STA 640 — Causal Inference

Chapter 7. Regression Discontinuity Designs

Fan Li

Department of Statistical Science
Duke University

Regression discontinuity design (RDD)

- The regression discontinuity designs (RDD) is a quasi-experimental design
- In this designs, the treatment status changes discontinuously according to some underlying pre-treatment variable (so-called *forcing variable* or *running variable*)
- Basic idea: in these studies, comparing units with similar values of this variable, but different levels of treatment would lead to causal effect of the treatment at the threshold
- The discontinuity is often created by a pre-fixed, artificial threshold of a pre-treatment variable

RDD: Examples

Example 1

- An educational program where the eligibility of a student depends solely on whether his/her test score of an exam is above or below a threshold
- Arguably students whose score are just above and students whose score are just below the threshold are comparable in their background (e.g., learning abilities and attitudes)

Example 2

- College financial aid where the eligibility of a student depends solely on whether his/her family income is above or below a threshold
- Arguably students whose family income are just above and students whose family income are just below the threshold are comparable in their background

RDD Formulation

- RDD: the probability of receiving the treatment changes discontinuously at a cutoff point s_0 of a covariate S (the running variable):

$$\Pr(Z = 1 \mid s_0^+) \neq \Pr(Z = 1 \mid s_0^-),$$

  where $\Pr(Z = 1 \mid s_0^+) = \lim_{s \downarrow s_0} \Pr(Z = 1 \mid S = s)$, and $\Pr(Z = 1 \mid s_0^-) = \lim_{s \uparrow s_0} \Pr(Z = 1 \mid S = s)$

- First introduced by Thistlewaite and Campbell (1960) in psychology

- Regain popularity since the 90s, especially in economics (reviews in Imbens and Lemieux, 2008, Lee and Lemieux, 2010)

- Two types: Sharp RD (SRD) and Fuzzy RD (FRD)

Sharp RD

- Treatment status is a deterministic step function of a running variable: $\Pr(Z = 1 \mid s_0^+) = 0, \quad \Pr(Z = 1 \mid s_0^-) = 1$ (or 0-1 reversed)

G.W. Imbens, T. Lemieux / Journal of Econometrics 142 (2008) 615–635

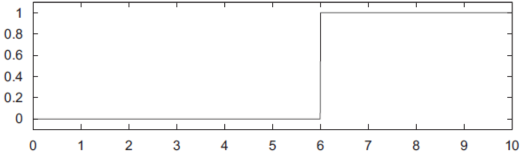


Fig. 1. Assignment probabilities (SRD).

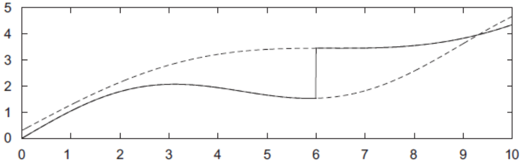


Fig. 2. Potential and observed outcome regression functions.

- Original form of RDDs, both previous examples are sharp RDD

Fuzzy RD

- Treatment status is a monotone function of the running variable with a jump (but not a 0-1 jump) at the threshold:

$$\Pr(Z = 1 \mid s_0^+) > \Pr(Z = 1 \mid s_0^-)$$

(or reversed sign)

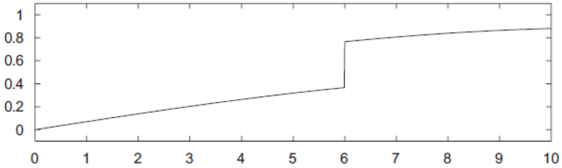
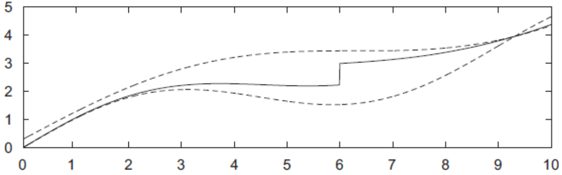


Fig. 3. Assignment probabilities (FRD).



Fuzzy RD: Example

- Italian state universities offer some grants every year to eligible students
- Objective: give equal opportunity to achieve higher education to motivated students irrespective of their financial background
- Grant allocation rule: A student must (1) meet the eligibility criteria; and (2) apply for the grant
- Eligibility: an economic measurement of the student's family income and assets (running variable) falling below or above a pre-determined threshold (e.g. 15 000 euros).
- Application is voluntary, and also ineligible students may apply - raising self-selection (confounding) issues
- Policy questions: are the grants effective in preventing dropouts?

