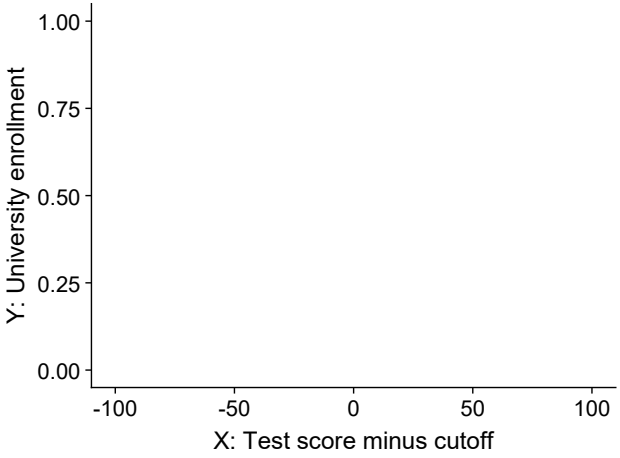
Composition Bias and Covariate Adjustment in Regression Discontinuity Designs

Tobias Großbölting

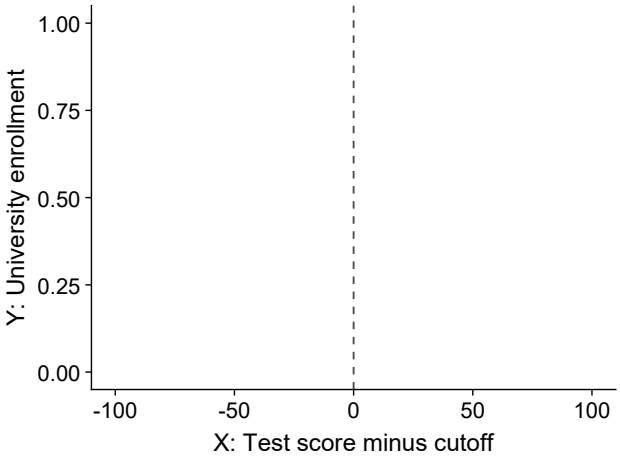
Universität Mannheim

RDD Illustration - Londoño-Vélez et al. (2020)



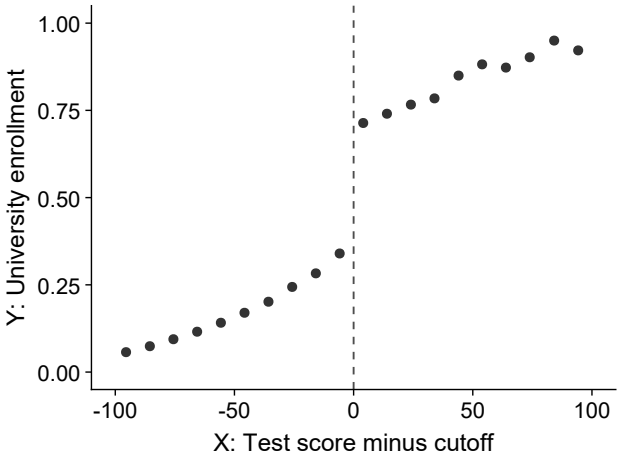
- Outcome variable: $Y_i \in \mathbb{R}$
- Score variable: $X_i \in \mathbb{R}$

RDD Illustration - Londoño-Vélez et al. (2020)



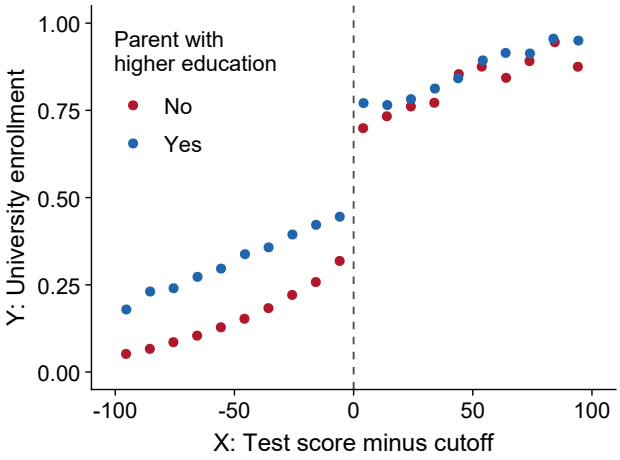
- Outcome variable: $Y_i \in \mathbb{R}$
- Score variable: $X_i \in \mathbb{R}$
- Treatment threshold: $c = 0$
- Binary treatment: $T_i := \mathbb{1}_{\{X_i \geq 0\}}$

RDD Illustration - Londoño-Vélez et al. (2020)



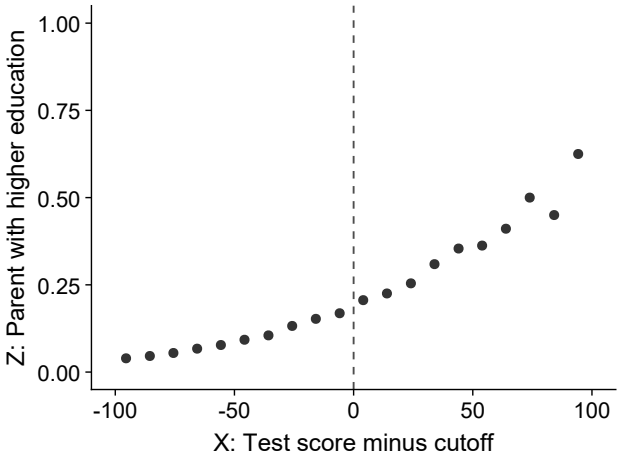
- Outcome variable: $Y_i \in \mathbb{R}$
- Score variable: $X_i \in \mathbb{R}$
- Treatment threshold: $c = 0$
- Binary treatment: $T_i := \mathbb{1}_{\{X_i \geq 0\}}$

RDD Illustration - Londoño-Vélez et al. (2020)



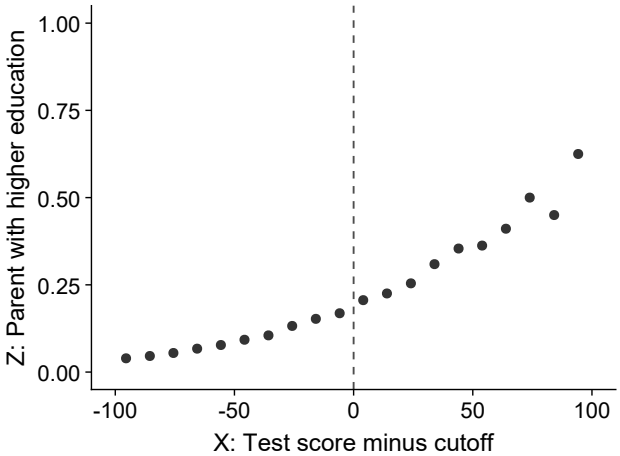
- Outcome variable: $Y_i \in \mathbb{R}$
- Score variable: $X_i \in \mathbb{R}$
- Treatment threshold: $c = 0$
- Binary treatment: $T_i := \mathbb{1}_{\{X_i \geq 0\}}$
- **This talk:** Binary covariate: $Z_i \in \{0, 1\}$

