Quant II

Lab 6: RDD extensions

Giacomo Lemoli

March 21, 2022

Today's plan

Additional tools and popular RDD typologies

- Validation tests
- Covariates
- Fuzzy RDD
- Geographic RDD
- DiD + RDD = Diff-in-Disc
- Local randomization approach

Sources and material

We are still using Cattaneo, Idrobo, and Titiunik (2019) as our main source.

- Text
- Full code

In the second part: Cattaneo, Idrobo, and Titiunik (forthcoming).

- Text

When are RDD assumptions violated?

Key identification assumption behind RDD: potential outcomes are continuous at the cutoff.

$$\lim_{x \to c^+} E[Y_i(j)|X = x] = \lim_{x \to c^-} E[Y_i(j)|X = x] = E[Y_i(j)|X = c], j \in (0, 1)$$

In other words, units just to the left of the cutoff are virtually identical to those just to the right.

There are instances where this assumption can be plausibly violated.

When are RDD assumptions violated?

Quite simply: units know the cutoff determines the treatment assignment and try to get just above it.

In the Meyersson example: Islamic parties may invest more in winning the election in some important municipalities. If they were successful, units with a bare victory would be different from cities with a bare defeat, i.e. a discontinuity at the cutoff.

We cannot directly test the continuity assumptions because of the fundamental problem etc etc. But we can check for discontinuities in observable variables that we would expect to be continuous.

When are RDD assumptions violated?

Quite simply: units know the cutoff determines the treatment assignment and try to get just above it.

In the Meyersson example: Islamic parties may invest more in winning the election in some important municipalities. If they were successful, units with a bare victory would be different from cities with a bare defeat, i.e. a discontinuity at the cutoff.

We cannot directly test the continuity assumptions because of the fundamental problem etc etc. But we can check for discontinuities in observable variables that we would expect to be continuous.

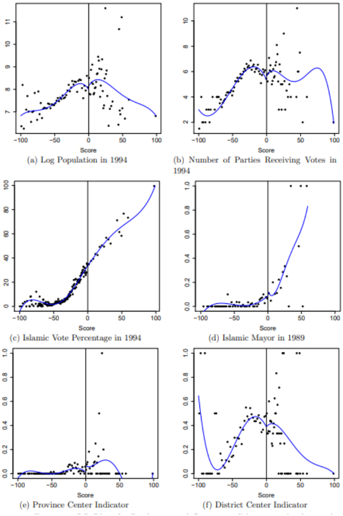
Placebos!

RDD with placebo outcomes

We run the same RDD design, but with outcomes that we know should not change discontinuously at the cutoff.

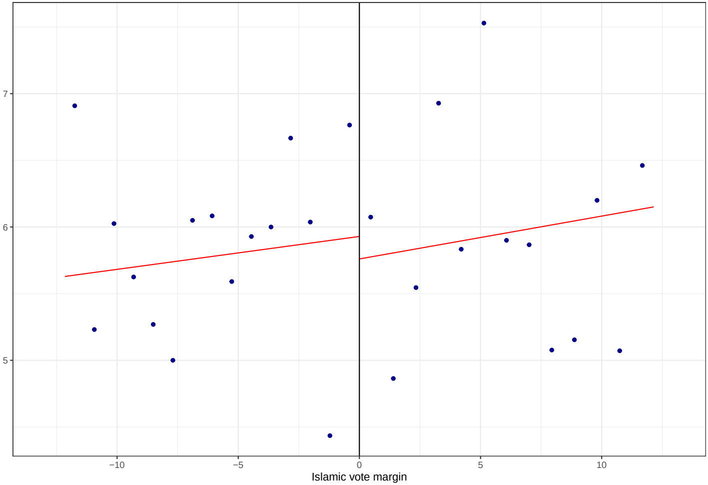
Pre-determined covariates or outcomes that the theory/design should predict to be unrelated to the treatment, but would be related if there was confounding.

RDD with placebo outcomes



RDD with placebo outcomes

```
bw <- rdrobust(data$partycount, X)$bws[1, 1]
xlim <- ceiling(bw)
rdplot(data$partycount[abs(X) <= bw], X[abs(X) <= bw],
      p = 1, kernel = "triangular", x.lim = c(-xlim, xlim), x.label = "Islamic vote margin", y.label = "", tit
```



Density of the running variable

Intuition: if units knew the position of the cutoff and could **exactly** manipulate their value of the running variable, we could observe sorting around the cutoff. It is natural to expect this would be associated to violations of continuity.