Fuzzy RD

- In fuzzy RD, a value of the running variable falling above or below the threshold acts as encouragement or incentive to take the treatment
- In fuzzy RD, the receipt of the treatment depends also on individual choices, possible confounding due to self-selection
- The observed outcome function is a weighted average of the $E[Y(1)|x]$ and $E[Y(0)|x]$ functions
- Fuzzy RDs are related to instrumental variable (Hahn, Todd, van der Klaauw (2001)): falling above or below the threshold can be viewed as an instrument

Two Frameworks to RDD

- Two frameworks to RDD in the literature:
  - Continuity-based: the dominant approach in the literature until recently
  - Local randomization: closer to potential outcome framework, newer and more flexible to extend to complex settings
- Different in conceptual and inferential procedures
- In practice, little difference in empirical results between the two methods (Mattei and Mealli, 2018)

Sharp RD: Continuity Approach

- S_i: the running variable; s_0 is the threshold/cutoff
- Z_i: binary treatment indicator for unit i, $Z_i = \mathbf{1}(S_i \le s_0)$
- Assumption 1. (Continuity of Conditional Regression Functions): $\mathbb{E}[Y(0)|S = s]$ and $\mathbb{E}[Y(1)|S = s]$ are continuous in s
- Under Assumption 1:

$$\mathbb{E}[Y(0)|S = s_0] = \lim_{s \uparrow s_0} \mathbb{E}[Y(0)|S = s] = \lim_{s \uparrow s_0} \mathbb{E}[Y(0)|Z = 0, S = s] = \lim_{s \uparrow s_0} \mathbb{E}[Y|S = s]$$

and similarly

$$\mathbb{E}[Y(1)|S = s_0] = \lim_{s \downarrow s_0} \mathbb{E}[Y|S = s]$$

- The estimand is the average treatment effect at s_0, $\tau_{\mathrm{RD}^{\mathrm{true}}}$,

$$\tau_{\mathrm{RD}^{\mathrm{true}}} = \lim_{s \uparrow s_0} \mathbb{E}[Y|S = s] - \lim_{s \downarrow s_0} \mathbb{E}[Y|S = s]$$

- The estimand is the causal effect at the threshold

Sharp RD: Estimation

- Fit nonparametric or local linear regressions (Fan and Gijbels, 1996) to the data at each side of the threshold
- For Sharp RD, only need to fit for Y: denote the regression line $E[Y^{obs}|S = s]$ on the left and right side of threshold as $f_-(s)$, $f_+(s)$, respectively
- The point estimate of the SRD estimand is the difference between the limits of the two regression lines of Y at the threshold

$$\hat{\tau}_{\mathrm{RD}^{\mathrm{true}}} = \hat{f}_+(s_0) - \hat{f}_-(s_0)$$

- Bandwidth selection? Later

Sharp RD: Remarks

- Sharp RD completely violates the overlap assumption: there is zero overlap in the treatment probability at the threshold
- Continuity is not directly testable, because it involves counterfactuals $Y(0)$ and $Y(1)$ across the cutoff, but we only observe each curve at one side
- Continuity can be indirectly tested, e.g. distributions of all pre-treatment variables should be very similar on either side of the cutoff
- So we should check continuity of all covariates around the cutoff
- A nice overview of sharp RD is given in Imbens and Lemieux (2008, JOE)

Manipulation of running variable

- A concern in RDD is the possibility of subjects manipulating the running variable around the cutoff to just above or below, and thus the treatment assignment is confounded
- This is a concern especially if the exact value of the cutoff is known to the subjects in advance
- This type of behavior, if it exists, may create a discontinuity in the distribution of S at the cutoff (i.e., “bunching” to the right or to the left of the cutoff)

Manipulation of running variable

- McCrary (2008) test (also see Lee (2008)):
  - Basic idea: the marginal density of running variable S should be continuous without manipulation, and hence we look for discontinuities in the density around the threshold
- In principle, RD design does not require continuity of the density of S at s_0, but a discontinuity is suggestive of possible manipulation

Continuity of other covariates

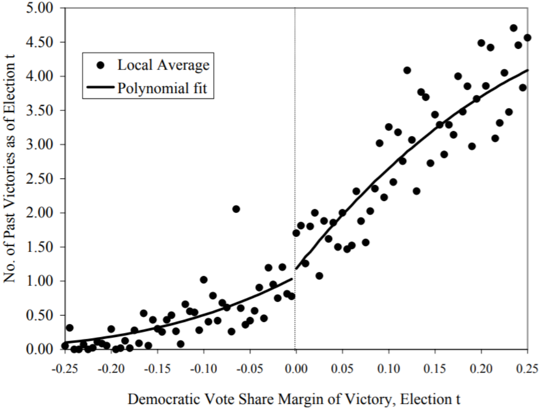
- Also need to check continuity of other covariates around the cutoff – should be continuous
- Test the null hypothesis of a zero average effect on pseudo-outcomes (falsification test) known not to be affected by the treatment
- Such variables include pre-treatment covariates that are by definition not affected by the treatment
- Although not required for the validity of the design, in most cases, the reason for the discontinuity in the probability of the treatment does not suggest a discontinuity in the average value of covariates. If we find such a discontinuity, it typically casts doubt on the assumptions underlying the RD design