Group Composition - Londoño-Vélez et al. (2020)



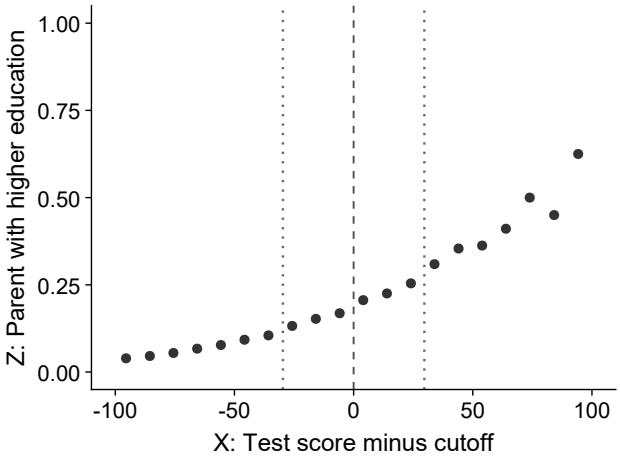
- Share of students with a parent with higher education changes smoothly with the test score

Group Composition - Londoño-Vélez et al. (2020)



- Share of students with a parent with higher education changes smoothly with the test score
- Covariates are balanced at the cutoff

Group Composition - Londoño-Vélez et al. (2020)



- Share of students with a parent with higher education changes smoothly with the test score
- Covariates are balanced at the cutoff
- But estimation uses students away from the cutoff, where covariates are imbalanced

Motivation

- In many empirical RD studies covariates are observed
  - For example education (discrete) or income (continuous)
- Covariates can be predictive of the outcome
- Covariates can be imbalanced within the sample used for estimation
  - Even if their distribution is continuous in the running variable

Research Question

Should a researcher adjust for observed covariates to reduce bias in RDDs?

The Puzzle

Observational Studies

- Adjusting for imbalanced covariates reduces bias (Imbens and Rubin 2015)
- RDDs are observational studies (Gelman and Imbens 2025; Sekhon and Titiunik 2017)

The Puzzle

Observational Studies

- Adjusting for imbalanced covariates reduces bias (Imbens and Rubin 2015)
- RDDs are observational studies (Gelman and Imbens 2025; Sekhon and Titiunik 2017)

Regression Discontinuity Designs

- Theoretical literature on covariate adjustment in RDDs emphasizes variance reduction (Calonico, Cattaneo, Farrell, et al. 2019; Kreiß and Rothe 2023; Noack et al. 2026)
- Guidelines for empirical practice (Cattaneo et al. 2023, p. 164) state:

  "The main and only goal of [covariate adjustment] is to improve efficiency [...]."

This Paper

- Covariate imbalance within the estimation window is common
  - 30 out of 95 tests for covariate balance reject in 9 replication datasets also considered in Noack et al. (2026)

▸ Balance Tests

- The local linear RD (base RD) estimator compares units with different covariate values
- This introduces composition bias if expected outcomes differ across covariate values
- Covariate adjustment reduces composition bias - Stratified RD

Recommendation

- A researcher should adjust for covariates to reduce bias
- Unless she is willing to assume that covariates are
  i) not predictive of the outcome given the running variable, or
  ii) balanced within the sample used for estimation

Covariate Adjustment under Pointwise Asymptotics

- **Need to mention that bandwidths shrink with the sample size!!!!**
- Typically bandwidths are chosen to trade off bias against variance
- Hence base RD has first-order bias (Calonico, Cattaneo, and Titiunik 2014)

Pointwise Asymptotics

- *Treat the unknown regression functions as fixed*
- *Impose differentiability assumptions*

Covariate Adjustment under Pointwise Asymptotics

- **Need to mention that bandwidths shrink with the sample size!!!!**
- Typically bandwidths are chosen to trade off bias against variance
- Hence base RD has first-order bias (Calonico, Cattaneo, and Titiunik 2014)

Pointwise Asymptotics

- *Treat the unknown regression functions as fixed*
- *Impose differentiability assumptions*

- Variance reduction through covariate adjustment (Calonico, Cattaneo, Farrell, et al. 2019; Kreiß and Rothe 2023; Ma and Yu 2024; Noack et al. 2026)
- Covariate adjustment leaves first-order bias of local linear RD estimator unchanged

Covariate Adjustment under Uniform Asymptotics

- In finite samples composition bias can be substantial

Covariate Adjustment under Uniform Asymptotics

- In finite samples composition bias can be substantial

Uniform Asymptotics

- *Impose differentiability assumptions and explicit bounds on derivatives*
- *Evaluate worst-case bias over functions satisfying the assumptions*
  (Armstrong and Kolesár 2018, 2020)

Covariate Adjustment under Uniform Asymptotics

- In finite samples composition bias can be substantial

Uniform Asymptotics

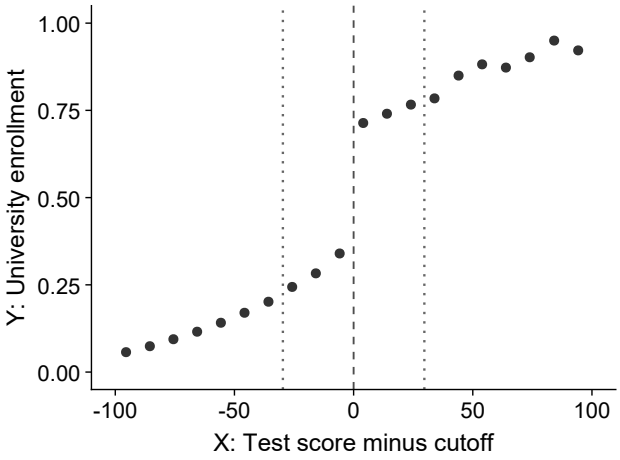
- *Impose differentiability assumptions and explicit bounds on derivatives*
- *Evaluate worst-case bias over functions satisfying the assumptions*
  (Armstrong and Kolesár 2018, 2020)

This Paper:

- The first-order worst-case bias of base RD includes a composition bias term
- Covariate adjustment can reduce first-order worst-case bias by removing composition bias
- Compared to base RD, Stratified RD has weakly lower
  - first-order worst-case bias and
  - asymptotic variance

Base RD & Composition Bias

Illustration - Base RD Estimator



Pooled Regression Functions:

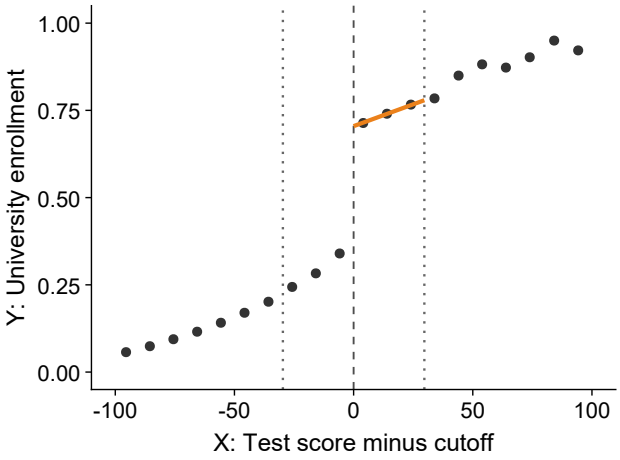
$$\bar{m}(x) := E[Y_i(0)|X_i = x]$$
$$\bar{\tau}(x) := E[Y_i(1) - Y_i(0)|X_i = x]$$

