Regression Discontinuity Designs in Stata

Matias D. Cattaneo

University of Michigan

July 30, 2015

Overview

- Main goal: learn about treatment effect of policy or intervention.
- If treatment randomization available, easy to estimate treatment effects.
- If treatment randomization not available, turn to observational studies.
  - Instrumental variables.
  - Selection on observables.
- **Regression discontinuity (RD) designs**.
  - Simple and objective. Requires little information, if design available.
  - Might be viewed as a “local” randomized trial.
  - Easy to falsify, easy to interpret.
  - *Careful*: very local!

Overview of RD packages

`https://sites.google.com/site/rdpackages`

- **rdrobust package**: estimation, inference and graphical presentation using local polynomials, partitioning, and spacings estimators.
  - `rdrobust`: RD inference (point estimation and CI; classic, bias-corrected, robust).
  - `rdbwselect`: bandwidth or window selection (IK, CV, CCT).
  - `rdplot`: plots data (with "optimal" block length).
- **rddensity package**: discontinuity in density test at cutoff (a.k.a. manipulation testing) using novel local polynomial density estimator.
  - `rddensity`: manipulation testing using local polynomial density estimation.
  - `rdbwdensity`: bandwidth or window selection.
- **rdlocrand package**: covariate balance, binomial tests, randomization inference methods (window selection & inference).
  - `rdrandinf`: inference using randomization inference methods.
  - `rdwinselect`: falsification testing and window selection.
  - `rdsensitivity`: treatment effect models over grid of windows, CI inversion.
  - `rdrbounds`: Rosenbaum bounds.

Randomized Control Trials

- **Notation**: $(Y_i(0), Y_i(1), X_i)$, $i = 1, 2, \ldots, n$.

- **Treatment**: $T_i \in \{0, 1\}$, $\qquad T_i$ independent of $(Y_i(0), Y_i(1), X_i)$.

- **Data**: (Y_i, T_i, X_i), $i = 1, 2, \ldots, n$, with

$$Y_i = \begin{cases} Y_i(0) & \text{if } T_i = 0 \\ Y_i(1) & \text{if } T_i = 1 \end{cases}$$

- **Average Treatment Effect**:

$$\tau_{\texttt{ATE}} = \mathbb{E}[Y_i(1) - Y_i(0)] = \mathbb{E}[Y_i | T = 1] - \mathbb{E}[Y_i | T = 0]$$

- **Experimental Design**.

Sharp RD design

- **Notation**: $(Y_i(0), Y_i(1), X_i)$, $i = 1, 2, \ldots, n$, $\qquad X_i$ continuous

- **Treatment**: $T_i \in \{0, 1\}$, $\qquad T_i = \mathbf{1}(X_i \geq \bar{x})$.

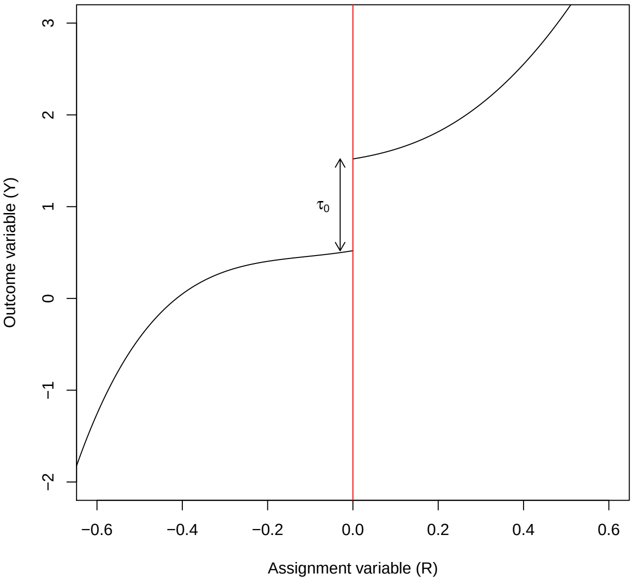
- **Data**: (Y_i, T_i, X_i), $i = 1, 2, \ldots, n$, with

$$Y_i = \begin{cases} Y_i(0) & \text{if } T_i = 0 \\ Y_i(1) & \text{if } T_i = 1 \end{cases}$$

