Regression Discontinuity Design

Lukáš Lafférs

Matej Bel University, Dept. of Mathematics

MUNI Brno

~~2.12.2021~~ 6.1.2022

Nature does not make jumps.

(Natura non facit saltus)

Some things do not occur naturally and requires some explanation.

Example: Campbell (1960)

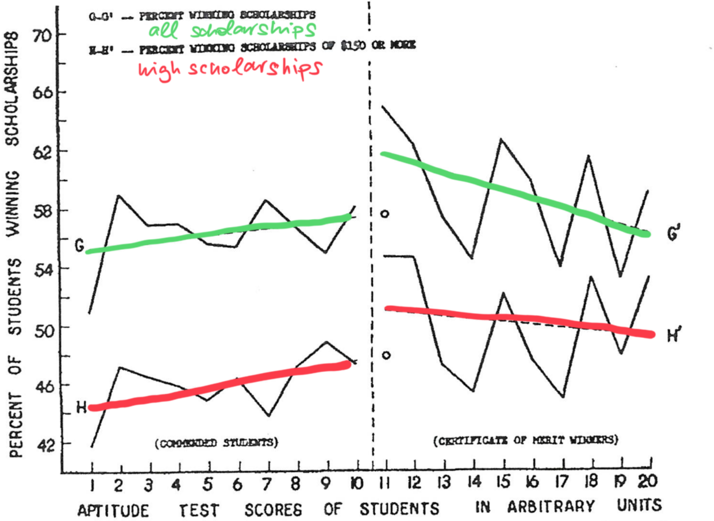


FIG. 2. Regression of success in winning scholarships on exposure determiner.

- ”Certificate of merit” is given to students above some **threshold** in an aptitude test.
- High-scoring students went for ”National merit scholarship” which is not included on the vertical axis $\implies$ downward slope.

Students right below and right above the thresholds are similar.

It is **as if they were assigned** to group below and the group above **randomly**.

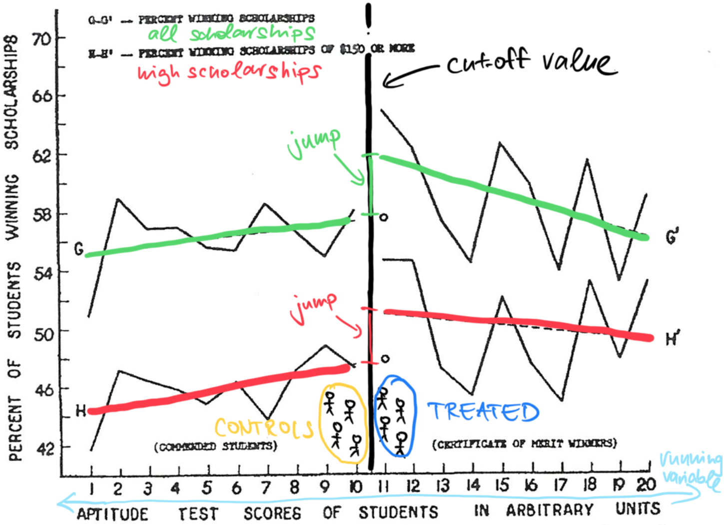
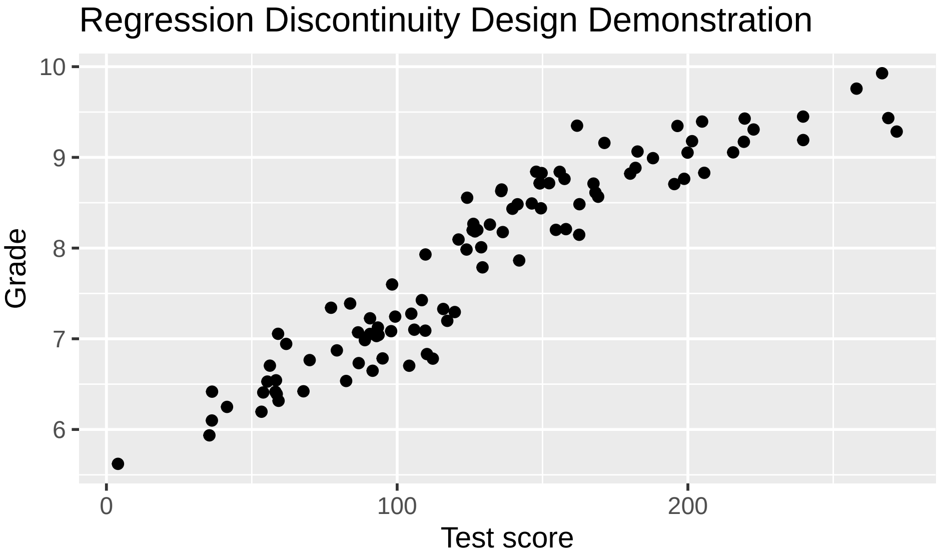
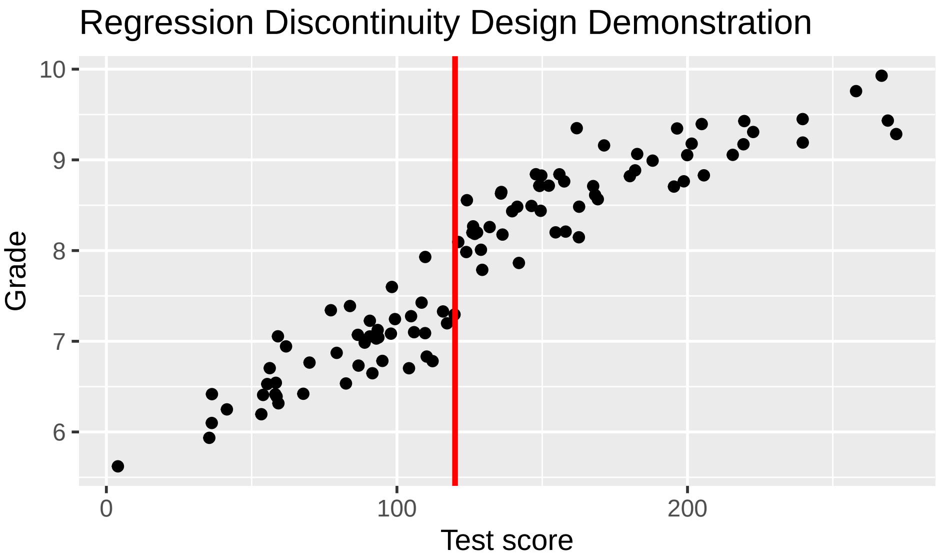