Target Parameter:

$$\tau := \bar{\tau}(0)$$

- Local linear regression on each side of the cutoff within bandwidth h

Illustration - Base RD Estimator



Pooled Regression Functions:

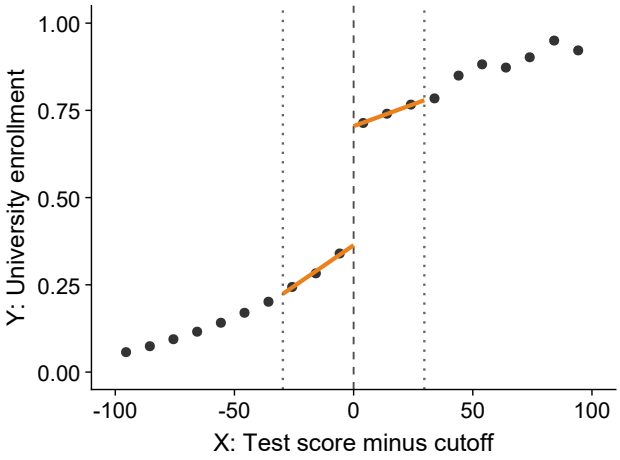
$$\bar{m}(x) := E[Y_i(0)|X_i = x]$$
$$\bar{\tau}(x) := E[Y_i(1) - Y_i(0)|X_i = x]$$

Target Parameter:

$$\tau := \bar{\tau}(0)$$

- Local linear regression on each side of the cutoff within bandwidth h
- $\hat{\tau}(h)$ - jump at c Base RD Definition

Illustration - Base RD Estimator



Pooled Regression Functions:

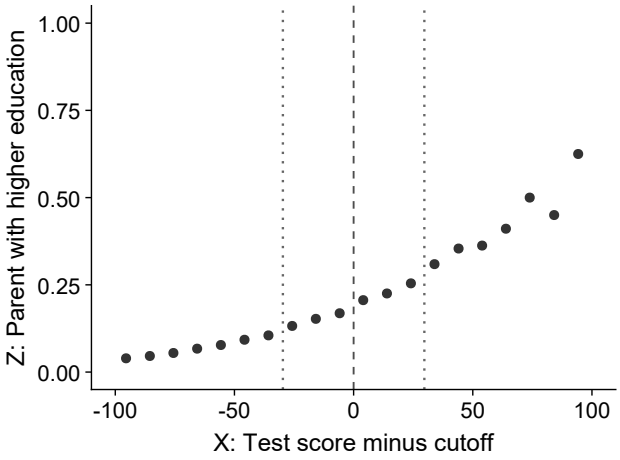
$$\bar{m}(x) := E[Y_i(0)|X_i = x]$$
$$\bar{\tau}(x) := E[Y_i(1) - Y_i(0)|X_i = x]$$

Target Parameter:

$$\tau := \bar{\tau}(0)$$

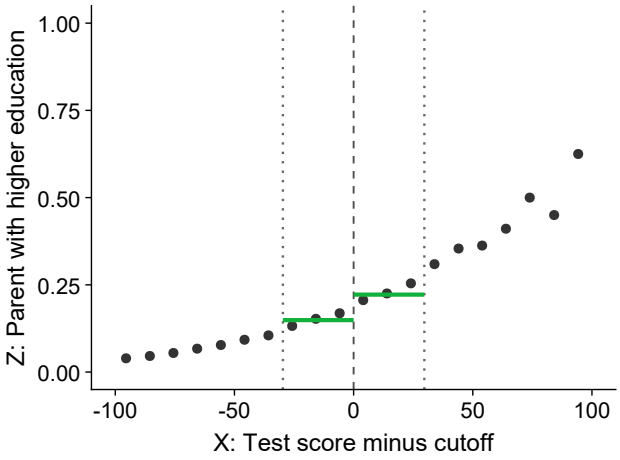
- Local linear regression on each side of the cutoff within bandwidth h
- $\hat{\tau}(h)$ - jump at c Base RD Definition

Covariate Imbalance within the Estimation Window



- Share of students with a parent with higher education
- $g(x) := E[Z|X = x]$ continuous in x

Covariate Imbalance within the Estimation Window



- Share of students with a parent with higher education
- $g(x) := E[Z|X = x]$ continuous in x
- But imbalanced within the bandwidth:
  - Treated: 22%
  - Control: 15%

Bias of Base RD

- Base RD is a weighted sum of outcomes
- Weights depend only on the running variable

$$\hat{\tau}(h) = \sum_{X_i \geq 0} w_i^+(h)\, Y_i - \sum_{X_i < 0} w_i^-(h)\, Y_i$$

Bias of Base RD

- Base RD is a weighted sum of outcomes
- Weights depend only on the running variable

$$\hat{\tau}(h) = \sum_{X_i \geq 0} w_i^+(h)\, Y_i - \sum_{X_i < 0} w_i^-(h)\, Y_i$$

Exact finite-sample bias (local linear)

$$E[\hat{\tau}(h)|\mathbf{X}] - \tau = \sum_{X_i \geq 0} w_i^+(h) \int_0^{X_i} [\bar{m}''(s) + \bar{\tau}''(s)](X_i - s)\, ds$$
$$- \sum_{X_i < 0} w_i^-(h) \int_0^{X_i} \bar{m}''(s)(X_i - s)\, ds$$

Bias of Base RD

- Base RD is a weighted sum of outcomes
- Weights depend only on the running variable

$$\hat{\tau}(h) = \sum_{X_i \geq 0} w_i^+(h)\, Y_i - \sum_{X_i < 0} w_i^-(h)\, Y_i$$

Exact finite-sample bias (local linear)

$$E[\hat{\tau}(h)|\mathbf{X}] - \tau = \sum_{X_i \geq 0} w_i^+(h) \int_0^{X_i} [\bar{m}''(s) + \bar{\tau}''(s)](X_i - s)\, ds$$
$$- \sum_{X_i < 0} w_i^-(h) \int_0^{X_i} \bar{m}''(s)(X_i - s)\, ds$$

- Bias depends on the curvature of the pooled regression functions within the bandwidth