SRD Example: Incumbency advantage in Elections

Lee (2008, JOE)

Lee (2008, JOE) uses the RDD to estimate party incumbency advantage in U.S. House elections

Figure IIb: Candidate's Accumulated Number of Past Election Victories, by Margin of Victory in Election t: local averages and parametric fit



SRD Example: Incumbency advantage in Elections

Lee (2008, JOE)

DO VOTERS AFFECT OR ELECT POLICIES? 835

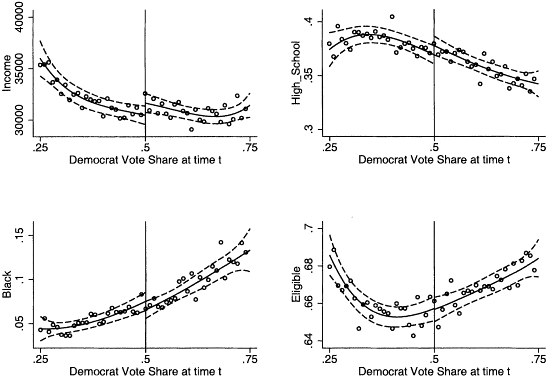


FIGURE III
Similarity of Constituents' Characteristics in Bare Democrat and Republican Districts–Part 1

Figure: Continuity of other covariates in Lee (2008)

SRD Example: Incumbency advantage in Elections

Lee (2008, JOE)

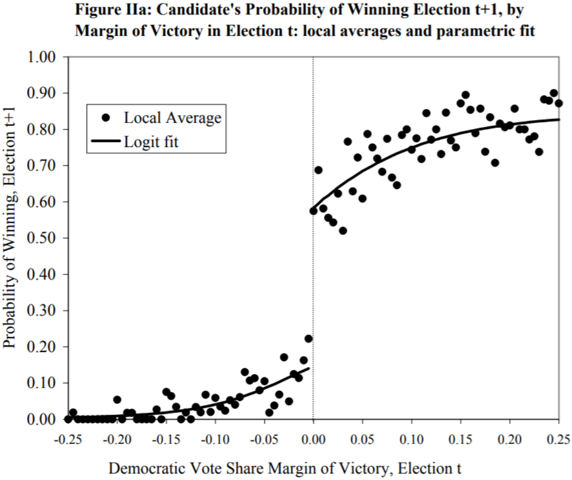


Figure: Discontinuity in outcome in Lee (2008)

Continuity Approach: Bandwidth selection

- The key of implementation is to select a bandwidth h around the cutoff (inference is impossible at a single point)
- In local linear regression, choice of bandwidth boils down to a bias-variance tradeoff
  - Smaller h: closer to threshold so less bias, but smaller sample size so more variance
  - Larger h: further away from threshold so more bias, but bigger sample size so less variance
- Choice of *optimal* bandwidth: achieve balance in the bias-variance tradeoff – minimizing the mean squared error in estimating the treatment effects
- Imbens and Kalyanaraman, 2012: method for setting optimal bandwidth in local linear regression (R, STATA packages are available)
- Standard errors: asymptotic variances

Fuzzy RD: Continuity Approach

- Z_i: assignment (eligibility) indicator for unit i, $Z_i = \mathbf{1}(S_i \leq s_0)$
- D_i: treatment received - a post-assignment variable
- FRD: Smaller than the $0-1$ jump in the probability of assignment to the treatment at the threshold
- FRD estimand: τ_{FRD}

$$\tau_{\text{FRD}} = \frac{\lim_{s \uparrow s_0} \mathbb{E}[Y|S=s] - \lim_{s \downarrow s_0} \mathbb{E}[Y|S=s]}{\lim_{s \uparrow s_0} \mathbb{E}[D|S=s] - \lim_{s \downarrow s_0} \mathbb{E}[D|S=s]}$$

- Hahn, Todd, van der Klaauw (2001) make the connection between FRD and IV: FRD is analogous to randomized experiment with noncompliance at cutoff
- Assumption 2 (Monotonicity): $D_i(s)$ is monotone in s at $s = s_0$
- Under monotonicity and continuity: $\tau_{\text{FRD}} = \tau_{\text{CACE}}(s_0)$

Fuzzy RD: Estimation

- Fit nonparametric or local linear regressions (Fan and Gijbels, 1996) to the data at each side of the threshold, need to fit separately for Y and D:
  - Fit for observed Y: denote the regression line $E[Y|S=s]$ on the left and right side of threshold as $f_-(s)$, $f_+(s)$, respectively
  - Fit for observed D: denote the regression line $E[D|S=s]$ on the left and right side of threshold as $g_-(s)$, $g_+(s)$ respectively

- The point estimate of the FRD estimand is ratio of the difference between the limits of the two regression lines of Y and of D at the threshold

$$\hat{\tau}_{\text{FRD}} = \frac{\hat{f}_+(s_0) - \hat{f}_-(s_0)}{\hat{g}_+(s_0) - \hat{g}_-(s_0)}$$