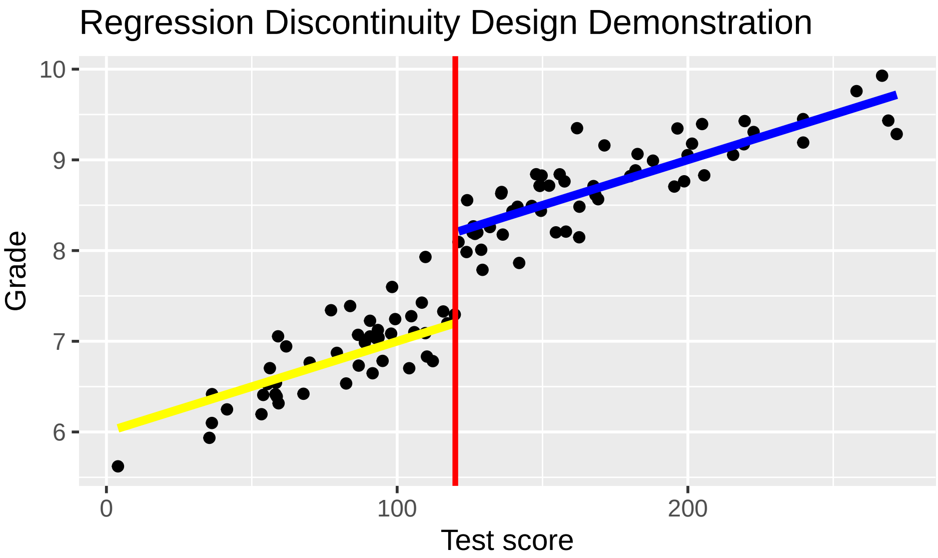
FIG. 2. Regression of success in winning scholarships on exposure determiner.

Sharp RDD



Regression Discontinuity Design Demonstration
10
9
8
7
6
Grade
0
100
200
Test score


Regression Discontinuity Design Demonstration
10
9
8
7
6
Grade
0
100
200
Test score


Regression Discontinuity Design Demonstration
Grade
10
9
8
7
6
0
100
200
Test score

