Regression Discontinuity - Part 2

API 210 Section - Advanced Quantitative Methods II: Econometric Methods

Caterina Chiopris

Harvard Kennedy School, MPA/ID

April 15, 2022

More on Bandwidths

More on Bandwidths

- Why does the choice of bandwidth matter?
- Let's simulate some data

```
n =1000
set.seed(210)
epsilon <- rnorm(n, mean=0, sd =300)
x <- runif(n, min=-15, max=15)
d <- ifelse(x >0, 1,0)
y <- 100 + 2*x + 0.5*x^3+ 1000*d + epsilon
data <- as.data.frame(cbind(y,x,d,epsilon))
```

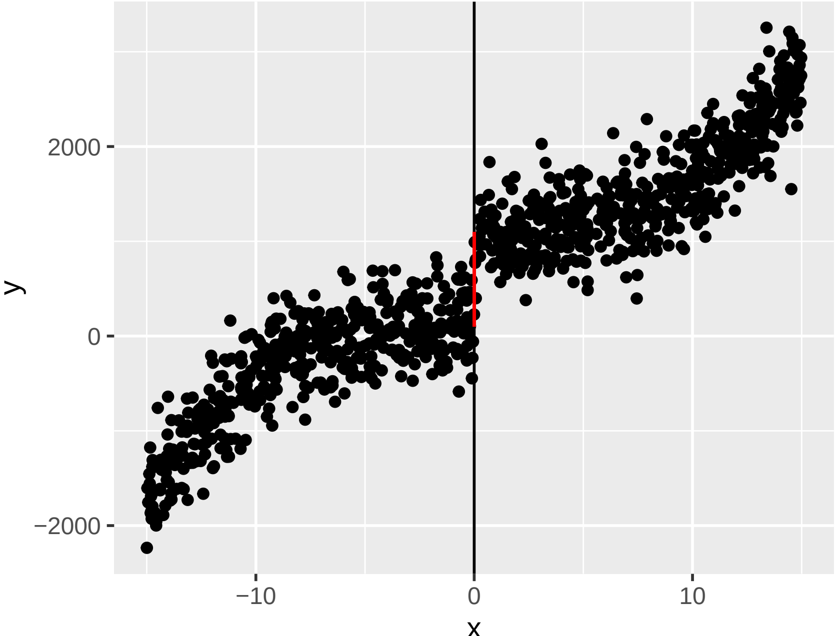
More on Bandwidths

- Why does the choice of bandwidth matter?
- Let's simulate some data

```
n =1000
set.seed(210)
epsilon <- rnorm(n, mean=0, sd =300)
x <- runif(n, min=-15, max=15)
d <- ifelse(x >0, 1,0)
y <- 100 + 2*x + 0.5*x^3+ 1000*d + epsilon
data <- as.data.frame(cbind(y,x,d,epsilon))
```

What 'number' are we trying to estimate here?

Plot



Binning

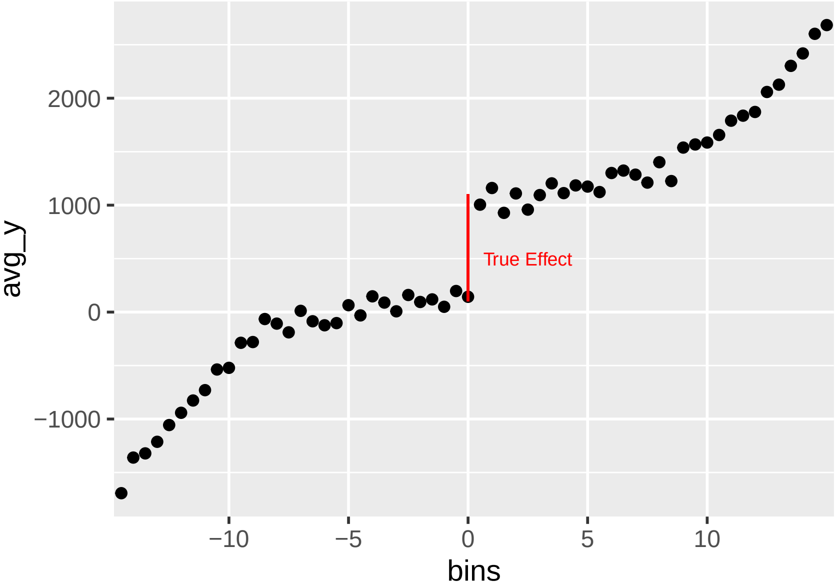
- Define bins, say: 1 unit per bin
- Find average y for each bin

