Regression Discontinuity Designs: theory and simulations

EMERG, University of Exeter

May 18, 2023

Carlotta Montorsi [1,2]

[1] Luxembourg Institute of Socio-Economic Research

[2] University of Luxembourg

Readings

- Cattaneo, M. D., & Titiunik, R. (2022). Regression discontinuity designs. Annual Review of Economics, 14, 821-851.
- Cattaneo, M. D., Idrobo, N., & Titiunik, R. (2019). A practical introduction to regression discontinuity designs: Foundations. Cambridge University Press. (chapters 2 and 4)
- Cattaneo, M. D., Idrobo, N., & Titiunik, R. (2023). A practical introduction to regression discontinuity designs: Extensions. arXiv preprint arXiv:2301.08958. (chapter 3)

Table of contents

Design and Framework

Causal Inference and Program Evaluation

- Main goal: learn about treatment effect of policy or intervention
- If treatment randomization available → easy to estimate causal effects
- If treatment randomization is not available → observational studies
  - Selection on observables
  - Instrumental variables, Diff-in-diff, etc.
- **Regression Discontinuity (RD) design**
  - Simple assignment, based on known exogenous factors
  - Objective basis to evaluate assumptions
  - Easy to falsify and interpret
  - *Careful*: very local!

Regression Discontinuity Design

- Observational units receive a **score** (X_i)
- A treatment is assigned based on the score and a *known* **cutoff** (c)
- The **treatment** is:
  - given to units whose score is greater than the cutoff
  - withheld form units whose score is smaller than the cutoff
- Under certain assumptions, the sharp change in the probability of treatment assignment allows us to learn about the effect of the treatment

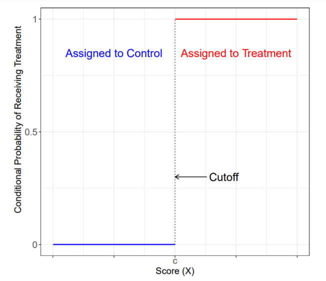

Figure 1: Source: *A Practical Introduction to Regression Discontinuity Designs: Foundations*, Cattaneo M., Idrobo N., Titiunik R. (2019)

RD Designs: Taxonomy

- **Frameworks**:
  - Identification: **Continuity based**, Local Randomization
  - Score: **Continuous**, Many Repeated, Few Repeated
- **Settings**:
  - **Sharp, Fuzzy**, Kink, Kink Fuzzy
  - Multiple Cutoff, Multiple Scores, Geographic RD
  - Dynamic, Continuous Treatments, Time, etc.
- **Parameters of Interest**
  - **Average Effects**, Distributional Effects, Partial Effects
  - Heterogeneity, Covariate-Adjustment, Differences, Time
  - Extrapolation

Sharp RDD

RCT vs (Sharp) RD Designs

- **Notation**: $(Y_i(0), Y_i(1), X_i), i = 1, 2, \ldots, n$
- **Treatment**: $T_i \in 0, 1, \quad T_i$ independent of $(Y_i(0), Y_i(1), X_i)$
- **Data**: $(Y_i, T_i, X_i), \quad i = 1, 2, \ldots, n$ 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**:

$$\tau_{\mathsf{ATE}} \equiv E[Y_i(1) - Y_i(0)] = E[Y_i \mid T = 1] - E[Y_i \mid T = 0]$$

RCT vs (Sharp) RD Designs

- **Notation**: $(Y_i(0), Y_i(1), X_i), \quad i = 1, 2, \ldots, n, \quad X_i$ score
- **Treatment**: $T_i \in 0, 1, \quad T_i = \mathbb{1}(X_i \geq c) \quad c$ cutoff
- **Data**: $(Y_i, T_i, X_i),\ i = 1, 2, \ldots, n$ with:

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

- **Average Treatment Effect at the cutoff (continuity-based)**:

$$\tau_{\mathsf{srd}} \equiv E[Y_i(1) - Y_i(0) \mid X_i = c] = \lim_{x \to c^+} E[Y_i \mid X_i = x] - \lim_{x \to c^-} E[Y_i \mid X_i = x]$$

- **Identifying Assumptions (continuity-based)** :
    - Comparability between units with very similar values of the score but on opposite sides of the cutoff → assume smoothness of other covariates at the cutoff
    - $E[Y_i(1) \mid X_i = x], \quad E[Y_i(0) \mid X_i = x]$ continuous at c

RCT vs (Sharp) RD Designs

$$\tau_{\text{srd}} \equiv \underbrace{E[Y_i(1) - Y_i(0) \mid X_i = c]}_{\text{unobservable}} = \underbrace{\lim_{x \to c^+} E[Y_i \mid X_i = x]}_{\color{red}{\text{estimable}}} - \underbrace{\lim_{x \to c^-} E[Y_i \mid X_i = x]}_{\color{blue}{\text{estimable}}}$$

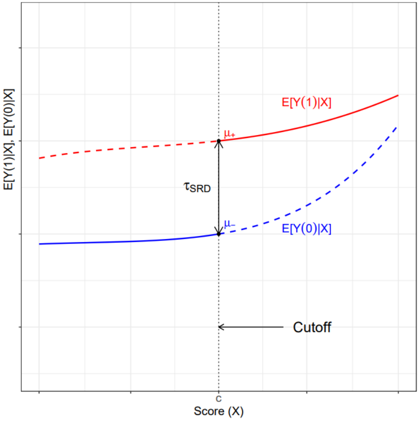

Figure 2: RD Treatment Effect in Sharp RD Design

Figure 2: Source: *A Practical Introduction to Regression Discontinuity Designs: Foundations*, Cattaneo M., Idrobo N., Titiunik R. (2019)

Sharp RDD

- **Parametric estimation**:
  - vanilla OLS:
    - $E(Y \mid X)$ is linear
    - Estimation employs data away from the discontinuity → convergence rate $(n^{1/2})$
    - $Y = \alpha + \tau_{\text{srd}} T + \beta(X - c) + \epsilon$
    - $\hat{\tau}_{\text{SRD}}$ unbiased and consistent if all assumptions are met
- **Non parametric estimation**
  - Local polynomial methods
    - Relax assumptions on $E(Y \mid X)$
    - Estimation uses only observations near the discontinuity
    - Estimation requires selecting three main hyper-parameters: polynomial order (p), Kernel function $K(.)$, and bandwidth (h)
    - Usually $p = 1$ (Local Linear Regression), triangular Kernel, and cross-validated bandwidth
    - Slower convergence rate than vanilla OLS, $(n^r, r < 1/2)$

Fuzzy RDD

Fuzzy RD designs

- **Treatment assignment**: $T_i \in (0,1)$, $\quad T_i = 1(X_i \geq c)$ c cutoff
- **Imperfect compliance**
  - probability of *receiving* treatment changes at cutoff, but not from 0 to 1
- **Treatment take up**: $D_i(T) \in (0,1)$ are potential treatments:

$$D_i = (1 - T_i) \cdot D_i(0) + T_i \cdot D_i(1) = \begin{cases} D_i(0) & \text{if } T_i = 0 \\ D_i(1) & \text{if } T_i = 1 \end{cases}$$

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

$$Y_i = T_i \cdot Y_i(1, D_i(1)) + (1 - T_i) Y_i(0, D_i(0)) = \begin{cases} Y_i(0,0) & \text{if } T_i = 0 & D_i(0) = 0 & \text{compliers} \\ Y_i(0,1) & \text{if } T_i = 0 & D_i(0) = 1 & \text{non-compliers} \\ Y_i(1,0) & \text{if } T_i = 1 & D_i(1) = 0 & \text{non-compliers} \\ Y_i(1,1) & \text{if } T_i = 1 & D_i(1) = 1 & \text{compliers} \end{cases}$$

