Causal Inference: Revisiting Instrumental Variables and Fuzzy Regression Discontinuity Design

Ken Stiller

Part 9

April 2025

Overview

Strategies for estimating effects of treatments so far:

- ▶ Randomize treatment and take the DIGM
- ▶ Identify and control for confounding variables such that the CIA holds
- ▶ Identify a plausible counterfactual using time and unit variation in panel data
- ▶ Identify an instrumental variable and use two-stage-least-squares to estimate average treatment effect for compliers
- ▶ Identify special situations where treatment is applied based on a cutoff to estimate a local average treatment effect

Now: Make use of special situations where treatment becomes **likelier** based on a cutoff: combining IV+RDD

Overview

Table of Contents

Generalizing from Wald Estimator to 2SLS

Wald estimator is limited to binary D_i and Z_i:

$$\lambda = \frac{E\big[Y_i|Z_i = 1\big] - E\big[Y_i|Z_i = 0\big]}{E\big[D_i|Z_i = 1\big] - E\big[D_i|Z_i = 0\big]} = \frac{\text{ITT}_Y}{\text{ITT}_D}$$

Two-stage least squares is a much more general procedure:

$$
\begin{aligned}
\text{First stage:} \quad D_i &= \alpha_1 + \phi Z_i + \beta_1 X_{1i} + \gamma_1 X_{2i} + e_{1i} \\
\text{Second stage:} \quad Y_i &= \alpha_2 + \lambda \hat{D}_i + \beta_2 X_{1i} + \gamma_2 X_{2i} + e_{2i}
\end{aligned}
$$

where Z_i and D_i might not be binary and you can include covariates e.g. X_{1i}, X_{2i}. same covariates must be included in the first and second stage

Introduction
oo

IV Review
oo●ooooo

Fuzzy RDD
oooooooooooooooo

(Fuzzy) Checks
o
ooooooooooooo

Review
oooooo

Two-stage least squares: terminology

Terminology:

$$
\begin{aligned}
\text{Reduced form:} \quad Y_i &= \alpha_0 + \rho Z_i + \beta_0 X_{1i} + \gamma_0 X_{2i} + e_{0i} \\
\text{First stage:} \quad D_i &= \alpha_1 + \phi Z_i + \beta_1 X_{1i} + \gamma_1 X_{2i} + e_{1i} \\
\text{Second stage:} \quad Y_i &= \alpha_2 + \lambda \hat{D}_i + \beta_2 X_{1i} + \gamma_2 X_{2i} + e_{2i}
\end{aligned}
$$

NB: λ is the LATE. Must use same covariates in first stage and second stage.

Two-stage least squares: assumptions

Key assumptions (Wald assumptions with covariates and without "complier" terminology):

- ▶ **Non-zero first-stage**: instrument affects treatment, conditional on covariates ($\phi \neq 0$ in first stage)
- ▶ **Independence** (exogeneity, ignorability): instrument unrelated to potential outcomes, conditional on covariates (no OVB on ρ in reduced form or ϕ in first stage)
- ▶ **Exclusion restriction**: instrument only affects outcome through treatment, conditional on covariates
- ▶ **Monotonicity**: instrument's effect on treatment is weakly positive or weakly negative for all units (no defiers)

Why we should be skeptical of most IV designs

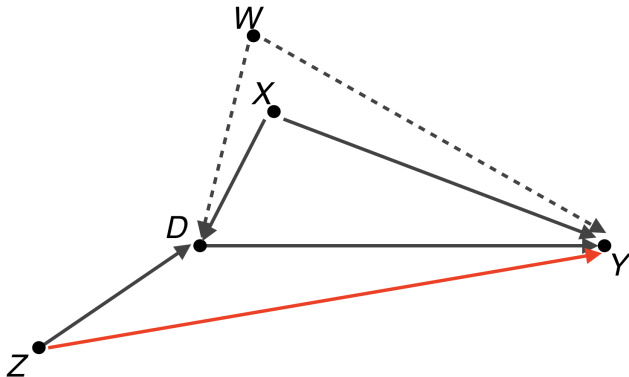IV designs must convince us of two key untestable assumptions:

▶ The instrument Z_i satisfies **independence**, i.e. the CIA is met with respect to D_i and Y_i, e.g. because Z_i is random

▶ The instrument Z_i satisfies **exclusion**, i.e. it only affects Y_i through D_i

When Z_i is randomly determined in an experiment, it's easier to accept **independence** and think hard and discuss the **exclusion** assumption.
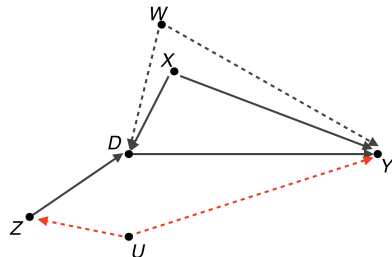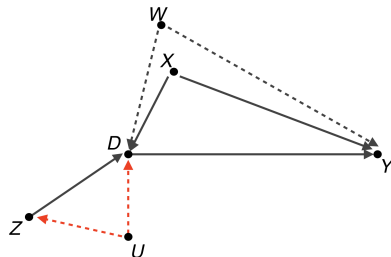
In an observational study, one should be skeptical about both.

▶ Is the CIA really satisfied in the reduced form?

▶ Is D_i really the only channel through which Z_i affects Y_i?

Introduction
oo

IV Review
ooooo●ooo

Fuzzy RDD
ooooooooooooooooo

(Fuzzy) Checks
o
ooooooooooooooo

Review
oooooo

Exclusion restriction violation DAG

Introduction
○○

IV Review
○○○○○○●○○

Fuzzy RDD
○○○○○○○○○○○○○○○○

(Fuzzy) Checks
○
○○○○○○○○○○○○○○

Review
○○○○○○

Independence violations DAGs

(where U is an unobserved covariate)

Can we test the exclusion restriction?

What about

- ▶ regress Y on D and Z

- ▶ conclude exclusion is valid if coefficient on Z is 0

Unfortunately this doesn't work, because D is also affected by X and W:

- ▶ if X and W are observed, you don't need IV to estimate effect of D on Y

- ▶ if they are *not* observed, by controlling for D you induce an association between Z and X and/or W, leading to **collider bias**
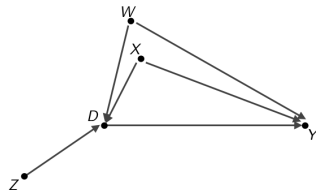
Further thoughts on IV

- ▶ A good instrument is hard to find – another reason to start by looking for randomness.

- ▶ **In observational studies, a variable that satisfies independence (CIA) is a rare and wonderful thing. Usually exclusion is doubtful, but you can gauge its effect and speculate about channels.**

- ▶ Now that you know about instrumental variables, you should always define what you refer to as "IV": instead, you could say "treatment", "control variable", "covariate", "regressor"

- ▶ The recent econometrics literature is very skeptical of the internal validity of IVs, see for example: *Lal, Apoorva, Mackenzie William Lockhart, Yiqing Xu, and Ziwen Zu. (2021) "How Much Should We Trust Instrumental Variable Estimates in Political Science? Practical Advice based on Over 60 Replicated Studies."*

Table of Contents
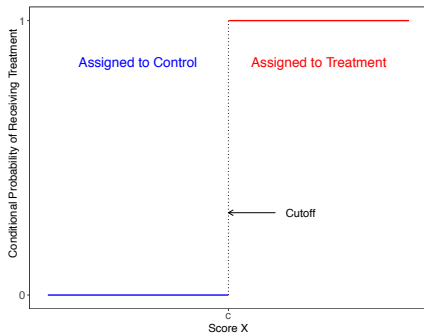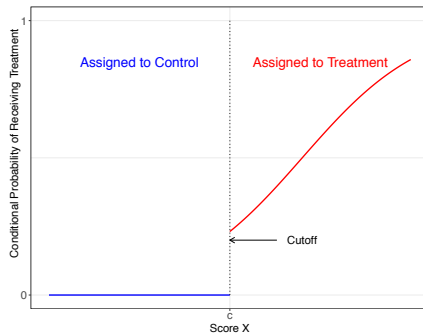
Fuzzy Regression Discontinuity Design

