Regression Discontinuity Design

School of Economics, QAU, Islamabad

Zahid Asghar

12/6/22

Outline

- Introduction to Regression Discontinuity Designs
- Sharp and Fuzzy Regression Discontinuity Designs
- Simulated Sharp and Fuzzy Regression Discontinuity Designs using R
- Encourage the use of this technique and R

Regression Discontinuity Designs (RDD)

When randomization is not feasible, how can we exploit implementation features of the program to measure its impact?

- Answer: Quasi-experiments - Example: Regression Discontinuity Design.

Regression discontinuity design (RDD) is an important method that can be used for programs that have a continuous eligibility index with a clearly defined eligibility threshold (cuttoff score) to determine who is eligible and who is not.

- The regression discontinuity design is a statistical approach that utilizes threshold based decision making to estimate causal estimates of different interventions.

- Regression discontinuity is relatively simple to implement, transparent, and provides “real world” effects of treatments and policies.
- Graphical presentation and statistical analysis of data.
- Developed by Psychologists but later on in much use with labor economists
- Despite frequent use threshold based decision making, regression discontinuity is underutilized.

Discontinuity Design

Many social programs select beneficiaries using an index or score

- **Anti-poverty Programs**
- **Pensions**
- **Education**
- **Agriculture**

- Targeted to households below a given poverty index/income
- Targeted to population above a certain age
- Scholarships targeted to students with high scores on standarized text
- Fertilizer program targeted to small farms less than given number of hectares)

Example: Effect of fertilizer program on agriculture production

Goal

Improve school attendance for poor students

Method

- **Households with a score (Pa) of assets $\leq$ are poor**
- **Households with a score (Pa) of assets $>$ are not poor**

Intervention

Poor households receive scholarships to send to children to school

Types of RDD

- **With regard to Treatment Assignment:**
  - **Sharp Design**: Where the cut-point perfectly predicts who does/doesn't receive intervention
    - Characteristics cant be perfectly manipulated by invidual
  - **Fuzzy Design**: Where there exists "cross-overs" (for various reasons)
- **With regard to Analytic Approach**:
  - **Non-parametric: local randomization approach**
  - Parametric: uses every observation in the sample

- Ideal for situations when there is a **cut-score** in which subjects above/below a threshold receive program resources or an intervention
- The variable used to determine the threshold is called the **“rating” or “forcing”** variable. Examples: GPA, household income
- Quasi-experimental design used to estimate the rigor and validity of a **Randomized Control Trial (RCT)**
- In many fields, regression discontinuity are considered to be one of the most rigorous quasi-experimental methods

Internal Validity

- **Imprecise Control**
  - No one single entity or factor should determine a rating score. For a RDD to yield an unbiased LATE estimate at the cuttoff, it is important that the eligibility index not be manipulated around the cutoff so that an individual can change treatment control status.
- Clear discontinuity for treatment status at the cut-point

The Scenario

- Ehsaas scholarship criteria for grant of support is Rs.45,000 income of parents *prior to viewing any actual data*. A threshold of Rs.45,000/ month income is established: students below this cut-point will receive scholarshio, those above will not.
- Besides assistance, one may want to determine if scholarship proves effective in relation to final grade.

Model

$$Y_i = \alpha + \beta_0 T_i + \beta_1 r_i + \epsilon$$

α is Estimated marginal mean ,β_0 is marginal impact of program, T_i is Treatment status (0,1), β_1 is Relationship between scholarship and Final Grade, r_i is Individual scores on the rating variable (Scholarship amount; centered)

