RIDGE

RESEARCH INSTITUTE FOR DEVELOPMENT, GROWTH AND ECONOMICS

SINCE 2014

LACEA

Recent Advances in Regression Discontinuity Designs: Bridging Methods and Policy in Latin America

Sebastian Calonico

RIDGE - LACEA Working Paper 008

Inequality

This paper was presented at the 2025 RIDGE May Forum jointly organized with the LACEA Networks

Recent Advances in Regression Discontinuity Designs: Bridging Methods and Policy in Latin America

Sebastian Calonico
UC Davis

2025 Ridge May Forum

Overview

- Today, I'll discuss recent advances in Regression Discontinuity (RD) designs, with a focus on connecting methodological innovations to policy evaluation practice in Latin America.
- I'll start by looking at how RD designs are currently used across the region:
    - What types of policies are being evaluated.
    - How researchers and practitioners implement RD in real-world settings.
    - What challenges commonly arise — including issues with data, design choices, and interpretation.

Overview

- Understanding this context is important because Latin America has become an increasingly active setting for applied causal inference.
- Across areas such as education reforms, social protection programs, and health interventions, researchers and policymakers are applying a range of empirical strategies — including RD, RCTs, DiD, matching — to evaluate the impacts of programs and policies.
- Among these approaches, RD designs are the natural choice when eligibility rules or thresholds create natural opportunities for identification.
- However, as RD applications have grown, so too have the complexities and limitations encountered in practice.

Overview

- In the second part of this talk, I'll turn to recent methodological developments that respond to these challenges — innovations in RD design, estimation, and inference — that aim to make applications more credible, robust, and informative for policymakers.
- My goal is to build a bridge between method and application, showing how these new tools can improve both the credibility and the usability of RD-based evaluations in the Latin American policy context.
- I'll focus on recent methodological developments in the estimation and inference of **Heterogeneous Treatment Effects (HTE)** in RD designs — a topic of growing interest as policymakers and researchers aim to go beyond average effects and understand who benefits most, under what conditions, and by how much.
- These new tools aim to make RD analyses more informative and more policy-relevant.
- Let's begin by briefly reviewing the standard RDD setup.

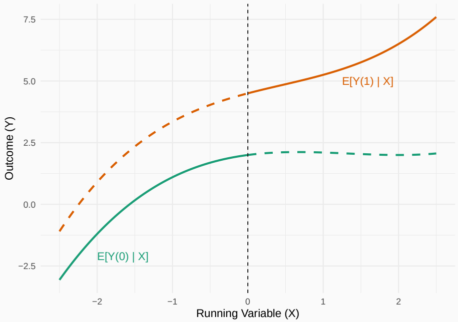
Outline

Regression Discontinuity Design

- Main goal: learn about the causal effect of a policy or intervention
- Units receive a **score**, X_i.
- Treatment is assigned based on the score and a *known* **cutoff** c:
  - given to units whose score is greater than the cutoff.
  - withheld from units whose score is less than the cutoff.
- Under continuity-type assumptions, the abrupt change in the probability of treatment assignment allows us to learn about the effect of the treatment on an **outcome** Y_i.

Regression Discontinuity Design

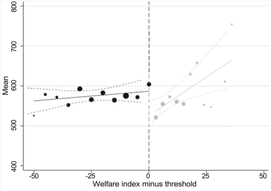
- **Data**: Random Sample (Y_i, T_i, X_i), $i = 1, 2, \ldots, n$, with $Y_i = \begin{cases} Y_i(0) & \text{if } T_i = 0 \\ Y_i(1) & \text{if } T_i = 1 \end{cases}$
- **Treatment**: $T_i \in \{0, 1\}$, $\quad T_i = \mathbf{1}(X_i \geq c) \quad X_i$ continuous



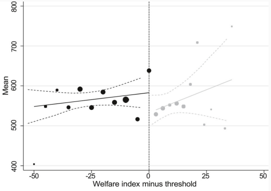
Regression Discontinuity Design

Example: Impact of **Social Health Insurance on Student Performance in Peru**: Carpio, Gomez & Lavado (2025), Health Economics.

- X_i: welfare index based on household characteristics.
- T_i: indicates whether households i is below a specific threshold
- c: welfare index cutoff.
- Y_i measures of student's performance.

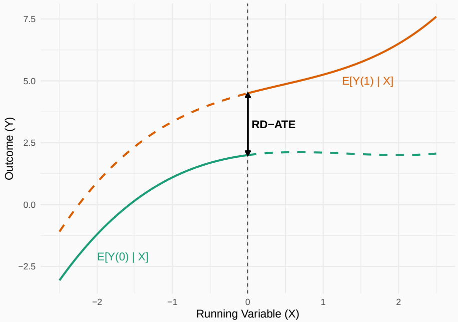


Panel A

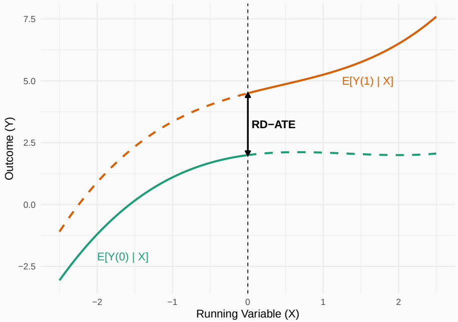


Panel B

Regression Discontinuity Design



Regression Discontinuity Design

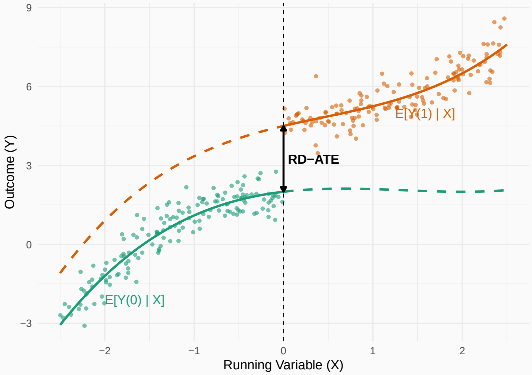


- **Average Treatment Effect at the cutoff**:

$$\tau_{\text{SRD}} = \underbrace{\mathbb{E}[Y_{1i} - Y_{0i}|X_i = x_0]}_{\text{Unobservable}} = \underbrace{\lim_{x \downarrow x_0} \mathbb{E}[Y_i|X_i = x]}_{\text{Observable}} - \underbrace{\lim_{x \uparrow x_0} \mathbb{E}[Y_i|X_i = x]}_{\text{Observable}}$$