▶ So far we have focused on cases where there is perfect compliance, all those that were assigned to treatment, end up taking the treatment

▶ A threshold may not perfectly determine treatment exposure, but may create a discontinuity in the probability of treatment exposure

▶ For example, incentives to participate in a program may change discontinuously at a threshold, but the incentives are not powerful enough to move all units from nonparticipation to participation

▶ We can use such discontinuities to produce instrumental variable estimators of the effect of the treatment (close to the discontinuity)

Sharp vs Fuzzy RDD

(a) Sharp RD

(b) Fuzzy RD (one-sided compliance)

Fuzzy Regression Discontinuity Design

▶ Probability of being offered a scholarship may jump at a certain SAT score threshold (when applicants are given "special consideration")

▶ We shouldn't compare recipients with non-recipients (even close to threshold) since they are likely differ along unobservables related to outcome (e.g., letters of recommendation, motivation)

▶ But for applicants with scores close to the threshold we can exploit the discontinuity as an instrument to estimate the LATE for the subgroup of applicants for whom scholarship receipt depends on the difference between their score and the threshold

    ▶ A *complier* in this framework is a student who switches from non-recipient to recipient if their score crosses the threshold

Potential Outcome Model for Instrumental Variables

Instrument Z_i: (Assignment status) - Binary instrument for *unit i*.

$$Z_i = \begin{cases} 1 & \text{if unit } i \text{ "encouraged" to receive treatment} \\ 0 & \text{if unit } i \text{ "encouraged" to receive control} \end{cases}$$

Potential Treatments D_z indicates the *'assigned' - or potential -* treatment status given $Z = z$

- $D_1 = 1$ encouraged to take treatment and takes treatment
- Note that $Z_i \neq D_i$ (Not all units take-up the treatment)

Potential Outcome Model for Instrumental Variables

Following the terminology of Angrist, Imbens and Rubin (1996), we can define:

Definition

- ▶ *Compliers*: $D_1 > D_0$ ($D_0 = 0$ and $D_1 = 1$).
- ▶ *Always-takers*: $D_1 = D_0 = 1$.
- ▶ *Never-takers*: $D_1 = D_0 = 0$.
- ▶ *Defiers*: $D_1 < D_0$ ($D_0 = 1$ and $D_1 = 0$).

Problem

Only one of the potential treatment indicators (D_0, D_1) is observed, so we cannot identify which group any particular individual belongs to

Identification Assumptions

▶ Independence of the Instrument: $(Y_0, Y_1) \perp\!\!\!\perp Z$ (often called
  ignorability - strong form of assumption: $(Y_0, Y_1, D_0, D_1) \perp\!\!\!\perp Z$)

  ▶ Implies that the instrument Z is exogenously assigned
  ▶ $Y(z, d)$ implies the *exclusion restriction*: Z can have no effect on Y
    except through its effect on D
  ▶ Allows to attribute correlation between Z and Y to the effect of D
    alone; assumption is not testable

▶ First Stage: $0 < P(Z = 1) < 1$ and $P(D_1 = 1) \neq P(D_0 = 1)$

  ▶ Implies that the instrument Z induces variation in D
  ▶ Testable by regressing D on Z (usually an RD of D on Z)

▶ Monotonicity: $D_1 \geq D_0$

  ▶ Rules out defiers - doing the opposite of encouragement
  ▶ Refusal would have been the same if the control units were offered
    the treatment
  ▶ Often easy to assess from institutional knowledge

Fuzzy RDD: Discontinuity in $E[D|X]$ - First stage
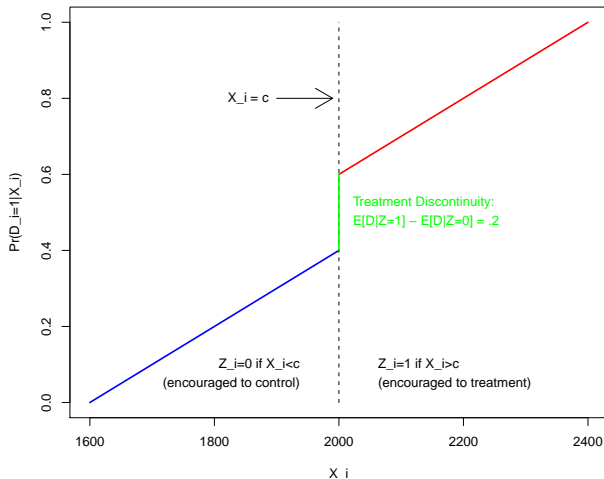
Fuzzy RDD: Identification