Fuzzy RD designs

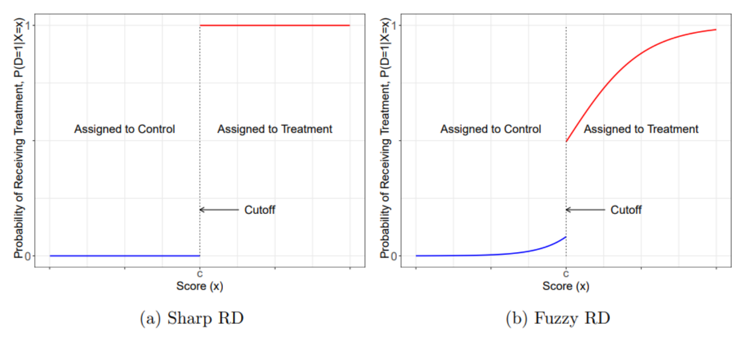

Figure 3: Source: *A Practical Introduction to Regression Discontinuity Designs: Foundations*, Cattaneo M., Idrobo N., Titiunik R. (2019)

Estimands and Inference

- **Intention-to-treat effects**:

$$\tau_{\text{itt}} \equiv E[Y_i(1, D_i(1) - Y_i(0, D_i(0)) \mid X_i = c]$$
$$= \lim_{x \to c^+} E[Y_i(1, D_i(1)) \mid X_i = x] - \lim_{x \to c^-} E[Y_i(0, D_i(0)) \mid X_i = x]$$

- **First-Stage effects**:

$$\tau_{\text{fs}} \equiv E[D_i(1) - D_i(0) \mid X_i = c]$$
$$= \lim_{x \to c^+} E[D_i \mid X_i = x] - \lim_{x \to c^-} E[D_i \mid X_i = x]$$

- **Identifying Assumptions**:
  - $E[Y(1, D_i(1)) \mid X_i = x]$, $E[Y(0, D_i(0)) \mid X_i = x]$ continuous in X at c
  - $E[D_i(1) \mid X_i = x]$, $E[D_i(0) \mid X_i = x]$ continuous in X at c

Estimands and Inference

- **Treatment Effects for Subpopulations**:

$$\tau_{\text{frd}} \equiv \frac{\tau_{\text{itt}}}{\tau_{\text{fs}}} \quad \text{(same interpretation as IV literature)}$$

- **Identifying Assumptions for $\tau_{\textbf{frd}} = \tau_{\textbf{LATE}}$**:
  - Exclusion restriction
  - Monotonicity: probability of receiving treatment increases with running variable, i.e. there are no defiers
  - Relevant instrument: $\tau_{\text{fs}} \neq 0$, rule of thumb $\text{F}_{\text{stat}} > 20$

Estimands and Inference

- **Parametric Estimation**:
  - Two-stage-least square:

$$\begin{aligned} Y_i &= \beta + \tau_{\text{itt}} T_i + \delta X + \nu \quad \text{intention-to-treat} \\ D_i &= \alpha + \tau_{\text{fs}} T_i + \gamma X + \epsilon \quad \text{first-stage} \\ Y_i &= \beta + \tau_{\text{frd}} \hat{D}_i + \omega X + \eta \quad \text{second stage} \end{aligned}$$

- **Non-parametric Estimation**:
  - Local polynomial methods at each stage

Simulations

Sharp RDD

- **Objectives**:
    - Understanding properties of Parametric and Non-parametric estimators under different conditions:
    - How well the estimators work as you vary how many observations you have near the cutoff?
    - How well estimators work in case of mi-specified functional forms of $E(Y \mid X)$?
- **Simulations set-up**
    - Number of simulations: 8000
    - cutoff $= 0$
    - $X \sim N(0, \sigma_x)$ running variable, with $\sigma_x = \{0.01, 10\}$
    - $T = I(X \geq 0)$ and perfect compliance
    - $Y(0) = 0.3 + 0.3 \cdot X + \epsilon, \quad \epsilon \sim N(0,1)$ (linear)
    - $Y(1) = 2 + 0.3 \cdot X + \epsilon, \quad \epsilon \sim N(0,1)$ (linear)
    - $\tau_{\text{srd}} = 2 - 0.3 = 1.7$

