# Regression Discontinuity Designs

Hansen (2022, Chapter 21)

Yasuyuki Matsumura (Kyoto University)

Last Updated: June 3, 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.

# Literature

- Here we review Chapter 21 of Hansen (2022) [25].
- Excellent reviews/textbooks on regression discontinuity designs include Abadie and Cattaneo (2018) [1], and Cattaneo, Idrobo and Titiunik (2021, 2024) [13] [14].
- The common software package is `rdrobust` by Calonico, Cattaneo, Farrell and Titiunik.
- 日本語の文献：
    - 川口・澤田（2024）[49]
    - 末石（2024）[50]
    - 高野（2025）[51]

# Contents

# Identification

## Setup: Rubin Causal Model

- $Y_i(d) \in \mathbb{R}$, $d \in \{0,1\}$ : potential outcome
- $Y_i$ : observed outcome
- $D_i \in \{0,1\}$ : treatment, which may be endogenous in that some unobserved factors may affect both $D_i$ and $Y_i$.
- $\theta = Y_i(1) - Y_i(0)$ : treatment effect for an individual
  $\Rightarrow$ We cannot identify $\theta$ without restrictive assumptions, because either $Y_i(1)$ or $Y_i(0)$ is unobservable.
- Instead, we are often interested in causal parameters such as ATE$= \mathbb{E}[Y_i(1) - Y_i(0)]$, ATT$= \mathbb{E}[Y_i(1) - Y_i(0)|D = 1]$, etc.

# Sharp RD

- Suppose that treatment is determined by

$$D_i = 1(X_i \geq c), \tag{1}$$

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

## Identification in Sharp RD

*Assume that, for each* $d \in \{0, 1\}$,

$$\mathbb{E}[Y_i(d)|X_i = x] \text{ is continuous at } x = c. \qquad (2)$$

*Under Assumptions (1) and (2), the average causal effect at the cutoff point* $\tau_{SRD} \equiv \mathbb{E}[Y_i(1) - Y_i(0)|X_i = c]$ *is identified by*

$$\tau_{SRD} = \lim_{x \downarrow c} \mathbb{E}[Y_i|X_i = x] - \lim_{x \uparrow c} \mathbb{E}[Y_i|X_i = x], \qquad (3)$$

*where* $\lim_{x \downarrow c}$ *and* $\lim_{x \uparrow c}$ *denote the right and left limits at* $x = c$, *respectively.*

**Proof** By construction,

$$Y_i = Y_i(0) \cdot 1(X_i < c) + Y_i(1) \cdot 1(X_i \geq c).$$

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

$$\begin{aligned}
&\mathbb{E}[Y_i|X_i = x] \\
&= \mathbb{E}[Y_i(0)|X_i = x]1(X_i < c) + \mathbb{E}[Y_i(1)|X_i = x]1(X_i \geq c)
\end{aligned}$$

Since $\mathbb{E}[Y_i(0)|X_i = x]$ and $\mathbb{E}[Y_i(1)|X_i = x]$ are continuous at $x = c$, they are identified by

$$\begin{aligned}
\mathbb{E}[Y_i(0)|X_i = x] &= \lim_{x \uparrow c} \mathbb{E}[Y_i|X_i = x], \\
\mathbb{E}[Y_i(1)|X_i = x] &= \lim_{x \downarrow c} \mathbb{E}[Y_i|X_i = x],
\end{aligned}$$

which completes the proof.

- Assumption (2) means 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 $\mathbb{E}[Y_i(d)|X_i = x]$ at $x = c$ cannot be directly examined, since $Y_i(0)$ and $Y_i(1)$ are unobservable under $X_i \geq c$ and $X_i < c$, respectively.

# Fuzzy RD

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

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

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

# Identification in Fuzzy RD

- Define $Z_i = 1(X_i \geq c)$.
- Let $D_i(z), z \in \{0, 1\}$ be the potential treatment status when $Z_i = z$. By construction, $D = Z_i D_i(1) + (1 - Z_i) D_i(0)$.
- Consider the following causal parameter for the "compliers:"

$$\tau_{\textsf{FRD}} \equiv \mathbb{E}[Y_i(1) - Y_i(0) \mid D_i(1) > D_i(0), X_i \in \{c - \epsilon, c + \epsilon\}].$$

- Under several assumptions, $\tau_{\textsf{FRD}}$ can be identified by the local Wald estimand:

$$\tau_{\textsf{FRD}} = \frac{\lim_{x \downarrow c} \mathbb{E}[Y_i | X_i = x] - \lim_{x \uparrow c} \mathbb{E}[Y_i | X_i = x]}{\lim_{x \downarrow c} \mathbb{E}[D_i | X_i = x] - \lim_{x \uparrow c} \mathbb{E}[D_i | X = x]}. \tag{4}$$

- 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) [24], Dong (2018) [20], and Hansen (2022, Sections 21.10-11) [25] for details.