Fuzzy RD: implementation

- For a given bandwidth h, use the following procedure to estimate the FRD estimand (essentially a complier average causal effect)
  1. Create a left variable L: $L_i = S_i$ if $Z_i = 0$ and $L_i = 0$ if $Z_i = 1$; and a right variable R: $R_i = 0$ if $Z_i = 0$ and $R_i = S_i$ if $Z_i = 1$.
  2. Run a linear regression of $D \sim lm(1 + Z + L + R)$ with intercept using units within bandwidth h, and get the fitted values of D, $\widehat{D}$
  3. Run a linear regression of $Y \sim lm(1 + \widehat{D} + L + R)$ using units within bandwidth h. The coefficient of $\widehat{D}$ is the FRD estimate with bandwidth h.
- Perform Step 2-3 for each h value with corresponding 95% confidence interval. Draw the estimated effect with CI as a function of h, for the whole range of h.
- Note: Step 2 and 3 is equivalent to conduct a separate linear regression on each side of the threshold. Think about why.

R package: *rdrobust*

- A nice R package on RD analysis: *rdrobust*, by Calonico, Cattaneo, Farrell, Titiunik
- Tools for data-driven graphical and analytical statistical inference in RD
- Function `rdrobust()`: conduct local-polynomial point estimators and robust confidence interval:
- Function `rdbwselect()`: provide bandwidth selection

Example: Italian university grant – revisited

Li, Mattei, Mealli, 2015, AOAS

- Administrative data of 1st year students in Universities of Florence and Pisa in year 2004 – 2006
- assignment rule: only applicants with the economic measure S below a fixed threshold s_0 get the grant
- Running variable: A combined economic measure of family income and asset. Threshold: 15 000 euros
- The threshold is set by the administration and the running variable is hard to manipulate
- Covariates: Sex, hight school type, high school grade, enrollment year, university, major in university
- There is a discontinuity in the treatment probability at s_0

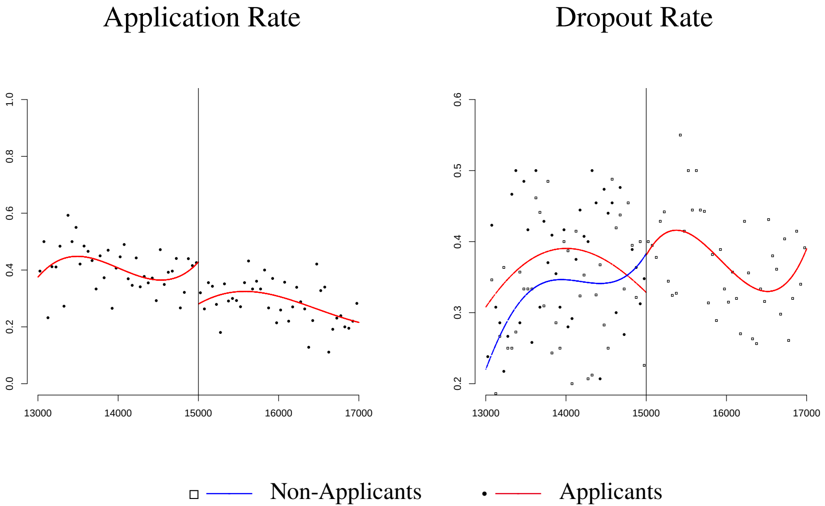
Italian university grants: SRD vs. FRD

- This can be analyzed as either SRD or FRD
- The key: not everyone who is eligible applied for the grant, but indeed students who are eligible are more likely to apply and thus receive the grant
  - Sharp RD: If we restrict to the sample who applied for the grants – below and above 15000 euro leads to treatment and control
  - Fuzzy RD: If we analyze all the sample, regardless of application status
- SRD estimation is straightforward, we focus on FRD to illustrate
- Eligibility: Z (1 yes; 0 no); Application: A (1 apply; 0 not)
- Treatment status: $D = ZA$, one has to both be eligible and apply to receive the aid

Summary Statistics

			$Z = 0$		$Z = 1$	
Variable	$Z = 0$	$Z = 1$	$A = 0$	$A = 1$	$A = 0$	$A = 1$
Sample size	4 281	12 080	3 215	1 066	5 759	6 321
Dropout	0.36	0.39	0.33	0.44	0.36	0.42
S (Euro)	17 373	8 942	17 445	17 157	9 818	8 144
Female	0.58	0.60	0.57	0.59	0.61	0.60
HS Grade	81.94	80.84	81.06	84.57	79.16	82.38