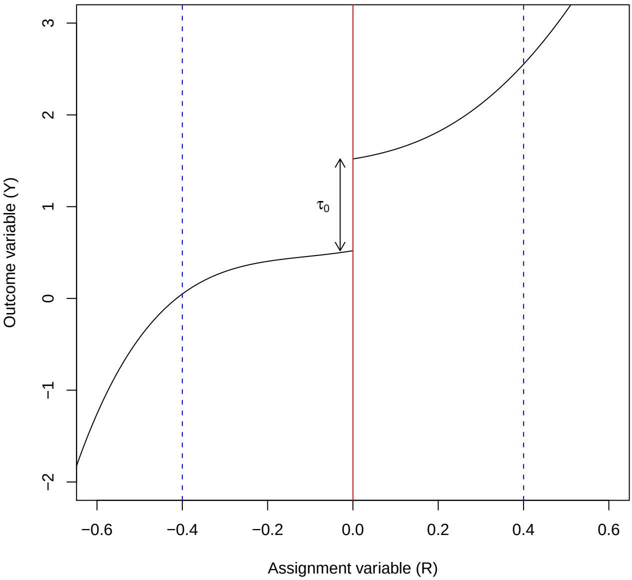
- **Average Treatment Effect at the cutoff**:

$$\tau_{\texttt{SRD}} = \mathbb{E}[Y_i(1) - Y_i(0) | X_i = \bar{x}] = \lim_{x \downarrow \bar{x}} \mathbb{E}[Y_i | X_i = x] - \lim_{x \uparrow \bar{x}} \mathbb{E}[Y_i | X_i = x]$$

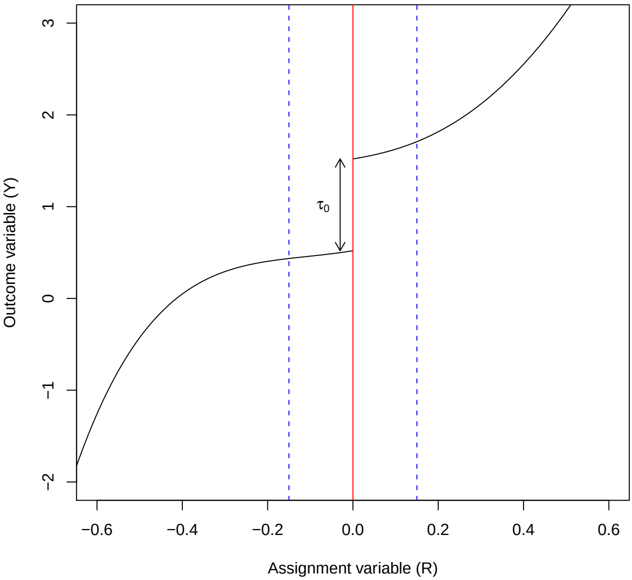
- **Quasi-Experimental Design**: “local randomization” (more later)



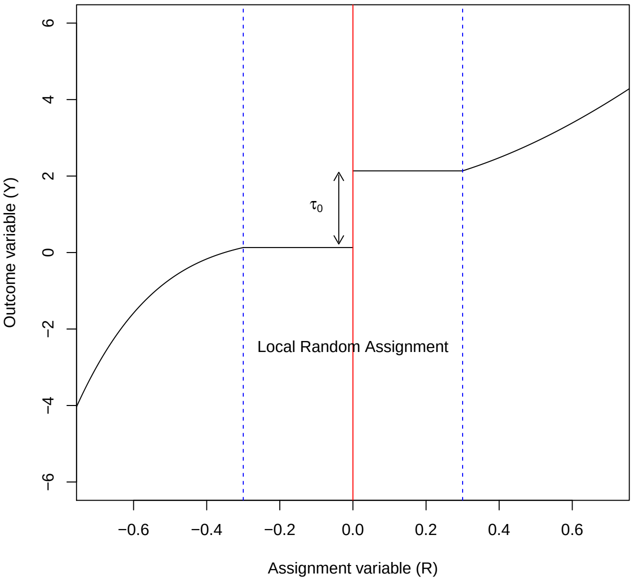
Outcome variable (Y)
Assignment variable (R)
τ0
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6
−2
−1
0
1
2
3