# Estimation

# Identification Results (re)

- Recall that the causal parameters $\tau_{\textsf{SRD}}$ and $\tau_{\textsf{FRD}}$ are identified respectively by (3) and (4):

$$\tau_{\textsf{SRD}} = \lim_{x \downarrow c} \mathbb{E}[Y_i | X_i = x] - \lim_{x \uparrow c} \mathbb{E}[Y_i | X_i = x],$$

$$\tau_{\textsf{FRD}} = \frac{\lim_{x \downarrow c} \mathbb{E}[Y_i | X_i = x] - \lim_{x \uparrow c} \mathbb{E}[Y_i | X_i = x]}{\lim_{x \downarrow c} \mathbb{E}[D_i | X_i = x] - \lim_{x \uparrow c} \mathbb{E}[D_i | X = x]}.$$

- For expositional purpose, denote one of $Y_i$ and $D_i$ by $A_i$, and define the following one-sided limits:

$$\mu_A^+ \equiv \lim_{x \downarrow c} \mathbb{E}[A_i | X_i = x], \quad \mu_A^- \equiv \lim_{x \uparrow c} \mathbb{E}[A_i | X_i = x].$$

- For estimating $\tau_{\textsf{SRD}}$ and $\tau_{\textsf{FRD}}$, it suffices to estimate $\mu_Y^+, \mu_Y^-, \mu_D^+$, and $\mu_D^-$.

# Estimation Procedures: Local Polynomial Regressions

- The quantities $\mu_Y^+, \mu_Y^-, \mu_D^+$, and $\mu_D^-$ are commonly estimated by local polynomial regressions (LPRs).
- As we studied before, the local constant (Nadaraya-Watson) regression causes the boundary bias. LPRs can circumvent such boundary bias (Hahn, Todd and van der Klaauw, 2001 [24]; Porter, 2003 [41]).
- A recommended choice of local polynomial order $p$ is 1 or 2. [1]
- On the other hand, there are several recommendations on bandwidth selection, which we study later.

[1] Gelman and Imbens (2019) [22] argue that controlling for global high-order polynomials in regression discontinuity analysis is a flawed approach with three major problems: it leads to noisy estimates, sensitivity to the degree of the polynomial, and poor coverage of confidence intervals.

# Estimation of $\tau_{\mathsf{SRD}}$ and $\tau_{\mathsf{FRD}}$

- Consider the following $p$-th order LPRs:

$$\hat{\beta}_A^+ \equiv \arg\min_{\beta \in \mathbb{R}^{p+1}} \sum_{i=1}^{n} 1(X_i \geq c) \left(A_i - [r_p(X_i - c)]^T \beta\right)^2 K\left(\frac{X_i - c}{h}\right),$$

$$\hat{\beta}_A^- \equiv \arg\min_{\beta \in \mathbb{R}^{p+1}} \sum_{i=1}^{n} 1(X_i < c) \left(A_i - [r_p(X_i - c)]^T \beta\right)^2 K\left(\frac{X_i - c}{h}\right),$$

where $p \geq 1$, $r_p(x) \equiv (1, x, \cdots, x^p)$ is a vector of polynomials, $h > 0$ is bandwidth, and $K(\cdot)$ is a kernel function.

- The LPR estimators of $\mu_A^+$ and $\mu_A^-$ are the first elements of $\hat\beta_A^+$ and $\hat\beta_A^-$, respectively:

$$\hat\mu_A^+ \equiv e_1^T \hat\beta_A^+, \quad \hat\mu_A^- \equiv e_1^T \hat\beta_A^-,$$

where $e_1 \equiv (1, 0, \cdots, 0)^T$ is the first unit vector.

- The causal parameters $\tau_{\mathsf{SRD}}$ and $\tau_{\mathsf{FRD}}$ can be estimated respectively by

$$\hat\tau_{\mathsf{SRD}} = \hat\mu_Y^+ - \hat\mu_Y^-, \quad \hat\tau_{\mathsf{FRD}} = \frac{\hat\mu_Y^+ - \hat\mu_Y^-}{\hat\mu_D^+ - \hat\mu_D^-}.$$

# Bandwidth Selection

- Under certain regularity conditions, we can show that

$$\mathbb{E}[\hat{\tau}_{\mathsf{SRD}} - \tau_{\mathsf{SRD}} | X_1, \cdots, X_n] = h^{p+1}\mathcal{B},$$
$$\mathsf{Var}[\hat{\tau}_{\mathsf{SRD}} | X_1, \cdots, X_n] = \frac{1}{nh}\mathcal{V},$$
$$\sqrt{nh}\left(\hat{\tau}_{\mathsf{SRD}} - \tau_{\mathsf{SRD}} - h^{p+1}\mathcal{B}\right) \xrightarrow{d} \mathsf{Normal}(0, \mathcal{V}).$$

- The AMSE is given by

$$\mathsf{AMSE}(\hat{\tau}_{\mathsf{SRD}}) = h^{2(p+1)}\mathcal{B}^2 + \frac{1}{nh}\mathcal{V}.$$

- Imbens and Kalyanaraman (2012) [27] propose the (A)MSE optimal bandwidth, which satisfies that

$$h_{\mathsf{IK2012}} \propto n^{-\frac{1}{2p+3}}.$$

# Inference

# Asymptotic Bias

- If we are interested in the point estimations of $\mu_Y^+$ and $\mu_Y^-$, it is sufficient to select IK2012's MSE optimal bandwidth.
- In practice, however, we are also interested in inference: IK2012's bandwidth does not satisfy that $nh^5 \to 0$. Then, we cannot eliminate the asymptotic bias.

How to select bandwidth?