Sharp RDD

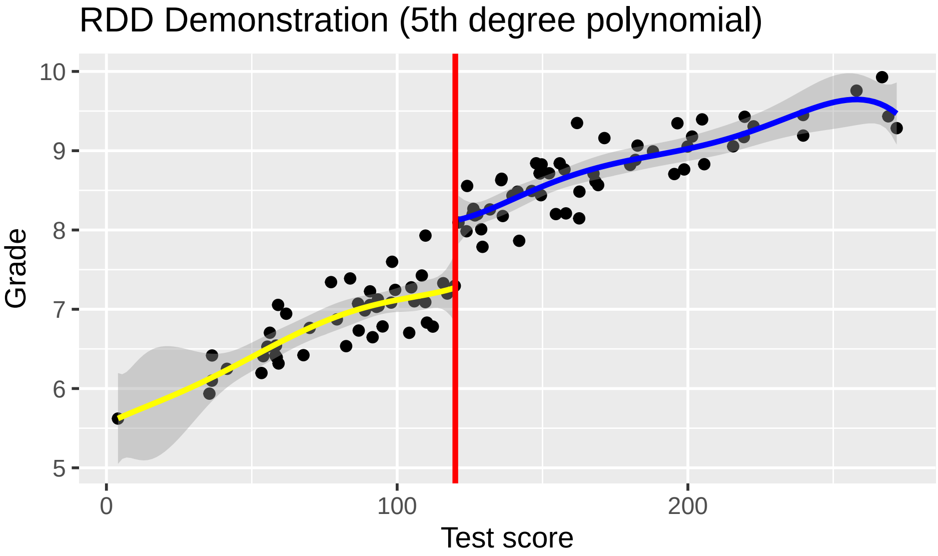
$$\hat{\delta} = \underbrace{\lim_{z \downarrow 0} \hat{E}[Y|Z=z]}_{\text{treated}} - \underbrace{\lim_{z \uparrow 0} \hat{E}[Y|Z=z]}_{\text{control}}$$

- Random variation is coming from the real world constraints/rules
- Treatment $D = 1$ or $D = 0$ if defined in terms of **running variable** Z:
  $D = 1 \iff Z > 0$

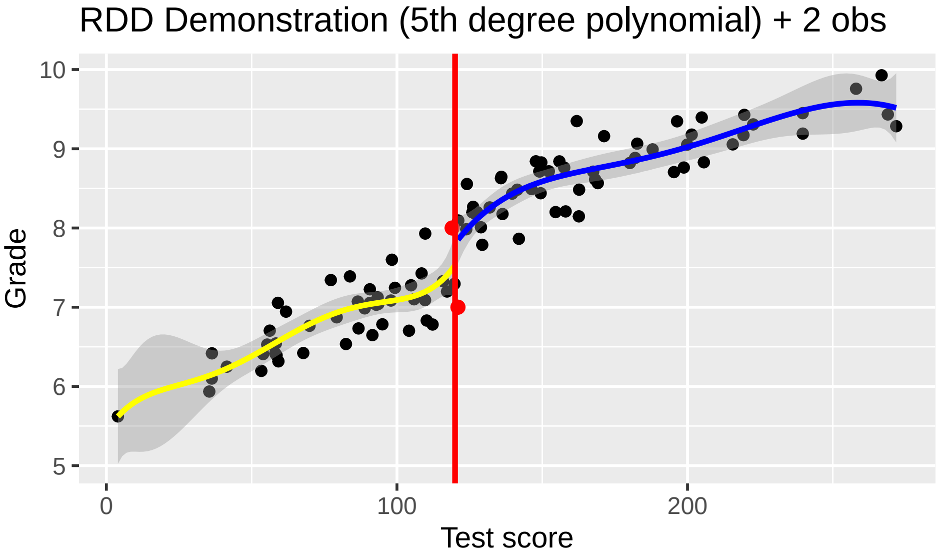
Sharp RDD

- In the previous example we used linear model.
- Yet in practice we have only seldom reasons to believe that such model is correct
- We wish to have a sufficiently flexible model.

Too flexible?



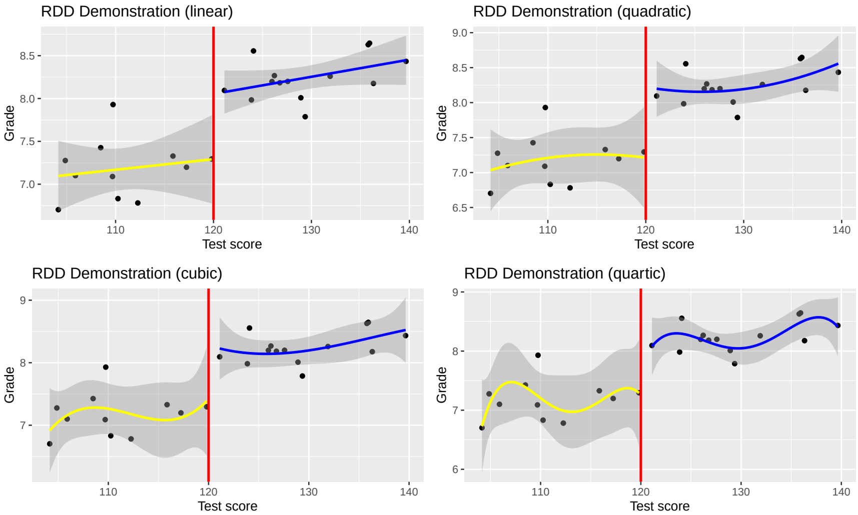
Sensitive? (add 2 points)