Graphical Illustration: Italian University Grants



Example of Fuzzy RD: Italian university grant

Results from two separate OLS on each side for whole h

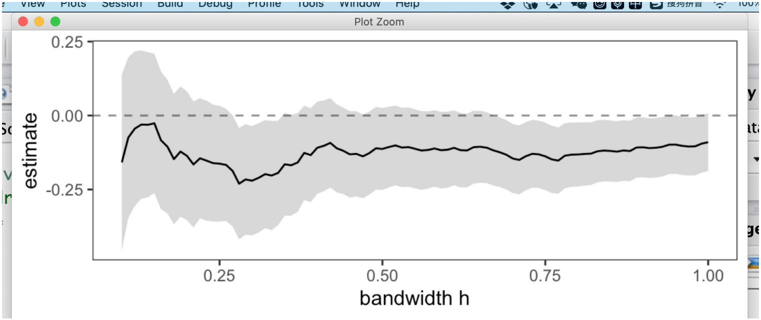


Figure: treatment effects as a function of h for the whole range of h in the Italian college aid example (Fuzzy setting)

Example of Fuzzy RD: Italian university grant

Results from the `rdrobust` package

```r
## data-driven bandwidth
library("rdrobust")
frd_italy = with(italy,
                 {
                   rdrobust(y = outcome,
                            x = rv0,
                            c = 0,
                            fuzzy = D)
                 })
cbind(frd_italy$coef, frd_italy$se)
```

Table 1. Characteristics of the Patients at Baseline.*

Characteristic	81-mg Group (N=7540)	325-mg Group (N=7536)
Median age (IQR) — yr	67.7 (60.7–73.6)	67.5 (60.7–73.5)
Female sex — no. (%)	2307 (30.6)	2417 (32.1)
Median weight (IQR) — kg	90.0 (78.6–103.6)	90.0 (78.2–104.1)
Race — no. (%)		
White	6014 (79.8)	5976 (79.3)
Black	664 (8.8)	647 (8.6)
Hispanic ethnic group — no. (%)	249 (3.3)	232 (3.1)
Medical history — no. (%)†		
Previous myocardial infarction	2674 (35.6)	2631 (35.0)
Previous coronary revascularization	4034 (53.6)	3943 (52.4)
Previous percutaneous coronary intervention	3005 (40.0)	2941 (39.1)
Previous coronary-artery bypass grafting	1786 (23.8)	1741 (23.2)
Hypertension	6264 (83.3)	6248 (83.1)
Dyslipidemia	6472 (86.1)	6474 (86.1)
Diabetes mellitus	2820 (37.5)	2856 (38.0)
Atrial fibrillation	605 (8.0)	628 (8.4)
Congestive heart failure	1718 (22.8)	1786 (23.8)
Chronic kidney disease	1315 (17.5)	1333 (17.7)
Peripheral artery disease	1706 (22.7)	1787 (23.8)
Previous clinically significant gastrointestinal bleeding	455 (6.1)	495 (6.6)
Previous intracranial hemorrhage	98 (1.3)	110 (1.5)

Local Randomization Approach to RDD

- Framework 2: Local randomization (Lee and Card, 2008; Catteneo et al. 2015):
  - Idea: treatment status is *as good as randomly assigned* among the subsample of observations that fall just above and just below the threshold *near the boundary* of the running variable
  - Implementation: find the neighborhood around the threshold where (local) randomization is plausible, and proceed as in a RCT within the neighbor
  - Closer to potential outcome framework, easier to extend to complex settings (e.g. ordinal running variables, spatial discontinuity)

Sharp RD: Local Randomization Approach

Catteneo, Frandsen, Titiunik, 2015; Li, Mattei, Mealli, 2015)

- Alternative view: the running variable as a random variable
- Assumption 1. Local overlap: There exists a subset of units, $\mathcal{U}_{s_0}$, in the the random sample (or population) in the study, such that for each $i \in \mathcal{U}_{s_0}$, $\Pr(S_i \leq s_0) > \epsilon$ and $\Pr(S_i > s_0) > \epsilon$ for some sufficiently large $\epsilon > 0$.
- Assumption 2. Local SRD-SUTVA: For each $i \in \mathcal{U}_{s_0}$, consider two treatment levels $D_i^{'} = \mathbf{1}(S_i^{\prime} \leq s_0)$ and $D_i^{''} = \mathbf{1}(S_i^{''} \leq s_0)$, with possibly $S_i^{'} \neq S_i^{''}$. If $D_i^{'} = D_i^{''}$, that is, if either $S_i^{'} \leq s_0$ and $S_i^{''} \leq s_0$, or $S_i^{'} > s_0$ and $S_i^{''} > s_0$, then $Y_i(\mathbf{D}^{'}) = Y_i(\mathbf{D}^{''})$.
- Under local SRD-SUTVA, for each unit within $\mathcal{U}_{s_0}$ there exist only two potential outcomes for each post-treatment variable
- We can define local versions of common estimands within $\mathcal{U}_{s_0}$,

$$\tau_{s_0} \equiv \mathbb{E}\left[Y_i(1) - Y_i(0) \mid i \in \mathcal{U}_{s_0}\right].$$