▶ Binary instrument Z with $Z = 1\{X > c\}$
▶ Restrict sample to observations close to discontinuity where $E[Y|D, X]$ jumps so that $X \approx c$ and thus $E[X|Z = 1] - E[X|Z = 0] \approx 0$.
▶ With IV assumptions: independence, first stage, monotonicity

$$
\begin{aligned}
\alpha_{FRDD} &= E[Y_1 - Y_0 | X = c \text{ and } i \text{ is a complier}] \\
&= \frac{\lim_{x\downarrow c} E[Y|X = c] - \lim_{x\uparrow c} E[Y|X = c]}{\lim_{x\downarrow c} E[D|X = c] - \lim_{x\uparrow c} E[D|X = c]} \\
&= \frac{\text{outcome discontinuity}}{\text{treatment discontinuity}} \\
&\approx \frac{E[Y|Z = 1] - E[Y|Z = 0]}{E[D|Z = 1] - E[D|Z = 0]}
\end{aligned}
$$

Introduction
oo

IV Review
ooooooooo

Fuzzy RDD
ooooooooo●ooooooo

(Fuzzy) Checks
o
oooooooooooooooo

Review
oooooo

Fuzzy RDD: Discontinuity in $E[Y|X]$
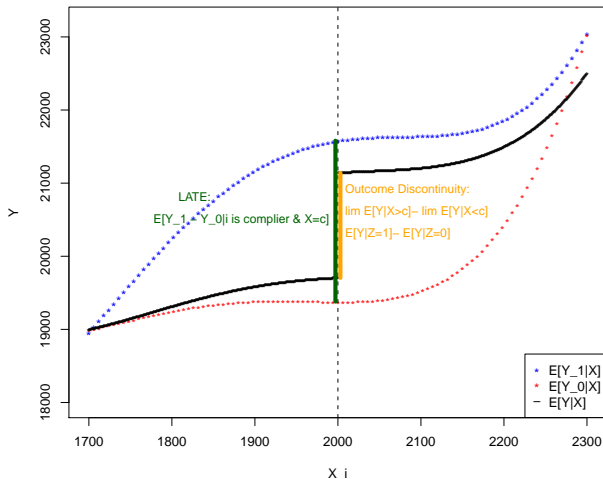
Fuzzy RDD: Estimation

- ▶ Estimate the fuzzy-rdd small window h above and below the threshold (discontinuity sample)
- ▶ Code instrument $Z = 1\{X > c\}$
- ▶ Fit 2SLS: $Y = \beta_0 + \alpha D + \beta_1(X - c) + \beta_2(D \cdot (X - c))$
- ▶ where D is instrumented with Z
- ▶ specification can be more flexible by adding polynomials
- ▶ Also helpful to separately plot (and estimate) the outcome discontinuity and treatment discontinuity

"Power of Money" Ruiz (2017) Fuzzy-RDD

▶ Study the effect of electing donor-funded politicians

▶ Focus on <u>similar</u> municipalities where there are tighter **or** looser campaign limits by a narrow registered voter margin

▶ Treatment: Share of donor funds of the elected politician

▶ Outcome: Number of contracts awared to donors

▶ Instrument: Discontinuity in rules around campaign limits

Empirical approach: Fuzzy-RDD

- ▶ Count registered voters:
- ▶ V_i = number of registered voters centered around 0 (25,000 cut-off)
- ▶ When $CampaignLimit_i = 1$ then $(V_i > 0)$

$$CampaignLimit_i = \begin{cases} CampaignLimit_i = 1 & \text{if } V_i > 0 \\ CampaignLimit_i = 0 & \text{if } V_i < 0 \end{cases} \tag{1}$$