Outcome variable (Y)
3
2
1
0
−1
−2
τ0
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6
Assignment variable (R)



Outcome variable (Y)
3
2
1
0
−1
−2
τ0
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6
Assignment variable (R)



Outcome variable (Y)
6
4
2
0
−2
−4
−6
τ_0
Local Random Assignment
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6
Assignment variable (R)

Empirical Illustration: Cattaneo, Frandsen & Titiunik (2015, JCI)

- **Problem**: incumbency advantage (U.S. senate).

- **Data**:

  Y_i = election outcome.

  T_i = whether incumbent.

  X_i = vote share previous election $\quad (\bar{x} = 0)$.

  Z_i = covariates (*demvoteshlag1*, *demvoteshlag2*, *dopen*, etc.).

- **Potential outcomes**:

  $Y_i(0)$ = election outcome if **had not been** incumbent.

  $Y_i(1)$ = election outcome if **had been** incumbent.

- **Causal Inference**:

  $$Y_i(0) \neq Y_i | T_i = 0 \qquad \text{and} \qquad Y_i(1) \neq Y_i | T_i = 1$$

Graphical and Falsification Methods

- Always plot data: main advantage of RD designs!
- Plot regression functions to assess treatment effect and validity.
- Plot density of X_i for assessing validity; test for continuity at cutoff and elsewhere.
- **Important**: use also estimators that do not "smooth-out" data.
- **RD Plots (Calonico, Cattaneo & Titiunik, JASA)**:
  - Two ingredients: (i) Smoothed global polynomial fit & (ii) binned discontinuous local-means fit.
  - Two goals: (i) detention of discontinuities, & (ii) representation of variability.
  - Two tuning parameters:
    - Global polynomial degree (k_n).
    - Location (ES or QS) and number of bins (J_n).

Manipulation Tests & Covariate Balance and Placebo Tests

- Density tests near cutoff:
  - **Idea**: distribution of running variable should be similar at either side of cutoff.
  - **Method 1**: Histograms & Binomial count test.
  - **Method 2**: Density Estimator at boundary.
    - Pre-binned local polynomial method – McCrary (2008).
    - New tuning-parameter-free method – **Cattaneo, Jansson and Ma (2015)**.
- Placebo tests on pre-determined/exogenous covariates.
  - **Idea**: zero RD treatment effect for pre-determined/exogenous covariates.
  - Methods: global polynomial, local polynomial, randomization-based.
- Placebo tests on outcomes.
  - **Idea**: zero RD treatment effect for outcome at values other than cutoff.
  - Methods: global polynomial, local polynomial, randomization-based.

Estimation and Inference Methods

- Global polynomial approach (*not recommended*).
- Robust local polynomial inference methods.
  - Bandwidth selection.
  - Bias-correction.
  - Confidence intervals.
- Local randomization and randomization inference methods.
  - Window selection.
  - Estimation and Inference methods.
  - Falsification, sensitivity and related methods

Conventional Local-polynomial Approach

- **Idea**: approximate regression functions for control and treatment units *locally*.

- "Local-linear" estimator (w/ weights $K(\cdot)$):

$-h_n \leq X_i < \bar{x}$:	$\bar{x} \leq X_i \leq h_n$:
$Y_i = \alpha_- + (X_i - \bar{x}) \cdot \beta_- + \varepsilon_{-,i}$	$Y_i = \alpha_+ + (X_i - \bar{x}) \cdot \beta_+ + \varepsilon_{+,i}$

  - Treatment effect (at the cutoff): $\hat{\tau}_{\mathtt{SRD}} = \hat{\alpha}_+ - \hat{\alpha}_-$

- Can be estimated using linear models (w/ weights $K(\cdot)$):

$$Y_i = \alpha + \tau_{\mathtt{SRD}} \cdot T_i + (X_i - \bar{x}) \cdot \beta_1 + T_i \cdot (X_i - \bar{x}) \cdot \gamma_1 + \varepsilon_i, \qquad -h_n \leq X_i \leq h_n$$

- Once h_n chosen, inference is "standard": weighted linear models.
  - Details coming up next.