What are settings where units can manipulate their score?

Density tests

A formal test: estimate the density of the running variable with a local polynomial density estimator on the two sides of the cutoff. Then test the null that the density is the same on the two sides.

The classical test is the one proposed by McCrary (2008).

- Stata: user-written program `DCdensity` (to be downloaded)
- R: `DCdensity` in the package `rdd`

Another test is proposed by Cattaneo, Jansson and Ma (2020)

- `rddensity` in R and Stata

Density tests

```r
library(rddensity)

out <- rddensity(X)
summary(out)
```

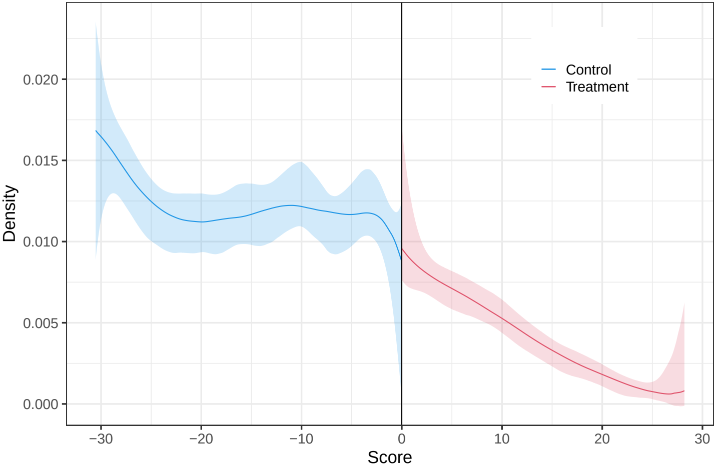
```
##
## Manipulation testing using local polynomial density estimation.
##
## Number of obs =       2629
## Model =               unrestricted
## Kernel =              triangular
## BW method =           estimated
## VCE method =          jackknife
##
## c = 0                 Left of c           Right of c
## Number of obs         2314                315
## Eff. Number of obs    965                 301
## Order est. (p)        2                   2
## Order bias (q)        3                   3
## BW est. (h)           30.539              28.287
##
## Method                T                   P > |T|
## Robust                -1.3937             0.1634
##
##
## P-values of binomial tests (H0: p=0.5).
##
## Window Length / 2          <c     >=c    P>|T|
## 0.426                      11       9    0.8238
## 0.852                      18      26    0.2912
## 1.278                      32      34    0.9022
## 1.704                      42      48    0.5984
## 2.130                      52      57    0.7018
```

Density tests

```r
library(lpdensity)

bw_left <- as.numeric(rddensity(X)$h[1])
bw_right <- as.numeric(rddensity(X)$h[2])

est1 <- lpdensity(data = X[X < 0 & X >= -bw_left], grid = seq(-bw_left, 0,
                                                               0.1),
                  bwselect = "IMSE", scale = sum(X < 0 & X >= -bw_left) / length(X))

est2 <- lpdensity(data = X[X >= 0 & X <= bw_right], grid = seq(0, bw_right,
                                                                0.1),
                  bwselect = "IMSE", scale = sum(X >= 0 & X <= bw_right) / length(X))
```

Density tests

```
library(ggplot2)
plot1 <- lpdensity.plot(est1, est2, CIshade = 0.2, lcol = c(4, 2), CIcol = c(4, 2), legendGroups = c("Control",
  labs(x = "Score", y = "Density") + geom_vline(xintercept = 0, color = "black") +
  theme_bw(base_size = 17)+theme(legend.position = c(0.8, 0.85))
plot1
```



RDD with placebo cutoffs

Idea of placebo treatments: replace the treatment with one that has no effect by construction but could suffer from the same confounding. If you find an effect, there is probably confounding also in the original treatment effect.

In RDD we can vary artificially the cutoff to see if there are discontinuities in the CEF at points where there is no real treatment.

Since we just need continuity *at the cutoff*, discontinuities away from it are not proof of bias (Cattaneo, Idrobo, and Titiunik 2019). But they do raise a flag.