1. Undersmoothing
2. Robust bias-corrected inference (CCT2014 [9], CCF2020 [8])
3. Uniformly honest inference (KR2018 [32])

# Undersmoothing

- A naive solution to mitigate asymptotic bias is to use undersmoothing.
- By using undersmoothing bandwidth such that

$$\sqrt{nh}h^{p+1} \to 0,$$

the standard $t$ statistic satisfies

$$\frac{\sqrt{nh}(\hat{\tau}_{\mathsf{SRD}} - \tau_{\mathsf{SRD}})}{\sqrt{\mathcal{V}_{\mathsf{SRD}}}} \xrightarrow{d} \mathsf{Normal}(0, 1).$$

# Robust Bias-Corrected Inference (CCT2014 & CCF2020)

- Calonico, Cattaneo and Titiunik (2014) [9], and Calonico, Cattaneo and Farrell (2020) [8] propose to correct asymptotic bias for valid statistical inference, not to eliminate the bias by undersmoothing.
- Let $\hat{\mathcal{B}}$ denote the estimator of $\mathcal{B}$ based on LPRs using bandwidth $b$, which can differ from $h$. The bias-corrected estimator of $\tau_{\mathsf{SRD}}$ is given by

$$\hat{\tau}^{\mathsf{bc}}_{\mathsf{SRD}} \equiv \hat{\tau}_{\mathsf{SRD}} - h^{p+1}\hat{\mathcal{B}}.$$

- Under certain conditions on $h$ and $b$ and regularity conditions, CCT2014 [9] show that the robust bias-corrected $t$ statistic satisfies

$$T^{\mathsf{bc}}_{\mathsf{SRD}} \equiv \frac{\sqrt{nh}\left(\hat{\tau}^{\mathsf{bc}}_{\mathsf{SRD}} - \tau_{\mathsf{SRD}}\right)}{\sqrt{\mathcal{V}^{\mathsf{bc}}_{\mathsf{SRD}}}} \xrightarrow{d} \text{Normal}(0,1),$$

where $\mathcal{V}^{\mathsf{bc}}_{\mathsf{SRD}} \equiv \mathcal{V}_{\mathsf{SRD}} + \mathcal{C}^{\mathsf{bc}}_{\mathsf{SRD}}$ and $\mathcal{C}^{\mathsf{bc}}_{\mathsf{SRD}}$ is a correction term.

- Using an estimator of $\mathcal{V}^{\mathsf{bc}}_{\mathsf{SRD}}$, the $1-\alpha$ confidence interval of $\tau_{\mathsf{SRD}}$ based on $T^{\mathsf{bc}}_{\mathsf{SRD}}$ is given by

$$\left[\hat{\tau}^{\mathsf{bc}}_{\mathsf{SRD}} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{\mathcal{V}}^{\mathsf{bc}}_{\mathsf{SRD}}}{nh}}, \hat{\tau}^{\mathsf{bc}}_{\mathsf{SRD}} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{\mathcal{V}}^{\mathsf{bc}}_{\mathsf{SRD}}}{nh}}\right]. \tag{5}$$

- CCF2020 [8] propose to select a bandwidth which minimizes the coverage rate error of (5).

# Uniformly Honest Inference (KR2018)

- Kolesar and Rothe (2018) [32] propose another bias correction.
- Assuming that the true regression function is in a known functional class and analyzing the upper bound of the bias, they construct a confidence interval uniformly robust to any DGPs.
- Some characteristics of KR2018 are that they do not require $h \to 0$, and thus they allow the runnning variable to be discretely distributed.

- Consider the following LLR with the uniform kernel:

$$\min_{\alpha,\tau_h,\beta,\gamma} \sum_{i=1}^{n} 1(|X_i - c| \leq h) \left[Y_i - \alpha - \tau_h D_i - \beta(X_i - c) - \gamma D_i(X_i - c)\right]^2 .$$

- Define

$$n_h = \sum_{i=1}^{n} 1(|X_i - c| \leq h),$$

$$\tilde{\tau}_h = \mathbb{E}[\hat{\tau}_h | X_1, \cdots, X_n].$$

- Let $\dfrac{\hat{\sigma}^2}{n_h}$ be the estimator of $\text{Var}(\hat{\tau}_h | X_1, \cdots, X_n)$.

- By simple calculation, we obtain

$$\frac{\sqrt{n_h}(\hat{\tau}_h - \tau_{\mathsf{SRD}})}{\hat{\sigma}} = \frac{\sqrt{n_h}(\hat{\tau}_h - \tilde{\tau}_{\mathsf{SRD}})}{\hat{\sigma}} + \frac{\sqrt{n_h}(\tilde{\tau}_h - \tau_{\mathsf{SRD}})}{\hat{\sigma}}$$

- For any $h$ (even if $h$ violates $h \to 0$), the first term of RHS satisfies that

$$\frac{\sqrt{n_h}(\hat{\tau}_h - \tilde{\tau}_{\mathsf{SRD}})}{\hat{\sigma}} \xrightarrow{d} \mathsf{Normal}(0, 1).$$

- The second term is the bias term. Let us analyze its upper bound.

- Assume that, for a known constant $K$,[2]

$$\mu(x) = \mathbb{E}[Y_i | X_i = x] \in \mathcal{M},$$

where

$$\mathcal{M} = \{\mu : |\mu'(a) - \mu'(b)| \leq K|a - b| \text{ for all } a, b < c \text{ and all } a, b > c\}.$$

- Then, for $\hat{\sigma}$ that is estimated with certain methods, we can obtain

$$\gamma_{sup} \equiv \sup_{\mu \in \mathcal{M}} \frac{\sqrt{n_h}|\tilde{\tau}_h - \tau_{\mathsf{SRD}}|}{\hat{\sigma}}.$$

[2] We cannot choose $K$ by data-driven procedures.

- Based on this result, we can construct the following confidence interval:

$$\mathsf{CI} = \left[\hat{\tau}_h - \mathsf{cv}_{1-\alpha}(\gamma_{sup})\frac{\hat{\sigma}}{\sqrt{n_h}}, \hat{\tau}_h + \mathsf{cv}_{1-\alpha}(\gamma_{sup})\frac{\hat{\sigma}}{\sqrt{n_h}}\right],$$

where $\mathsf{cv}_{1-\alpha}(\gamma)$ denotes the $1-\alpha$ quantile of $|\mathsf{Normal}(\gamma, 1)|$.