Conventional Local-polynomial Approach

- How to choose h_n?

- Imbens & Kalyanaraman (2012, ReStud): "optimal" plug-in,

$$\hat{h}_{\mathtt{IK}} = \hat{C}_{\mathtt{IK}} \cdot n^{-1/5}$$

- Calonico, Cattaneo & Titiunik (2014, ECMA): refinement of IK

$$\hat{h}_{\mathtt{CCT}} = \hat{C}_{\mathtt{CCT}} \cdot n^{-1/5}$$

- Ludwig & Miller (2007, QJE): cross-validation,

$$\hat{h}_{\mathtt{CV}} = \arg\min_{h} \sum_{i=1}^{n} w(X_i)\left(Y_i - \hat{\mu}_1(X_i, h)\right)^2$$

- **Key idea**: trade-off bias and variance of $\hat{\tau}_{\mathtt{SRD}}(h_n)$. Heuristically:

$$\uparrow \ \mathsf{Bias}(\hat{\tau}_{\mathtt{SRD}}) \quad \Longrightarrow \quad \downarrow \hat{h} \qquad \text{and} \qquad \uparrow \ \mathsf{Var}(\hat{\tau}_{\mathtt{SRD}}) \quad \Longrightarrow \quad \uparrow \hat{h}$$

Local-Polynomial Methods: Bandwidth Selection

- Two main methods: plug-in & cross-validation. Both MSE-optimal in some sense.

- Imbens & Kalyanaraman (2012, ReStud): propose MSE-optimal rule,

$$h_{\texttt{MSE}} = C_{\texttt{MSE}}^{1/5} \cdot n^{-1/5} \qquad\qquad C_{\texttt{MSE}} = C(K) \cdot \frac{\mathsf{Var}(\hat{\tau}_{\texttt{SRD}})}{\mathsf{Bias}(\hat{\tau}_{\texttt{SRD}})^2}$$

  - IK implementation: first-generation plug-in rule.
  - CCT implementation: second-generation plug-in rule.
  - They differ in the way $\mathsf{Var}(\hat{\tau}_{\texttt{SRD}})$ and $\mathsf{Bias}(\hat{\tau}_{\texttt{SRD}})$ are estimated.

- Imbens & Kalyanaraman (2012, ReStud): discuss cross-validation approach,

$$\hat{h}_{\texttt{CV}} = \arg\min_{h>0} \mathsf{CV}_\delta(h), \qquad \mathsf{CV}_\delta(h) = \sum_{i=1}^{n} \mathbf{1}(X_{-,[\delta]} \le X_i \le X_{+,[\delta]})\,(Y_i - \hat{\mu}(X_i;h))^2,$$

where

  - $\hat{\mu}_{+,p}(x;h)$ and $\hat{\mu}_{-,p}(x;h)$ are local polynomials estimates.
  - $\delta \in (0,1)$, $X_{-,[\delta]}$ and $X_{+,[\delta]}$ denote δ-th quantile of $\{X_i : X_i < \bar{x}\}$ and $\{X_i : X_i \ge \bar{x}\}$.
  - Our implementation uses $\delta = 0.5$; but this is a tuning parameter!