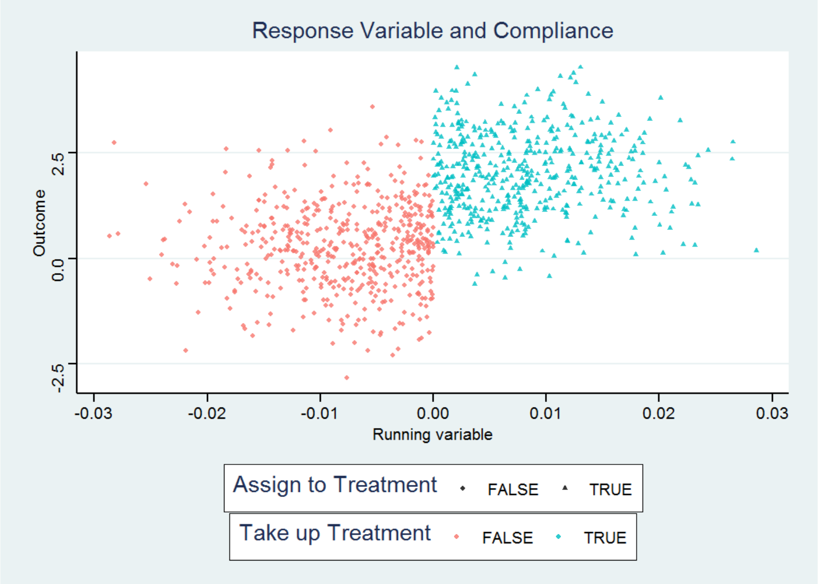

Figure 4: Case with $\sigma_x = 0.01$

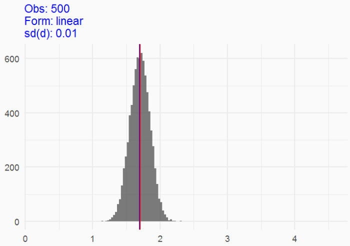

Figure 5: Vanilla OLS

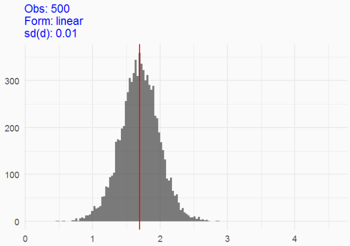

Figure 6: Local Linear Regression

- RMSE: 0.12
- Bias: 0.00076
- Coverage: 95%

- RMSE: 0.23
- Bias: -0.00266
- Coverage: 94%

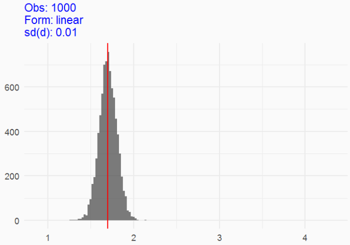

Figure 7: Vanilla OLS

- RMSE: 0.08
- Bias: 0.00034
- Coverage: 95%

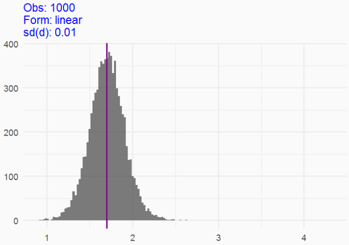

Figure 8: Local Linear Regression

- RMSE: 0.16
- Bias: 0.00203
- Coverage: 94%

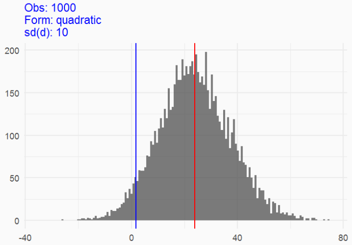

Figure 9: Vanilla OLS

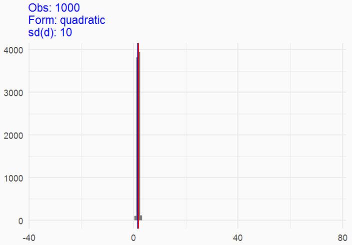

Figure 10: Local Linear Regression

- RMSE: 22.75
- Bias: 22.2743
- Coverage: 32%

- RMSE: 0.3
- Bias: 0.1664
- Coverage: 90%

Fuzzy RDD

- **Objectives**:
  - Understanding properties of Parametric and Non-parametric estimators under different compliers distributions around cutoff.
  - Understanding properties of Parametric and Non-parametric estimators under violation of theoretical assumptions
- **Simulations set-up**
  - $X \sim N(0, \sigma_x)$ running variable, with $\sigma_x = 1$
  - $\text{D}_i = \begin{cases} \text{Binomial}(1, 0.2), & \text{if } \text{T}_i < c \\ \text{Binomial}(1, 0.8), & \text{if } \text{T}_i \geq c \end{cases}$
  - $\text{D}_i = \begin{cases} 0 \quad \text{if } \text{T}_i < 0, \quad X_i < -2 \\ \text{Binomial}(1, 0.2) \quad \text{if } \text{T}_i < 0, \text{-2} < \text{X}_i < 0 \\ \text{Binomial}(1, 0.8) \quad \text{if } \text{T}_i \geq 0, \quad 0 \leq X_i < 2 \\ 1 \quad \text{if } \text{T}_i \geq 0, \quad X_i \geq 2 \end{cases}$
  - $\text{D}_i = \text{Binomial}\left(1, \Phi\left(\frac{X_i - \mu_x}{\sigma_d}\right)\right)$, treatment take-up, with $\sigma_d = \{0.2, 1\}$