Sharp RDD

- It is the threshold that is important
- This is where all the action takes place
- You have to defend that there is no manipulation around the threshold

It’s all about the threshold


RDD Demonstration (linear)
RDD Demonstration (quadratic)
RDD Demonstration (cubic)
RDD Demonstration (quartic)
Grade
Test score
110
120
130
140

Sharp RDD

- But how do we choose the model for $\hat{E}[Y|Z = z]$ on the left and on the right?
- Clearly, there are many many ways how we can do this!
- Gelman and Imbens (2019) warns against the higher order polynomials. Obviously: you can cook up the results according to your liking.
- And that is always a bad thing.
- Just look at the title: **"Why high-order polynomials should not be used in regression discontinuity designs."**

Gelman and Imbens (2019)

- Issue 1: Some observations are given excessive weights → Check the weights.
- Issue 2: Results are sensitive to the degree of polynomial (see below) → Use local regression.
- Issue 3: Confidence intervals are too narrow. → Global regressions are simply not precise enough at the cut-off point, use local regression instead.

	Order of polynomial	estimate	(se)
global	1	0.167	(0.008)
global	2	0.079	(0.010)
global	3	0.112	(0.011)
global	4	0.077	(0.013)
global	5	0.069	(0.016)
global	6	0.104	(0.018)
local	1	0.080	(0.012)
local	2	0.063	(0.017)

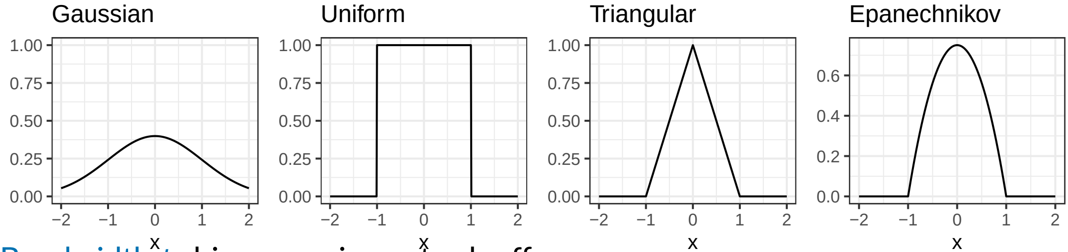
	Order of polynomial	estimate	(se)
global	1	0.024	(0.008)
global	2	0.176	(0.012)
global	3	0.209	(0.015)
global	4	0.174	(0.018)
global	5	0.164	(0.021)
global	6	0.197	(0.025)
local	1	0.196	(0.018)
local	2	0.176	(0.027)

Local linear regression (loess)

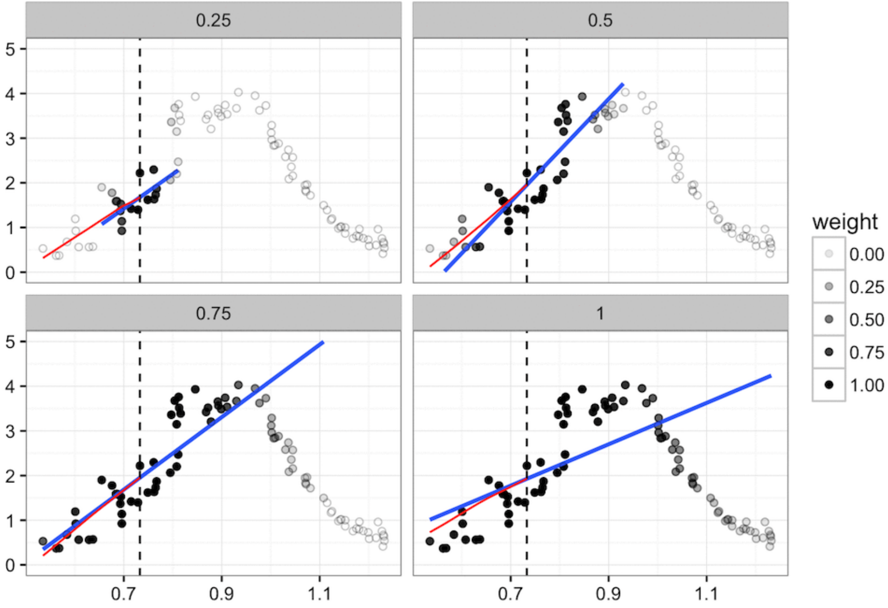
$$\left(\hat\beta_0(x), \hat\beta_1(x)\right) = \arg\min_{\beta_0,\beta_1} \underbrace{\sum_{i=1}^{n}}_{\text{Sum of}} \underbrace{K\left(\frac{x_i - x}{h}\right)}_{\text{weighted}} \cdot \underbrace{(y_i - (\beta_0 + \beta_1 x_i))^2}_{\text{squared errors}}$$