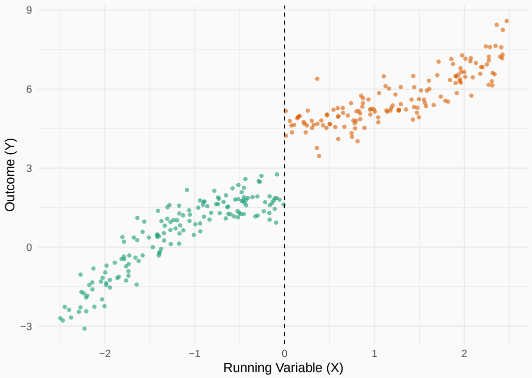
- We need to estimate $\lim_{x \downarrow x_0} \mathbb{E}[Y_i|X_i = x]$ and $\lim_{x \uparrow x_0} \mathbb{E}[Y_i|X_i = x]$.

Regression Discontinuity Design: Estimation



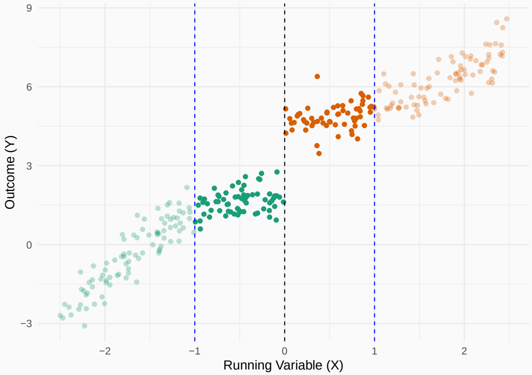
- Fundamental feature: no observations with score X_i exactly equal to the cutoff.
- Thus, extrapolation is unavoidable: we must rely on observations away from the cutoff.
- The unknown regression functions can be approximated by a **polynomial function**.

Regression Discontinuity Design: Estimation



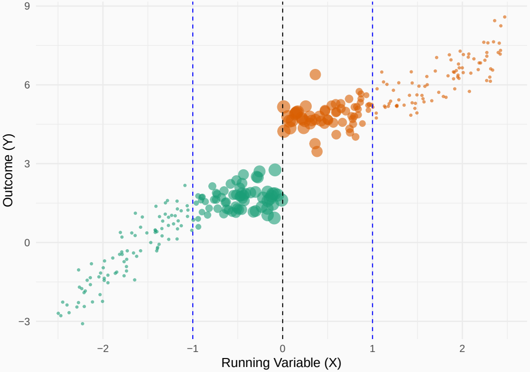
- Note: the approximation occurs at the cutoff, a **boundary point**.
- Global polynomial methods can lead to unreliable point estimators.
- Modern RD empirical work employs local (near the cutoff) polynomial methods.

Estimation: Local Polynomial Regression



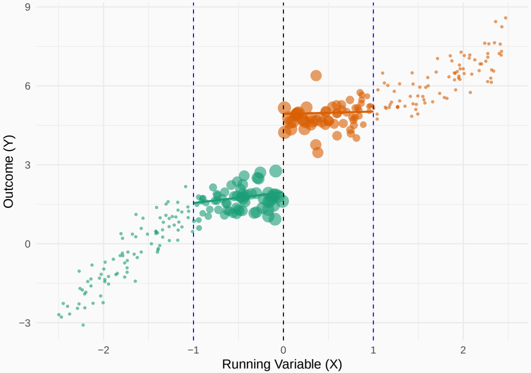
- We only use observations between $c - h$ and $c + h$, where $h > 0$ is a **bandwidth**.
- Use data-driven (optimal) bandwidth selector. h_{MSE}: bias/variance trade-off.
- Statistical properties depend on the accuracy of the approximation.

Estimation: Local Polynomial Regression



- Observations closer to c receive more weight: **kernel** function $K()$.

Estimation: Local Polynomial Regression



- Finally, we fit a polynomial of order p to construct point estimator $\hat{\tau}_n$ (optimal).
- Via joint or separate regressions at each side of the cutoff.

Choice of Bandwidth

- Given p, find h to ensure optimal properties of the point estimator $\hat{\tau_{RD}}$
- MSE-optimal plug-in rule:

$$MSE(\hat{\tau_{RD}}) = Bias^2 + Variance \approx h^{2(p+1)}\mathcal{B}^2 + \frac{1}{nh}\mathcal{V}$$

$$h_{\texttt{MSE}} = C_{\texttt{MSE}}^{1/(2p+3)} \cdot n^{-1/(2p+3)} \qquad C_{\texttt{MSE}} = C(K) \cdot \frac{\mathrm{Var}(\hat{\tau}_{\texttt{SRD}})}{\mathrm{Bias}(\hat{\tau}_{\texttt{SRD}})^2}$$

- **Key idea**: trade-off bias and variance of point estimator $\hat{\tau}$

$$\uparrow \ \mathrm{Bias}(\hat{\tau}) \Longrightarrow \downarrow \hat{h} \qquad \text{and} \qquad \uparrow \ \mathrm{Var}(\hat{\tau}) \Longrightarrow \uparrow \hat{h}$$

Inference

Given above steps, **how do we make inference** about τ?

- Conventional approach: invert Wald test statistic under the null to form CI.

$$T(h) = \frac{\hat{\tau}_\nu(h) - \tau_\nu}{\sqrt{\hat{\mathcal{V}}(\hat{\tau}_\nu(h))}} \sim^a \mathcal{N}(0,1)$$

$$I_{\text{US}}(h) = \left[\hat{\tau}_\nu(h) - 1.96 \times \sqrt{\hat{\mathcal{V}}(\hat{\tau}_\nu(h))} \quad ; \quad \hat{\tau}_\nu(h) + 1.96 \times \sqrt{\hat{\mathcal{V}}(\hat{\tau}_\nu(h))}\right]$$

Inference

Given above steps, **how do we make inference** about τ?