Binning

- Define bins, say: 1 unit per bin
- Find average y for each bin

```
data$bins <- cut(data$x, breaks = seq(-15,15, by=0.5),
                 labels = seq(-14.5,15,by=0.5))
binned_data <- data %>% group_by(bins) %>%
  summarize(avg_y = mean(y))
```

Binning



Regression

```
library(rdd)
reg1 <- RDestimate(y ~ x, kernel = "rectangular")
```

Regression

```
library(rdd)
reg1 <- RDestimate(y ~ x, kernel = "rectangular")
```

Question: Consider the optimal Bandwidth, half the optimal bandwidth, and double the optimal bandwidth. Which one of the following will be the most biased?

- The optimal Bandwidth
- Half the optimal bandwidth
- Double the optimal bandwidth

Regression

```r
library(rdd)
reg1 <- RDestimate(y ~ x, kernel = "rectangular")
```

Question: Consider the optimal Bandwidth, half the optimal bandwidth, and double the optimal bandwidth. Which one of the following will be the most biased?

- The optimal Bandwidth
- Half the optimal bandwidth
- Double the optimal bandwidth

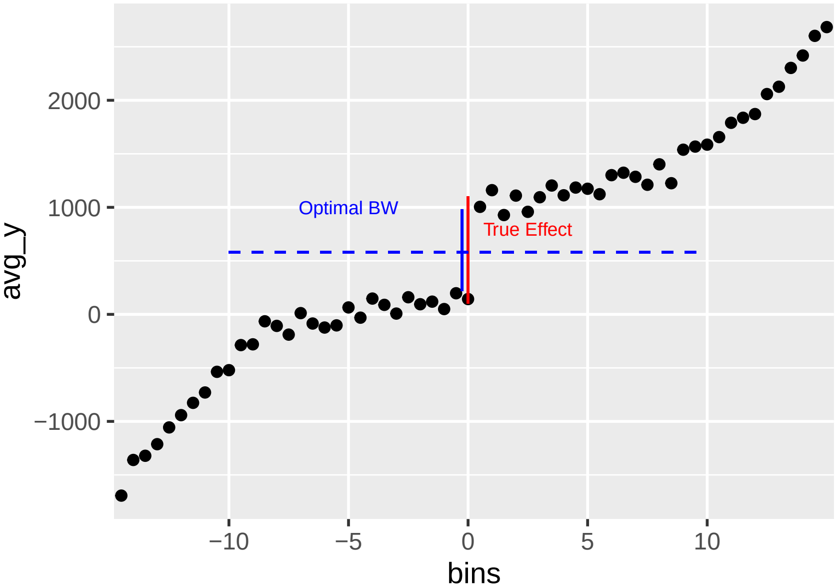
Which one of the following will be the most precise?

- The optimal Bandwidth
- Half the optimal bandwidth
- Double the optimal bandwidth