$$\hat{E}[Y|X = x] = \hat\beta_0(x) + \hat\beta_1(x) \cdot x$$

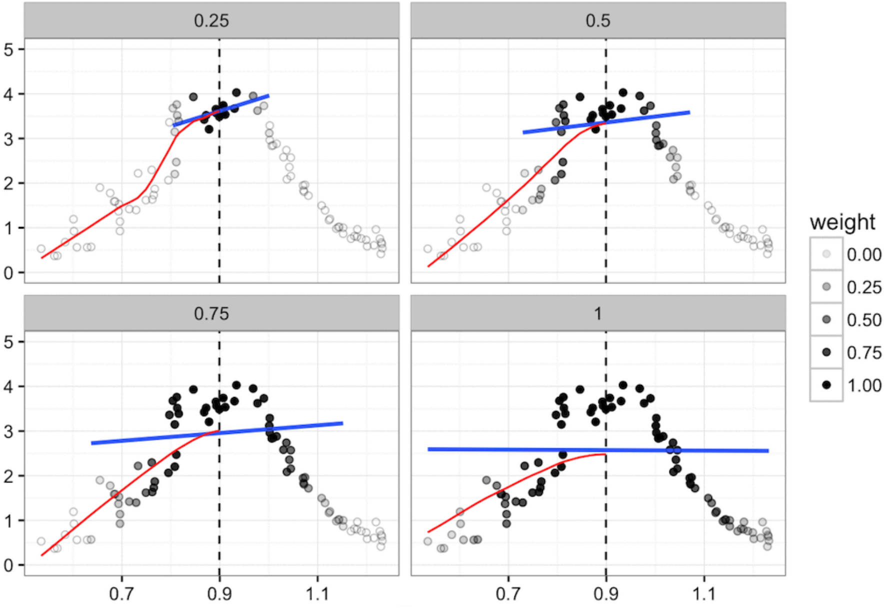
Kernel functions:



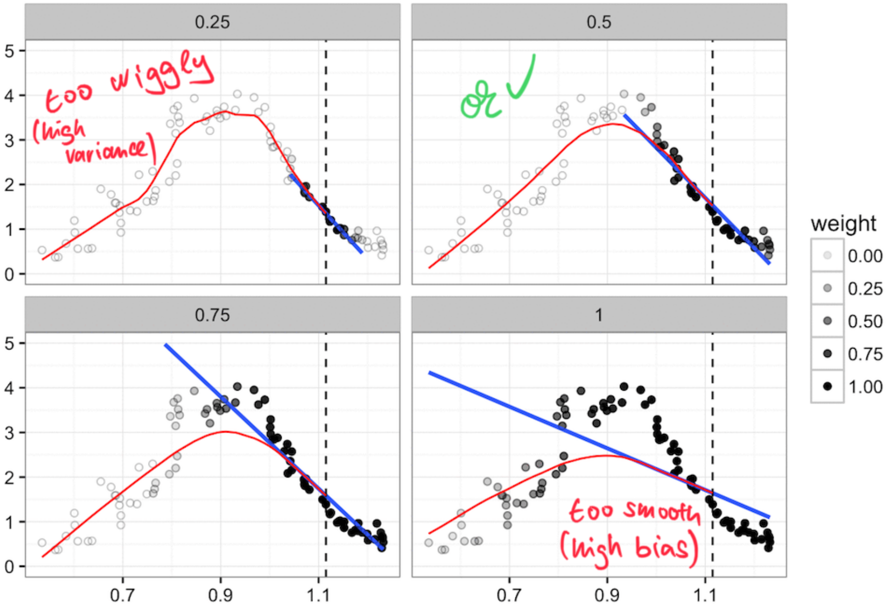
Bandwidth h: bias vs variance tradeoff.



0.25
0.5
0.75
1
weight
0.00
0.25
0.50
0.75
1.00
0
1
2
3
4
5
0.7
0.9
1.1



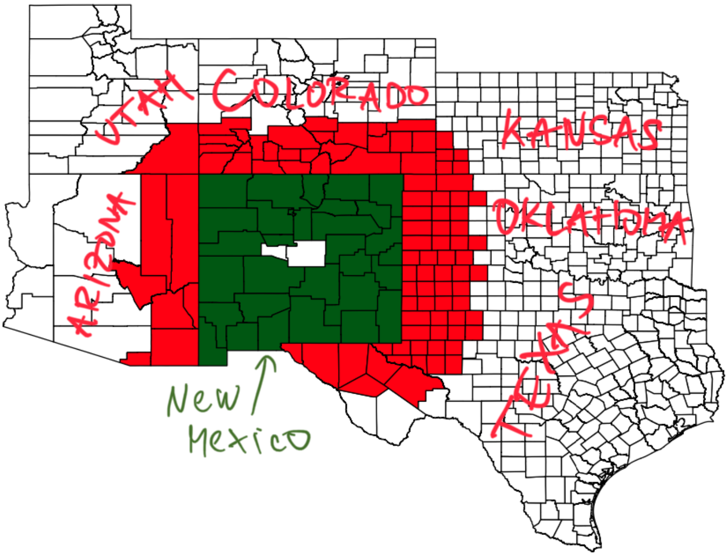
0.25
0.5
0.75
1
0
1
2
3
4
5
0.7
0.9
1.1
weight
0.00
0.25
0.50
0.75
1.00



