Regression Discontinuity Designs

Hansen (2022, Chapter 21)

Yasuyuki Matsumura (Kyoto University)

Last Updated: June 2, 2025

https://yasu0704xx.github.io

Introduction

- Regression discontinuity designs (RDDs) are quasi-experimental designs which allow researchers to identify the causal effect of endogenous treatment on an outcome based on discontinuous policy rules.
- Local randomization is a key idea.
    - Consider a certain discontinuous rule under which treatment (e.g. college scholarship) is determined by whether a continuous covariate (e.g. admission score) is greater than a known threshold.
    - If all factors determined prior to the treatment are balanced just above and just below the threshold, the average causal effect can be estimated by comparing the mean outcome just above the threshold with that just below the threshold.
- Software package: `rdrobust`

Literature

- Here we review Chapter 21 of Hansen (2022) [20].
- Excellent reviews/textbooks on regression discontinuity designs include Abadie and Cattaneo (2018) [1], and Cattaneo, Idrobo and Titiunik (2021, 2024) [14] [15].
- 日本語の文献：
  - 川口・澤田（2024）[32]
  - 末石（2024）[33]
  - 高野（2025）[34]

Contents

Identification

Setup

- $Y_d \in \mathbb{R}$, $d \in \{0,1\}$: potential outcome
- Y : observed outcome
- $D \in \{0,1\}$: treatment, which may be endogenous in that some unobserved factors may affect both D and Y.
- $\theta = Y_1 - Y_0$: treatment effect for an individual
  $\Rightarrow$ We cannot observe θ without restrictive assumptions, because either Y_1 or Y_0 is unobservable.
- Instead, we are interested in the conditional average treatment effect (conditional ATE)

$$\theta(x) = \mathbb{E}[\theta|X = x] = \mathbb{E}[Y_1 - Y_0|X = x],$$

where X is an observable covariate.

Sharp RD

- Suppose that treatment is determined by

$$D = 1(X \geq c),$$

  where the cut-off c is determined by policy or rule and common to all individuals.
- The covariate X is called the score, forcing variable, running variable, assignment variable, etc.
- In a standard RD setting, X is assumed to be continuously distributed on a subset of $\mathbb{R}$.

Identification in Sharp RD

- Set $\bar{\theta} = \theta(c) \equiv \mathbb{E}[\theta | X = c]$, which is the parameter of interest.
- Let $m(x) = \mathbb{E}[Y | X = x]$ and $m_d(x) = \mathbb{E}[Y_d | X = x]$ for each $d \in \{0, 1\}$. Note that $\theta(x) = m_1(x) - m_0(x)$.
- Set $m(x+) = \lim_{z \downarrow x} m(z)$ and $m(x-) = \lim_{z \uparrow x} m(z)$.

Theorem 21.1

Assume that treatent is assigned as $D = 1(X \geq c)$. Suppose that $m_0(x)$ and $m_1(x)$ are continuous at $x = c$. Then,

$$\bar{\theta} = m(c+) - m(c-).$$

Proof By construction,

$$Y = Y_0 \cdot 1(X < c) + Y_1 \cdot 1(X \geq c).$$

Taking expectations conditional on $X = x$, we obtain

$$m(x) = m_0(x)1(x < c) + m_1(x)1(x \geq c).$$

Since $m_0(x)$ and $m_1(x)$ are continuous at $x = c$,

$$m(c+) = m_1(c), \text{ and } m(c-) = m_0(c),$$

which completes the proof.

- The conditions for Theorem 21.1 mean that the conditional expectation of the untreated and treated outcome are continuously affected by the running variable.
- It is implied that the distributions of confounding factors, including observable covariates determined prior to the treatment, are balanced near the cutoff.
- In particular, there should be no policy/legal/experimental changes at the cutoff, except for the treatment assignment.
- Counterfactual: The continuity of $m_0(x)$ and $m_1(x)$ cannot be directly examined, since Y_0 and Y_1 are unobservable under $X \geq c$ and $X < c$, respectively.

Fuzzy RD

- In the fuzzy RD, D is partially determined by whether X is no less than a known fixed cutoff c, such that

$$\lim_{x\downarrow c}\mathbb{E}[D|X=x] \neq \lim_{x\uparrow c}\mathbb{E}[D|X=x],$$

  where $\lim_{x\downarrow c}$ and $\lim_{x\uparrow c}$ denote the right and left limits at $x=c$, respectively.
- Notice that $\mathbb{E}[D|X] = \mathbb{P}(D=1|X)$.