Composition Bias

Pooled functions average the long regression functions over groups

$$\bar{m}(x) = m(x,1)\,g(x) + m(x,0)\,(1 - g(x)),$$

where $m(x,z) := E[Y_i(0)|X_i = x, Z_i = z]$

Composition Bias

Pooled functions average the long regression functions over groups

$$\bar{m}(x) = m(x,1)\,g(x) + m(x,0)\,(1-g(x)),$$

where $m(x,z) := E[Y_i(0)|X_i = x, Z_i = z]$

Curvature of the pooled function

$$\bar{m}''(x) = \underbrace{m''(x,1)g(x) + m''(x,0)(1-g(x))}_{\text{(A) within groups}} + \underbrace{2\Delta m'(x)g'(x) + \Delta m(x)g''(x)}_{\text{(B) composition}}$$

where $\Delta m(x) := m(x,1) - m(x,0)$

Composition Bias

Pooled functions average the long regression functions over groups

$$\bar{m}(x) = m(x,1)\,g(x) + m(x,0)\,(1-g(x)),$$

where $m(x,z) := E[Y_i(0)|X_i = x, Z_i = z]$

Curvature of the pooled function

$$\bar{m}''(x) = \underbrace{m''(x,1)g(x) + m''(x,0)(1-g(x))}_{\text{(A) within groups}} + \underbrace{2\Delta m'(x)g'(x) + \Delta m(x)g''(x)}_{\text{(B) composition}}$$

where $\Delta m(x) := m(x,1) - m(x,0)$

- Likewise for $\bar{\tau}''$

Example 1 - Composition Bias with Linear Primitives

Primitives: Long regression functions and group share are linear

$$m(x,z) = \alpha + \beta x, \qquad \tau(x,z) = (\tau_1 + \delta x)\, z, \qquad g(x) = g + \eta x$$

- Both groups share the same baseline
- Only group 1 responds to treatment
- Isolates composition bias: $\bar{m}''(x) = 0$ and $\bar{\tau}''(x) = 2\Delta\tau'(x)\, g'(x) = 2\delta\eta$

Example 1 - Composition Bias with Linear Primitives

Primitives: Long regression functions and group share are linear

$$m(x,z) = \alpha + \beta x, \qquad \tau(x,z) = (\tau_1 + \delta x)\, z, \qquad g(x) = g + \eta x$$

- Both groups share the same baseline
- Only group 1 responds to treatment
- Isolates composition bias: $\bar{m}''(x) = 0$ and $\bar{\tau}''(x) = 2\Delta\tau'(x)\, g'(x) = 2\delta\eta$

Bias of local linear base RD (uniform kernel)

$$E[\hat{\tau}(h)|\mathbf{X}] - \tau = \delta\eta \sum\nolimits_{X_i \geq 0} w_i^+(h)\, X_i^2 = -\frac{\delta\eta}{6}\, h^2 + o_p(h^2)$$

Example 1 - Composition Bias with Linear Primitives

Primitives: Long regression functions and group share are linear

$$m(x,z) = \alpha + \beta x, \qquad \tau(x,z) = (\tau_1 + \delta x)\, z, \qquad g(x) = g + \eta x$$

- Both groups share the same baseline
- Only group 1 responds to treatment
- Isolates composition bias: $\bar{m}''(x) = 0$ and $\bar{\tau}''(x) = 2\Delta\tau'(x)\, g'(x) = 2\delta\eta$

Bias of local linear base RD (uniform kernel)

$$E[\hat{\tau}(h)|\mathbf{X}] - \tau = \delta\eta \sum\nolimits_{X_i \geq 0} w_i^+(h)\, X_i^2 = -\frac{\delta\eta}{6}\, h^2 + o_p(h^2)$$

- Base RD has first-order composition bias

Example 2 - Pointwise Asymptotics

Primitives: Constant long regression functions and nonparametric group composition

$$m(x,z) = \mu_0 + \Delta m\, z, \qquad \tau(x,z) = \tau, \qquad g \in C^2 \text{ with } |g''(x)| \leq L$$

- homogeneous and constant treatment effect
- Isolates composition bias: $\bar{m}''(x) = \Delta m\, g''(x)$ and $\bar{\tau}''(x) = 0$

Example 2 - Pointwise Asymptotics

Primitives: Constant long regression functions and nonparametric group composition

$$m(x,z) = \mu_0 + \Delta m\, z, \qquad \tau(x,z) = \tau, \qquad g \in C^2 \text{ with } |g''(x)| \leq L$$

- homogeneous and constant treatment effect
- Isolates composition bias: $\bar{m}''(x) = \Delta m\, g''(x)$ and $\bar{\tau}''(x) = 0$

$$E[\hat{\tau}(h)|\mathbf{X}] - \tau = \Delta m \left(\underbrace{\sum_{X_i \geq 0} w_i^+(h) \int_0^{X_i} g''(s)\,(X_i - s)\,ds}_{\text{I)}} - \underbrace{\sum_{X_i < 0} w_i^-(h) \int_0^{X_i} g''(s)\,(X_i - s)\,ds}_{\text{II)}} \right)$$

Example 2 - Pointwise Asymptotics

Primitives: Constant long regression functions and nonparametric group composition

$$m(x,z) = \mu_0 + \Delta m\, z, \qquad \tau(x,z) = \tau, \qquad g \in C^2 \text{ with } |g''(x)| \leq L$$

- homogeneous and constant treatment effect
- Isolates composition bias: $\bar{m}''(x) = \Delta m\, g''(x)$ and $\bar{\tau}''(x) = 0$

$$E[\hat{\tau}(h)|\mathbf{X}] - \tau = \Delta m \left(\underbrace{\sum_{X_i \geq 0} w_i^+(h) \int_0^{X_i} g''(s)\,(X_i - s)\,ds}_{\text{I)}} - \underbrace{\sum_{X_i < 0} w_i^-(h) \int_0^{X_i} g''(s)\,(X_i - s)\,ds}_{\text{II)}} \right)$$

- Fix g and let $n \to \infty$
- I) and II) converge to the same constant
- Cancellation yields $E[\hat{\tau}(h)|\mathbf{X}] - \tau = o_p(h^2)$ (Superefficiency (Brown et al. 1997))

Example 2 - Uniform Asymptotics

- Evaluate worst case over all g that satisfy the assumptions
- Then let $n \to \infty$

Example 2 - Uniform Asymptotics

- Evaluate worst case over all g that satisfy the assumptions
- Then let $n \to \infty$

$$\sup_{g(x)} E[\hat{\tau}(h)|\mathbf{X}] - \tau = h^2 L\, 2|\Delta m| \frac{1}{12} + o_p(h^2)$$

▸ Worst-Case Attainment

Example 2 - Uniform Asymptotics

- Evaluate worst case over all g that satisfy the assumptions
- Then let $n \to \infty$

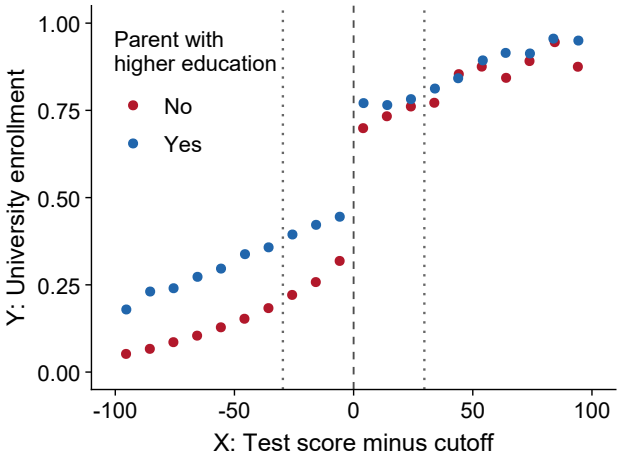
$$\sup_{g(x)} E[\hat{\tau}(h)|\mathbf{X}] - \tau = h^2 L \, 2|\Delta m| \, \frac{1}{12} + o_p(h^2)$$

▸ Worst-Case Attainment

- Base RD compares treated and control samples with different group compositions
- First-order worst-case composition bias depending on heterogeneity between groups
- **Solution:** Stratify by group