- Conventional approach: invert Wald test statistic under the null to form CI.

$$T(h) = \frac{\hat{\tau}_\nu(h) - \tau_\nu}{\sqrt{\hat{\mathcal{V}}(\hat{\tau}_\nu(h))}} \sim^a \mathcal{N}(0,1)$$

$$I_{\mathtt{US}}(h) = \left[\hat{\tau}_\nu(h) - 1.96 \times \sqrt{\hat{\mathcal{V}}(\hat{\tau}_\nu(h))} \quad ; \quad \hat{\tau}_\nu(h) + 1.96 \times \sqrt{\hat{\mathcal{V}}(\hat{\tau}_\nu(h))}\right]$$

- It will only have correct asymptotic coverage if $nh^{2p+3} \to 0$, bandwidth "small enough".
- MSE-optimal bandwidth is "too large" $\mathbb{P}[\tau_\nu \in I_{\mathtt{US}}(h_{\mathtt{MSE}})] \not\to 1-\alpha$.
- Alternative: **undersmoothing** by selecting a bandwidth "smaller" than $h_{\mathtt{MSE}}$ for inference.

Modern Approaches to Inference

- **Robust Bias Correction** inference (Calonico, Cattaneo, Titiunik, 2014):

$$T_{\mathrm{RBC}}(h) = \frac{\hat{\tau}_{\nu,\mathrm{BC}}(h) - \tau_\nu}{\sqrt{\hat{\mathcal{V}}_{\mathrm{BC}}(\hat{\tau}_{\nu,\mathrm{BC}}(h))}}, \qquad \hat{\tau}_{\nu,\mathrm{BC}}(h) = \hat{\tau}_\nu(h) - h^{1+p-\nu}\hat{\mathcal{B}}(h),$$

Modern Approaches to Inference

- **Robust Bias Correction** inference (Calonico, Cattaneo, Titiunik, 2014):

$$T_{\mathrm{RBC}}(h) = \frac{\hat{\tau}_{\nu,\mathrm{BC}}(h) - \tau_\nu}{\sqrt{\hat{\mathcal{V}}_{\mathrm{BC}}(\hat{\tau}_{\nu,\mathrm{BC}}(h))}}, \qquad \hat{\tau}_{\nu,\mathrm{BC}}(h) = \hat{\tau}_\nu(h) - h^{1+p-\nu}\hat{\mathcal{B}}(h),$$

  - Key feature: $\hat{\mathcal{V}}_{\mathrm{BC}}(h)$ is an estimator of the variance of $\hat{\tau}_{\nu,\mathrm{BC}}(h)$, not $\hat{\tau}_\nu(h)$.
  - $I_{\mathrm{RBC}}(h)$ remains valid under a wider set of bandwidth sequences $\mathbb{P}[\tau_\nu \in I_{\mathrm{RBC}}(h_{\mathrm{MSE}})] \to 1-\alpha$
  - Calonico, Cattaneo & Farrell (2020): coverage error of $I_{\mathrm{RBC}}(h)$ vanishes faster than $I_{\mathrm{US}}(h)$.

Several Extensions

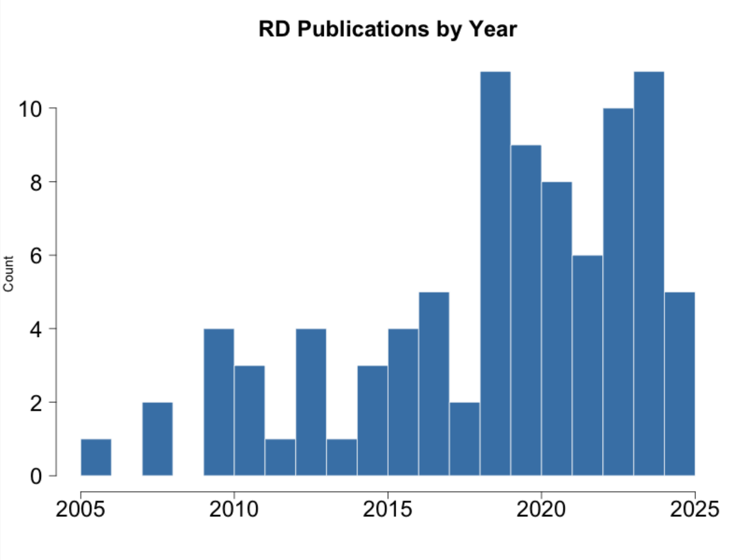
- Imperfect compliance: Fuzzy RD.
- Derivatives of the regression functions: Kink RD.
- Multi-cutoff RD.
- Multi-score RD / Geographic RD.
- Dynamic RD / Differences-in-Discontinuities.
- RD with Covariates.

Outline

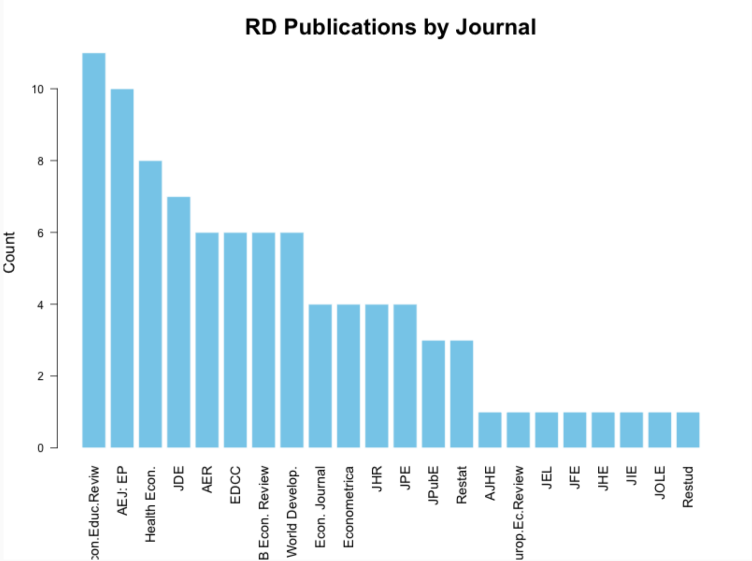
RD Policy Analysis in Latin America

- Goal: understand how RD designs are currently used across the region:
  - What types of policies are being evaluated.
  - How researchers and practitioners implement RD in real-world settings.
  - What challenges commonly arise — including issues with data, design choices, and interpretation.
- Review over the last 20 years of empirical papers using RDD as their main tool to evaluate policies in Latin America.
- Almost 30 journals, including top 5s and field journals.
- Found 90 articles.

