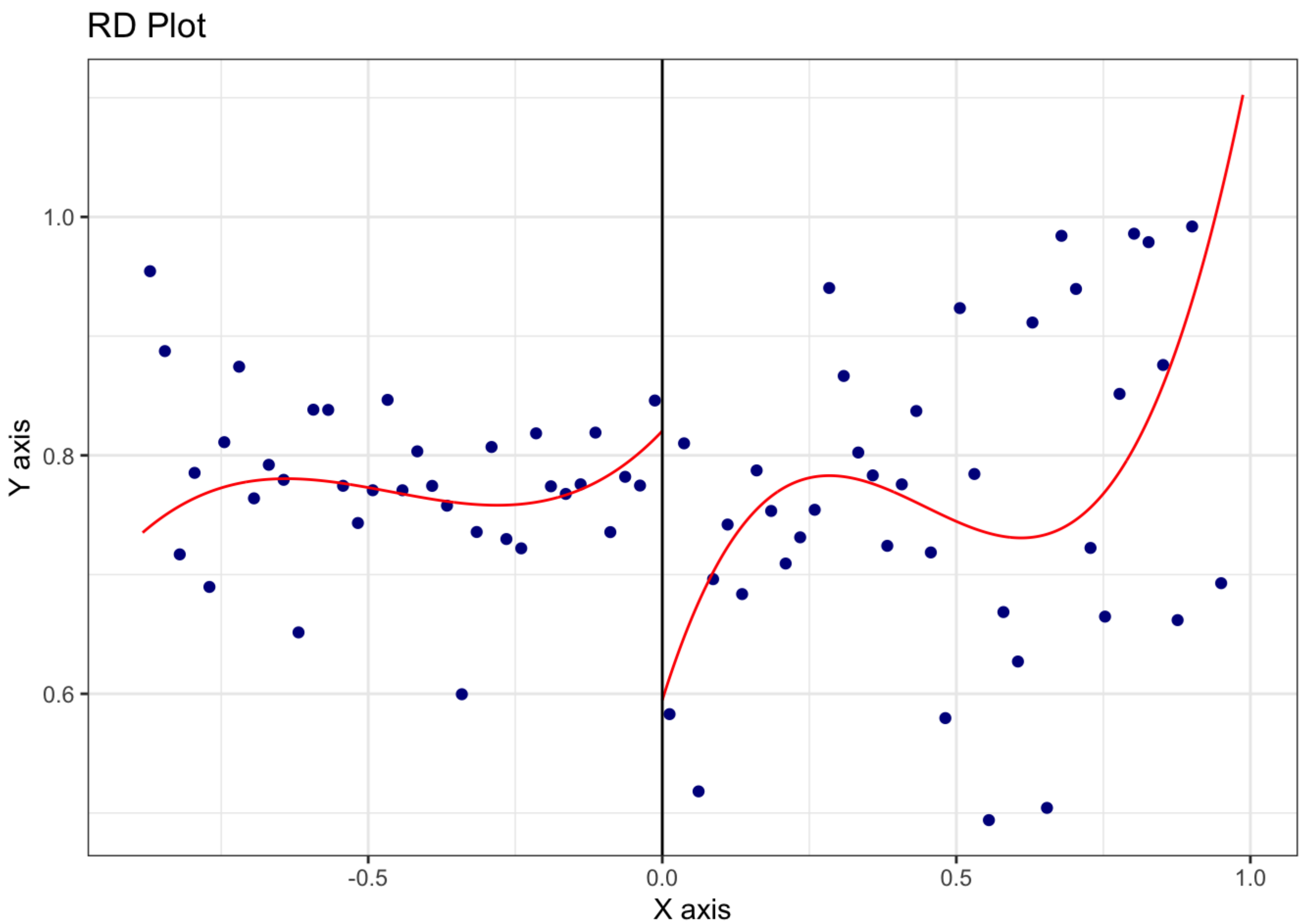
```
attach(vrn_b)
rdplot(y=competitive_construction,x=businesswinmargin,
       covs=cbind(factor(unit_type),totalexpend_log_year0),
       kernel='uni',p=1)
```




```
rdplot(y=competitive_construction,x=businesswinmargin,
       covs=cbind(factor(unit_type),totalexpend_log_year0),
       kernel='uni',p=2)
```




```
rdplot(y=competitive_construction,x=businesswinmargin,
       covs=cbind(factor(unit_type),totalexpend_log_year0),
       kernel='uni',p=3)
```




```
m_test1<-with(vrn_b, rdrobust(y=competitive_construction,
                              x=businesswinmargin,
                              bwselect="mserd",
                              covs=cbind(factor(unit_type),totalexpend_log_year0),
                              all=TRUE, kernel='uni',p=1,cluster=region,h=.05))

m_test2<-with(vrn_b, rdrobust(y=competitive_construction,
                              x=businesswinmargin,
                              bwselect="mserd",
                              covs=cbind(factor(unit_type),totalexpend_log_year0),
                              all=TRUE, kernel='uni',p=2,cluster=region,h=.05))
m_test3<-with(vrn_b, rdrobust(y=competitive_construction,
                              x=businesswinmargin,
                              bwselect="mserd",
                              covs=cbind(factor(unit_type),totalexpend_log_year0),
                              all=TRUE, kernel='uni',p=3,cluster=region,h=.05))

summary(m_test1)
```

```
## Covariate-adjusted Sharp RD estimates using local polynomial regression.
##
## Number of Obs.                 1662
## BW type                      Manual
## Kernel                      Uniform
## VCE method                       NN
##
## Number of Obs.                 1299          363
## Eff. Number of Obs.              58           43
## Order est. (p)                    1            1
## Order bias  (q)                   2            2
## BW est. (h)                   0.050        0.050
## BW bias (b)                   0.050        0.050
## rho (h/b)                     1.000        1.000
## Unique Obs.                    1299          363
##
## =============================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## =============================================================================
##   Conventional    -0.323     0.104    -3.096     0.002    [-0.528 , -0.119]
## Bias-Corrected    -0.356     0.104    -3.408     0.001    [-0.561 , -0.151]
##         Robust    -0.356     0.197    -1.808     0.071    [-0.742 , 0.030]
## =============================================================================
```

```
summary(m_test2)
```