0.25
0.5
0.75
1
too wiggly
(high variance)
or ✓
too smooth
(high bias)
weight
0.00
0.25
0.50
0.75
1.00
0
1
2
3
4
5
0.7
0.9
1.1

- In practice, the choice of the bandwidth is more important than the choice of the kernel (weighting function)
- Local linear regression methods (and other semi/non-parametrics methods) have problems at the boundary, fewer points $\rightarrow$ more bias. There are different ways (fixes) how to deal with this (Noack and Rothe, 2021).
- If the running variable is discrete, one should account for it (Kolesar and Rothe, 2018)

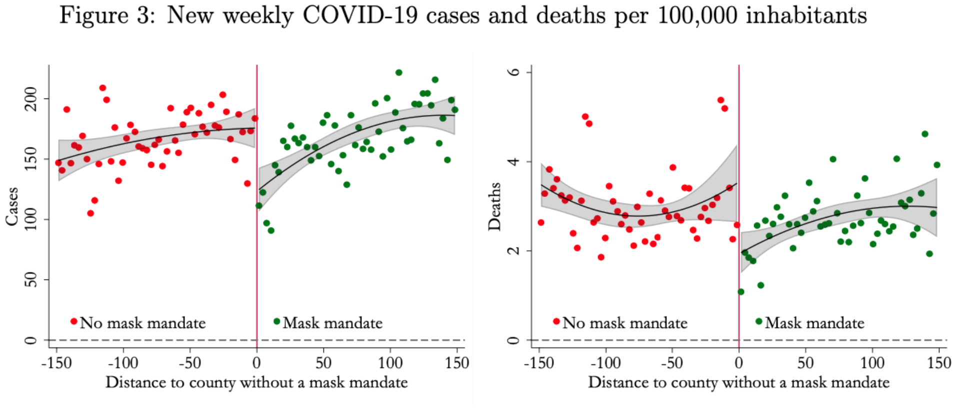
Example: Mask mandates



$$y_{c,s,t+k} = \delta mask_{st} + \beta dist_{cst} + \gamma dist^2_{cst} + \theta X_{c,s,t} + \tau_t + \tau_c + \epsilon_{cst}$$

- c - county
- s - state
- t - week
- k lag ($k = 1$ for cases, $k = 2$ for hospital admissions, $k = 4$ for deaths)
- $mask_{st}$ - mask mandate in state s at time t
- $dist_{cst}$ - minimum distance of country c is state s to a county with a different mask policy at time t
- $X_{c,s,t}$ - controls using lagged mobility

Discontinuity?



(a) Cases (b) Deaths

Niels-Jakob H Hansen ; Rui C. Mano: ”Mask Mandates Save Lives.” IMF Working Papers 2021.205 (2021).

Results

Table 1: Results per 100,000 inhabitants

(a) Unconditional results

	(1) Cases	(2) Admissions	(3) Deaths
State mask mandate	-55.22***	-11.46***	-0.74**
	[17.68]	[4.24]	[0.34]
Observations	45577	22042	41034
R2	0.487	0.417	0.215
Mean of dep. variable	166.44	23.57	2.64
Linear term	Yes	Yes	Yes
Quadratic term	Yes	Yes	Yes
Lagged mobility	Yes	Yes	Yes

Niels-Jakob H Hansen ; Rui C. Mano: "Mask Mandates Save Lives." IMF Working Papers 2021.205 (2021).

Fuzzy RDD

- Random variation is coming from the real world constraints/rules
- Treatment probability $Pr(D = 1|Z = z)$ here discontinuously changes at the cut-off value $Z = 0$ (it is just an IV!!!)
- Very specific subpopulation - those marginal people at the cut-off (compliers)
- The question is: How do we extrapolate??