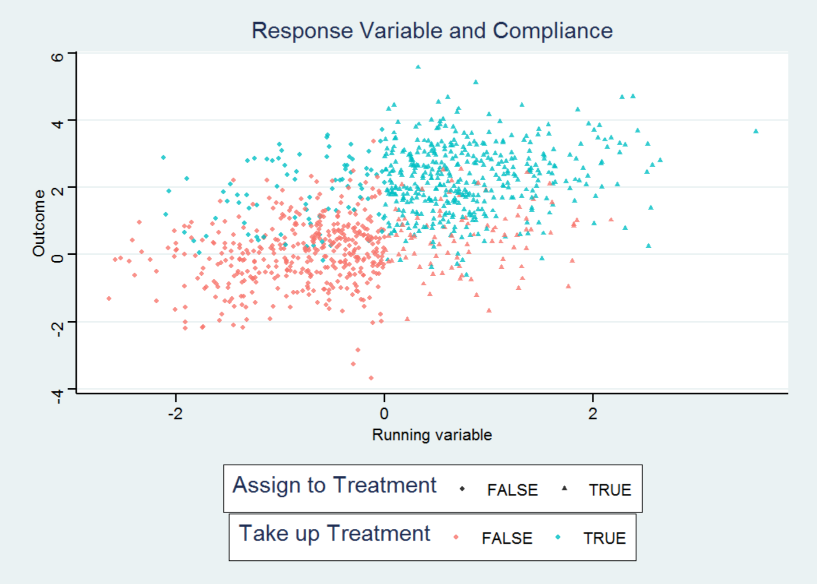

Figure 11: Outcomes in fuzzy Design

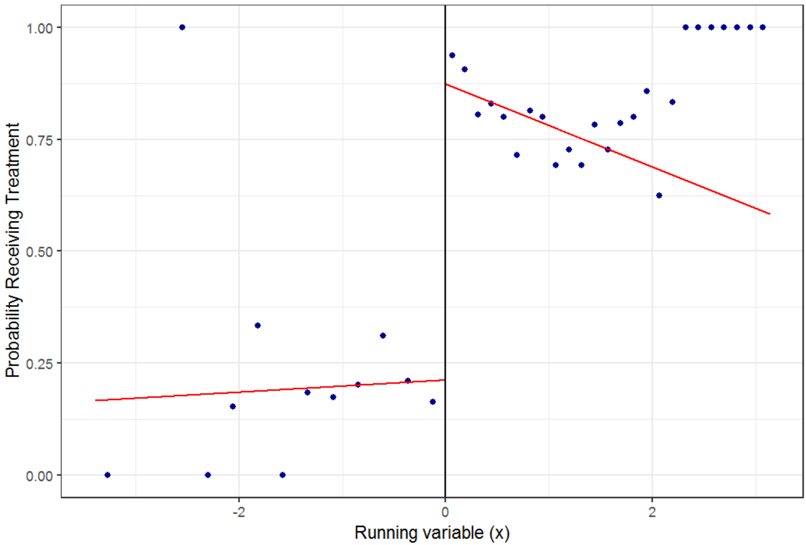

Figure 12: Case with "Uniform" probability, i.e. 20% before cutoff, 80% after

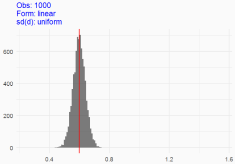

Figure 13: First-stage (2sls)

- Bias: -0.0006

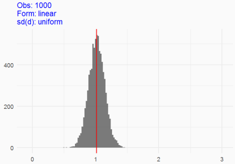

Figure 14: ITT (2sls)

- Bias: 0.0003