Conventional Approach to RD

- "Local-linear" estimator (w/ weights $K(\cdot)$):

$-h_n \leq X_i < \bar{x}$:	$\bar{x} \leq X_i \leq h_n$:
$Y_i = \alpha_- + (X_i - \bar{x}) \cdot \beta_- + \varepsilon_{-,i}$	$Y_i = \alpha_+ + (X_i - \bar{x}) \cdot \beta_+ + \varepsilon_{+,i}$

  - Treatment effect (at the cutoff): $\hat{\tau}_{\texttt{SRD}} = \hat{\alpha}_+ - \hat{\alpha}_-$

- Construct usual t-test. For $\mathsf{H}_0 : \tau_{\texttt{SRD}} = 0$,

$$\hat{T}(h_n) = \frac{\hat{\tau}_{\texttt{SRD}}}{\sqrt{\hat{\mathsf{V}}_n}} = \frac{\hat{\alpha}_+ - \hat{\alpha}_-}{\sqrt{\hat{\mathsf{V}}_{+,n} + \hat{\mathsf{V}}_{-,n}}} \approx_d \mathcal{N}(0,1)$$

- 95% Confidence interval:

$$\hat{I}(h_n) = \left[\ \hat{\tau}_{\texttt{SRD}} \ \pm \ 1.96 \cdot \sqrt{\hat{\mathsf{V}}_n}\ \right]$$

Bias-Correction Approach to RD

- Note well: for usual t-test,

$$\hat{T}(h_{\mathtt{MSE}}) = \frac{\hat{\tau}_{\mathtt{SRD}}}{\sqrt{\hat{\mathsf{V}}_n}} \approx_d \mathcal{N}(\mathsf{B},1) \quad \neq \quad \mathcal{N}(0,1), \qquad \mathsf{B} > 0$$

  - Bias B in RD estimator captures "curvature" of regression functions.

- Undersmoothing/"Small Bias" Approach: Choose "smaller" h_n... Perhaps $\hat{h}_n = 0.5 \cdot \hat{h}_{\mathtt{IK}}$?

  $\Longrightarrow$ Not clear guidance & power loss!

- Bias-correction Approach:

$$\hat{T}^{\mathsf{bc}}(h_n, b_n) = \frac{\hat{\tau}_{\mathtt{SRD}} - \hat{\mathsf{B}}_n}{\sqrt{\hat{\mathsf{V}}_n}} \approx_d \mathcal{N}(0,1)$$

$\Longrightarrow$ 95% Confidence Interval: $\hat{I}^{\mathsf{bc}}(h_n, b_n) = \left[\ \left(\hat{\tau}_{\mathtt{SRD}} - \hat{\mathsf{B}}_n\right) \ \pm \ 1.96 \cdot \sqrt{\hat{\mathsf{V}}_n} \ \right]$

- How to choose b_n? Same ideas as before... $\hat{b}_n = \hat{C} \cdot n^{-1/7}$

Robust Bias-Correction Approach to RD

- **Recall**:

$$\hat{T}(h_n) = \frac{\hat{\tau}_{\mathsf{SRD}}}{\sqrt{\hat{\mathsf{V}}_n}} \approx_d \mathcal{N}(0,1) \qquad \text{and} \qquad \hat{T}^{\mathsf{bc}}(h_n, b_n) = \frac{\hat{\tau}_{\mathsf{SRD}} - \hat{\mathsf{B}}_n}{\sqrt{\hat{\mathsf{V}}_n}} \approx_d \mathcal{N}(0,1)$$

  - $\hat{\mathsf{B}}_n$ is constructed to estimate leading bias B.

- **Robust approach**:

$$\hat{T}^{\mathsf{bc}}(h_n, b_n) = \frac{\hat{\tau}_{\mathsf{SRD}} - \hat{\mathsf{B}}_n}{\sqrt{\hat{\mathsf{V}}_n}} = \underbrace{\frac{\hat{\tau}_{\mathsf{SRD}} - \mathsf{B}_n}{\sqrt{\hat{\mathsf{V}}_n}}}_{\approx_d \ \mathcal{N}(0,1)} + \underbrace{\frac{\mathsf{B}_n - \hat{\mathsf{B}}_n}{\sqrt{\hat{\mathsf{V}}_n}}}_{\approx_d \ \mathcal{N}(0,\gamma)}$$