Simulation Data

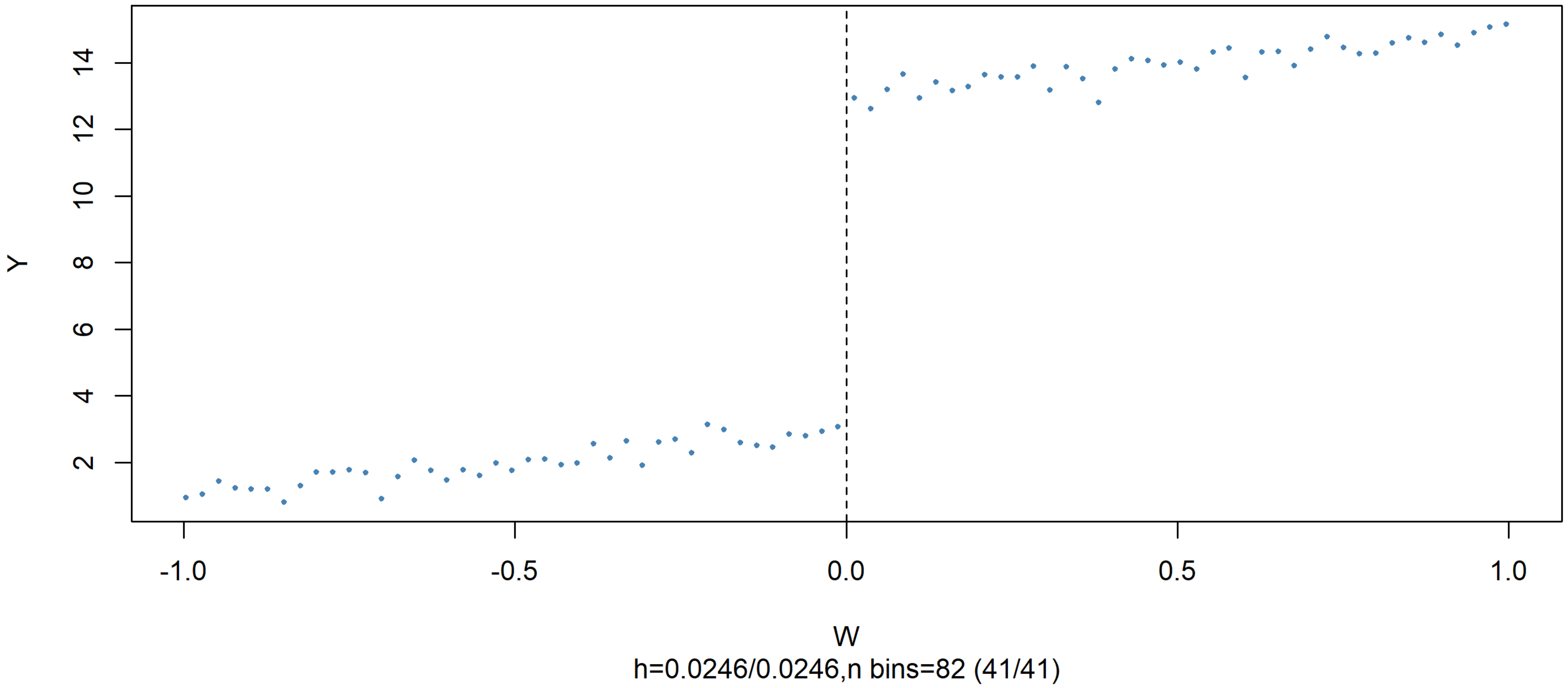
- Institutional Data:
  - Taken from supplemental instruction STEM report, N=589
  - Scores are adjusted on outcome for demonstration purposes