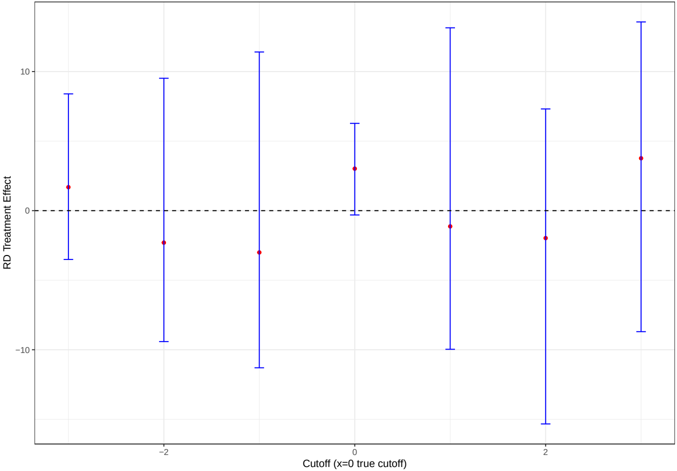
RDD with placebo cutoffs

```
placebo <- function(Y, X, new_cutoff){
  if (new_cutoff > 0){
    Y <- Y[X>=0]; X <- X[X>=0]
  }
  if (new_cutoff < 0){
    Y <- Y[X<0]; X <- X[X<0]
  }
  else{
    Y <- Y; X <- X
  }

  out <- rdrobust(Y, X, c = new_cutoff)
  coef <- out$coef["Conventional",]
  ll <- out$ci["Robust",1]
  ul <- out$ci["Robust",2]

  cbind(coef, ll, ul)
}

cutoffs <- as.list(c(-3:3))
(placebos <- do.call("rbind", lapply(cutoffs, function(i) placebo(Y, X, i))))
```

```
##           coef          ll        ul
## [1,]  1.687817  -3.5083421  8.397096
## [2,] -2.300012  -9.4137423  9.517862
## [3,] -3.004159 -11.2961682 11.407925
## [4,]  3.019526  -0.3092892  6.275769
## [5,] -1.130650  -9.9671538 13.146914
## [6,] -1.972790 -15.3333665  7.313371
## [7,]  3.766433  -8.6998298 13.569346
```

RDD with placebo cutoffs

```r
library(dplyr)
placebos %>% as.data.frame() %>% mutate(cutoff = -3:3) %>%
  ggplot(aes(x=cutoff, y=coef)) + geom_point(col="red") +
  geom_errorbar(aes(ymin=ll, ymax=ul), col="blue",width=0.1) +
  labs(y = "RD Treatment Effect", x = "Cutoff (x=0 true cutoff)") +
  geom_hline(yintercept=0, col="black", linetype = "dashed") + theme_bw()
```



Sensitivity checks

Cattaneo, Idrobo and Titiunik recommend to perform sensitivity analyses:

- Exclude points closer to the cutoff and check stability of the results
- Vary the bandwidth and check stability of the results
  - This should be done in a neighborhood of the optimal bandwidth, e.g. by varying the bandwidth selection criterion, without selecting manually too large or small values

Summing up

As for other research designs, RDD assumptions need to be defended. Common practice is to include

- Graphical and formal placebo tests with covariates and other outcomes
- Density tests for sorting around the cutoff
- Perturbate the cutoff values
- Exclude observations near the cutoff
- Vary the bandwidth choice

Covariates

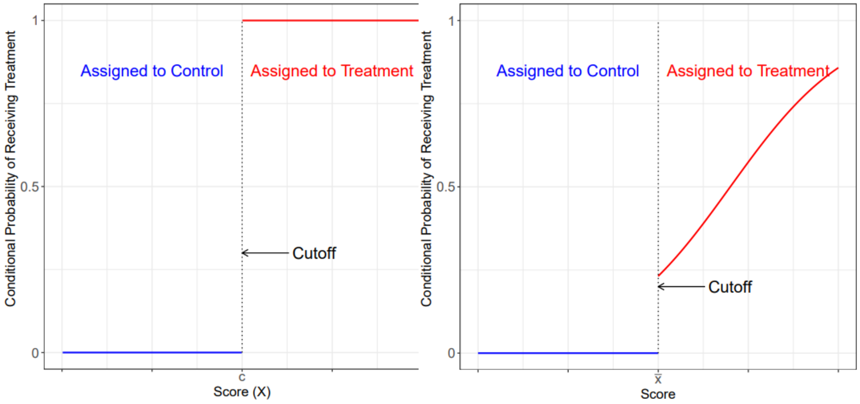
Can we add covariates to RDD estimation?