- ▶ Note that the treatment implies **looser campaign limits**

"Power of Money" Ruiz (2017) Fuzzy-RDD

▶ **First stage:**

$$DF_i = \beta_1 CampaignLimit_i + \beta_2 f(V_i) + \beta_3 CampaignLimit_i \times f(V_i) + \varepsilon_i \tag{2}$$

▶ **Second-stage:**

$$Y_i = \alpha + \beta_1 \widehat{DF_i} + \beta_2 f(V_i) + \beta_3 \widehat{DF_i} \times f(V_i) + \varepsilon_i \tag{3}$$

▶ Where DF which is the percent of the winning candidate's campaign funds that came from donors

▶ Y_i is the number of contracts awarded to donors.

Looser limits and rewards for donors

	(1)	(2)	(3)	(4)
		Panel A: Second stage		
		Total contracts for donors		
Looser Campaign Limits	50.531*	47.996*	31.631*	24.115
Robust p-value	0.084	0.098	0.075	0.140
CI 95%	[-7.146, 112.912]	[-9.092, 108.499]	[-3.277, 68.190]	[-7.796, 55.518]
Controls		✓		✓
		Panel B: First stage variables		
	Donor income(% of total)		Candidate was donor funded	
Looser Campaign Limits	0.272*	0.312*	0.430**	0.616***
Robust p-value	0.092	0.089	0.012	0.005
CI 95%	[-0.045, 0.598]	[-0.049, 0.697]	[0.101, 0.824]	[0.201, 1.108]
Controls		✓		✓
Observations	921	843	921	843
Bandwith Obs.	137	124	147	91
Mean	2.243	2.243	2.243	2.243
Effect Mean (%)	2252.83	2139.81	1410.21	1075.12
Bandwidth	7569.892	7938.083	8427.081	6339.717

Note: Local linear estimates of average treatment effects at cutoff estimated with triangular kernel weights and optimal MSE bandwidth. 95% robust confidence intervals and robust p-values are computed following Calonico, Cattaneo, and Tituinik (2013). Bandwith Obs. denotes number of observations in the optimal MSE bandwidth for each dependent variable. Controls included: Royalty income as a % of total municipality income and indigenous *** p<0.01, ** p<0.05, * p<0.1

Introduction
00

IV Review
000000000

Fuzzy RDD
00000000000000●

(Fuzzy) Checks
0
0000000000000

Review
000000

Similar characteristics of winner politicians

Table: Mayors' characteristics across campaign contribution limits cutoff

	Mean	Std. Dev.	Looser limits	CI 95%	Obs.	Band. Obs.	Bandwidth	p-valu
	(1)	(2)	(1)	(2)	(3)	(4)	(5)	(6)
Women	0.098	0.298	-0.010	[-0.129,0.134]	999	81	4869	0.969
Age	44.863	9.740	-3.693	[-15.066,7.071]	927	99	6448	0.479
Black	0.046	0.210	-0.090	[-0.320,0.089]	927	67	4782	0.268
Indigenous	0.112	0.315	-0.317	[-0.653,-0.069]	927	114	7070	0.015
Leftist party	0.028	0.165	-0.023	[-0.134,0.088]	999	92	5268	0.681
Rightwing	0.240	0.427	0.261	[-0.238,0.957]	999	106	5943	0.239
Sanctioned	0.116	0.320	0.047	[-0.326,0.344]	999	79	4759	0.958
Illegal Registration of ID.	0.007	0.086	0.000	[-0.002,0.009]	999	50	3171	0.202
Political experience	0.458	0.498	-0.159	[-0.644,0.201]	999	153	7854	0.304
Held office before	0.369	0.483	-0.125	[-0.684,0.274]	999	131	6939	0.401

Note: 1 and 2 report descriptive statistics. Column 3 reports local linear estimates of average treatment effects at cutoff estimated with triangular kernel weights and optimal MSE bandwidth (reported in column 7). Columns 4 and 8 report 95% robust confidence intervals and robust p-values computed following Cattaneo, Calonico, Tituinik (2013). Columns 5 and 6 report total observations and observations in optimal MSE bandwidth.