Stratified RD

Illustration - Stratified RD Estimator

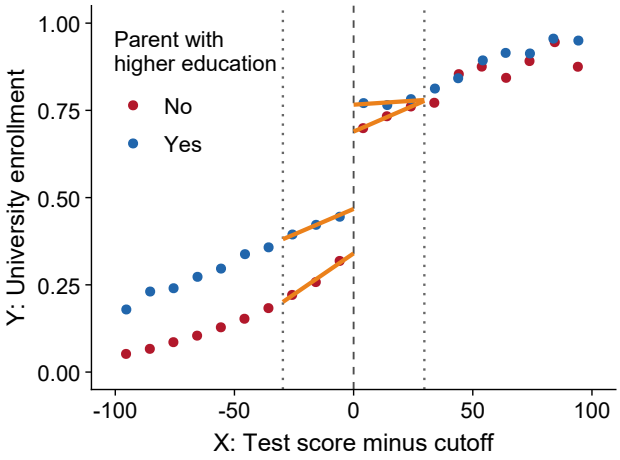


Target Parameter:

$$\tau \overset{\text{LIE}}{=} g\,\tau_1 + (1-g)\,\tau_0$$

Stratification:

Illustration - Stratified RD Estimator



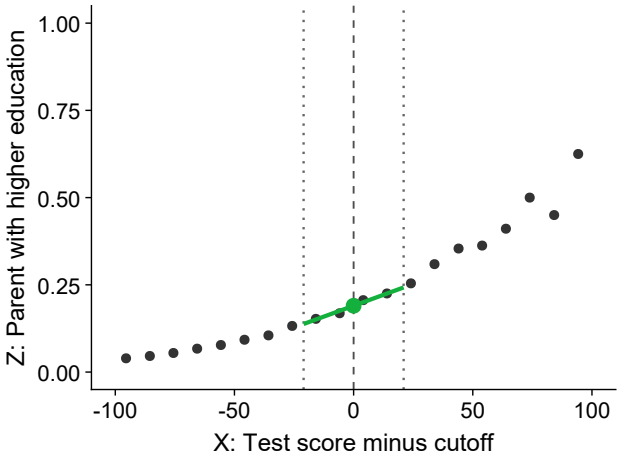
Target Parameter:

$$\tau \overset{\text{LIE}}{=} g\,\tau_1 + (1-g)\,\tau_0$$

Stratification:

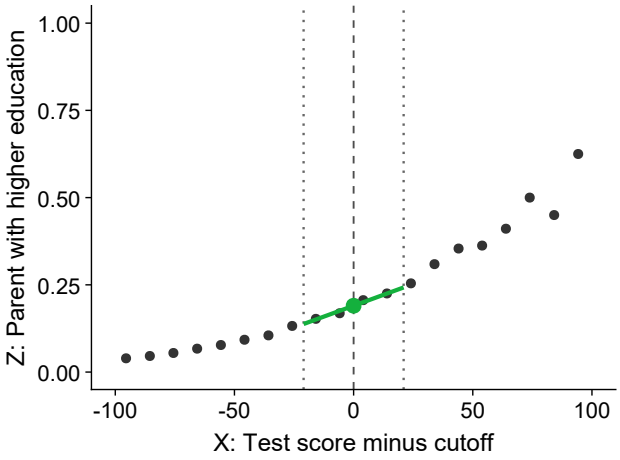
- $\hat{\tau}_j(h)$ - Estimate effect within each group
- using the same bandwidth h

Estimating the Group Share at the Cutoff



- $g(x)$ continuous
- Estimation at an interior point
- $\hat{g}(h_z)$ - Local linear regression of Z on X across the cutoff with bandwidth $h_z = \kappa h$

Estimating the Group Share at the Cutoff



- $g(x)$ continuous
- Estimation at an interior point
- $\hat{g}(h_z)$ - Local linear regression of Z on X across the cutoff with bandwidth $h_z = \kappa h$

Stratified RD Estimator:

- $\hat{\tau}_S = \hat{\tau}_1 \hat{g} + \hat{\tau}_0 (1 - \hat{g})$

▸ Stratified RD Definition

Stratified RD in the Examples

Leading bias of Stratified RD

$$LB = \underbrace{g\,E[\hat{\tau}_1(h) - \tau_1 | \mathbf{X}, \mathbf{Z}] + (1-g)\,E[\hat{\tau}_0(h) - \tau_0 | \mathbf{X}, \mathbf{Z}]}_{\text{(A) within groups}} + \underbrace{(\tau_1 - \tau_0)\,E[\hat{g}(h_z) - g | \mathbf{X}]}_{\text{(C) weight estimation}}$$

- No composition bias (B)

Stratified RD in the Examples

Leading bias of Stratified RD

$$LB = \underbrace{g\, E[\hat{\tau}_1(h) - \tau_1 | \mathbf{X}, \mathbf{Z}] + (1-g)\, E[\hat{\tau}_0(h) - \tau_0 | \mathbf{X}, \mathbf{Z}]}_{\text{(A) within groups}} + \underbrace{(\tau_1 - \tau_0)\, E[\hat{g}(h_z) - g | \mathbf{X}]}_{\text{(C) weight estimation}}$$

- No composition bias (B)

Example 1

- Linear long regression functions and group share, so all terms are zero at any h, h_z

Stratified RD in the Examples

Leading bias of Stratified RD

$$LB = \underbrace{g\, E[\hat{\tau}_1(h) - \tau_1 | \mathbf{X}, \mathbf{Z}] + (1-g)\, E[\hat{\tau}_0(h) - \tau_0 | \mathbf{X}, \mathbf{Z}]}_{\text{(A) within groups}} + \underbrace{(\tau_1 - \tau_0)\, E[\hat{g}(h_z) - g | \mathbf{X}]}_{\text{(C) weight estimation}}$$

- No composition bias (B)

Example 1

- Linear long regression functions and group share, so all terms are zero at any h, h_z

Example 2

- Constant long regression functions remove (A), and $\tau_1 = \tau_0$ removes (C)

Estimator Comparison

Smoothness Class

Definition 1 (Hölder Class of Order p)

Let $f : \mathbb{R} \to \mathbb{R}$ and $p \geq 1$ be an integer, then the (integer) Hölder class of order p is defined as:

$$\mathcal{F}_p(L) := \left\{ f : |f^{(p-1)}(x) - f^{(p-1)}(y)| \leq L|x - y| \quad \forall x, y \in \mathbb{R} \right\}$$

- Includes all functions f:
  i) that are p times continuously differentiable with $|f^{(p)}|_\infty \leq L$
  ii) or that are arbitrarily well approximated by functions that satisfy i)

Smoothness Assumptions on the Primitives

Assumption 1 (Smoothness on the Primitives)

For $p \in \mathbb{N}^+$, across all x and $z \in \{0,1\}$, and nonnegative constants:

Stratified RD

i. $m(x,z),\ m(x,z)+\tau(x,z) \in \mathcal{F}_p(M)$

ii. $g(x) \in \mathcal{F}_p(L)$

iii. $|\Delta\tau(x)| \leq C_{\Delta\tau}$

Base RD, in addition

iv. $|\Delta m(x)| \leq C_{\Delta m}$

v. *For* $j = 1, \ldots, p-1$: $|g^{(j)}(x)| \leq L_j$, $|\Delta m^{(j)}(x)| \leq C_{\Delta m,j}$, $|\Delta\tau^{(j)}(x)| \leq C_{\Delta\tau,j}$

- Place smoothness assumptions on primitives: long and group regression functions
- Smoothness of the pooled regression function derived using triangle inequality

$$|(\bar{m}+\bar{\tau})'(x)| \leq M + \underbrace{(C_{\Delta\tau}+C_{\Delta m})L}_{\text{composition}}$$

- Same class for both estimators

Worst-Case Bias

Worst-case first order bias of both estimators over the class of primitives:

Proposition 2 (Stratified RD)

$$\overline{\mathrm{Bias}}\big(\hat{\tau}_S(h, h_z)\big) = h^p\big(b_S + o_p(1)\big), \qquad b_S := \frac{1}{p!}\Big[2M\, B_{p,q}(k) + \kappa^p\, C_{\Delta\tau} L\, B^{\mathrm{int}}_{p,q}(k)\Big]$$

Proposition 3 (Base RD, Lower Bound)

$$\overline{\mathrm{Bias}}\big(\hat{\tau}(h)\big) \geq h^p\big(b_B^{\mathrm{low}} + o_p(1)\big), \qquad b_B^{\mathrm{low}} := \frac{1}{p!}\Big[2M\, B_{p,q}(k) + 2C_{\Delta m} L\, B_{p,q}(k) + C_{\Delta\tau} L\, B_{p,q}(k)\Big]$$

Worst-Case Bias

Worst-case first order bias of both estimators over the class of primitives:

Proposition 2 (Stratified RD)

$$\overline{\mathrm{Bias}}\big(\hat{\tau}_S(h, h_z)\big) = h^p\big(b_S + o_p(1)\big), \qquad b_S := \frac{1}{p!}\Big[2M\, B_{p,q}(k) + \kappa^p\, C_{\Delta\tau} L\, B^{\mathrm{int}}_{p,q}(k)\Big]$$

Proposition 3 (Base RD, Lower Bound)

$$\overline{\mathrm{Bias}}\big(\hat{\tau}(h)\big) \geq h^p\big(b_B^{\mathrm{low}} + o_p(1)\big), \qquad b_B^{\mathrm{low}} := \frac{1}{p!}\Big[2M\, B_{p,q}(k) + 2C_{\Delta m} L\, B_{p,q}(k) + C_{\Delta\tau} L\, B_{p,q}(k)\Big]$$

- Stratified RD has no composition bias $2C_{\Delta m}L$

Worst-Case Bias

Worst-case first order bias of both estimators over the class of primitives:

Proposition 2 (Stratified RD)

$$\overline{\mathrm{Bias}}\big(\hat{\tau}_S(h, h_z)\big) = h^p\big(b_S + o_p(1)\big), \qquad b_S := \frac{1}{p!}\Big[2M\, B_{p,q}(k) + \kappa^p\, C_{\Delta\tau} L\, B^{\mathrm{int}}_{p,q}(k)\Big]$$

Proposition 3 (Base RD, Lower Bound)

$$\overline{\mathrm{Bias}}\big(\hat{\tau}(h)\big) \geq h^p\big(b_B^{\mathrm{low}} + o_p(1)\big), \qquad b_B^{\mathrm{low}} := \frac{1}{p!}\Big[2M\, B_{p,q}(k) + 2C_{\Delta m} L\, B_{p,q}(k) + C_{\Delta\tau} L\, B_{p,q}(k)\Big]$$

- Stratified RD has no composition bias $2C_{\Delta m}L$
- Kernel constants $B_{p,q}(k)$, $B^{\mathrm{int}}_{p,q}(k)$ from Armstrong and Kolesár (2020, Thm. 3.1)
- If $\kappa \leq U_{p,q}(k) := \big(B_{p,q}(k)/B^{\mathrm{int}}_{p,q}(k)\big)^{1/p}$ its weight estimation bias is weakly smaller than that of base RD

Main Result

Asymptotic worst-case MSE of Stratified RD and its lower bound for base RD

$$\overline{\mathrm{AMSE}}\big(\hat{\tau}_S(h, h_z)\big) := \big(h^p\, b_S\big)^2 + \frac{\mathcal{V}_S}{nh} \qquad \text{and} \qquad \overline{\mathrm{AMSE}}^{\mathrm{low}}\big(\hat{\tau}(h)\big) := \big(h^p\, b_B^{\mathrm{low}}\big)^2 + \frac{\mathcal{V}_B}{nh}$$

Main Result

Asymptotic worst-case MSE of Stratified RD and its lower bound for base RD

$$\overline{\mathrm{AMSE}}\big(\hat{\tau}_S(h,h_z)\big) := \big(h^p\, b_S\big)^2 + \frac{\mathcal{V}_S}{nh} \qquad \text{and} \qquad \overline{\mathrm{AMSE}}^{\mathrm{low}}\big(\hat{\tau}(h)\big) := \big(h^p\, b_B^{\mathrm{low}}\big)^2 + \frac{\mathcal{V}_B}{nh}$$

Theorem 4 (Comparison of the Stratified RD and Base RD Estimators)

Under regularity conditions

(i) *If* $\kappa \le U_{p,q}(k)$, *then* $b_S \le b_B^{\mathrm{low}}$.

(ii) *If* $\kappa \ge L_{p,q}(k)$, *then* $\mathcal{V}_S \le \mathcal{V}_B$.

(iii) *If* $\kappa \in I_{p,q}(k) := [L_{p,q}(k), U_{p,q}(k)]$, *then* $\overline{\mathrm{AMSE}}\big(\hat{\tau}_S(h,h_z)\big) \le \overline{\mathrm{AMSE}}^{\mathrm{low}}\big(\hat{\tau}(h)\big)$.

Main Result

Asymptotic worst-case MSE of Stratified RD and its lower bound for base RD

$$\overline{\mathrm{AMSE}}\big(\hat{\tau}_S(h,h_z)\big) := \big(h^p\, b_S\big)^2 + \frac{\mathcal{V}_S}{nh} \qquad \text{and} \qquad \overline{\mathrm{AMSE}}^{\mathrm{low}}\big(\hat{\tau}(h)\big) := \big(h^p\, b_B^{\mathrm{low}}\big)^2 + \frac{\mathcal{V}_B}{nh}$$

Theorem 4 (Comparison of the Stratified RD and Base RD Estimators)

Under regularity conditions

(i) *If* $\kappa \le U_{p,q}(k)$, *then* $b_S \le b_B^{\mathrm{low}}$.

(ii) *If* $\kappa \ge L_{p,q}(k)$, *then* $\mathcal{V}_S \le \mathcal{V}_B$.

(iii) *If* $\kappa \in I_{p,q}(k) := [L_{p,q}(k), U_{p,q}(k)]$, *then* $\overline{\mathrm{AMSE}}\big(\hat{\tau}_S(h,h_z)\big) \le \overline{\mathrm{AMSE}}^{\mathrm{low}}\big(\hat{\tau}(h)\big)$.