Calonico, Cattaneo, Farrell, and Titiunik (2019): study properties of RDD estimators with covariate adjustment.

- There can be an efficiency argument in favor of adjusting for covariates ($\sim$ randomized experiments)
- You can also invoke covariate adjustment for achieving identification, although this requires stronger parametric assumptions
  - In empirical work, often people include group fixed effects or unbalanced covariates
- CCFT recommend not to interact covariates with treatment indicator (allowing for different effects on the two sides requires stronger assumptions)
- In `rdrobust`: use the `covs` argument

Fuzzy RDD

Idea: endogenous treatment take-up or non-compliance.



(a) Sharp RD

(b) Fuzzy RD (One-Sided)

Fuzzy RDD

This is just an IV model where the treatment D is endogenous and the cutoff rule gives the instrument:

$$\frac{\lim_{x\to c^+} E[Y_i|X=x] - \lim_{x\to c^-} E[Y_i|X=x]}{\lim_{x\to c^+} E[D_i|X=x] - \lim_{x\to c^-} E[D_i|X=x]} = \tau_{FRD}$$

under the usual IV assumptions.

If the treatment effects are heterogeneous, τ_{FRD} can also have a LATE interpretation.

Once people could have used IV regression. In `rdrobust`: argument `fuzzy`.

Geographic RDD

There are cases where treatment assignment depends on more than one running variable: e.g. admissions based on multiple test scores or policy eligibility based on more than one criteria.

We focus here on Geographic RDD, which is a popular and widely used one.

In essence, treatment is defined as being on a given side of a geographic boundary (intuition with a long history in CP).

Why do we have more than one running variable?

Geographic RDD

Basic differences with respect to standard RDD:

- Continuity along both dimensions
- GRD identifies an effect at every boundary point: a *treatment effect curve*
- One could estimate the effect at different points of the boundary to characterize heterogeneity
    - See also Frey (2019)

Dell (2010)

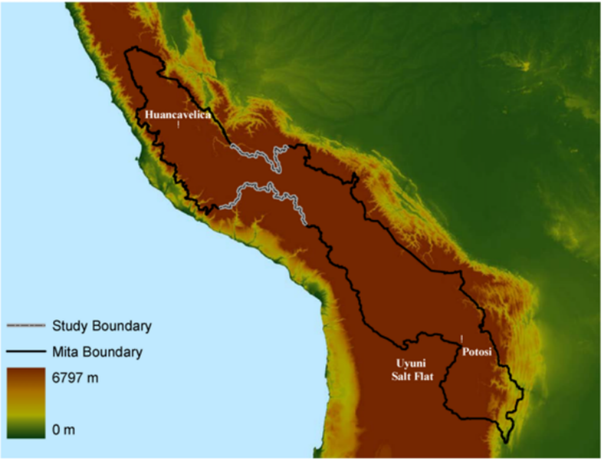
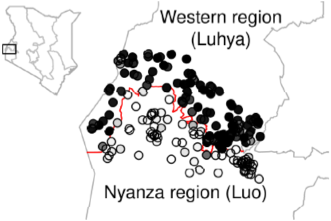


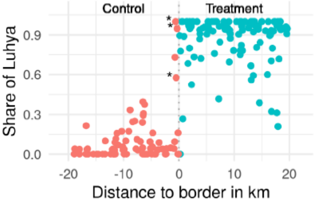
FIGURE 1.—The *mita* boundary is in black and the study boundary in light gray. Districts falling inside the contiguous area formed by the *mita* boundary contributed to the *mita*. Elevation is shown in the background.