Table of Contents

Sharp and Fuzzy RDD: Falsification Checks

1. Sensitivity: Are results sensitive to alternative specifications?

2. Continuity Checks: Do predetermined covariates X jump at the threshold?
   ▶ Do placebo outcomes jump at the threshold?

3. Sorting: Do units sort around the threshold?

4. Check if jumps occur at placebo thresholds $c^\star$ (?)

RDD: Sensitivity to Specification

C. Nonlinearity mistaken for discontinuity
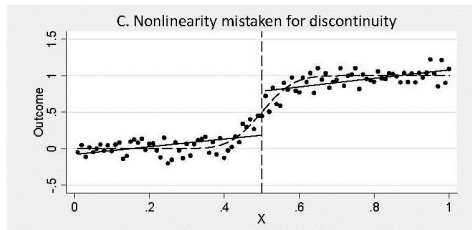
▶ $Y = f(X) + \alpha D + \varepsilon$: A miss-specified control function $f(X)$ can lead to a spurious jump: Take care not to confuse a nonlinear relation with a discontinuity

▶ More flexibility reduces bias, decreases efficiency. But higher flexibility leads to disappearance of actual discontinuities

▶ Local linear and local polynomials are standard. Gelmans and Imbens (2015)

▶ Check sensitivity to size of bandwidth (i.e. estimation window)

Example of graphical representation of result with different polynomials

Figure: Ruiz (2021): Effect of barely electing a politician on probability of donor getting a contract
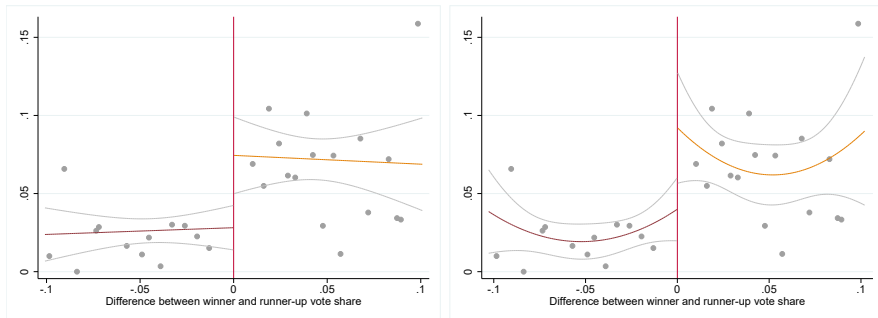
Left: linear fit. Right: quadratic fit.

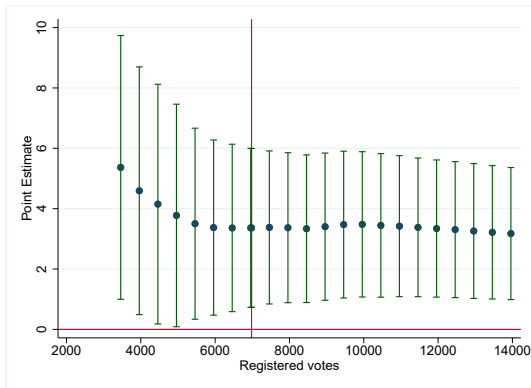# Estimate for different bandwidth sizes

Figure: Gulzar, Rueda and Ruiz (2021): Effect of loser campaign limits on number of contracts for donors

- ▶ Note: Red line denotes the optimal bandwidth
- ▶ Usually you estimate the effect for half of the optimal bandwidth to double the optimal bandwidth (or in the vicinity of the h)

Continuity Checks Test

- ▶ Test for comparability of agents around the cut-off:
  - ▶ What covariates (W) could vary at the threshold and correlated to the outcome?
  - ▶ Run the RDD regression using W as the outcome

- ▶ Finding a discontinuity in W does not necessarily invalidate the RDD
  - ▶ Can be a multiple hypothesis testing problem
  - ▶ Can incorporate W as additional controls into our main RDD regression. Ideally, this should only impact statistical significance, not the treatment effect

- ▶ Smoothness checks address only observables, not unobservables
  - ▶ On expectation, unobservables should be similar across the cut-off, if many observables do not jump discontinuously