- The CI above is an asymptotically uniform confidence interval with respect to $\mathcal{M}$:

$$\lim_{n\to\infty} \inf_{\mu\in\mathcal{M}} \mathbb{P}_\mu(\tau_{\mathsf{SRD}} \in \mathsf{CI}) \geq 1-\alpha,$$

which KR2018 [32] call the honest confidence interval.

# Covariates

## RD with Covariates

- Robinson (1988) [42]
- Calonico, Cataneo, Farrell and Titiunik (2019) [10]
- Noack, Olma and Rothe (2025) [38]

## RD with High-Dimensional Covariates

- Kreiss and Rothe (2023) [33]
- Arai, Otsu and Seo (2024) [5]
- Review: Chernozhukov et al. (2025, Chapter 17) [16]

# Covariate Adjustment Relying on Robinson (1988)

- Researchers can adjust covariates to RD analysis for the purpose of efficiency gain by implementing the following 2 step LPR procedures, which are based on Robinson's (1988) [42] estimation of semiparametric partially linear model.

1. Use the RDD local linear estimator to regress $Y_i$ on $X_i$ to obtain the first-step fitted values $\widehat{m}_i = \widehat{m}(X_i)$.
2. Using LL regression, regress $Z_{i1}$ on $X_i$, $Z_{i2}$ on $X_i$, ..., and $Z_{ik}$ on $X_i$, obtaining the fitted values for the covariates, say $\widehat{g}_{1i}, \ldots, \widehat{g}_{ki}$.
3. Regress $Y_i - \widehat{m}_i$ on $Z_{1i} - \widehat{g}_{1i}, \ldots, Z_{ki} - \widehat{g}_{ki}$ to obtain the coefficient estimate $\widehat{\beta}$ and standard errors.
4. Construct the residual $\widehat{e}_i = Y_i - Z_i'\widehat{\beta}$.
5. Use the RDD local linear estimator to regress $\widehat{e}_i$ on $X_i$ to obtain the nonparametric estimator $\widehat{m}(x)$, conditional ATE $\widehat{\theta}$, and associated standard errors.

- The estimator reaches the semiparametric efficiency bound.

# Covariate-Adjusted RD Estimation (CCFT2019)

- Calonico, Cattaneo, Farrell and Titiunik (2019) [10] propose to include covariates in RD for the purpose of efficiency gain.
- CCFT2019 examine the WLS estimators with weights $K(X_i/h)$ computed only for observations with $X_i \in [-h, h]$ in the five different specifications, where the cutoff is normalized as $c = 0$, and $\bar{Z}$, $\bar{Z}^-$, $\bar{Z}^+$ correspond, respectively, to the sample averages of $Z_i$ for $X_i \in [-h, h],\ X_i \in [-h, 0),$ and $X_i \in [0, h]$:

$$\tilde{\tau} \;:\; \tilde{Y}_i = \tilde{\alpha} + T_i\tilde{\tau} + X_i\tilde{\beta}_- + T_iX_i\tilde{\beta}_+ + \mathbf{Z}_i'\tilde{\boldsymbol{\gamma}}, \tag{2}$$

$$\check{\tau} \;:\; \check{Y}_i = \check{\alpha} + T_i\check{\tau} + X_i\check{\beta}_- + T_iX_i\check{\beta}_+ + (1-T_i)\mathbf{Z}_i'\check{\boldsymbol{\gamma}}_- + T_i\mathbf{Z}_i'\check{\boldsymbol{\gamma}}_+, \tag{3}$$

$$\dot{\tau} \;:\; \dot{Y}_i = \dot{\alpha} + T_i\dot{\tau} + X_i\dot{\beta}_- + T_iX_i\dot{\beta}_+ + (\mathbf{Z}_i - \bar{\mathbf{Z}})'\dot{\boldsymbol{\gamma}}, \tag{4}$$

$$\ddot{\tau} \;:\; \ddot{Y}_i = \ddot{\alpha} + T_i\ddot{\tau} + X_i\ddot{\beta}_- + T_iX_i\ddot{\beta}_+ + (1-T_i)(\mathbf{Z}_i - \bar{\mathbf{Z}})'\ddot{\boldsymbol{\gamma}}_- + T_i(\mathbf{Z}_i - \bar{\mathbf{Z}})'\ddot{\boldsymbol{\gamma}}_+, \tag{5}$$

$$\dddot{\tau} \;:\; \dddot{Y}_i = \dddot{\alpha} + T_i\dddot{\tau} + X_i\dddot{\beta}_- + T_iX_i\dddot{\beta}_+ + (1-T_i)(\mathbf{Z}_i - \bar{\mathbf{Z}}_-)'\,\dddot{\boldsymbol{\gamma}}_- + T_i(\mathbf{Z}_i - \bar{\mathbf{Z}}_+)'\,\dddot{\boldsymbol{\gamma}}_+, \tag{6}$$

**Assumption SRD.** For $t \in \{0, 1\}$ and all $x \in [x_l, x_u]$, where $x_l, x_u \in \mathbb{R}$ such that $x_l < \bar{x} < x_u$:

a. The Lebesgue density of $X_i$, denoted $f(x)$, is continuous and bounded away from zero.
b. $\mu_{Y-}(x) = \mathbb{E}[Y_i(0)|X_i = x]$ and $\mu_{Y+}(x) = \mathbb{E}[Y_i(1)|X_i = x]$ are thrice continuously differentiable.
c. $\boldsymbol{\mu}_{Z-}(x) = \mathbb{E}[\mathbf{Z}_i(0)|X_i = x]$ and $\boldsymbol{\mu}_{Z+}(x) = \mathbb{E}[\mathbf{Z}_i(1)|X_i = x]$ are thrice continuously differentiable, and $\mathbb{E}[\mathbf{Z}_i(t)Y_i(t)|X_i = x]$ is continuously differentiable.
d. $\mathbb{V}[(Y_i(t), \mathbf{Z}_i(t)')|X_i = x]$ is continuously differentiable and invertible.
e. $\mathbb{E}[|(Y_i(t), \mathbf{Z}_i(t)')|^4|X_i = x]$, is continuous, where $|\cdot|$ denotes the Euclidean norm.