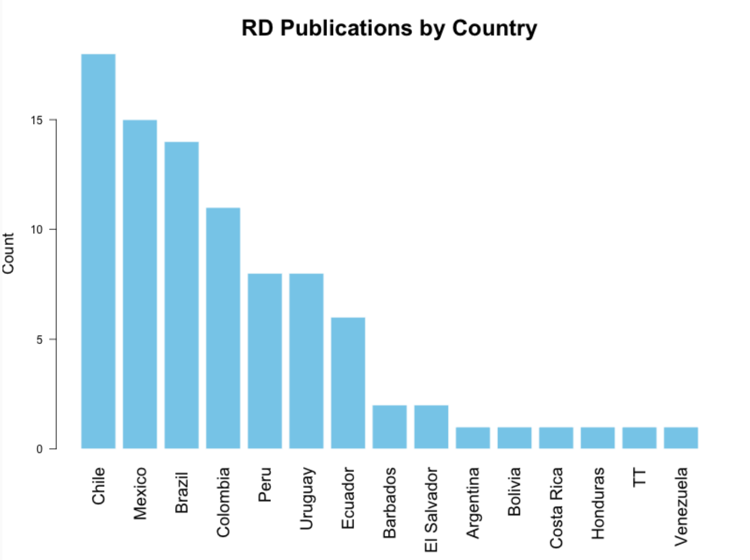
RD Policy Analysis in Latin America



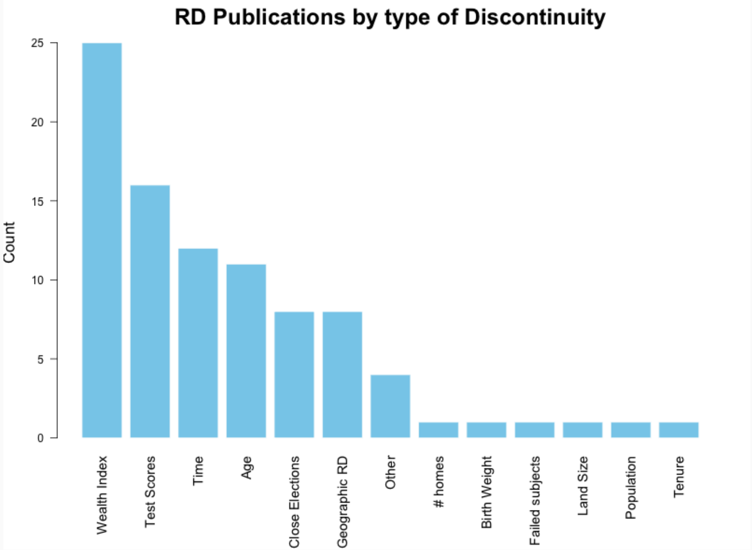
RD Policy Analysis in Latin America



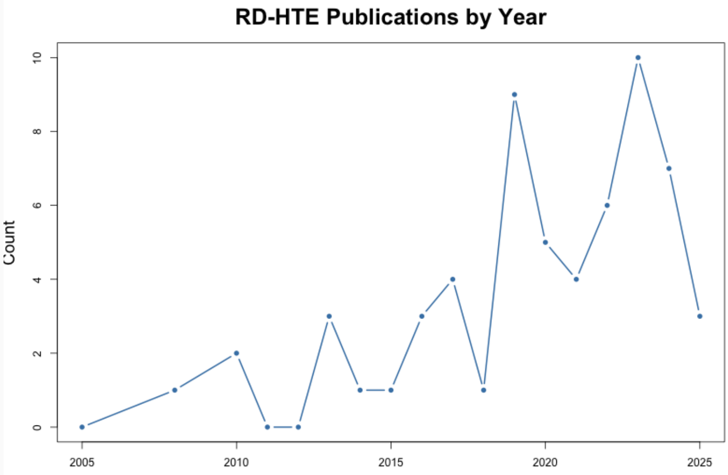
RD Policy Analysis in Latin America



RD Policy Analysis in Latin America



RD Policy Analysis in Latin America



RD Policy Analysis in Latin America

Overview of empirical practice from recent RD publications.

- Most recent papers employ local polynomial methods, with data-driven (optimally) selected bandwidths.
- Need for extensions compared to the standard model: adjust for covariates, time trends, fixed effects, clustered standard errors, etc.
- This has implications for inference that are not typically accounted for.
- Growing interest in HTE analyses: mostly by subgroups, including **discretize a continuous variable**, and estimate **separate** or **interacted** RDDs.

Outline

Covariates in RDD

- Standard RD model fits the outcome on the score alone.
- Sometimes we may want to augment it by including other covariates. Many possible goals:

1. **Efficiency**: to capture finite sample variability as in RCTs.
2. Alternative: to increase the **plausibility of the design**.
   - In general, covariate adjustment cannot be used to restore identification of standard RDD when treated and control observations differ systematically at the cutoff.
   - Continuity assumption of the potential outcomes is implausible.
   - Non-parametric framework no longer appropriate without further assumptions about dgp.
3. A final use of covariates is to study **heterogeneous effects**.

Overview

- Heterogeneous treatment effects (HTE) is a key part of modern causal inference.
- Large methodological literature for estimation and inference on ATE, but comparatively little work has been done analyzing heterogeneity.
- This is particularly true for Regression Discontinuity Designs (RDD).
- RDD has become a widely used tool for causal inference in economics and social science, but most work focuses on ATE.
- This has lead to a proliferation of ad-hoc HTE analyses in applications.
- Goal: to provide practical and theoretically-grounded framework for HTE in RDD, yielding a unified methodology for empirical research.

Heterogeneous RD Effects via Covariates

The use of covariates in RDD to study **heterogeneous effects** can take several forms:

1. Multi-cutoff/Multivalued Treatments
2. Dynamic RD
3. Extrapolation based on CIA
4. Subgroup analysis

Heterogeneous RD Effects via Covariates

1. Multi-cutoff RD: Cattaneo, Keele, Titiunik, Vazquez-Bare (2016), Bertanha (2020).
   - Subgroups defined by the cutoff units are exposed to.
   - HTE: performing RD-ATE separately for each cutoff subgroup.
   - ATE at a particular cutoff for a subpopulation of units exposed to a different one.
2. Dynamic RD:
   - Units are ranked and assigned to treatment repeatedly over time.
     - School districts referenda on bond issues, Cellini, Ferreira, & Rothstein (2010).
   - Hsu & Shen (2024) recently provided a formal framework.
3. Extrapolation of TE based on CIA: Angrist, Rokkanen (2015), Keele et al (2015)
   - Effects far from the cutoff can be recovered with *auxiliary pre-determined covariates*.