Introduction
oo

IV Review
oooooooooo

Fuzzy RDD
ooooooooooooooooooo

(Fuzzy) Checks
o
ooooo●ooooooooo

Review
oooooo

Testing of discontinuities - Smooth covariates

Table: Municipality characteristics around campaign contribution limits cutoff

	Looser limits (1)	CI 95% (2)	Obs. (3)	Band. Obs. (4)	Bandwidth (5)	p-value (6)
Discretionary revenue	592.716	[-9000,8867.212]	970	76	4518.17	0.986
Municipal category	0.111	[-0.196,0.479]	999	61	3528.11	0.412
Mayor wages	-0.222	[-0.955,0.396]	999	61	3524.19	0.417
Council size	-0.354	[-1.134,0.286]	999	62	3563.62	0.241
Population	-448.213	[-4.5000,4011.520]	999	171	8786.33	0.907
Schools	60.152	[-34.789,176.287]	999	103	5767.05	0.189
Contracts	-87.087	[-880.423,686.785]	992	106	5989.31	0.809

Column 1 reports local linear estimates of average treatment effects at cutoff estimated with triangular kernel weights and optimal MSE bandwidth (reported in column 7). Columns 2 and 6 report 95% robust confidence intervals and robust p-values computed following Calonico, Cattaneo and Titiunik (2014). Columns 4 and 5 report total observations and observations in optimal MSE bandwidth. Discretionary income scaled in # of minimum monthly wages. Schools denotes all educational establishments.
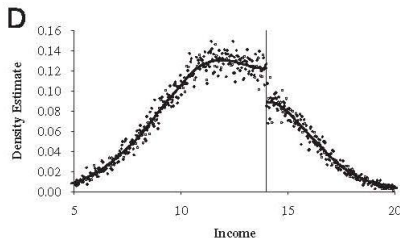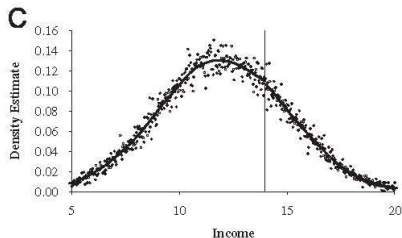
Figure: Gulzar, Rueda and Ruiz (2021)

Sorting Around the Threshold

▶ Can subjects behavior invalidate the local continuity assumption?
  ▶ Can they exercise control over their values of the assignment variable?
  ▶ Can administrators strategically choose what assignment variable to use or which cut-off point to pick?
  ▶ Either can invalidate the comparability of subjects near the threshold because of sorting of agents around the cut-off, where those below may differ form those just above
  ▶ Does not necessarily invalidate the design unless sorting is very widespread and very precise
  ▶ We look at the continuity of the distribution of the running variable at the cut-off

▶ What else changes at c? Continuity violated in the presence of other programs that use a discontinuous assignment rule with the exact same assignment variable and cut-off
  ▶ This can lead to a compound treatment

Sorting Around the Threshold

Example: Beneficial job training program offered to agents with income $< c$. Concern, people will withhold labor to lower their income below the cut-off to gain access to the program.

Introduction
oo

IV Review
000000000

Fuzzy RDD
0000000000000000

(Fuzzy) Checks
o
00000000●00000

Review
000000

Sorting Around the Threshold

▶ Test for discontinuity in density of forcing variable:

  ▶ Visual Histogram Inspection:

    ▶ Construct equal-sized non-overlapping bins of the forcing variable
      such that no bin includes points to both the left and right of the
      cut-off

    ▶ For each bin, compute the number of observations and plot the bins
      to see if there is a discontinuity at the cut-off

  ▶ Formal tests (e.g. McCrary, 2008)