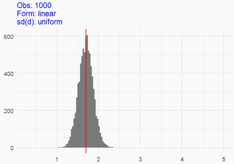

Figure 15: FRD (2sls)

- Bias: -0.0003

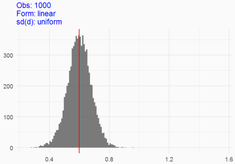

Figure 16: First-stage (LLR)

- Bias: 0.00047

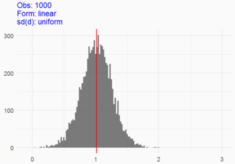

Figure 17: ITT (LLR)

- Bias: 0.0006

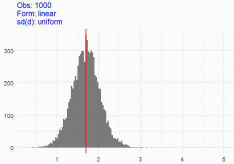

Figure 18: FRD (LLR)

- Bias: 0.0018

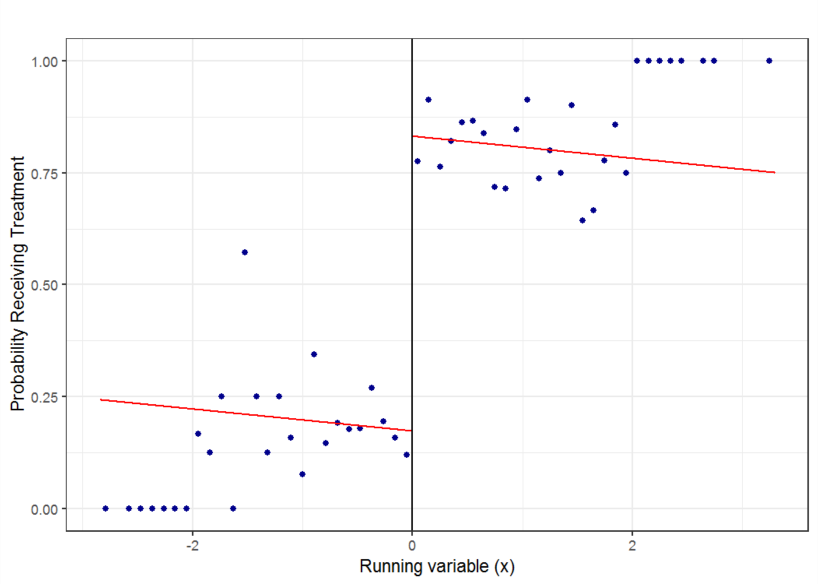

Figure 19: Case with "Jump"