Heterogeneous RD Effects via Covariates

Subgroup analysis

- Groups typically defined in terms of observed values of pre-determined covariates.
- The analysis changes the parameter of interest.

$$\tau_{Y|Z}(z) = E[Y_i(1) - Y_i(0)|X_i = c; Z_i = z]$$

- Typical approach: conduct RD analysis for each subgroup defined by $Z_i = z$, **separately**.
    - Example: TE patients for different age groups, e.g., 45-64 versus 65+.
- Very popular in practice, but no formal guidance on how to do it.

Heterogeneous RD Effects via Covariates

Subgroup analysis

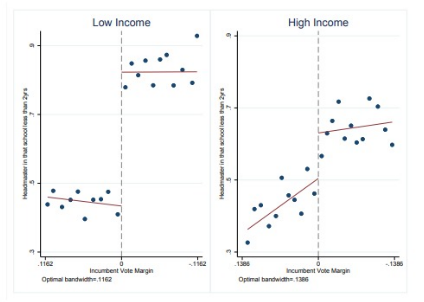
- Groups typically defined in terms of observed values of pre-determined covariates.
- The analysis changes the parameter of interest.

$$\tau_{Y|Z}(z) = E[Y_i(1) - Y_i(0)|X_i = c; Z_i = z]$$

- Typical approach: conduct RD analysis for each subgroup defined by $Z_i = z$, **separately**.
    - Example: TE patients for different age groups, e.g., 45-64 versus 65+.
- Very popular in practice, but no formal guidance on how to do it.
- Hsu & Shen (2019, 2021) propose formal tests for RD treatment effect heterogeneity among individuals with different observed pre-treatment characteristics.
- Reguly (2023), Alcantara et al (2024): tree- and forest- based methods for recovering RD heterogeneity, without specifying subgroups a priori.

Heterogeneous RD Effects in Applied Work

Akhtari, Moreira, Trucco (2022): political turnover in Brazil and public service provision.

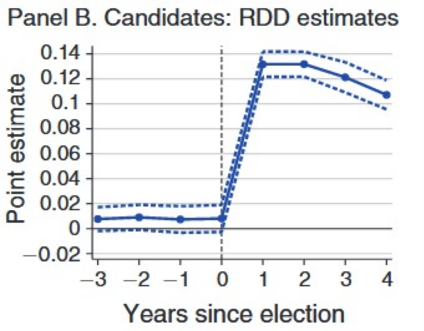


$$Y_{jmt+1} = \left(\alpha + \beta \mathbf{1}\{IVM_{mt} < 0\} + \delta \mathbf{1}\{IVM_{mt} < 0\} \times IVM_{mt} + \gamma IVM_{mt}\right)$$

$$\otimes\ Charact_{jm} + X'_{jmt}\Lambda + \epsilon_{jmt}.$$

Heterogeneous RD Effects in Applied Work

Colonnelli, Prem, and Teso (2020): effect over time of supporting the winning party on probability of employment in the public sector.



$$(2) \qquad y_{ikpmt} = \sum_{s=-3}^{+4} \beta_s Mayor_{pmt} \mathbf{1}(s = k) + \theta_k MV_{pmt} + \gamma_{kmt} + \epsilon_{ikpmt}.$$

Heterogeneous RD Effects in Applied Work

Overview of empirical practice from recent AEA publications.

Paper	Outcome	Treatment	Running Variable	Heterogeneity			Estimation					Inference	
				Disc.	Time	Cont.	BW	HTE BW	Local	Joint	Covs/ FE	Robust	Cluster SE
Adams et al (2022)	Healthcare utilization	Financial Assistance	Poverty level		Yes		None						
Akhtari et al (2022)	Municipal Bureaucracy	Political turnover	Close Elections	Yes			MSE	Yes	Yes		Yes		Yes
Asher, Novosad (2020)	Economic development	Road construction	Population size	Yes			MSE		Yes		Yes		Yes
Brollo et al (2013)	Political corruption, quality	Government revenues	Population size	Yes		Yes	None			Yes	Yes		Yes
Dell (2015)	Drug-related violence	Drug enforcement	Close Elections	Yes	Yes		Manual		Yes	Yes	Yes		Yes
Miralles, Leganza (2024)	Savings, Retirement behavior	Public pension benefits	Age	Yes	Yes		MSE		Yes		Yes	Yes	
Han et al (2020)	Healthcare Utilization	Patient Cost-Sharing	Age	Yes			Manual		Yes				Yes
Huh, reif (2021)	Mortality, risky behaviors	Teenage Driving	Age	Yes			MSE		Yes			Yes	
Jones et al (2022)	Irrigation Adoption, Profits	Access to water	Spatial Disc.	Yes			Manual		Yes		Yes		Yes
Lindo et al (2010)	Academic Performance	Academic probation	Test Scores	Yes			Manual		Yes		Yes		Yes
McEwan et al (2021)	Economics Major Choice	Higher letter grade	Test Scores	Yes			MSE		Yes	Yes	Yes	Yes	
Miglino et al (2023)	Health outcomes	Financial Assistance	Age	Yes	Yes		Manual		Yes		Yes		Yes
Eleches, Urquiola (2013)	Academic Performance	Access to better schools	Test Scores	Yes	Yes		Manual		Yes		Yes		Yes
Shigeoka (2014)	Utilization, health	Patient Cost Sharing	Age	Yes	Yes		Manual		Yes		Yes		Yes
Silliman virtanen (2021)	Labor market returns	Vocational Education	Test Scores	Yes	Yes		Manual		Yes		Yes		Yes
Zimmerman (2019)	Better jobs, Income	Access to better schools	Test Scores	Yes			Manual		Yes	Yes	Yes		Yes

- Most HTE analyses **discretize a continuous variable**, and estimate **separate** RDDs.
- Same bandwidth used for all analyses.
- Most inference is incorrect, unless willing to assume correctly specified model.

Outline

Introduction

- Analyzing Heterogeneous Treatment Effects (HTE) is crucial in modern causal inference.
- While the ATE provides a broad measure of a treatment's effectiveness and informs general policy decisions, uncovering heterogeneity patterns is essential for evaluating substantive hypotheses and specific policy interventions.
- HTE can help understand fairness and differential impacts across subpopulations, develop targeted interventions, and optimize social policies.