- **Robust bias-corrected t-test**:

$$\hat{T}^{\mathsf{rbc}}(h_n, b_n) = \frac{\hat{\tau}_{\mathsf{SRD}} - \hat{\mathsf{B}}_n}{\sqrt{\hat{\mathsf{V}}_n + \hat{\mathsf{W}}_n}} = \frac{\hat{\tau}_{\mathsf{SRD}} - \hat{\mathsf{B}}_n}{\sqrt{\hat{\mathsf{V}}_n^{\mathsf{bc}}}} \approx_d \mathcal{N}(0,1)$$

$\Longrightarrow$ 95% Confidence Interval:

$$\hat{I}^{\mathsf{rbc}}(h_n, b_n) = \left[\ \left(\hat{\tau}_{\mathsf{SRD}} - \hat{\mathsf{B}}_n\right) \ \pm \ 1.96 \cdot \sqrt{\hat{\mathsf{V}}_n^{\mathsf{bc}}} \ \right], \qquad \hat{\mathsf{V}}_n^{\mathsf{bc}} = \hat{\mathsf{V}}_n + \hat{\mathsf{W}}_n$$

Local-Polynomial Methods: Robust Inference

- Approach 1: *Undersmoothing/"Small Bias"*.

$$\hat{I}(h_n) = \left[\; \hat{\tau}_{\texttt{SRD}} \; \pm \; 1.96 \cdot \sqrt{\hat{\mathsf{V}}_n} \; \right]$$

- Approach 2: *Bias correction* (**not recommended**).

$$\hat{I}^{\mathsf{bc}}(h_n, b_n) = \left[\; \left(\hat{\tau}_{\texttt{SRD}} - \hat{\mathsf{B}}_n\right) \; \pm \; 1.96 \cdot \sqrt{\hat{\mathsf{V}}_n} \; \right]$$

- Approach 3: *Robust Bias correction.*

$$\hat{I}^{\mathsf{rbc}}(h_n, b_n) = \left[\; \left(\hat{\tau}_{\texttt{SRD}} - \hat{\mathsf{B}}_n\right) \; \pm \; 1.96 \cdot \sqrt{\hat{\mathsf{V}}_n + \hat{\mathsf{W}}_n} \; \right]$$

Local-randomization approach and finite-sample inference

- Popular approach: local-polynomial methods.
  - Approximates regression function and relies on continuity assumptions.
  - *Requires*: choosing weights, bandwidth and polynomial order.

- Alternative approach: local-randomization + randomization-inference
  - Gives an alternative that can be used as a robustness check.
  - **Key assumption**: exists window $W = [-h_n, h_n]$ around cutoff $(-h_n < \bar{x} < h_n)$ where

$$T_i \text{ independent of } (Y_i(0), Y_i(1)) \qquad\qquad (\text{for all } X_i \in W)$$

  - In words: treatment is randomly assigned within W.
  - Good news: if plausible, then RCT ideas/methods apply.
  - Not-so-good news: most plausible for very small windows (very few observations).
  - One solution: employ small window but use randomization-inference methods.
  - *Requires*: choosing randomization rule, window and statistic.

Local-randomization approach and finite-sample inference

- **Recall key assumption**: exists $W = [-h_n, h_n]$ around cutoff $(-h_n < \bar{x} < h_n)$ where

$$T_i \text{ independent of } (Y_i(0), Y_i(1)) \qquad\qquad (\text{for all } X_i \in W)$$

- How to choose window?
  - Use balance tests on pre-determined/exogenous covariates.
  - Very intuitive, easy to implement.

- How to conduct inference? Use randomization-inference methods.
  1. Choose statistic of interest. E.g., t-stat for difference-in-means.
  2. Choose randomization rule. E.g., number of treatments and controls given.
  3. Compute finite-sample distribution of statistics by permuting treatment assignments.

Local-randomization approach and finite-sample inference

- Do not forget to validate & falsify the empirical strategy.
  1. Plot data to make sure local-randomization is plausible.
  2. Conduct placebo tests.

     (e.g., use pre-intervention outcomes or other covariates not used select W)
  3. Do sensitivity analysis.
- See **Cattaneo, Frandsen and Titiunik (2015)** for introduction.
- See **Cattaneo, Titiunik and Vazquez-Bare (2015)** for further results and implementation.