All limits are taken as $n \to \infty$, unless otherwise noted.

**Lemma 1** *(Sharp RD with covariates). Let assumption SRD hold, and assume the weights obey $K(u) = \mathbb{1}(u < 0)k(-u) + \mathbb{1}(u \geq 0)k(u)$, with $k(\cdot) : [0, 1] \mapsto \mathbb{R}_+$ bounded, zero outside its support, and positive and continuous on (0,1). If $nh \to \infty$ and $h \to 0$, then*

$$\tilde{\tau} \to_{\mathbb{P}} \tau - \left[\boldsymbol{\mu}_{Z+} - \boldsymbol{\mu}_{Z-}\right]' \boldsymbol{\gamma}_Y,$$

$$\check{\tau} \to_{\mathbb{P}} \tau - \left[\boldsymbol{\mu}_{Z+}'\boldsymbol{\gamma}_{Y+} - \boldsymbol{\mu}_{Z-}'\boldsymbol{\gamma}_{Y-}\right],$$

$$\dot{\tau} \to_{\mathbb{P}} \tau - \left[\boldsymbol{\mu}_{Z+} - \boldsymbol{\mu}_{Z-}\right]' \boldsymbol{\gamma}_Y,$$

$$\ddot{\tau} \to_{\mathbb{P}} \tau - \left[(\boldsymbol{\mu}_{Z+} - \bar{\boldsymbol{\mu}}_Z)'\boldsymbol{\gamma}_{Y+} - (\boldsymbol{\mu}_{Z-} - \bar{\boldsymbol{\mu}}_Z)'\boldsymbol{\gamma}_{Y-}\right],$$

$$\dddot{\tau} \to_{\mathbb{P}} \tau,$$

*where* $\boldsymbol{\gamma}_Y = (\boldsymbol{\sigma}_{Z-}^2 + \boldsymbol{\sigma}_{Z+}^2)^{-1}\mathbb{E}[(\mathbf{Z}_i(0) - \boldsymbol{\mu}_{Z-}(X_i))\,Y_i(0) + (\mathbf{Z}_i(1) - \boldsymbol{\mu}_{Z+}(X_i))\,Y_i(1)|X_i = \bar{x}]$, $\boldsymbol{\mu}_{Z-} = \boldsymbol{\mu}_{Z-}(\bar{x})$, $\boldsymbol{\gamma}_{Y-} = (\boldsymbol{\sigma}_{Z-}^2)^{-1}\mathbb{E}[(\mathbf{Z}_i(0) - \boldsymbol{\mu}_{Z-}(X_i))\,Y_i(0)|X_i = \bar{x}]$, $\boldsymbol{\sigma}_{Z-}^2 = \mathbb{V}[\mathbf{Z}_i(0)|X_i = \bar{x}]$, *and similarly for* $\boldsymbol{\mu}_{Z+}$, $\boldsymbol{\gamma}_{Y+}$ *and* $\boldsymbol{\sigma}_{Z+}^2$, *and* $\bar{\boldsymbol{\mu}}_Z = \boldsymbol{\mu}_{Z+}/2 + \boldsymbol{\mu}_{Z-}/2$.

Specifications (2)-(6), Assumption SRD, Lemma 1 in CCFT2019 [10]

- These results imply that only $\dddot{\tau}$ is consistent for $\tau$ without additional conditions.
- If there are no treatment effects on the covariates so that $\mu_{Z+} = \mu_{Z-}$, the estimators $\tilde{\tau}$ and $\dot{\tau}$ are also consistent to $\tau$.
- Since the treatment should not affect the covariates in a valid sharp RD, $\mu_{Z+} = \mu_{Z-}$ is a plausible assumption.
- CCFT2019 recommend to employ $\tilde{\tau}$, which they call the covariate-adjusted RD estimator.

# High-Dimensional Covariates

- Kreiss and Rothe (2023) [33] and Arai, Otsu and Seo (2024) [5] propose to implement certain Lasso-type selections to chosse active covariates and address the bias caused by these regularizations.
- KR2023 : a "localized" version of Lasso regression
- AOS2024 : the debiased Lasso (Zhang and Zhang, 2014 [48])

# Falsification Test

# Identification Assumptions

**Identification Assumption (2) in Sharp RD**

*For each* $d \in \{0, 1\}$,

$$\mathbb{E}[Y_i(d)|X_i = x] \text{ is continuous at } x = c.$$

- Counterfactual: The continuity of $\mathbb{E}[Y_i(d)|X_i = x]$ at $x = c$ cannot be directly examined, since $Y_i(0)$ and $Y_i(1)$ are unobservable under $X_i \geq c$ and $X_i < c$, respectively.
- Instead, researchers often examine certain necessary conditions.
- No manipulation: Under the above assumptions, it is necessary for the density of running variable $X_i$ to be continuous at the cutoff point.