Identification in Fuzzy RD

- Define $Z = 1(X \geq c)$.
- Let $D_z, z \in \{0,1\}$ be the potential treatment status when $Z = z$. By construction, $D = ZD_1 + (1-Z)D_0$.
- Consider the following causal parameter for the "compliers," sometimes called the local Wald estimand:

$$\tau_{\mathsf{FRD}} \equiv \mathbb{E}[Y_1 - Y_0 \mid D_1 > D_0, X \in \{c-\epsilon, c+\epsilon\}].$$

- Under several assumptions, τ_{FRD} can be identified by

$$\tau_{\mathsf{FRD}} = \frac{\lim_{x \downarrow c} \mathbb{E}[Y|X=x] - \lim_{x \uparrow c} \mathbb{E}[Y|X=x]}{\lim_{x \downarrow c} \mathbb{E}[D|X=x] - \lim_{x \uparrow c} \mathbb{E}[D|X=x]}.$$

- The arguments are quite similar to the identification of LATE parameter in the IV estimations (so skiped in the class).
- See Hahn, Todd and van der Klaauw (2001) [19] and Hansen (2022, Sections 21.10-11) [20] for details.

Estimation

Notes

LLR

Bandwidth: Imbens and Kalyanaraman (2012) [21]

`rdrobust`

LPR ($p > 2$) should be avoided: Gelman and Imbens (2019) [18]

Regression Specification

Inference

Bandwidth

A naive solution: undersmoothing

Robust bias-corrected:
Calonico, Cattaneo and Titiunik (2014) [7] & Calonico, Cattaneo and Farrell (2020) [6]

`rdrobust`

Unifromly honest CI: Kolesar and Rothe (2018) [23] `rdhonest`

Covariates

Notes

Calonico et al. (2019) [8]

Kreiss and Rothe (2023) [24]

Arai, Otsu and Seo (2024) [4]

Chernozhukov et al. (2025) [16]

Noack, Olma and Rothe (2025) [27]

Falsification Test

Identification Assumptions

Theorem 21.1 (re)

Assume that treatent is assigned as $D = 1(X \geq c)$. Suppose that $m_0(x)$ and $m_1(x)$ are continuous at $x = c$. Then,

$$\bar{\theta} = m(c+) - m(c-).$$

- Counterfactual: The continuity of $m_0(x)$ and $m_1(x)$ cannot be directly examined, since Y_0 and Y_1 are unobservable under $X \geq c$ and $X < c$, respectively.
- Instead, researchers often examine certain necessary conditions.
- No manipulation: Under the above assumptions, it is necessary for the density of running variable X to be continuous at the cutoff point.

Manipulation Test

- Testing the continuity of the density of the assignment variable:
  - McCrary (2008) [26]
  - Otsu, Xu and Matsushita (2013) [28]
  - Cattaneo, Jansson and Ma (2020) [10]
- Testing the continuity of the conditional distribution of covariates:
  - Lee (2008) [25]
  - Canay and Kamat (2018) [9]
  - Fusejima, Ishihara and Sawada (2025) [17]

Placebo Test

- Take several placebo cutoff points.
- At a placebo cutoff point $c' < c$, researchers can observe whether the dendity of untreated potential outcome Y_0 is continuous or not.
- Similarly, at $c' > c$, they can observe whether the density of treated potential outcome Y_1 is continuous or not.
- If the continuity is observed, then there arises some plausibility of the identification assumptions in Theorem 21.1.
- Howevere, it is just a plausibility. Note that the continuity at placebo cutoff points is neither necessary nor sufficient for the identification assumptions.

Caveats on Pretesting

- Recent works argue that, in general, researchers should not implement such pretesting.
  - Roth (2022) [29] : Pre-trend test
  - Sueishi (2023) [30] : Hausman test
- See Section 5.2 of Fusejima, Ishihara and Sawada (2025) [17] for pretesting analysis in the RD setting.

Empirical Application

“Waiting for Life”

- The first study relying on RD can be found in Thistlehwaite and Cambell (1960) [31].
- As Cook (2008) [11] says that RD was “waiting for life,” RD was not popular until 1999, the year when Angrist and Lavy (1999) [3] and Black (1999) [5] were published.
- Hahn, Todd, and Klaauw (2001) [19] formalize general RDDs and establish identification results for treatment effects.