Introduction

- RD design has become a prominent tool for causal inference, leading to extensive methodological advancements for identification, estimation, and inference for Average Treatment Effects (ATE) in various contexts.
- However, despite its widespread adoption, there is a notable lack of rigorous methods for studying heterogeneity, resulting in many ad-hoc empirical approaches.
- We address this gap by developing practical, theoretically-grounded tools for heterogeneity analysis in RD designs.

https://rdpackages.github.io/rdhte/

Treatment Effect Heterogeneity in Regression Discontinuity Designs*

Sebastian Calonico† Matias D. Cattaneo‡ Max H. Farrell§ Filippo Palomba¶
Rocío Titiunik‖

March 26, 2025

Introduction

- Extending RD estimation and inference to encompass HTE presents a high-dimensional, nonparametric challenge: theoretically feasible, but often impractical without restrictions.
- Consequently, empirical researchers typically use semiparametric models, and thus explore heterogeneity in limited dimensions.
- Following empirical practice, we study the most common approach for RD-HTE analysis: **local least squares regression with interactions** with pretreatment covariates.
- This model balances practical applicability with flexible (nonparametric) nature of RDD.

Outline

Setup

- We consider a RD design with covariates **interacted** with treatment.
  - Popular approach in practice, but no formal results have been developed.
- **Contribution 1**: we show when this model correctly recovers useful causal objects.
  - When potential outcomes obey a linear functional coefficient model in covariates.
  - This structure holds without loss of generality for discrete variables.
  - Without such structure, linear interactions do not have a useful probability limit.
- **Contribution 2**: formalize how to conduct valid inference for HTE.
  - MSE expansion, central limit theorem, and standard error consistency.
  - We develop robust inference and optimal bandwidth selection methods.
  - Optimal bandwidth depends on specific inference target: subgroups, group differences, etc.

Setup

- Standard **Sharp RD** design based on a continuous running variable.
- Binary treatment $T_i \in \{0, 1\}$, assigned according to $T_i = \mathbb{1}\{X_i \geq c\}$.
- In addition to (Y_i, T_i, X_i), we observe a vector $\mathbf{W}_i \in \mathbb{R}^d$ of covariates.
- Matching empirical practice, we assume $\mathbf{W}_i$ are pre-treatment covariates.
- Usual parameter of interest in Sharp RD: ATE at the cutoff.

$$\tau = \mathbb{E}[Y_i(1) - Y_i(0) \mid X_i = 0]. \tag{1}$$

- The point estimate $\dot{\tau}$ is the coefficient on T_i in a (local, weighted) least squares regression:

$$\dot{Y}_i = \dot{\alpha} + \dot{\tau} T_i + \dot{\beta}_- X_i + \dot{\beta}_+ T_i X_i \tag{2}$$

Setup

- Calonico et al, 2019 added covariates to (2), but aiming for **efficiency**, not heterogeneity.
- Pre-treatment covariates $\mathbf{Z}_i$ enter linearly with "common" effect:

$$\widetilde{Y}_i = \widetilde{\alpha} + \widetilde{\tau} T_i + \widetilde{\beta}_- X_i + \widetilde{\beta}_+ T_i X_i + \widetilde{\gamma}' \mathbf{Z}_i. \tag{3}$$

- Retain the *same* RD-ATE under "no RD effect on covariates" $\tilde{\tau} \to_{\mathbb{P}} \tau - [\mu_{Z+} - \mu_{Z-}]' \gamma_Y$
- Local linear projection, not assumption about functional form of regression functions.
- The exclusion of interactions with covariates is crucial for correct estimation of τ, but by construction does not allow for learning heterogeneity.

Setup

- One might try to recover the (local) CATE function fully flexibly in $\mathbf{W}_i$:

$$\tau(\mathbf{w}) = \mathbb{E}[Y_i(1) - Y_i(0) \mid X_i = 0, \mathbf{W}_i = \mathbf{w}]. \tag{4}$$

- $\tau(\mathbf{w})$: difference between two $(d+1)$ dimensional nonparametric regressions evaluated at $(0, \mathbf{w})$.
- Challenging problem: curse of dimensionality, many tuning parameters.
- We aim for something more tractable.
- Thus, we focus on specifications with the same RD practical appeal: *one* dimensional nonparametric regression at a single point.
- Central question: to what extent, and under what assumptions, we can still obtain estimation and inference for heterogeneous causal effects.

Setup

- We study a local weighted least squares regression with full interactions.

$$\widehat{Y}_i = \widehat{\alpha} + \widehat{\tau} T_i + \widehat{\beta}_- X_i + \widehat{\beta}_+ T_i X_i + \left(\widehat{\boldsymbol{\delta}}'_- + \widehat{\boldsymbol{\delta}}'_+ T_i + \widehat{\boldsymbol{\rho}}'_- X_i + \widehat{\boldsymbol{\rho}}'_+ T_i X_i \right) \mathbf{W}_i, \tag{5}$$

- Interpretation of coefficients changed.
- Without further assumptions on the dgp, the probability limit of the RD estimator has a generic best linear MSE predictor interpretation.
- We need to leverage extra structure to attach a causal interpretation.

Outline

Main Results

- We derive theoretical properties of the interacted regression fit.
- We first characterize exactly when the true CATE function is recovered.
- Next, we provide feasible inference with optimal tuning parameter selection.

Main Results

- Besides standard RD regularity assumptions (modified to accommodate $\mathbf{W}$), we impose two main conditions to obtain a causal interpretation of the estimands:

Assumption: Pre-determined covariates

At least in a neighborhood of the cutoff, $\mathbf{W}_i$ is a *pre-treatment* or *exogenous* covariate,

$$\{\mathbf{W}_i(1), \mathbf{W}_i(0)\} \perp\!\!\!\perp X_i$$

Assumption: CATE Structure

$$E[Y(t) \mid X = x, \mathbf{W} = \mathbf{w}] = \mu_t(x) + \boldsymbol{\lambda}_t(x)'\mathbf{w} \quad \text{for} \quad t \in \{0, 1\}$$

where $\mu_t(x)$ and $\lambda_{t,k}(x)$, $k = 1, \ldots, d$ are thrice continuously differentiable for all $x \in [x_l, x_u]$, $x_l < c < x_i$.