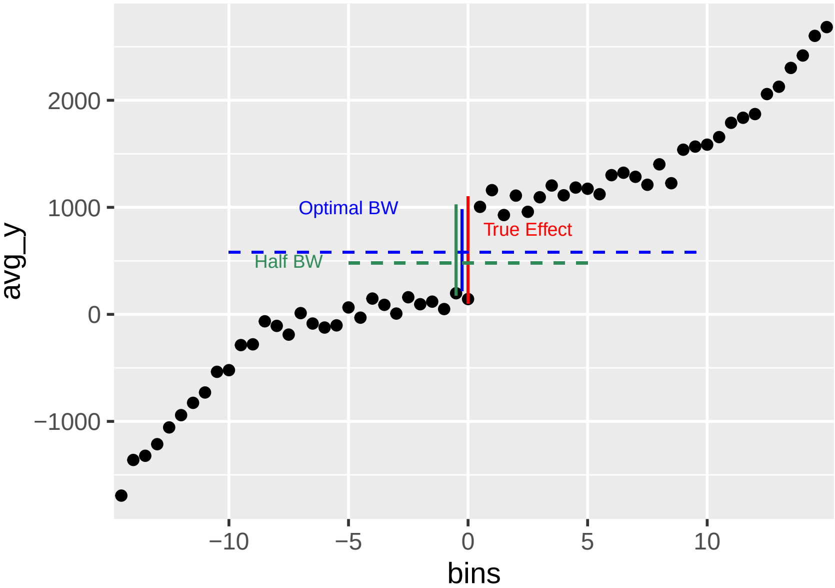
Regression

```
reg <- rbind(reg1$est, reg1$se, reg1$obs, reg1$bw)
rownames(reg) <- c("Estimate", "St.Err", "Obs", "Bandwidth")
round(reg,2)
```

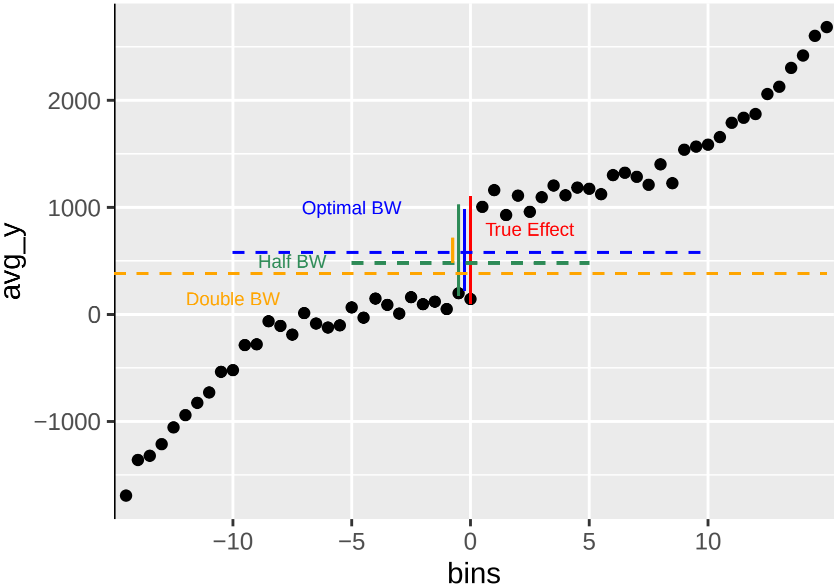
```
##              LATE Half-BW Double-BW
## Estimate   765.75  852.42    233.63
## St.Err      46.62   66.16     44.51
## Obs        628.00  325.00   1000.00
## Bandwidth    9.56    4.78     19.13
```



Note: these are estimates, *not* confidence intervals!