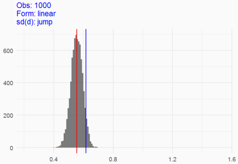

Figure 20: First-stage (2sls)

- Bias: -0.063

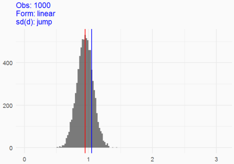

Figure 21: ITT (2sls)

- Bias: -0.1038

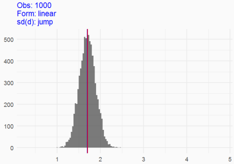

Figure 22: FRD (2sls)

- Bias: 0.005

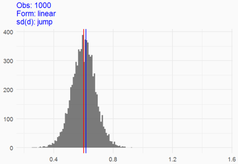

Figure 23: First-stage

- Bias: -0.017

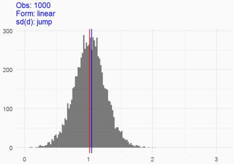

Figure 24: ITT (LLR)

- Bias: -0.027

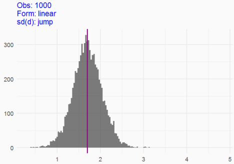

Figure 25: FRD (LLR)

- Bias: 0.00035

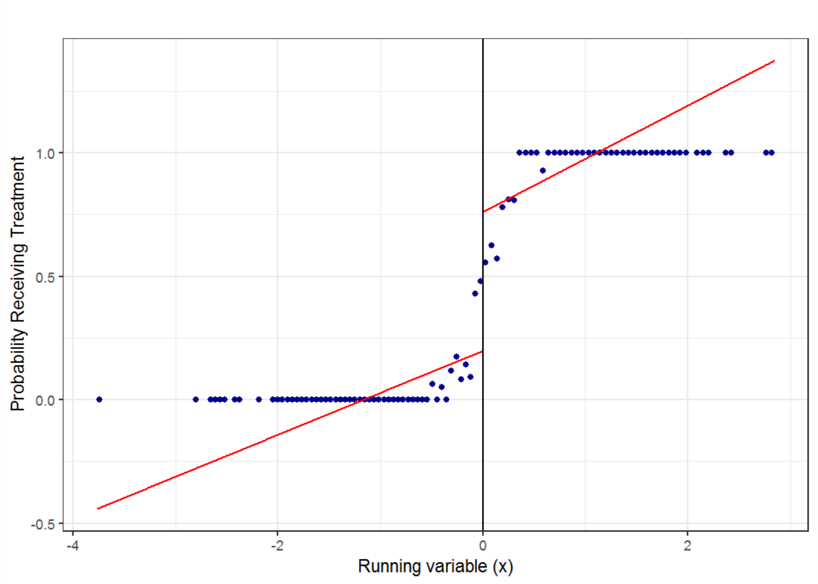

Figure 26: Case with $\sigma_d = 0.2$

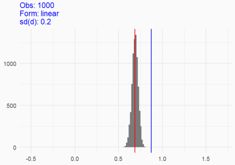

Figure 27: First-stage (2sls)

- Bias: -0.186

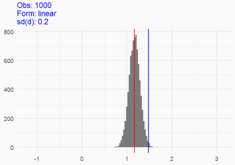

Figure 28: ITT (2sls)

- Bias: -0.315

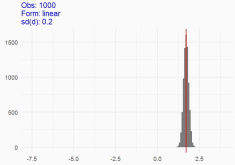

Figure 29: FRD (2sls)