- $\kappa \in I_{p,q}(k)$ a kernel- and order-specific interval
- e.g. Local linear, uniform kernel: $I_{2,1} = [1/4, \sqrt{1/2}] \approx [0.25, 0.71]$

Main Results (informal)

Using appropriate bandwidths the Stratified RD estimator has asymptotically weakly

- lower first-order worst-case bias
- lower variance
- lower worst-case MSE
- shorter honest confidence intervals

than the base RD estimator.

Main Results (informal)

Using appropriate bandwidths the Stratified RD estimator has asymptotically weakly

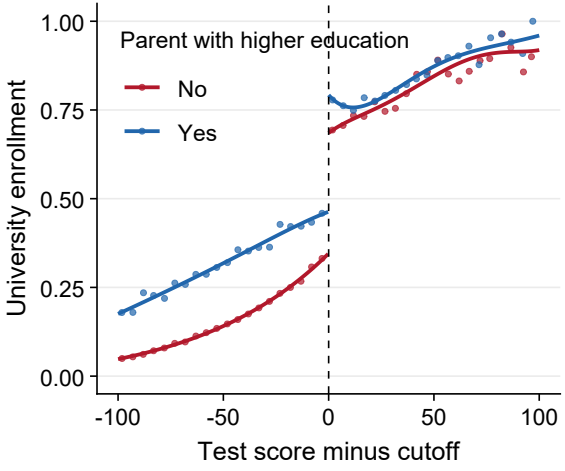
- lower first-order worst-case bias
- lower variance
- lower worst-case MSE
- shorter honest confidence intervals

than the base RD estimator.

- The reduction in variance is strict if groups are heterogeneous
- The reduction in first-order worst-case bias is strict unless composition bias is ruled out by assumption

Empirical Application

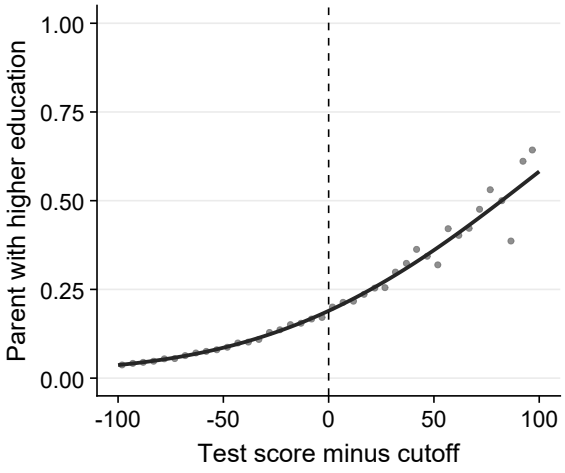
Calibrated Simulation - Enrollment by Group



- Calibrate simulation using quintic polynomials to estimate $m(x, z)$ (Imbens and Kalyanaraman 2012)
- One fit per group and side of the cutoff
- Normal distribution for X_i matched to the data near the cutoff

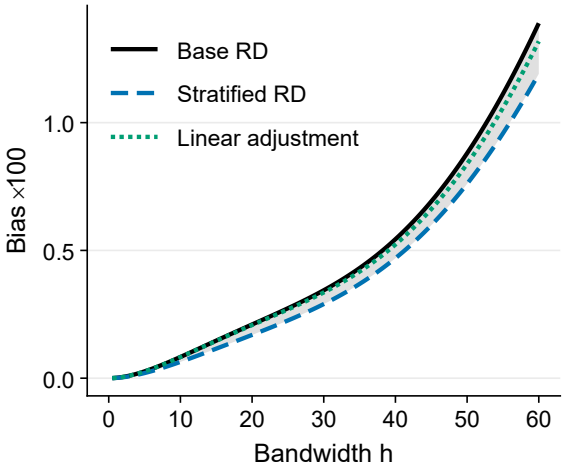
▸ Running Variable

Calibrated Simulation - Group Composition



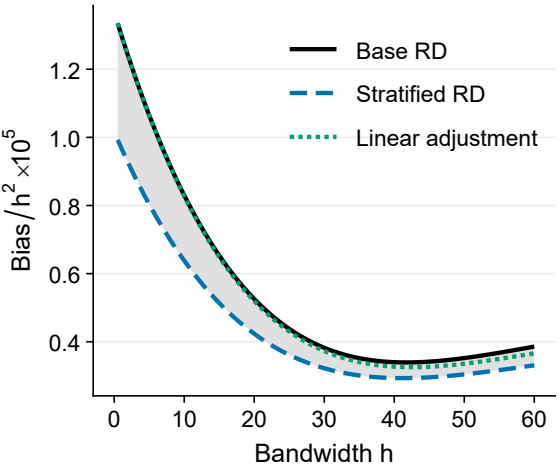
- Quintic polynomial fitted across the cutoff for the group share $g(x)$
- Heterogeneity and changing composition: composition bias in base RD

Specification Bias



- Specification bias at fixed bandwidth (Kolesár and Rothe 2018)
- $\text{plim}_{n\to\infty} \hat{\tau}(h) - \tau$
- Same h for all estimators
- Weight estimation using $h_z = \kappa h$ with $\kappa = \sqrt{1/2}$
- Bias shrinks with bandwidth
- Stratified RD has the lowest bias at every h

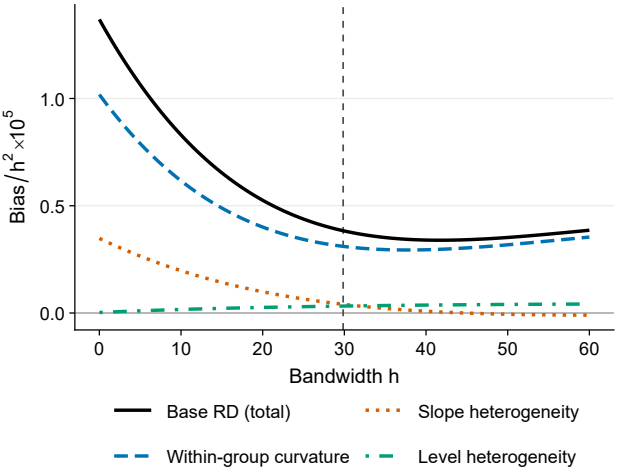
First-Order Specification Bias



- Bias divided by h^2
- $h \to 0$ gives the leading bias constant
- composition bias is first order
- 26% of the leading bias constant of base RD
- Linear adjustment leaves the leading constant unchanged

▸ Decomposition

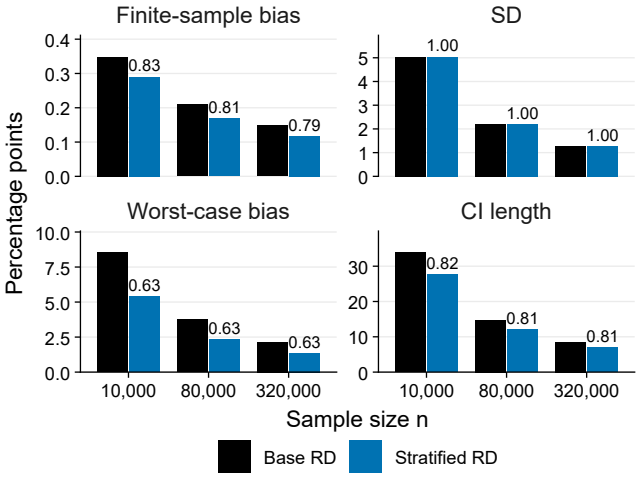
Decomposition of the Base RD Specification Bias