Main Results

- Potential outcomes obey a linear-in-**W** structure (i.e. a local partially linear functional coefficient model).
- Assumption is placed on the potential outcomes, to match the two-sided nature of RDD.
- It does imply that the CATE function (defined at the cutoff) is:

$$\tau(\mathbf{w}) = \mathbb{E}[Y_i(1) - Y_i(0)|X_i = 0, \mathbf{W}_i = \mathbf{w}] = \tau(\mathbf{0}) + \boldsymbol{\delta}_\star'\mathbf{w}$$

  with $\tau(\mathbf{0}) = \mu_1(c) - \mu_0(c)$ and $\boldsymbol{\delta}_\star = \boldsymbol{\lambda}_1(c) - \boldsymbol{\lambda}_0(c)$.
- The estimation of CATE function thus remains a one dimensional nonparametric problem.
- Bias can be removed up to the same order as learning the RD-ATE.
- Important for maintaining the empirical tractability of RD analyses.

Main Results

- **Good News**: it holds without loss of generality for discrete covariates.
- Under these assumptions, $\widehat{\tau}(\mathbf{w}) \rightarrow_p \tau(\mathbf{w})$.
- Thus, the coefficients $(\widehat{\tau}, \widehat{\boldsymbol{\delta}}'_{+})$, correctly capture the heterogeneous causal effects.
- Without it, the probability limit of the coefficients in (6) do not appear to have a causal interpretation.

Main Results

$$\widehat{Y}_i = \widehat{\alpha} + \widehat{\tau} T_i + \widehat{\beta}_- X_i + \widehat{\beta}_+ T_i X_i + \left(\widehat{\boldsymbol{\delta}}'_- + \widehat{\boldsymbol{\delta}}'_+ T_i + \widehat{\boldsymbol{\rho}}'_- X_i + \widehat{\boldsymbol{\rho}}'_+ T_i X_i \right) \mathbf{W}_i, \tag{6}$$

- For discrete covariates, the corresponding estimate of the CATE function $\tau(\mathbf{w})$ given by

$$\widehat{\tau}(\mathbf{w}) = \widehat{\tau} + \widehat{\boldsymbol{\delta}}'_+ \mathbf{w}. \tag{7}$$

- The coefficient on T_i no longer recovers the ATE, but rather $\widehat{\tau} = \widehat{\tau}(\mathbf{0})$, the CATE for the baseline group $\mathbf{w} = \mathbf{0}$.
- The coefficients $\widehat{\boldsymbol{\delta}}_+$ give the difference from the baseline group:
  - as discrete change for binary covariates, or
  - as slope of $\tau(\mathbf{w})$ with respect to continuous $\mathbf{w}$ (under a functional coefficient assumption).

Mean Squared Error & Bandwidth Selection

- Estimation and inference relies crucially on selecting a bandwidth h that appropriately localizes to the cutoff.
- We develop MSE-optimal bandwidth selection.
- This yields the best possible point estimates of heterogeneity.
- However, such a bandwidth yields invalid inference in general, and thus we also develop RBC inference methods.

Mean Squared Error & Bandwidth Selection

- To characterize the optimal bandwidth we provide a MSE expansion of the coefficients $\widehat{\boldsymbol{\theta}} = (\widehat{\tau}, \widehat{\boldsymbol{\delta}}_+')'$ around their population counterparts.
- Main difference: we now have several parameters of interest.
- Linear combinations of these yield the CATE function for different groups.
- Let $\mathbf{s}$ be the vector that selects the appropriate elements.
  - Binary Z_i: specific subgroup can be studied as $\widehat{\tau}(0) = \mathbf{s}'\widehat{\boldsymbol{\theta}}$ with $\mathbf{s} = (1, 0)'$
  - Testing group differences: $\widehat{\tau}(1) - \widehat{\tau}(0)$, which uses $\mathbf{s} = (-1, 1)'$

Mean Squared Error & Bandwidth Selection

- MSE expansion:

$$\mathrm{MSE}\left[\mathbf{s}'\widehat{\boldsymbol{\theta}} - \mathbf{s}'\boldsymbol{\theta}_{\star}\right] = \frac{1}{nh}\mathrm{V}_{\mathbf{s}} + h^4\mathrm{B}_{\mathbf{s}} + o_p\left((nh)^{-1} + h^4\right).$$

- $\mathrm{V}_{\mathbf{s}}$, $\mathrm{B}_{\mathbf{s}}$ are the constant portions of the (asymptotic) variance and bias of $\mathbf{s}'\widehat{\boldsymbol{\theta}}$.
- Balancing the two terms yields the optimal bandwidth

$$h_{\mathbf{s}}^{\star} = \left(\frac{\mathrm{V}_{\mathbf{s}}}{\mathrm{B}_{\mathbf{s}}^2}\right) n^{-1/5} \tag{8}$$

Mean Squared Error & Bandwidth Selection

- $h_{\mathbf{s}}^{\star}$ depends on the specific inference target: each estimand has its own bandwidth.
- This is not necessarily practical. However, note:
  - The rate is not affected: this is a one-dimensional nonparametric problem (in X_i).
  - The constants are often less important empirically.
- Thus, we expect that results should not be too sensitive to the specific choice, allowing some flexibility in bandwidth selection.
- A common bandwidth could be the same used for obtaining the RD-ATE.

Inference

- Asymptotic distribution for inference on RD-HTE:

$$\sqrt{nh}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_\star - h^2\mathrm{B}\right) \to_d \mathcal{N}(0, \mathrm{V}),$$

- h-MSE does not remove sufficient bias, rendering inference invalid.
- We propose RBC inference to deal with this problem. Two key ingredients:
  1. an estimate of $h^2\mathrm{B}$ is subtracted from the point estimate $\widehat{\boldsymbol{\theta}}$.
  2. standard errors adjusted to account for additional estimation error of the bias correction.
- We thus base inference on

$$\widehat{\boldsymbol{\theta}}_{\mathrm{rbc}} = \widehat{\boldsymbol{\theta}} - h^2\widehat{\mathrm{B}}, \qquad \widehat{V} = \widehat{\mathrm{var}}(\widehat{\boldsymbol{\theta}}_{\mathrm{rbc}}). \tag{9}$$

Setup

- Then, for any specific estimand defined by $\mathbf{s}$ and $h = h^{\star}_{\mathbf{s}}$:

$$\sqrt{nh}\widehat{V}^{-1/2}\left(\mathbf{s}'\widehat{\boldsymbol{\theta}}_{\mathrm{rbc}} - \mathbf{s}'\boldsymbol{\theta}_{\star}\right) \to_d \mathcal{N}(0,1).$$

- This result can be used to conduct inference on any estimand we discussed.
- For example, an asymptotic 95% confidence interval for $\tau(\mathbf{w})$ is given by

$$\widehat{\tau}_{\mathrm{rbc}} \pm 1.96\sqrt{\mathbf{s}'\widehat{V}\mathbf{s}/(nh)}.$$

- Hypothesis tests can be constructed similarly. For example, $\tau(1) = \tau(0)$ test of no heterogeneity for a binary covariate.