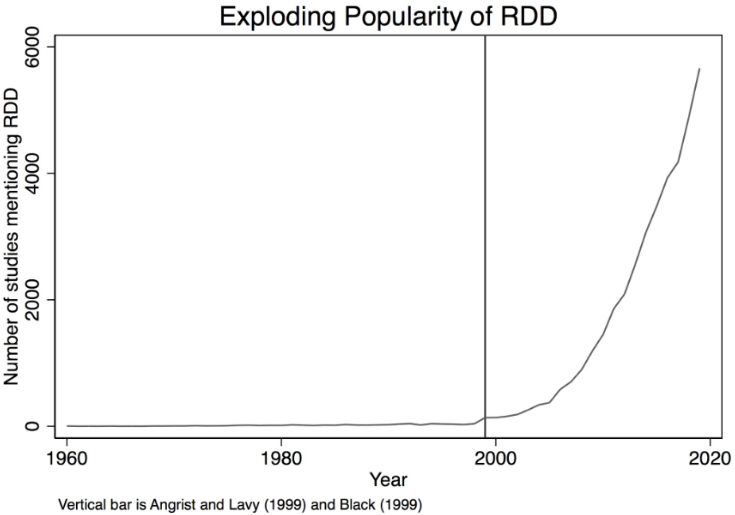

Figure 6.1 of Cunningham (2021) [12]

Anderson (2018, AER)

- Anderson (2018) [2] examines causal relationship between legal systems and female HIV infection rates in sub-Saharan Africa.
- RD, motiveted by Dell (2010) [13]:
  - As-if random borders can mitigate an endogeneity emerged within ethnicity level.
- Result 1 (HIV positive rates)
  - Female: common law countries $>$ civil law countries
  - Male: no significant difference
- Result 2 (Contraception use)
  - Female: common law countries $<$ civil law countries
  - Male: common law countries $<$ civil law countries
- Common Law $\Rightarrow$ Female bargaining power $\downarrow$
  $\Rightarrow$ Negotiation for safe sex practices $\times$ $\Rightarrow$ HIV prevalence $\uparrow$

Split Ethnic Groups with Different Legal Origins

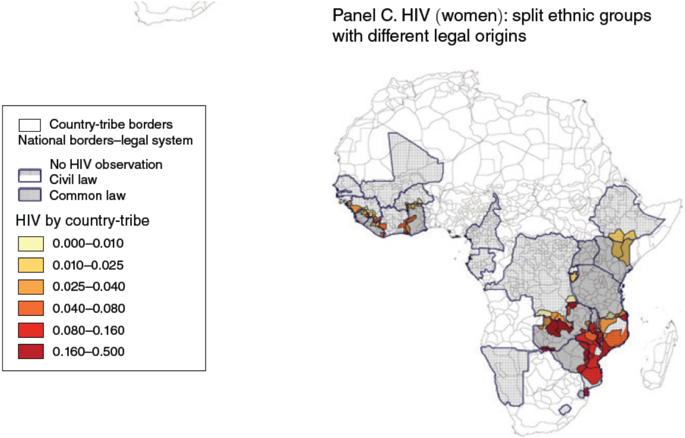

Figure 1. Female HIV by Ethnic Group

Empirical Strategy & Main Result

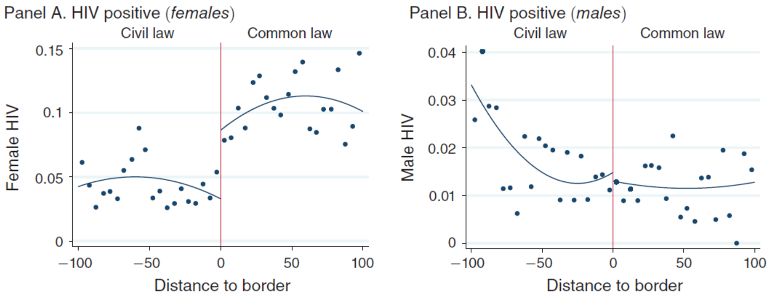

FIGURE 2. HIV POSITIVE

- Y_i = HIV Positive Rate (Proxy for Females Bargaining Power)
- D_i = Legal Origins (Civil Law vs Common Law)
- $X_i - c$ = Distance to Border
- As-if random borders of Scramble for Africa (1884)

  $\Longrightarrow$ Same cultures but different legal origins near a border!

References

- Abadie, A., and Cattaneo, M. D. (2018). Econometric methods for program evaluation. *Annual Review of Economics*, **10**, 465-503.
- Anderson, S. Legal origins and female HIV. *American Economic Review*, 108(6):1407-1439.
- Angrist, Joshua D.; Lavy, Victor. (1999). Using Maimonides ' Rule to Estimate the Effect of Class Size on Scholastic Achievement. *Quarterly Journal of Economics*, 114(2):533-575
- Arai, Yoichi; Otsu, Taisuke; Seo, Myung Hwan. (2024). Regression Discontinuity Design with Potentially Many Covariates. *arXiv preprint arXiv:2109.08351*

Black, Sandra E. (1999). " Do Better Schools Matter? Parental Valuation of Elementary Education. " *Quarterly Journal of Economics*, 114(2):577-599.