Geographic RDD

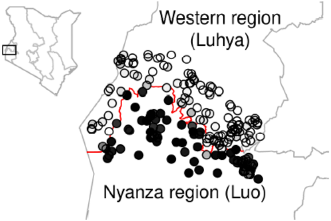
Müller-Crepon (2021)



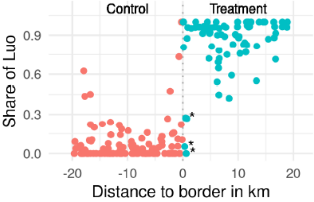
(a) Luhya share



(b) Luhya share



(c) Luo share



(d) Luo share

Geographic RDD

Rozenas, Schutte, and Zhukov (2017)

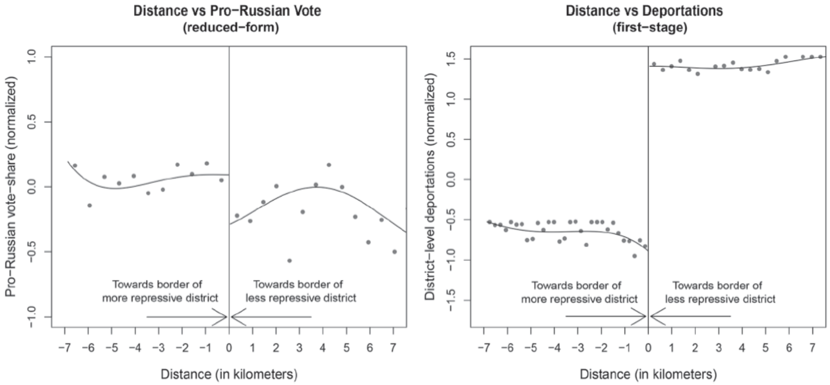


Figure 3. Reduced-form relationships between the instrument (distance from the contiguous district with more repression), deportations, and pro-Russian vote.

Geographic RDD: things to keep in mind

Discussion in Keele and Titiunik (2015)

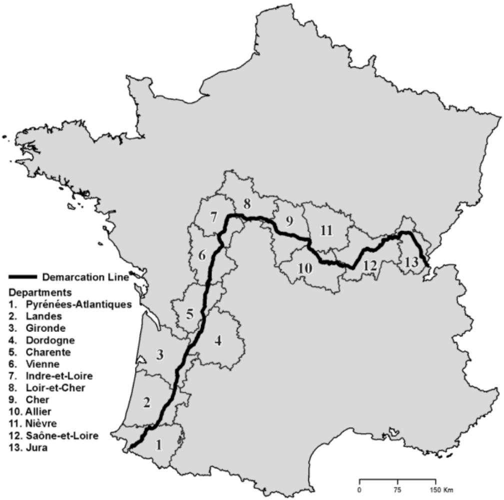
GRDD have distinct challenges:

- Possible compound treatments
- Sparsity of units around the boundary can be more serious (since we want continuity in two dimensions)
- Non-exogeneity of the boundary

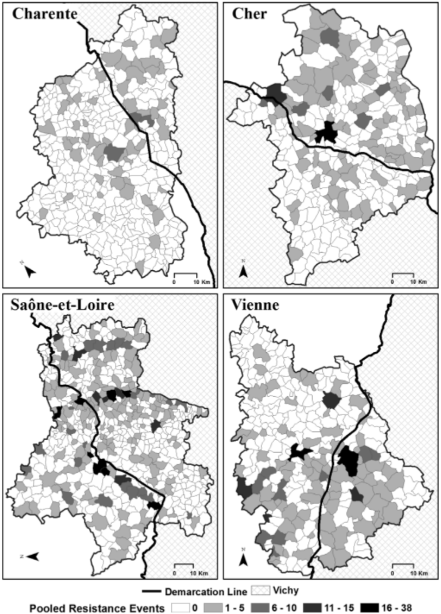
Non-exogeneity of the boundary

Ferwerda and Miller (2014)

FIGURE 1. Map of the Demarcation Line across Intersected French Departments



Non-exogeneity of the boundary



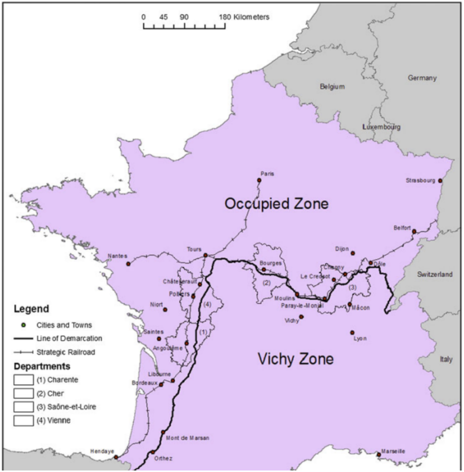
Non-exogeneity of the boundary

Kocher and Monteiro (2016)