# Manipulation Test

- Testing the continuity of the density of the assignment variable:
  - McCrary (2008) [36]
  - Otsu, Xu and Matsushita (2013) [40]
  - Cattaneo, Jansson and Ma (2020) [15]
  - Software packages by CJM2020: `lpdensity`, `rddensity`
  - Imai and Okamoto (2024) [26]
- Testing the continuity of the conditional distribution of covariates:
  - Lee (2008) [34]
  - Canay and Kamat (2018) [11]
  - Fusejima, Ishihara and Sawada (2025) [21]

# Placebo Test

- Take several placebo cutoff points.
- At a placebo cutoff point $c' < c$, researchers can observe whether the density of untreated potential outcome $Y_i(0)$ is continuous or not.
- Similarly, at $c' > c$, they can observe whether the density of treated potential outcome $Y_i(1)$ is continuous or not.
- If the continuity is observed, then there arises some plausibility of the identification assumptions in Theorem 21.1.
- However, 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) [43] : Pre-trend test
  - Sueishi (2023) [46] : Hausman test
- Pretesting analysis in the RD setting can be found in Section 5.2 of Fusejima, Ishihara and Sawada (2025) [21].

# Practical Recommendation

# Practical Recommendation in Sharp RD

- The polynomial order $p$ of LPRs should be 1 or 2.
- For point estimation, use IK2012's (A)MSE optimal bandwidth.
- For inference, use CCT2014 & CCF2020's bandwidth or KR2018's bandwidth.
- The common package `rdrobust` is equipped with point estimation relying on IK2012 and robust bias-corrected inference by CCT2014 & CCF2020.
- `RDHonest` provides KR2018's uniformly honest inference procedures.
- Observable covariates can be utilized to gain efficiency of the sharp RD estimator.
- Note that pretesting for Assumption (2) may distort the estimation and inference results.

# Empirical Application

# “Waiting for Life”

- The first study relying on RD can be found in Thistlehwaite and Cambell (1960) [47].
- As Cook (2008) [17] says that RD was “waiting for life,” RD was not popular until 1999, the year when Angrist and Lavy (1999) [4] and Black (1999) [7] were published.
- Hahn, Todd, and Klaauw (2001) [24] formalize general RDDs and establish identification results for treatment effects.

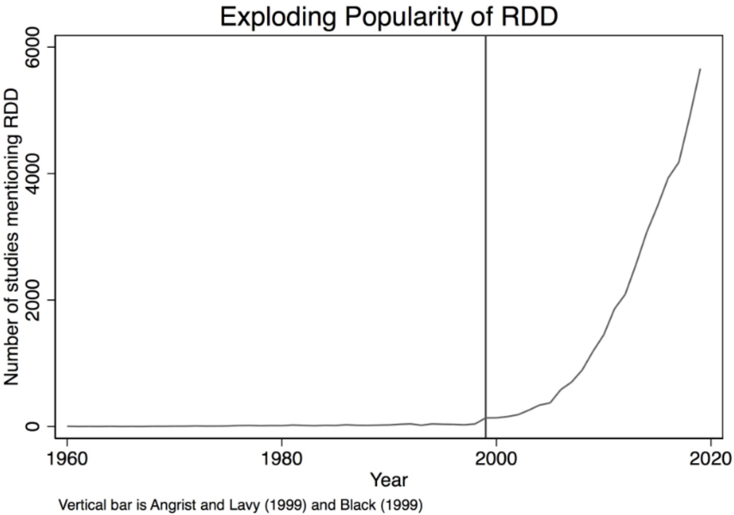


Figure 6.1 of Cunningham (2021) [18]

# Recent Empirical RD Studies

Ludwig and Miller (2007) [35],
Lee (2008) [34],
Matsudaira (2008) [37],
Battistin et al. (2009) [6],
Carpenter, Christopher and Carlos Dobkin (2009) [12],
Greenstone, Hornbeck and Moretti (2010) [23],
Abdulkadiroglu, Angrist and Pathak (2014) [2],
Ito (2014) [28],
Kleven et al. (2014) [31],
Shigeoka (2014) [44],
Shigeoka (2016) [45],
Ito and Sallee (2018) [29],
Oizer (2018) [39],
Kawai et al. (2023) [30], etc.

# Anderson (2018, AER)

- Anderson (2018) [3] examines causal relationship between legal systems and female HIV infection rates in sub-Saharan Africa.
- RD, motivated by Dell (2010) [19]:
  - As-if random borders can mitigate an endogeneity that emerges within ethnicities.
- 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

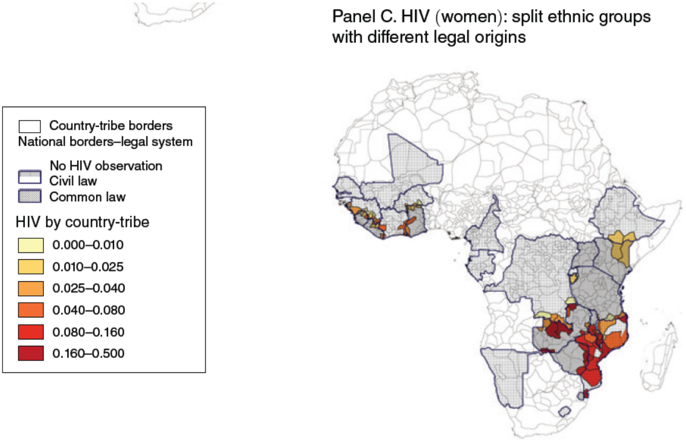


FIGURE 1. FEMALE HIV BY ETHNIC GROUP

## Specification

- Model:

$$\begin{aligned} Y_{rcepi} = & \alpha_0 + \alpha_1 L_{rc} + \alpha_2 X_{rc} + \alpha_3 X_{rcep} + \alpha_4 X_{rcepi} + f(BD_{rcep}) \\ & + \delta_e + \gamma_r + \lambda_t + \epsilon_{rcepi} \end{aligned}$$