Outline

Empirical Application

- To illustrate our methodological results, we revisit the analysis of Akhtari, Moreira, Trucco (2022) (AMT).
- Role of political turnover on public service provision by local governments in Brazil.
- Approach: compare outcomes in municipalities where the incumbent party barely loses (political turnover) to those where they barely win (no political turnover).
- Close election analysis, a canonical RD setting.
- Main findings: municipalities with a new party in office experience increases in bureaucracy, with new personnel appointed across multiple sectors and levels.
- Also, increases in the replacement rate of personnel in schools controlled by the municipal government, accompanied by lower test scores.
- We focus on heterogeneous effects on headmaster replacement (i.e., whether a headmaster is new to the school) by municipality income.

Empirical Application

The analysis of AMT highlights two key methodological difficulties researchers have faced when studying heterogeneous treatment effects in RD.

1. They use an MSE-optimal bandwidth selection method to localize at the cutoff.
    - However, both estimation and inference rely solely on local linear fitting.
    - Invalid inference: standard errors from the linear fit are invalid when paired with the optimal bandwidth.
2. The heterogeneity variable (income) is continuous.
    - AMT compare municipalities above and below the median income level.

Empirical Application

TABLE A.21: POLITICAL TURNOVER AND HEADMASTER REPLACEMENT IN LOW- AND HIGH-INCOME MUNICIPALITIES

Outcome:	Headmaster is new to the school (as Headmaster)					
Panel A	Low Income Municipalities (Below Median Income)					
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathbb{1}\{IncumbVoteMargin < 0\}$	0.389	0.389	0.371	0.370	0.379	0.379
	(0.038)	(0.037)	(0.047)	(0.045)	(0.039)	(0.038)
Observations	6,703	6,703	4,294	4,294	6,447	6,447
R-squared	0.151	0.154	0.160	0.167	0.156	0.159
Controls	No	Yes	No	Yes	No	Yes
Clusters	1073	1073	754	754	1030	1030
Mean Dep. Variable	0.447	0.447	0.447	0.447	0.445	0.445
Using Bandwidth	0.116	0.116	0.0700	0.0700	0.110	0.110
Optimal Bandwidth	0.116	0.116	0.116	0.116	0.116	0.116
Panel B	High Income Municipalities (Above Median Income)					
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathbb{1}\{IncumbVoteMargin < 0\}$	0.126	0.125	0.136	0.138	0.107	0.112
	(0.044)	(0.043)	(0.065)	(0.064)	(0.049)	(0.049)
Observations	5,809	5,809	3,114	3,114	4,560	4,560
R-squared	0.050	0.051	0.030	0.032	0.045	0.046
Controls	No	Yes	No	Yes	No	Yes
Clusters	1220	1220	764	764	1048	1048
Mean Dep. Variable	0.430	0.430	0.467	0.467	0.448	0.448
Using Bandwidth	0.139	0.139	0.0700	0.0700	0.110	0.110
Optimal Bandwidth	0.139	0.139	0.139	0.139	0.139	0.139

Empirical Application

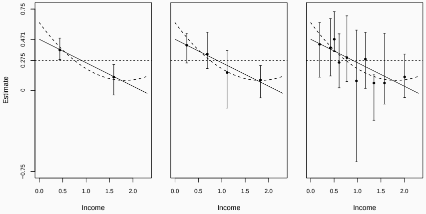
- We first replicate their findings, adding RBC measures of uncertainty.
- Our point estimates are nearly identical (0.379 and 0.120 versus 0.389 and 0.126).
- Next, we incorporate valid inference, revealing that the treatment effect is statistically significant only for below-median income municipalities.
- We formally test that TE are the same across income groups, finding strong evidence against it.
- We also extend this discretized analysis by studying treatment effects by income quartiles and deciles, rather than a binary median split.
- Finally, we analyze income directly, as a continuous variable.

Empirical Application

```
. rdhte Y X,  covs_hte(i.W_qrt)

RD Heterogeneous Treatment Effects Estimation.

Cutoff c          =          0
Number of obs     =     206337
Order est. (p)    =          1
Kernel            =  Triangular
VCE method        =        hc3

Outcome: Y. Running variable: X.
```

Group	Estimate	[95% Conf.	Interval]	P>|z|	Size	Bandwidth
1.W_qrt	.41487	0.317	0.451	0.000	27414	0.142
2.W_qrt	.34048	0.318	0.450	0.000	18074	0.156
3.W_qrt	.08712	-0.096	0.126	0.791	5372	0.098
4.W_qrt	.10157	0.029	0.188	0.008	9484	0.153

Application



- The continuous specifications highlight differences in income distribution within each group: top quartile has much wider range—an important finding for policy implications.
- These results further strengthen the original findings, showing that variation in treatment effects is more pronounced across the median than within the upper or lower groups.
- Ultimately, the decision of how to best capture heterogeneity is up to the researcher.
- The continuous approach, particularly when compared to a fine discretization, may yield substantial precision improvements, but requires an extra assumption.

Outline

Conclusion

- Motivated from empirical practice, we study the statistical properties of the RD estimator when an interaction term between a covariate and the treatment is included.
- We showed how RD-HTE can be recovered using a tractable linear-interactions model.
- Our approach does not require high dimensional nonparametric regression and fits into current RD practice.
- By providing bandwidth selection and valid inference, we hope to bring the same rigor and standardization to heterogeneity in RD effects that currently exists for ATE.
- We have focused exclusively on Sharp RD (including Kink).
- Extending our work to other cases, particularly Fuzzy RD, would be a valuable next step.
- Robust inference, implementation details, replications:

`https://rdpackages.github.io/`