- Bias: 0.002

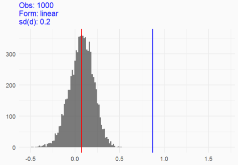

Figure 30: First-stage (LLR)

- Bias: -0.803

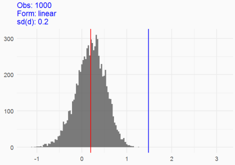

Figure 31: ITT (LLR)

- Bias: -1.292

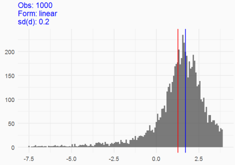

Figure 32: FRD (LLR)

- Bias: -0.430

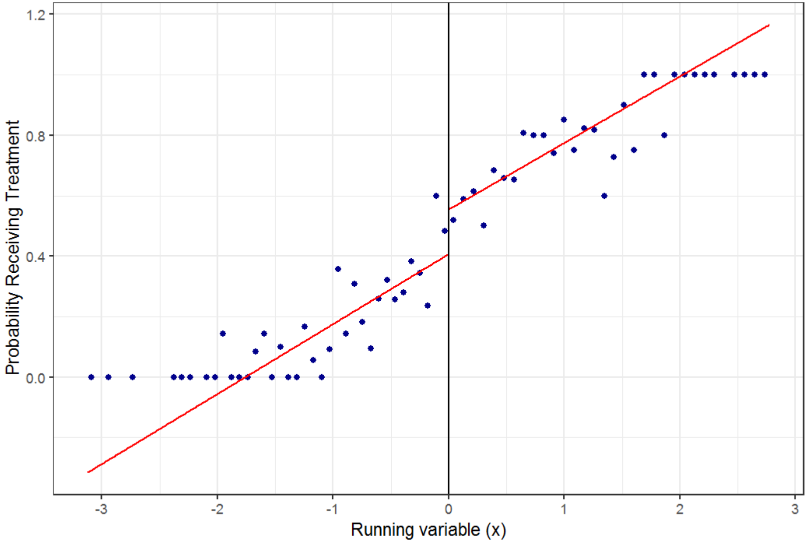

Figure 33: Case with $\sigma_d = 1$

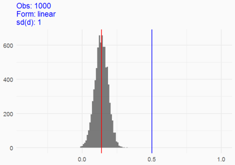

Figure 34: First-stage

- Bias: -0.363

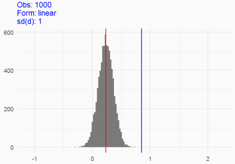

Figure 35: ITT

- Bias: -0.616

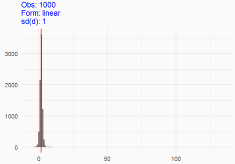

Figure 36: FRD

- Bias: 0.018

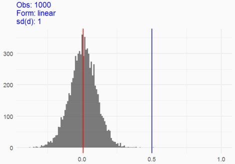

Figure 37: First-stage

- Bias: -0.494

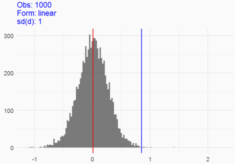

Figure 38: ITT

- Bias: -0.839

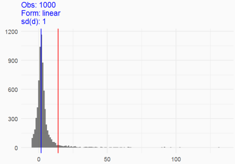

Figure 39: FRD

- Bias: 12.478

Conclusion

Conclusion

- **Sharp RDD**
  - Vanilla OLS estimator is the best model when functional form assumption is correct
  - Local linear regression is the best model when the functional form is unknown and the running variable has outliers, i.e., observations far from the cutoff with great values.
- **Fuzzy RDD**
  - Non-compliance to treatment assignment damage estimators' properties
  - Vanilla OLS estimators work better than the Local linear model when non-compliance is close to the cutoff
  - When the first stage is weak, estimates become unreliable