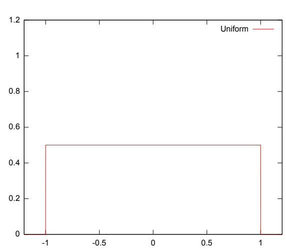
```
## Covariate-adjusted Sharp RD estimates using local polynomial regression.
##
## Number of Obs.                 1662
## BW type                      Manual
## Kernel                      Uniform
## VCE method                       NN
##
## Number of Obs.                 1299          363
## Eff. Number of Obs.              58           43
## Order est. (p)                    2            2
## Order bias  (q)                   3            3
## BW est. (h)                   0.050        0.050
## BW bias (b)                   0.050        0.050
## rho (h/b)                     1.000        1.000
## Unique Obs.                    1299          363
##
## =============================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## =============================================================================
##   Conventional    -0.357     0.200    -1.789     0.074    [-0.748 , 0.034]
## Bias-Corrected    -0.313     0.200    -1.570     0.116    [-0.705 , 0.078]
##         Robust    -0.313     0.306    -1.024     0.306    [-0.913 , 0.286]
## =============================================================================
```

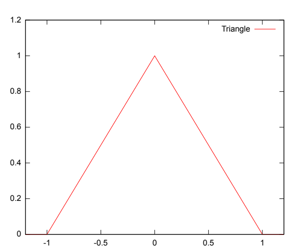
```r
summary(m_test3)
```

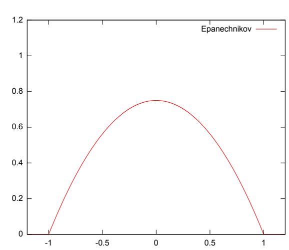
```
## Covariate-adjusted Sharp RD estimates using local polynomial regression.
##
## Number of Obs.                 1662
## BW type                      Manual
## Kernel                      Uniform
## VCE method                       NN
##
## Number of Obs.                 1299          363
## Eff. Number of Obs.              58           43
## Order est. (p)                    3            3
## Order bias  (q)                   4            4
## BW est. (h)                   0.050        0.050
## BW bias (b)                   0.050        0.050
## rho (h/b)                     1.000        1.000
## Unique Obs.                    1299          363
##
## ================================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## ================================================================================
##   Conventional    -0.299     0.307    -0.972     0.331    [-0.901 , 0.304]
## Bias-Corrected    -0.045     0.307    -0.147     0.883    [-0.647 , 0.557]
##         Robust    -0.045     0.425    -0.107     0.915    [-0.877 , 0.787]
## ================================================================================
```

Test of Sensitivity (Kernel)

- How to weight these obs within bandwidth?







```
## Sharp RD estimates using local polynomial regression.
##
## Number of Obs.                 1662
## BW type                      Manual
## Kernel                      Uniform
## VCE method                       NN
##
## Number of Obs.                 1299          363
## Eff. Number of Obs.              58           43
## Order est. (p)                    1            1
## Order bias  (q)                   2            2
## BW est. (h)                   0.050        0.050
## BW bias (b)                   0.050        0.050
## rho (h/b)                     1.000        1.000
## Unique Obs.                    1299          363
##
## ================================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## ================================================================================
##   Conventional    -0.349     0.134    -2.615     0.009    [-0.611 , -0.088]
##         Robust         -         -    -2.167     0.030    [-0.864 , -0.043]
## ================================================================================
```

```
## Sharp RD estimates using local polynomial regression.
##
## Number of Obs.                 1662
## BW type                      Manual
## Kernel                   Triangular
## VCE method                       NN
##
## Number of Obs.                 1299          363
## Eff. Number of Obs.              58           43
## Order est. (p)                    1            1
## Order bias  (q)                   2            2
## BW est. (h)                   0.050        0.050
## BW bias (b)                   0.050        0.050
## rho (h/b)                     1.000        1.000
## Unique Obs.                    1299          363
##
## ================================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## ================================================================================
##   Conventional    -0.396     0.153    -2.582     0.010    [-0.696 , -0.095]
##         Robust         -         -    -1.841     0.066    [-0.863 , 0.027]
## ================================================================================
```

```
## Sharp RD estimates using local polynomial regression.
##
## Number of Obs.                 1662
## BW type                      Manual
## Kernel                   Epanechnikov
## VCE method                       NN
##
## Number of Obs.                 1299          363
## Eff. Number of Obs.              58           43
## Order est. (p)                    1            1
## Order bias  (q)                   2            2
## BW est. (h)                   0.050        0.050
## BW bias (b)                   0.050        0.050
## rho (h/b)                     1.000        1.000
## Unique Obs.                    1299          363
##
## ================================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## ================================================================================
##   Conventional    -0.389     0.148    -2.621     0.009    [-0.680 , -0.098]
##         Robust         -         -    -1.944     0.052    [-0.864 , 0.004]
## ================================================================================
```