Non-exogeneity of the boundary

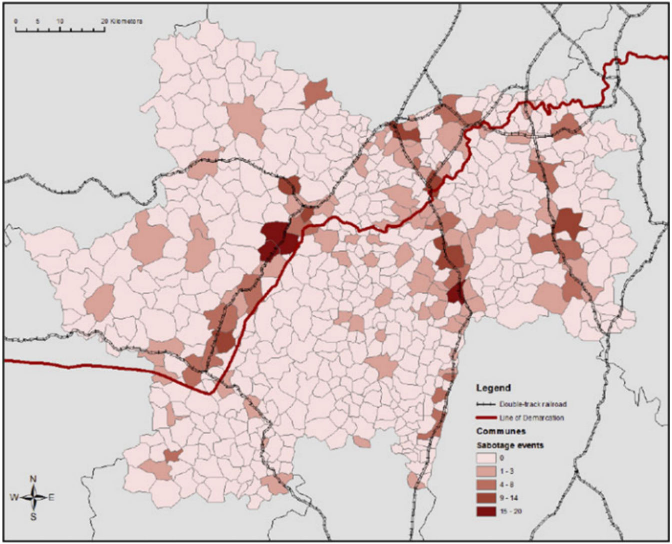
Kocher and Monteiro (2016)

Map 1
The Line of Demarcation (LoD) and Nantes-Tours-Belfort and Paris-Tours-Bordeaux railroads



Non-exogeneity of the boundary

Map 4
Double-track railways and sabotage in Saône-et-Loire, 1940–1944



Difference-in-discontinuities

Add a dimension of heterogeneity (usually time) in RDD.

- Resolve problems of compound treatment when one of the treatments is added after the other
- Estimate moderation effects

Usually "standard" regression models estimated within a bandwidth, more parametric assumptions.

Difference-in-discontinuities

Grembi, Nannicini, and Troiano (2016)

The diff-in-disc estimator can be implemented by estimating the boundary points of four regression functions of Y_{it} on P_{it}: two on both sides of P_c, both before and after t_0. We apply a local linear regression, following Gelman and Imbens (2014).[23] The method consists in fitting linear regression functions to the observations distributed within a distance h on either side of P_c, both before and after t_0. Formally, we restrict the sample to cities in the interval $P_{it} \in [P_c - h, P_c + h]$ and estimate the model

$$(1) \qquad Y_{it} = \delta_0 + \delta_1 P^*_{it} + S_i(\gamma_0 + \gamma_1 P^*_{it}) + T_t[\alpha_0 + \alpha_1 P^*_{it} + S_i(\beta_0 + \beta_1 P^*_{it})] + \xi_{it},$$

where S_i is a dummy for cities below 5,000 capturing treatment status, T_t an indicator for the posttreatment period, and $P^*_{it} = P_{it} - P_c$ the normalized population size. Standard errors are clustered at the city level. The coefficient β_0 is the diff-in-disc estimator and identifies the treatment effect of relaxing fiscal rules, as the treatment is $R_{it} = S_i \cdot T_t$. We present the robustness of our results to multiple bandwidths h, optimally computed first following the algorithm developed by Calonico, Cattaneo, and Titiunik (2014a, b), and then implementing the cross-validation method proposed by Ludwig and Miller (2007).[24]

Difference-in-discontinuities

Larreguy, Marshall, and Querubín (2016)

TABLE 4. Effect of Split Polling Station by Distance

	Turnout (1)	PRI Vote Share (2)	PAN Vote Share (3)	PRD Vote Share (4)
Split	0.0084***	0.0034***	0.0042**	0.0006
	(0.0014)	(0.0011)	(0.0018)	(0.0011)
Distance	−0.0067	0.0058	−0.0089***	−0.0017
	(0.0049)	(0.0036)	(0.0032)	(0.0020)
Distance squared	−0.0003	−0.0006	0.0003	0.0000
	(0.0007)	(0.0007)	(0.0006)	(0.0003)
Split × Distance	0.0014	0.0110***	−0.0035	−0.0002
	(0.0047)	(0.0036)	(0.0028)	(0.0031)
Split × Distance squared	−0.0009	−0.0030**	0.0010*	−0.0005
	(0.0014)	(0.0013)	(0.0006)	(0.0007)
Observations	27,420	27,420	27,420	27,420

Notes: All specifications include district-year fixed effects, and are estimated with OLS. All results are for a 20 voter bandwidth. Block-bootstrapped standard errors are clustered by state (1,000 resamples). Locality-weighted distance to the polling station was unavailable for 347 electoral precincts. *denotes $p < 0.1$, **denotes $p < 0.05$, ***denotes $p < 0.01$.