- Subscripts: region, country, ethnic homeland, pixel
- $Y_{rcepi}$ : an outcome of interest
- $L_{rc}$ : common law legal system indicator
- $X_{rc}, X_{rcep}, X_{rcepi}$ : vectors of controls
- $f(BD_{rcep})$: a second-order RD polynomial of the distance from the centroid of pixel to the nearest national border with different legal origins
- $\delta_e$, $\gamma_r$ : fixed effects w.r.t. ethnicity and region, respectively
- $\epsilon_{rcepi}$ : clustered at the ethnicity and country level
- $\lambda_t$ : years of survey

# HIV Prevalence Rates

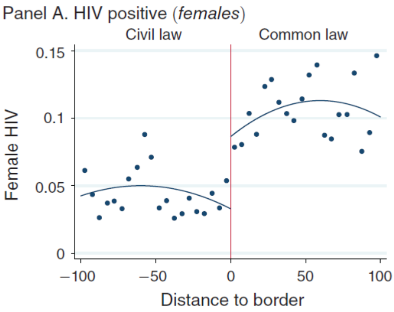


FIGURE 2. HIV POSITIVE

# HIV Prevalence Rates

TABLE 1—HIV POSITIVE: FEMALES AGED 15–49

| Variable | Whole sample $\leq$ 200 km | $\leq$ 150 km | $\leq$ 100 km | Non-Muslim Non-Polygynous $\leq$ 100 km | Muslim Polygynous $\leq$ 100 km |
|---|---|---|---|---|---|
| Common law | 0.016 | 0.018 | 0.019 | 0.016 | 0.007 |
| | (0.006) | (0.006) | (0.007) | (0.006) | (0.013) |
| Observations | 118,903 | 99,511 | 77,336 | 55,507 | 21,829 |

*Notes:* Standard errors are clustered at the ethnic and country level using the approach of Cameron, Gelbach, and Miller (2011). All estimations include: country, individual, and pixel controls; region fixed effects; ethnic fixed effects; second-order RD polynomial of distance to national border; and the year of the survey. Refer to the online Appendix for details on the data.

TABLE 2—HIV POSITIVE: MALES AGED 15–49

| Variable | Whole sample $\leq$ 200 km | $\leq$ 150 km | $\leq$ 100 km | Non-Muslim Non-Polygynous $\leq$ 100 km | Muslim Polygynous $\leq$ 100 km |
|---|---|---|---|---|---|
| Common law | 0.001 | 0.001 | −0.001 | −0.003 | 0.002 |
| | (0.006) | (0.005) | (0.005) | (0.005) | (0.01) |
| Observations | 50,754 | 40,780 | 31,189 | 24,261 | 6,928 |

*Notes:* Standard errors are clustered at the ethnic and country level using the approach of Cameron, Gelbach, and Miller (2011). All estimations include: country, individual, and pixel controls; region fixed effects; ethnic fixed effects; second-order RD polynomial of distance to national border; and the year of the survey. Refer to the online Appendix for details on the data.

# Protective Contraception

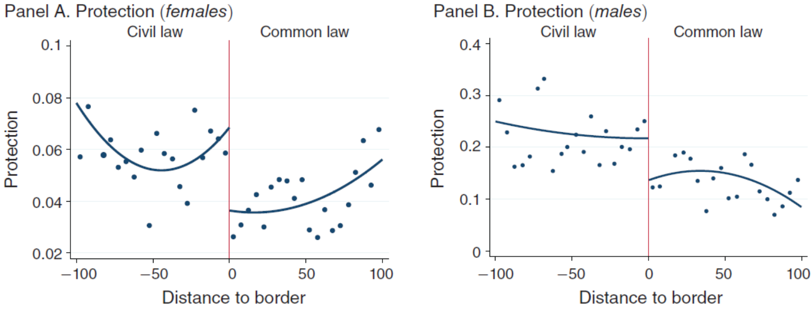


Figure 3. Protection

# Protective Contraception

Table 3—Protective Contraception: Females Aged 15–49

| Variable | Whole sample $\leq$ 200 km | $\leq$ 150 km | $\leq$ 100 km | Non-Muslim Non-Polygynous $\leq$ 100 km | Muslim Polygynous $\leq$ 100 km |
|---|---|---|---|---|---|
| Common law | –0.018 | –0.019 | –0.019 | –0.024 | –0.008 |
| | (0.006) | (0.006) | (0.007) | (0.01) | (0.007) |
| Observations | 117,263 | 97,285 | 76,698 | 55,261 | 21,437 |

*Notes:* Standard errors are clustered at the ethnic and country level using the approach of Cameron, Gelbach, and Miller (2011). All estimations include country, individual, and pixel controls; region fixed effects; ethnic fixed effects; second-order RD polynomial of distance to national border; and the year of the survey. Refer to the online Appendix for details on the data.