Fuzzy RDD as an IV

Wald estimator:

$$\hat{\beta}_{Wald} = \frac{\hat{E}[Y|Z=1] - \hat{E}[Y|Z=0]}{\hat{E}[D|Z=1] - \hat{E}[D|Z=0]}$$

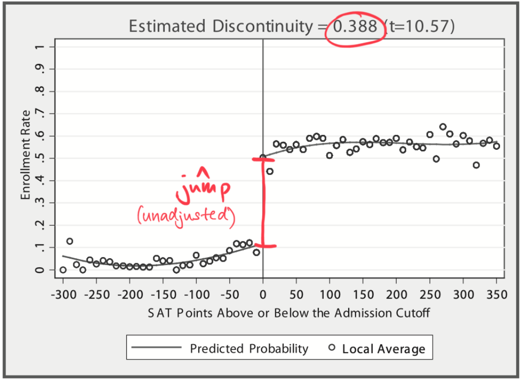
Or

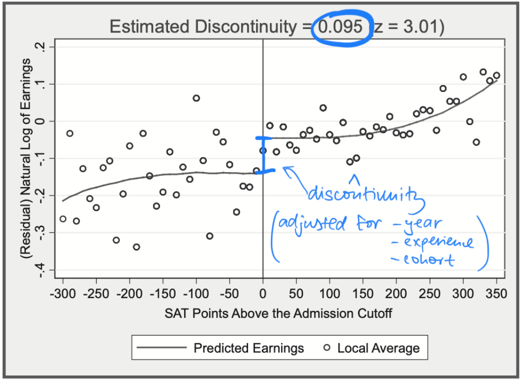
$$\hat{\beta}_{IV} = (\boldsymbol{X}^T P_{\boldsymbol{Z}} \boldsymbol{X})^{-1} \boldsymbol{X}^T P_{\boldsymbol{Z}} y$$

where $\boldsymbol{X} = [\mathbf{1}, X, D]$ $\boldsymbol{Z} = [\mathbf{1}, X, Z]$ $y_i = \boldsymbol{X}_i \beta + e_i$

Example: Hoekstra (2009)

$$\log(earnings) = \psi_{year} + \phi_{experience} + \theta_{cohort} + \varepsilon$$





We have exogenous variation in the prob. of enrollment rate. What is the effect on earnings 15 years later?

How to quantify the effect?

We wish to estimate this:

$$Outcome = \beta_0 + \beta_1 (Above) + \gamma_1 (\overbrace{h(\underbrace{AdjustedSATscore}_{\text{Cut-off set to 0.}})}^{\text{Some flexible function}}) + \gamma_2 GPA + \gamma_3 (SAT\ Score) + \varepsilon$$

But taking into account variation in year of admission, experience and cohort. Hoekstra did it in two steps:

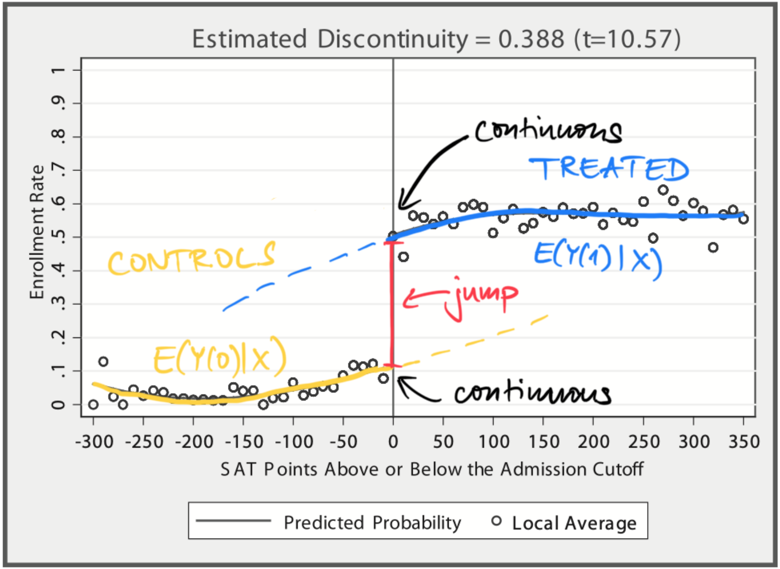
- **Step 1** For every individual, regress outcome and independent variables on ψ_{year}, $\phi_{experience}$, θ_{cohort} and obtain residuals. Average these residuals across different years for every single individual.
- **Step 2** Plug in residuals from these regressions in this equation and estimate it via OLS.
- **Step 3** Scale the estimate of coefficient of interest $\hat{\beta}_1$ by the size of the discontinuous jump.

Two stage estimator. Standard errors? Bootstrap.