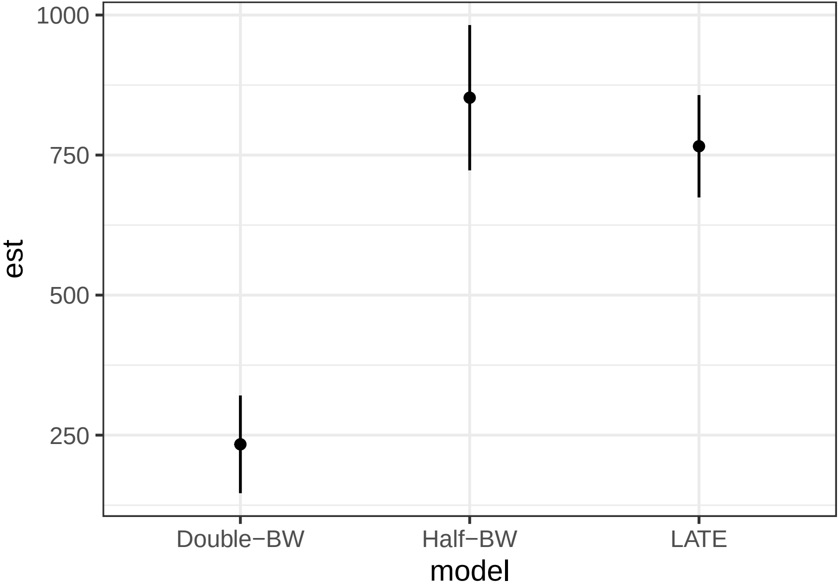


Note: these are estimates, *not* confidence intervals!



Note: these are estimates, *not* confidence intervals!

Visualization - Confidence Intervals



Comprehension Check

Question: What type of RD is this?

- Parametric
- Non-parametric
- Sharp
- Fuzzy

Comprehension Check

Question: What type of RD is this?

- Parametric
- Non-parametric
- Sharp
- Fuzzy

Question: How would you implement a parametric RD?

Comprehension Check

Question: What type of RD is this?

- Parametric
- Non-parametric
- Sharp
- Fuzzy

Question: How would you implement a parametric RD?

```
lm(y ~ x + I(x^2) + I(x^3) + d)
```

Parametric RD

```
par <- lm(y ~ x + I(x^3) + d)
```

term	estimate	std.error	statistic	p.value
(Intercept)	125.3199651	27.0440667	4.633917	0.0000041
x	6.6161721	5.5053546	1.201770	0.2297382
I(x^3)	0.5023396	0.0244772	20.522737	0.0000000
d	927.4180142	50.8351858	18.243624	0.0000000

Parametric RD

Robust to choice of polynomial?

```
par2 <- lm(y ~ x + I(x^2) + I(x^3) + I(x^4) + I(x^5)+ d)
```

term	estimate	std.error	statistic	p.value
(Intercept)	139.6856807	35.0039339	3.9905709	0.0000708
x	11.8432483	9.7262018	1.2176643	0.2236408
I(x^2)	0.4569200	0.4926238	0.9275233	0.3538803
I(x^3)	0.4291669	0.1115573	3.8470539	0.0001272
I(x^4)	-0.0035929	0.0024105	-1.4905035	0.1364095
I(x^5)	0.0002720	0.0003851	0.7063079	0.4801624
d	904.8431569	61.4815955	14.7173012	0.0000000

Fuzzy RD

Overview of RD

- Main assumption: continuity of potential outcomes
  - Institutional knowledge: no manipulation of running variable
  - Balance of pre-treatment covariates
  - No jumps in density of X_i around cutoff
- Parametric RD
  - Choice of polynomial
- Non-parametric RD
  - Kernel and bandwidth choice
- Sharp RD
  - Treatment status changes deterministically around the threshold
- Fuzzy RD
  - Likelihood of getting treated changes around the threshold
  - Same assumptions as IV

Review of Fuzzy RD

- Sometimes treatment assignment does not change perfectly around the threshold
- However, the probability of getting treated can jump at the threshold
  - Can think of such scenario as imperfect compliance, with the threshold as a predictor of treatment assignment
- Why is this similar to IV?
  - Instrument affects treatment status - although it does not define it deterministically
  - Discontinuity in fuzzy RD affects treatment status - but knowing which side of the threshold you're on does not tell you *for certain* if you're treated or not

LATE and ITT

- Similarly to IV, a fuzzy RD estimates the Local Average Treatment Effect
- Two ways in which the effect is 'local':

LATE and ITT

- Similarly to IV, a fuzzy RD estimates the Local Average Treatment Effect
- Two ways in which the effect is 'local':
  1. For the compliers (who are the compliers?)
  2. In the neighborhood of the cutoff x_0
- Remember in RCT setting, with imperfect compliance, we could estimate ITT
  - 'Reduced form': regress outcome on treatment offer
- Can you use the same principle in a fuzzy RD?