  ▶ Cattaneo, Jansson and Ma (2018). Not dependent on the binning

Gulzar, Rueda and Ruiz (2021): Sorting tests
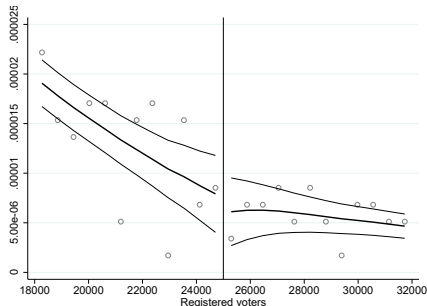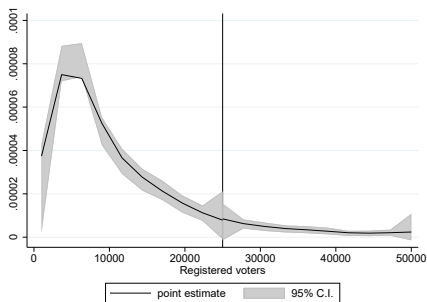
The left figure shows the density of the running variable. The test of no discontinuity at the cutoff (Cattaneo, Janson and Ma 2019) gives a statistic of -0.128 and a p-value of 0.98). The right figure presents the density graph in a narrower band around the cutoff. Dots represent averages of multiple observations.

Sorting Checks: The case of campaign finance limits
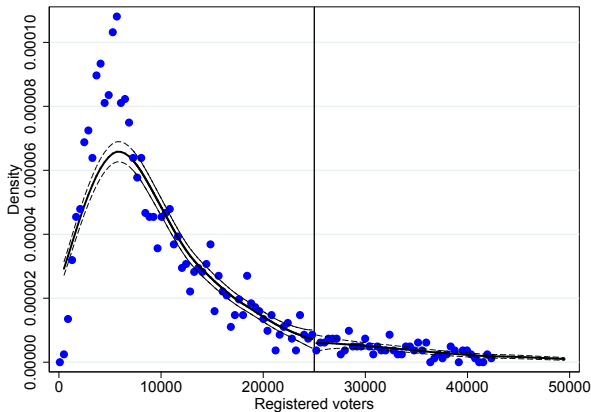
Note: Figure shows the density of the running variable on both sides of the threshold, binned averages and 95% confidence intervals. The discontinuity estimate (log difference in height) is -0.188 with standard error of 0.321

Placebo Threshold

▶ Test whether the treatment effect is zero when it should be (other $c^\star$ instead of policy cut-off c)

▶ Let $c^\star$ be a placebo threshold value. Run the regression of:
$E[Y|X, D] = \beta_0 + \beta_1(X - c^\star) + \alpha D + \beta_3((X - c^\star) \cdot D)$
and check if α large and significant?

  ▶ We split the sample to the left and the right of the actual threshold c in order to avoid miss-specification by imposing a zero jump at c

▶ The existence of large placebo jumps does not invalidate the RDD, but does require an explanation

▶ Concern is that the relation is fundamentally discontinuous and jump at cut-off is contaminated by other factors

RD check list

1. Understand the context: Is there an arbitrary cut-off?

2. Why was the cut-off designed? Is there compliance?

3. Can you obtain the data?

4. Are there enough observations close to the cut-off?

5. Are there possibilities for manipulation?
   - ▶ Check the manipulation tests

6. Plot the regression discontinuity - patterns should striking

7. Run the first regression discontinuity, are the results robust?
   - ▶ To different bandwidth sizes around the optimal bandwidth?
   - ▶ To different local polynomials? Usually order 1 and 2
   - ▶ Are other covariates jumping discontinously at the cut-off? Are they potentially related to treatment and outcome?

Beware bad RDDs

"Some political scientists define democracy as being 6 or higher in Polity score. I will study the effect of democracy on redistribution with an RDD using a 6 on Polity as the cutoff" (Eggers, 2020)

▶ What is democracy at 6?

▶ Can the coding be reproduced?

▶ What treatment does passing the 6 score implies?

▶ Are there enough units close to the cut-off?

▶ Is there sorting on the cut-off?

▶ Would this ever make sense?

Variation - Conditioning

Introduction
oo

IV Review
ooooooooo

Fuzzy RDD
ooooooooooooooooo

(Fuzzy) Checks
o
ooooooooooooooo

Review
o●oooo

Variation - Matching

Variation - FE

Variation - DiD

Variation - IV

Introduction
oo

IV Review
ooooooooo

Fuzzy RDD
oooooooooooooooo

(Fuzzy) Checks
o
oooooooooooooo

Review
ooooo●

Variation - RDD