Local Randomization/Unconfoundedness

Assumption 3. Local randomization: within $\mathcal{U}_{s_0}$

$$\Pr\left(S_i | Y_i(0), Y_i(1), \mathbf{X}_i\right) = \Pr\left(S_i\right).$$

- Under local randomization:
  $\tau_{s_0} = \mathbb{E}[Y_i | D_i = 1, i \in \mathcal{U}_{s_0}] - \mathbb{E}[Y_i | D_i = 0, i \in \mathcal{U}_{s_0}]$
- Estimation does not hinge on continuity at s_0: τ_{s_0} can be estimated directly from the difference in the mean outcome of the units in the two sides of s_0 in $\mathcal{U}_{s_0}$
- Local randomization is a stronger assumption than continuity; a special case of $\mathcal{U}_{s_0}$ is a single point s_0
- Local randomization can be relaxed to local unconfoundedness (allowing RHS conditional on covariates)

$$\Pr\left(S_i | Y_i(0), Y_i(1), \mathbf{X}_i\right) = \Pr\left(S_i | \mathbf{X}_i\right), \quad i \in \mathcal{U}_{s_0}.$$

Fuzzy RD: Basic Setup

- S_i: the running variable; s_0 is the threshold
- Z_i: assignment (eligibility) indicator for unit i, $Z_i = \mathbf{1}(S_i \leq s_0)$
- D_i: treatment received - a post-assignment variable
- Modify assumptions
  - Assumption 2'. Local FRD-SUTVA: For each $i \in \mathcal{U}_{s_0}$, consider two eligibility statuses $Z_i^{'} = \mathbf{1}(S_i^{\prime} \leq s_0)$ and $Z_i^{''} = \mathbf{1}(S_i^{''} \leq s_0)$, with possibly $S_i^{'} \neq S_i^{''}$. If $Z_i^{'} = Z_i^{''}$, that is, if either $S_i^{'} \leq s_0$ and $S_i^{''} \leq s_0$, or $S_i^{'} > s_0$ and $S_i^{''} > s_0$, then $D_i(\mathbf{Z}^{'}) = D_i(\mathbf{Z}^{''})$, and $Y_i(\mathbf{Z}^{'}) = Y_i(\mathbf{Z}^{''})$.
  - Assumption 3'. Local Randomization (unconfoundedness: within $\mathcal{U}_{s_0}$, $\Pr\left(S_i | D_i(0), D_i(1), Y_i(0), Y_i(1), \mathbf{X}_i\right) = \Pr\left(S_i | \mathbf{X}_i\right)$.

Fuzzy RD and IV

- Within $\mathcal{U}_{s_0}$ and under Assumption 2'-3', fuzzy RD is a special case of the noncompliance/IV: Eligibility Z is an IV
- Four principal strata based on the intermediate variable D, $G_i = (D_i(0), D_i(1))$: compliers, always-takers, never-takers, and defiers
- Similar as before, assume monotonicity and exclusion restriction
- Causal Estimand:

$$\tau_{g,s_0} \equiv \mathbb{E}\left[Y_i(1) - Y_i(0) \mid G_i = g, i \in \mathcal{U}_{s_0}\right], \quad (1)$$

- Compliers are the only group that is informative about the effect of receiving the treatment

Local Randomization: Bandwidth Selection

- Key of RD analysis in local randomization approach: Selection of the subpopulation $\mathcal{U}_{s_0}$ where the RDD assumptions
- There can be a diverse choice of the shape of $\mathcal{U}_{s_0}$
- A convenient choice is to limit to symmetric intervals about s_0, $(s_0 - h, s_0 + h)$, then selecting $\mathcal{U}_{s_0}$ reduces to determining plausible values of h
- Under randomization, covariates of units in $\mathcal{U}_{s_0}$ should be well balanced in the two subsamples defined by assignment Z
- Thus any test of the null hypothesis of no effect of assignment on covariates should fail to reject the null (Lee, 2008; Cattaneo et al. 2015)

Local Randomization: Bandwidth Selection

- Each h such that the sharp null hypothesis of no effect of assignment on covariates cannot be rejected defines a potential subpopulation, $\mathcal{U}_{s_0}$
- Rejection of the sharp null hypothesis can be interpreted as evidence against the local randomization assumption
- Procedure of randomization tests: start with a small bandwidth h around the threshold and then increase h until the null hypothesis is rejected (Cattaneo et al. 2015, JCI)
- Choice of the covariates: important to consider all variables known (believed) to be related to both assignment and the outcome