Calonico, S., M. D. Cattaneo and M. H. Farrell (2020). Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. *The Econometrics Journal 23*(2), 192-210.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica 82*(6), 2295-2326.

References iii

- Calonico, Sebastian; Cattaneo, Matias D.; Farrell, Max H.; Titiunik, Roc o. (2019). Regression Discontinuity Designs Using Covariates. *The Review of Economics and Statistics*, 101(3):442-451
- Canay, I. A. and V. Kamat (2018). Approximate permutation tests and induced order statistics in the regression discontinuity design. *Review of Economic Studies 85*, 1577-1608.
- Cattaneo, M. D., M. Jansson and X. Ma (2020). Simple local polynomial density estimators. *Journal of the American Statistical Association 115*(531), 1449-1455.

- Cook, Thomas D. (2008). Waiting for Life to Arrive ' : A History of the Regression-Discontinuity Design in Psychology, Statistics and Economics. *Journal of Econometrics*, 142(2):636-654.
- Cunningham, Scott. (2021). *Causal Inference: The Mixtape.* Yale University Press
- Dell, Melissa. (2010). The Persistent Effects of Peru's Mining Mita. *Econometrica*, 78(6):1863-1903.

Cattaneo, Matias D., Nicolas Idrobo, and Rocio Titiunik. (2019). *A Practical Introduction to Regression Discontinuity Designs: Foundations*. Elements in Quantitative and Computational Methods for the Social Sciences. Cambridge University Press.

Cattaneo, Matias D., Nicolas Idrobo, and Rocio Titiunik. (2024). *A Practical Introduction to Regression Discontinuity Designs: Extensions*. Elements in Quantitative and Computational Methods for the Social Sciences. Cambridge University Press.

- Chernozhukov, Victor; Hansen, Christian; Kallus, Nathan; Spindler, Martin; Syrgkanis, Vasilis. (2025). Applied Causal Inference: Machine Learning and AI. CausalML Book, Version 0.1.1. Available at: https://causalml-book.org.
- Fusejima, K., T. Ishihara and M. Sawada (2024). A unified diagnostic test for regression discontinuity designs. arXiv:2205.04345v4.
- Gelman, A.; Imbens, G.W. Why high-order polynomials should not be used in regression discontinuity designs. (2019). *Journal of Business & Economic Statistics*, 37(3):447-456.

- Hahn, Jinyong; Todd, Petra; van der Klaauw, Wilbert. (2001) Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design. *Econometrica*, 69(1):201-209.
- Hansen, B. E. (2022). *Econometrics.* Princeton.
- Imbens, G. W. and K. Kalyanaraman (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies 79*(3), 933-959.
- Imbens, G. W., and Rubin, D. B. (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences*. New York: Cambridge University Press.

- Kolesar, Michal; Rothe, Christoph. (2018). Inference in Regression Discontinuity Designs with a Discrete Running Variable. *American Economic Review*, 108(8):2277-2304
- Kreiss, Alexander; Rothe, Christoph. (2023). Inference in Regression Discontinuity Designs with High-dimensional Covariates. *The Econometrics Journal*, 26(2):105-123.
- Lee, D. S. (2008). Randomized experiments from non-random selection in US House elections. *Journal of Econometrics 142*(2), 675-697.
- McCrary (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics 142*(2), 698-714.

- Noack, C.; Olma, T.; Rothe, C. (2025). Flexible Covariate Adjustments in Regression Discontinuity Designs. *arXiv preprint arXiv:2107.07942*
- Otsu, T., K. L. Xu, and Y. Matsushita (2013). Estimation and inference of discontinuity in density. *Journal of Business and Economic Statistics 31*(4), 507-524.
- Roth, J. (2022). Pretest with caution: Event-study estimates after testing for parallel trends. *American Economic Review: Insights 4*(3), 305-322.
- Sueishi, N. (2023). A misuse of specification tests. arXiv:2211.11915v1.

- Thistlehwaite, Donald, and Donald Campbell. (1960). Regression-Discontinuity Analysis: An Alternative to the Ex-Post Facto Experiment. *Journal of Educational Psychology 51*, 309-317.
- 川口康平，澤田真行（2024）『因果推論の計量経済学』日本評論社.
- 末石直也（2024）『データ駆動型回帰分析』日本評論社.
- 高野久紀（2025）『開発経済学：実証経済学へのいざない』日本評論社.