Sharp Designs

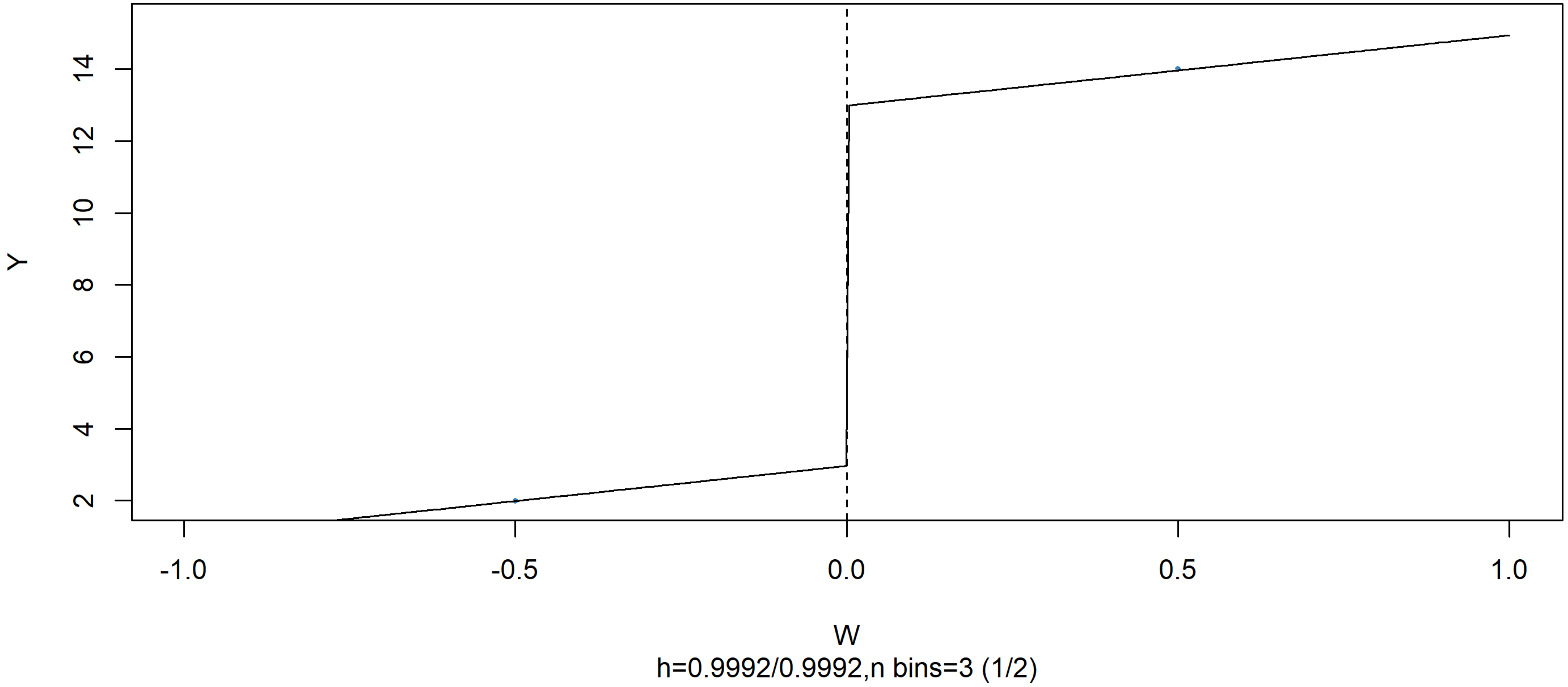
```r

# generate some sample data
W <- runif(1000, -1, 1)
y <- 3 + 2 * W + 10 * (W>=0) + rnorm(1000)
# load the package 'rddtools'
library(rddtools)

# construct rdd_data
data <- rdd_data(y, W, cutpoint = 0)
library(ggplot2)
ggplot(data)+aes(x=W,y=y)+geom_point()
```

```r
# estimate the sharp RDD model
rdd_mod <- rdd_reg_lm(rdd_object = data,
                      slope = "same")
summary(rdd_mod)
##
## Call:
## lm(formula = y ~ ., data = dat_step1, weights = weights)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.525 -0.664   0.001   0.654   3.150
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.964      0.071    41.7   <2e-16 ***
## D             10.024      0.130    77.0   <2e-16 ***
```

```r
# plot the sample data

plot(data,
     col = "steelblue",
     cex = 0.35,
     xlab = "W",
     ylab = "Y")
```



```r
# plot the RDD model along with binned observations
plot(rdd_mod,
     cex = 0.35,
     col = "steelblue",
     xlab = "W",
     ylab = "Y")
```



Fuzzy Designs

```r
library(MASS)
# generate sample data
mu <- c(0, 0)
sigma <- matrix(c(1, 0.7, 0.7, 1), ncol = 2)

set.seed(1234)
d <- as.data.frame(mvrnorm(2000, mu, sigma))
colnames(d) <- c("W", "Y")

# introduce fuzziness
d$treatProb <- ifelse(d$W < 0, 0, 0.8)

fuzz <- sapply(X = d$treatProb, FUN = function(x) rbinom(1, 1, prob = x))

# treatment effect
d$Y <- d$Y + fuzz * 2
```

```r
# generate a colored plot of treatment and control group
plot(d$W, d$Y,
     col = c("steelblue", "darkred")[factor(fuzz)],
     pch= 20,
     cex = 0.5,
     xlim = c(-3, 3),
     ylim = c(-3.5, 5),
     xlab = "W",
     ylab = "Y")
# add a dashed vertical line at cutoff
abline(v = 0, lty = 2)
```