LATE and ITT

- Similarly to IV, a fuzzy RD estimates the Local Average Treatment Effect
- Two ways in which the effect is 'local':
  1. For the compliers (who are the compliers?)
  2. In the neighborhood of the cutoff x_0
- Remember in RCT setting, with imperfect compliance, we could estimate ITT
  - 'Reduced form': regress outcome on treatment offer
- Can you use the same principle in a fuzzy RD?
  - Yes! Regress outcome on discontinuity - dummy for being above cutoff
  - Diluted effect compared to LATE
  - LATE: 'scaled' ITT by compliance

Example: Crost, Felter and Johnston (2014)

Reminder:

- Research question: Do development programs increase the likelihood of violent conflict (or deaths from conflict, etc.)?
- Findings: There was an increase in violent conflict and associated conflict casualties in the regions of the Philippines where the KALAHI-CIDSS community-driven development program.
- One likely reason for this increase in conflict is that insurgents in the Philippines, worried that the program would weaken their political support in the population, may have attempted to sabotage the early stages of the program.

Sharp RD

- Remember last time we estimated the sharp RD, with estimates:

phase	bandwidth	Estimate	SE
Entire program period	4	0.127	(0.046)
Entire program period	6	0.090	(0.042)
Entire program period	8	0.077	(0.038)

- Baseline estimate: bw of 6

ITT vs. LATE

- In this setting, we dot not have perfect compliance
  - Not all municipalities who were eligible for the program participated
- What effect did we estimate with the sharp RD?

ITT vs. LATE

- In this setting, we dot not have perfect compliance
  - Not all municipalities who were eligible for the program participated
- What effect did we estimate with the sharp RD?
  - ITT, where the actual effect of participation is diluted by non-participation
- How would you estimate the LATE?

ITT vs. LATE

- In this setting, we dot not have perfect compliance
  - Not all municipalities who were eligible for the program participated
- What effect did we estimate with the sharp RD?
  - ITT, where the actual effect of participation is diluted by non-participation
- How would you estimate the LATE?
  - We can use a fuzzy RD design

ITT vs. LATE

- In this setting, we dot not have perfect compliance
  - Not all municipalities who were eligible for the program participated
- What effect did we estimate with the sharp RD?
  - ITT, where the actual effect of participation is diluted by non-participation
- How would you estimate the LATE?
  - We can use a fuzzy RD design
  - Use eligibility as instrument for participation
  - Scale the effect of eligibility by take-up

Implementing Fuzzy RD

```r
ff_fuzzy <- totcas ~ rank_norm + el_rank | province + year |
  (kalahi ~ eligible) | mcode
frd <- felm(formula = ff_fuzzy, data=crost)
summary(frd)
```

Implementing Fuzzy RD

```
ff_fuzzy <- totcas ~ rank_norm + el_rank | province + year |
  (kalahi ~ eligible) | mcode
frd <- felm(formula = ff_fuzzy, data=crost)
summary(frd)
```

phase	Estimate	SE
Entire data	0.125	(0.083)

Interpretation

- Compare the sharp and fuzzy RD estimates. Do the magnitudes make intuitive sense?

Interpretation

- Compare the sharp and fuzzy RD estimates. Do the magnitudes make intuitive sense?
  - The fuzzy RD estimates are overall larger in magnitude than the effects of eligibility
  - This is intuitive because not all eligible municipalities participated, whereas no ineligible municipalities could have participated
  - We can think of the main RD results as an ITT, where the actual effect of participation is diluted by non-participation.
  - Note that the authors point out that these estimates of LATE may still be biased

Robustness Checks

Robustness Checks

1. How do you interpret the following table?

	Local linear regressions with bandwidth:				
	2	3	4	5	6
Panel A: Entire program period					
Poisson QMLE	0.176***	0.131***	0.115***	0.099***	0.091***
	(0.049)	(0.038)	(0.034)	(0.032)	(0.031)
OLS	0.221***	0.175***	0.127***	0.103**	0.090**
	(0.059)	(0.051)	(0.046)	(0.044)	(0.042)
Control Mean	0.090	0.083	0.077	0.075	0.076
	(0.016)	(0.012)	(0.010)	(0.009)	(0.008)