Local randomization

An alternative approach to RDD: instead of assuming continuity, we assume random assignment in a neighborhood of the cutoff.

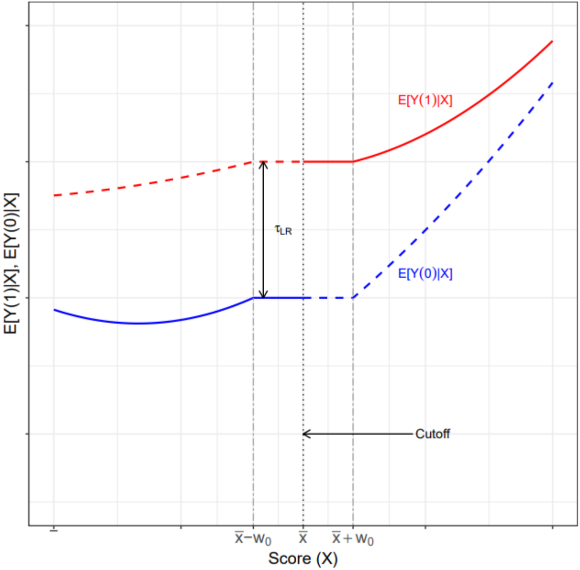
LRD assumptions (CIT Vol.2):

Given a window W_0 around the cutoff:

- The distribution of the running variable in the W_0 is known, is the same for all units, and does not depend on the potential outcomes: $F_{X_i|X_i \in W_0}(x) = F(x)$
- Inside W_0, the potential outcomes depend on the running variable solely through the treatment indicator $T_i = \mathbb{I}(X_i > c)$, but not directly: $Y_i(X_i, T_i) = Y_i(T_i)$ for all i such that $X_i \in W_0$

Relative to the continuity approach, it has more restrictive assumptions, but the advantage of finite sample inference methods.

Local randomization



Local randomization

Estimation and inference under local randomization approach is implemented by the function `rdrandinf` in the package `rdlocrand`.

Data-driven approach to specify the window: specify pre-determined covariates and choose larger and larger windows until T and C units are no longer balanced on them. Inference uses Fisherian methods (randomization inference for the sharp null).

Estimation

To choose the window, the researcher specifies:

- Covariates to be used for balance
- Test statistic to be used for assessing balance (e.g. difference in means)
- Randomization mechanism
- Minimum number of observations in the smallest window
- Significance level

Then: difference-in-means within the selected window with randomization inference.

Application to Islamic mayors

```r
library(rdlocrand)
Z <- cbind(data$i89, data$vshr_islam1994, data$partycount, data$lpop1994,
           data$merkezi, data$merkezp, data$subbuyuk, data$buyuk)
colnames(Z) = c("i89 ", "vshr_islam1994", "partycount", "lpop1994",
                "merkezi", "merkezp", "subbuyuk", "buyuk")
```

Application to Islamic mayors

```r
out <- rdrandinf(Y, X, covariates = Z, seed = 50, d = 3.019522)
```

```
Number of obs     =          2629
Order of poly     =             0
Kernel type       =       uniform
Reps              =          1000
Window            =   rdwinselect
H0:          tau  =         0.000
Randomization     = fixed margins

Cutoff c =    0.000   Left of c  Right of c
      Number of obs        2314         315
 Eff. number of obs          21          26
    Mean of outcome      12.933      16.049
    S.d. of outcome       8.752      10.815
             Window      -0.875       0.875

================================================================================

                                   Finite sample         Large sample
                                ------------------  -------------------------
----
          Statistic          T        P>|T|        P>|T|    Power vs d =
3.020
================================================================================

     Diff. in means      3.116        0.306        0.275
0.185
================================================================================
```