Bandwidth Selection

- Both continuity and local randomization approaches to RDD require bandwidth selection, but with different goals and procedures: the former tries to achieve balance in bias-variance tradeoff; the latter tries to identify a region where the local randomization hold
- Regardless of the approach, important points for bandwidth selection:
  - Try multiple choices of bandwidth, the optimal bandwidth is just a starting point; sensitivity to the bandwidth raises concerns of credibility of the results – In fact, I would recommend to repeat the RD analysis for all bandwidth and draw a graph of h vs. treatment effect – get a much better global view
  - The optimal bandwidth for testing discontinuities in covariates (e.g. McCrary test) may not be the same as the optimal bandwidth for the treatment

Mode of Inference

Once $\mathcal{U}_{s_0}$ is selected, statistical inference can be proceeded via several modes:

- Fisherian randomization inference of sharp null of no effects
- Frequentist inference: point estimate and confidence interval of an estimand
- Moment-based instrumental variable (IV) approach
- Bayesian (model-based) approach

Simplicity in implementation is important

Frequentist inference: Propensity Score

- Once $\mathcal{U}_{s_0}$ is selected, within $\mathcal{U}_{s_0}$ we can use standard methods learned before
- Estimand (local to $\mathcal{U}_{s_0}$): ATE , or weighted ATE such as ATT, ATO
- Estimator: moment estimators – difference-in-means estimator,
- What if some covariates are not balanced (e.g. conditioning set of the local unconfoundedness assumption)?
  - Propensity score weighting estimator: in sharp RD, propensity score is the probability of receiving treatment (i.e. running variable above (or below) the threshold (Li et al. 2020)
  - Consider combining outcome model with PS weighting: augmented (or double-robust) estimators, where the imbalanced covariates are predictors in both PS and outcome model
- Below we illustrate the randomization approach to FRD (SRD is straightforward)

Example: Italian university grant – revisited

Li, Mattei, Mealli, 2015, AOAS

- The key: not everyone who is eligible applied for the grant, but indeed students who are eligible are more likely to apply and thus receive the grant
  - Sharp RD: If we restrict to the sample who applied for the grants – below and above 15000 euro leads to treatment and control
  - Fuzzy RD: If we analyze all the sample, regardless of application. It has two principal strata:
    - Never-takers: Subjects who would not get a grant irrespective of their eligibility status

$$G = n = (0,0) \equiv \{i : D_i(0) = 0, D_i(1) = 0\}$$

    - Compliers: Subjects who would not get a grant if ineligible and would get a grant if eligible

$$G = c = (0,1) \equiv \{i : D_i(0) = 0, D_i(1) = 1\}$$

Randomization test: P-values of difference-in-means

bandwidth (h) Variable	500 (n = 1 042)	1 000 (n = 2 108)	1 500 (n = 3 166)	2 000 (n = 4 197)	ALL (n = 16 361)
Sex	.574	.629	.741	1.000	.003
HS Type (Other)					
Humanity	.563	.387	.632	.779	.308
Science	1.000	.436	.794	.143	.000
Tech	.846	.587	.974	.750	.000
HS Grade	.249	.879	.658	.686	.000
Year (2004)					
2005	.943	.174	.469	.847	.006
2006	.833	.517	.904	1.000	.000
University (PI versus FI)	.325	.623	.007	.001	.000
Major in University (Other)					
Humanity	.202	.023	.034	.146	.214
Science	1.000	.855	.376	.344	.429
Social Science	.942	.792	.933	1.000	.000
BioMed	.328	.074	.351	.389	.487
Tech	.196	.242	.455	.874	.000

0.05 h | 500 1 000 1 500 2 000 ALL

Posterior probabilities of zero mean differences between eligible groups

bandwidth (h) Variable	500 ($n=1\,042$)	1 000 ($n=2\,108$)	1 500 ($n=3\,166$)	2 000 ($n=4\,197$)	5 000 ($n=9\,846$)
Sex	0.95	0.96	0.98	0.97	0.80
HS Type (Other)					
Humanity	0.95	0.96	0.98	0.97	0.96
Science	0.91	0.93	0.95	0.89	0.00
Tech	0.81	0.81	0.82	0.62	0.00
HS Grade	0.96	0.98	0.97	0.98	0.99
Year(2004)					
2005	0.96	0.93	0.97	0.98	0.92
2006	0.92	0.91	0.96	0.93	0.88
University (PI versus FI)	0.92	0.98	0.69	0.10	0.00
Major in University (Other)					
Humanity	0.90	0.80	0.80	0.93	0.95
Science	0.86	0.75	0.78	0.90	0.91
Social Science	0.82	0.71	0.76	0.86	0.86
BioMed	0.78	0.68	0.74	0.84	0.83
Tech	0.63	0.62	0.70	0.79	0.45

FRD: Model Specification

- Model for principal strata (standard probit):

$$\begin{aligned} \Pr(G_i = c) &= \Pr(G_i^*(c) \leq 0), \\ \Pr(G_i = n) &= 1 - \Pr(G_i = c), \end{aligned}$$

where

$$G_i^*(c) = \alpha_0 + S_i \boldsymbol{\alpha}^{(S)} + \mathbf{X}_i' \boldsymbol{\alpha}^{(X)} + \epsilon_i, \quad \epsilon_i \sim N(0, 1)$$