Robustness Checks

1. How do you interpret the following table?

	Local linear regressions with bandwidth:				
	2	3	4	5	6
Panel A: Entire program period					
Poisson QMLE	0.176***	0.131***	0.115***	0.099***	0.091***
	(0.049)	(0.038)	(0.034)	(0.032)	(0.031)
OLS	0.221***	0.175***	0.127***	0.103**	0.090**
	(0.059)	(0.051)	(0.046)	(0.044)	(0.042)
Control Mean	0.090	0.083	0.077	0.075	0.076
	(0.016)	(0.012)	(0.010)	(0.009)	(0.008)

The main claims are not sensitive to the particular choice of the bandwidth.

Robustness Checks

2. Why is this useful?

Dependent variable: Total casualties (entire program period)				
	Poisson QMLE		OLS	
	Local Linear	Quadratic	Local Linear	Quadratic
	(1)	(2)	(3)	(4)
Pseudo-threshold at -2 (median of eligible municipalities)				
Eligible	0.029	-0.010	0.038	0.011
	(0.041)	(0.059)	(0.037)	(0.054)
Pseudo-threshold at 3 (median of ineligible municipalities)				
Eligible	-0.052	-0.022	-0.010	0.011
	(0.058)	(0.091)	(0.096)	(0.12)
Municipalities	222	222	222	222

Robustness Checks

2. Why is this useful?

Dependent variable: Total casualties (entire program period)				
	Poisson QMLE		OLS	
	Local Linear	Quadratic	Local Linear	Quadratic
	(1)	(2)	(3)	(4)
Pseudo-threshold at -2 (median of eligible municipalities)				
Eligible	0.029	-0.010	0.038	0.011
	(0.041)	(0.059)	(0.037)	(0.054)
Pseudo-threshold at 3 (median of ineligible municipalities)				
Eligible	-0.052	-0.022	-0.010	0.011
	(0.058)	(0.091)	(0.096)	(0.12)
Municipalities	222	222	222	222

To show that it is really the actual threshold, and not some linear trend of the ranking, that produces the main effect.

Robustness Checks - One Implementation

- A common robustness check in RD designs is to show that we don't see the same effect if the threshold were somewhere else
- Build some 'fake' threshold, and estimate the effect around this fake cutoff
- If we see significant effects, it's problematic!

Robustness Checks - One Implementation

```
crost_pseudo <- crost %>%
  mutate(rank_pseudo = rank_norm + 3,
         eligible_pseudo = rank_pseudo < 0,
         el_rank_pseudo = eligible_pseudo*rank_pseudo) %>%
  filter(abs(rank_norm) <= 6)

ff_pseudo <- totcas ~ eligible_pseudo + rank_pseudo + el_rank_pseudo +
  I(rank_pseudo^2) + I(el_rank_pseudo^2) | year | 0 | mcode

summary(felm(ff_pseudo, filter(crost_pseudo,
                  month_year >= ymd("2003-01-01"))))
```

Robustness Checks - One Implementation

```
##
## Call:
##    felm(formula = ff_pseudo, data = filter(crost_pseudo, month_year >=
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.1415 -0.1158 -0.0915 -0.0781 19.8823
##
## Coefficients:
##                      Estimate Cluster s.e. t value Pr(>|t|)
## eligible_pseudoTRUE  0.090782     0.187216   0.485    0.628
## rank_pseudo          0.005805     0.015502   0.374    0.708
## el_rank_pseudo       0.152528     0.234719   0.650    0.516
## I(rank_pseudo^2)    -0.001372     0.001643  -0.835    0.404
## I(el_rank_pseudo^2)  0.051372     0.074279   0.692    0.489
##
## Residual standard error: 0.6936 on 7984 degrees of freedom
```

Problem Set Questions?