TABLE 1.—EARNINGS DISCONTINUITIES AND CORRESPONDING INTENT-TO-TREAT AND ENROLLMENT ESTIMATES FOR WHITE MEN

				Discontinuity	Treatment Effect	
Regression Specification	Function of Adjusted SAT	Flexible Polynomial?	Additional Controls	Estimated Earnings Discontinuity	Intent-to-Treat Effect	Enrollment Effect
(1) Plotted in Figure 2	Cubic	No	No	0.095***	0.135***	0.223***
				(0.032)	(0.046)	(0.079)
				[0.003]	[0.004]	[0.005]
(2)	Cubic	No	Yes	0.092***	0.131***	0.216***
				(0.033)	(0.048)	(0.081)
				[0.005]	[0.006]	[0.008]
(3)	Quadratic	Yes	Yes	0.111**	0.170**	0.281**
				(0.045)	(0.073)	(0.121)
				[0.014]	[0.019]	[0.021]
(4) (includes only applicants within 200 points of cutoff)	Quadratic	No	Yes	0.081**	0.116**	0.192**
				(0.038)	(0.056)	(0.094)
				[0.034]	[0.038]	[0.041]
(5) (includes only applicants within 100 points of cutoff)	Linear	No	Yes	0.074**	0.110*	0.181*
				(0.038)	(0.058)	(0.099)
				[0.050]	[0.060]	[0.067]

Table 1 in Hoekstra (2009)

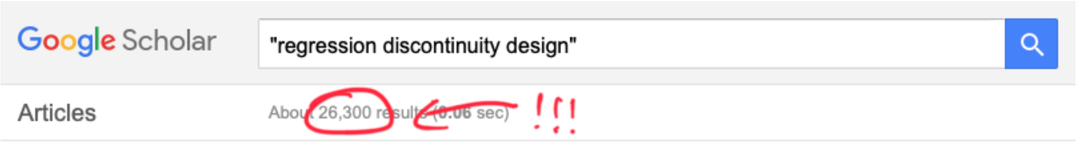


Estimated Discontinuity = 0.388 (t=10.57)
Enrollment Rate
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1
-300
-250
-200
-150
-100
-50
0
50
100
150
200
250
300
350
SAT Points Above or Below the Admission Cutoff
Predicted Probability
Local Average
continuous
TREATED
CONTROLS
E(Y(1)|X)
jump
E(Y(0)|X)
continuous

Three well-known papers

- Angrist and Lavy (1999) - Maimonides rule - what is the effect of classroom size on academic achievements. Classrooms cannot have more than 40 students.
- Black (1999) - used impact of school districts on the house prices. Houses at the district boundaries, similar but differ in terms of elementary school that the child attends. ”...parents are willing to pay 2.5 percent more for a 5 percent increase in test scores”
- Van Der Klaauw (2002) - effect of financial aid on student enrollments. Exploiting the discontinuity in financial aid eligibility in student’s test performance.

RDD



- Only popular since 2000s.
- Visually very appealing. The details matters. Work hard on compelling figures, they do matter a lot.
- You should not see a jump in the density of a running variable around the threshold - it may suggest some manipulation.
- There seems to be many choices to consider: estimation, bandwidth, kernel. Stick to defaults unless you have very good reasons to not to.
- There are many many situations with cut-offs.

RDD

- Data hungry - we need many points around the threshold to provide credible evidence on the jump.
- Cunningham (2021) suggests you to work on relationships with people who have access to this kind of data. These are built on trust and honesty.
- Korting et al. (2020) - people are quite good at spotting the discontinuities - they conducted some tests.

Thank you for your attention!

References

- Thistlethwaite, D. L., & Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. Journal of Educational Psychology, 51(6), 309–317. `https://doi.org/10.1037/h0044319`
- Angrist, Joshua D., and Victor Lavy. "Using Maimonides' rule to estimate the effect of class size on scholastic achievement." The Quarterly journal of economics 114.2 (1999): 533-575.
- Black, Sandra E. "Do better schools matter? Parental valuation of elementary education." The quarterly journal of economics 114.2 (1999): 577-599.
- Van der Klaauw, Wilbert. "Estimating the effect of financial aid offers on college enrollment: A regression–discontinuity approach." International Economic Review 43.4 (2002): 1249-1287.
- Gelman, Andrew, and Guido Imbens. "Why high-order polynomials should not be used in regression discontinuity designs." Journal of Business & Economic Statistics 37.3 (2019): 447-456.
- Niels-Jakob H Hansen ; Rui C. Mano: "Mask Mandates Save Lives." IMF Working Papers 2021.205 (2021).
- Noack, Claudia, and Christoph Rothe. Bias-Aware Inference in Fuzzy Regression Discontinuity Designs. arXiv. org, 2021.
- Kolesár, Michal, and Christoph Rothe. "Inference in regression discontinuity designs with a discrete running variable." American Economic Review 108.8 (2018): 2277-2304.
- Korting, Christina, et al. "Visual Inference and Graphical Representation in Regression Discontinuity Designs." (2020).
- The chapter in Cunningham, Scott. Causal Inference. Yale University Press, 2021 is very engaging and fun to read.