- Model for outcome (probit link): for $g = c, n$

$$\Pr(Y_i(z) = 1 | G_i = g, S_i, \mathbf{X}_i) = \Phi\left(\beta_{0,g,z} + S_i \boldsymbol{\beta}_{g,z}^{(S)} + \mathbf{X}_i' \boldsymbol{\beta}_{g,z}^{(X)}\right)$$

- We assume (1) parameters are *a priori* independent, and (2) multivariate normal priors for the coefficients with mean 0 and large variances

FRD: Results

Estimand	Population-average Median	Population-average 2.5%	Population-average 97.5%	Sample-average Median	Sample-average 2.5%	Sample-average 97.5%
$h = 500$						
$\Pr(G_i = c)$	.384	.291	.450	.363	.288	.426
$\Pr(G_i = n)$	.616	.570	.650	.637	.590	.651
CACE	−.116	−.253	−.005	−.120	−.265	−.009
$h = 1\,000$						
$\Pr(G_i = c)$	.377	.310	.451	.360	.316	.429
$\Pr(G_i = n)$	.623	.584	.652	.640	.599	.645
CACE	−.132	−.242	−.021	−.139	−.247	−.034
$h = 1\,500$						
$\Pr(G_i = a)$	.375	.285	.416	.358	.306	.403
$\Pr(G_i = n)$	.625	.591	.642	.642	.605	.644
CACE	−.153	−.256	−.040	−.165	−.266	−.057

Recap and comparisons of two frameworks to RDD

- Continuity
  - Main idea: the (potential) outcome function is continuous wrt the running variable *at* the threshold
  - Implementation: fit local regression or polynomials of the outcome model around the threshold, find a bandwidth to achieve the bias-variance tradeoff
- Local randomization/confoundedness
  - Implementation: find the neighborhood around the threshold where (local) randomization is plausible, and proceed as in a RCT within the neighborhood
  - Closer to potential outcome framework, easier to extend to complex settings
- In practice, little difference in empirical results between the two methods (Mattei and Mealli, 2018)

RDD Extensions

- Multiple running variables: e.g. test scores of two subjects (e.g. Imbens and Zajonc 2011; Cattaneo et al. 2016)
- Discrete running variable: rounding errors (Dong, 2015; Kolesar and Rothe, 2018; Imbens and Wager, 2019)
- Ordinal running variable: e.g. credit rating (Li et al. 2021)
- Spatial running variable: e.g. geographically implemented policies (see lecture slides of Christopher Moore 2009; Rischard et al. 2020)

A good recent review is Cattaneo, Idrobo, Titiunik (2017).

References

Hahn J. Todd P. E., Van der Klaauw W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. Econometrica, 69: 201-209.

Imbens G. W., Kalyanaraman K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. The Review of Economic Studies, 79: 933-959.

Imbens G. W., Lemieux T. (2008). Regression discontinuity designs: A guide to practice. Journal of Econometrics, 142: 615-635.

Lee, D. S. (2008). Randomized experiments from non-random selection in U.S. House elections. Journal of Econometrics 142: 675-697.

Lee, D. S. and Card, D. (2008). Regression discontinuity inference with specification error. Journal of Econometrics 142: 655-674.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. Journal of econometrics, 142(2), 698-714.

Thistlethwaite D., Campbell D. (1960). Regression-discontinuity analysis: an alternative to the ex-post facto experiment. Journal of Educational Psychology, 51: 309-317.

References

Dehejia, R. Regression Discontinuity: Advanced Topics. Lecture notes.

Imbens, G., Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. The Review of economic studies, 79(3), 933-959.

Cattaneo, M. D., Idrobo, N., Titiunik, R. (2017). A practical introduction to regression discontinuity designs. Cambridge Elements: Quantitative and Computational Methods for Social Science-Cambridge University Press I.

Cattaneo M. D., Frandsen B., Titiunik R. (2015). Randomization Inference in the Regression Discontinuity Design: An Application to Party Advantages in the U.S. Senate. Journal of Causal Inference, 3 1-24.

Li, F., Mattei, A. and Mealli, F. (2015). Evaluating the Causal Effect of University Grants on Student Dropout: Evidence from a Regression Discontinuity Design Using Principal Stratification. The Annals of Applied Statistics 9 1906-1931.

Mattei, A. and Mealli, F. (2016). Regression Discontinuity Designs as Local Randomized Experiments. Observational Studies 2 156-173.

Li F, Mercatanti A, Mäkinen T, Silvestrini, A. (2021). A regression discontinuity design for ordinal running variable: Evaluating Central Bank purchases of corporate bonds. *Annals of Applied Statistics*. 15(1), 304-322.