Table 4—Protective Contraception: Males Aged 15–49

| Variable | Whole sample $\leq$ 200 km | $\leq$ 150 km | $\leq$ 100 km | Non-Muslim Non-Polygynous $\leq$ 100 km | Muslim Polygynous $\leq$ 100 km |
|---|---|---|---|---|---|
| Common law | –0.07 | –0.07 | –0.07 | –0.08 | –0.003 |
| | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) |
| Observations | 81,873 | 67,887 | 52,902 | 46,016 | 6,886 |

*Notes:* Standard errors are clustered at the ethnic and country level using the approach of Cameron, Gelbach, and Miller (2011). All estimations include country, individual, and pixel controls; region fixed effects; ethnic fixed effects; second-order RD polynomial of distance to national border; and the year of the survey. Refer to the online Appendix for details on the data.

# References

# References i


- Abadie, A., and Cattaneo, M. D. (2018). Econometric methods for program evaluation. *Annual Review of Economics*, **10**, 465-503.
- Abdulkadiroglu, Atila; Angrist, Joshua D.; Pathak, Parag A. (2014). The Elite Illusion: Achievement Effects at Boston and New York Exam Schools. *Econometrica*, 82(1):137-196.
- 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

# References ii


- Arai, Yoichi; Otsu, Taisuke; Seo, Myung Hwan. (2024). Regression Discontinuity Design with Potentially Many Covariates. *arXiv preprint arXiv:2109.08351*
- Battistin, Erich; Brugiavini, Agar; Rettore, Enrico; Weber, Guglielmo. (2009). The Retirement Consumption Puzzle: Evidence from a Regression Discontinuity Approach. *American Economic Review*, 99(5):2209-2226.
- 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.
- Calonico, Sebastian; Cattaneo, Matias D.; Farrell, Max H.; Titiunik, Rocio. (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.
- Carpenter, Christopher; Dobkin, Carlos. The Effect of Alcohol Consumption on Mortality: Regression Discontinuity Evidence from the Minimum Drinking Age. *American Economic Journal: Applied Economics*, 1(1):164-182.
- 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.
- Cattaneo, M. D., M. Jansson and X. Ma (2020). Simple local polynomial density estimators. *Journal of the American Statistical Association 115*(531), 1449-1455.
- 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.

# References vi


- 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.
- Dong, Yingying. (2018). Alternative Assumptions to Identify LATE in Fuzzy Regression Discontinuity Designs. *Oxford Bulletin of Economics and Statistics*, 80(5):1020-1027.

- Fusejima, K., T. Ishihara and M. Sawada (2025). 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.
- Greenstone, Michael; Hornbeck, Richard; Moretti, Enrico. (2010). Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings. *Journal of Political Economy*, 118(3):536-598.

- 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.
- Imai, Shunsuke; Okamoto, Yuta. (2024). Kernel Choice Matters for Boundary Inference Using Local Polynomial Density: With Application to Manipulation Testing. *arXiv preprint arXiv:2306.07619*
- Imbens, G. W. and K. Kalyanaraman (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies 79*(3), 933-959.

- Ito, Koichiro. (2014). Do Consumers Respond to Marginal or Average Price? Evidence from Nonlinear Electricity Pricing. *American Economic Review*, 104(2):537-563.
- Ito, Koichiro; Sallee, James M. (2018). The Economics of Attribute-Based Regulation: Theory and Evidence from Fuel-Economy Standards. *Review of Economics and Statistics*, 100(2):319-336.
- Kawai, Kei; Nakabayashi, Jun; Ortner, Juan; Chassang, Sylvain. (2023). Using Bid Rotation and Incumbency to Detect Collusion: A Regression Discontinuity Approach. *The Review of Economic Studies*, 90(1):376-403.

- Kleven, Henrik J.; Landais, Camille; Saez, Emmanuel; Schultz, Esben A. Migration and Wage Effects of Taxing Top Earners: Evidence From the Foreigners ' Tax Scheme in Denmark. *Quarterly Journal of Economics*, 129(1):333-378.
- 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.
- Ludwig, Jens; Miller, Douglas L. (2007). Does Head Start Improve Children ' s Life Chances? Evidence from a Regression Discontinuity Design. *Quarterly Journal of Economics*, 122(1):159-208.
- McCrary (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics 142*(2), 698-714.

- Matsudaira, Jordan D. (2008). Mandatory summer school and student achievement. *Journal of Econometrics*, 142(2):829-850.
- Noack, C.; Olma, T.; Rothe, C. (2025). Flexible Covariate Adjustments in Regression Discontinuity Designs. *arXiv preprint arXiv:2107.07942*
- Ozier, Owen. (2018). Exploiting Externalities to Estimate the Long-Term Effects of Early Childhood Deworming. *American Economic Journal: Applied Economics*, 10(3):235-262.
- 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.

# References xiii


- Porter, Jack (2003). Estimation in the regression discontinuity model. *Unpublished Manuscript, Department of Economics, University of Wisconsin at Madison*, 5-19.
- Robinson, Peter M. (1988). Root-N-Consistent Semiparametric Regression. *Econometrica*, 56(4):931-954
- Roth, J. (2022). Pretest with caution: Event-study estimates after testing for parallel trends. *American Economic Review: Insights 4*(3), 305-322.
- Shigeoka, Hitoshi. (2014). The Effect of Patient Cost Sharing on Utilization, Health, and Risk Protection. *American Economic Review*, 104(7):2152-2184.

- Shigeoka, Hitoshi. (2015). School Entry Cutoff Date and the Timing of Births. NBER Working Paper No. 21402. Available at: https://www.nber.org/papers/w21402.pdf.
- 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.

- Zhang, Cun-Hui; Zhang, Stephanie S. (2014) Confidence Intervals for Low-Dimensional Parameters in High-Dimensional Linear Models. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 76(1):217-242.
- 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.