- Pointwise thought experiment
- As $h \to 0$:
  - slope heterogeneity remains
  - level heterogeneity cancels across the cutoff
- At the finite sample bandwidth used in the paper both are relevant:
  - $n = 299,475$ with bandwidth $h = 29.9$
  - Slope heterogeneity: 10.6%
  - Level heterogeneity: 8.4%

Back

Calibrated Simulation - Finite-Sample Performance



- 1,000 simulations using the calibrated model
- Following Armstrong and Kolesár (2020)
  - Rule-of-thumb bandwidth to pick smoothness bounds in full sample
  - Optimal bandwidth using sample size in simulation
- Finite-sample bias about 17-21% lower
- Worst-case bias 37% lower
- Honest CIs 18-19% shorter

▸ Placebo Study ▸ Full Table

Conclusion

Conclusion

Base RD has leading bias due to discrete covariates

Unless the researcher assumes...

- Group homogeneity
  - $\Delta m(x) = \Delta \tau(x) = 0$
- Or constant group-composition
  - Quasi-randomized group variable
  - Diagnostics: Test for within-window covariate balance

▸ Balance Tests

Conclusion

Base RD has leading bias due to discrete covariates

Unless the researcher assumes...

- Group homogeneity
  - $\Delta m(x) = \Delta \tau(x) = 0$
- Or constant group-composition
  - Quasi-randomized group variable
  - Diagnostics: Test for within-window covariate balance

▸ Balance Tests

Stratified RD

- Lower worst-case bias and variance over common smoothness class
- Shorter honest confidence intervals for the same target parameter
- Relies less on extrapolation and smoothness

Conclusion

Base RD has leading bias due to discrete covariates

Unless the researcher assumes...

- Group homogeneity
  - $\Delta m(x) = \Delta \tau(x) = 0$
- Or constant group-composition
  - Quasi-randomized group variable
  - Diagnostics: Test for within-window covariate balance

▸ Balance Tests

Stratified RD

- Lower worst-case bias and variance over common smoothness class
- Shorter honest confidence intervals for the same target parameter
- Relies less on extrapolation and smoothness

Extensions and Implementation

- Extends to continuous covariates (Lipschitz smoothness)
- Intuitive regression-based implementation - R package `RDStratified` (work in progress)

Discussion

Any comments welcome!

tgrossbo@mail.uni-mannheim.de

References i

- Armstrong, Timothy B. and Michal Kolesár (2018). **"Optimal Inference in a Class of Regression Models".** en. In: *Econometrica* 86.2, pp. 655–683.
- — (2020). **"Simple and Honest Confidence Intervals in Nonparametric Regression".** en. In: *Quantitative Economics* 11.1, pp. 1–39.
- Brown, Lawrence D. et al. (Dec. 1997). **"Superefficiency in Nonparametric Function Estimation".** In: *The Annals of Statistics* 25.6, pp. 2607–2625.
- Calonico, Sebastian, Matias D. Cattaneo, Max H. Farrell, et al. (July 2019). **"Regression Discontinuity Designs Using Covariates".** en. In: *The Review of Economics and Statistics* 101.3, pp. 442–451.
- Calonico, Sebastian, Matias D. Cattaneo, and Rocío Titiunik (2014). **"Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs".** en. In: *Econometrica* 82.6, pp. 2295–2326.

References ii

- Cattaneo, Matias D. et al. (Mar. 2023). **"Covariate Adjustment in Regression Discontinuity Designs".** en. In: *Handbook of Matching and Weighting Adjustments for Causal Inference*. Ed. by José R. Zubizarreta et al. Boca Raton: Chapman and Hall/CRC, pp. 153–168.
- Gelman, Andrew and Guido Imbens (Dec. 2025). **"Stable Adjustments in Regression Discontinuity. A Reply to Albada (Journal of Comments and Replications in Economics, 2025)".** en. In: *Journal of Comments and Replications in Economics* 4.
- Imbens, Guido and Karthik Kalyanaraman (July 2012). **"Optimal Bandwidth Choice for the Regression Discontinuity Estimator".** In: *The Review of Economic Studies* 79.3, pp. 933–959.
- Imbens, Guido and Donald Rubin (2015). **Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction.** Cambridge: Cambridge University Press.
- Kolesár, Michal and Christoph Rothe (Aug. 2018). **"Inference in Regression Discontinuity Designs with a Discrete Running Variable".** en. In: *American Economic Review* 108.8, pp. 2277–2304.

References iii

- Kreiß, Alexander and Christoph Rothe (May 2023). **"Inference in Regression Discontinuity Designs with High-Dimensional Covariates".** en. In: *The Econometrics Journal* 26.2, pp. 105–123.
- Londoño-Vélez, Juliana et al. (May 2020). **"Upstream and Downstream Impacts of College Merit-Based Financial Aid for Low-Income Students: Ser Pilo Paga in Colombia".** en. In: *American Economic Journal: Economic Policy* 12.2, pp. 193–227.
- Ma, Jun and Zhengfei Yu (May 2024). **Empirical Likelihood Covariate Adjustment for Regression Discontinuity Designs.**
- Noack, Claudia et al. (Sept. 2026). **"Flexible Covariate Adjustments in Regression Discontinuity Designs".** In: *Journal of Econometrics* 257, p. 106298.
- Sekhon, Jasjeet S. and Rocío Titiunik (2017). **"Understanding Regression Discontinuity Designs As Observational Studies".** In: *Observational Studies* 3.2, pp. 174–182.

Appendix

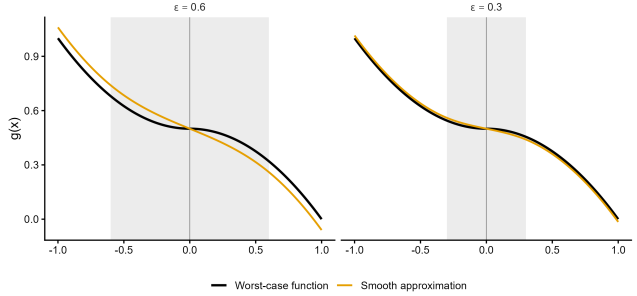
Base RD Estimator

$$\hat{\tau}(h) := e_1^\top \arg\min_{\beta \in \mathbb{R}^{2q+2}} \sum_{i=1}^{n} k(X_i/h)(Y_i - S_q(X_i)^\top \beta)^2$$

- k: symmetric, nonnegative, and bounded kernel with bounded support
- h: bandwidth
- $S_q(X_i) = (T_i, 1, X_i, T_i X_i, \ldots, X_i^q, X_i^q T_i)^\top$: q-th order polynomial in X_i fully interacted with T_i
- $e_1 := (1, 0, \ldots, 0)^\top$ selects the coefficient of T_i

Back

Worst-Case Function and Smooth Approximations



- Worst case $\tilde{g}(x) = -\frac{L}{2}\, x^2 \,\mathrm{sign}(x)$
- $\tilde{g}''$ jumps from L to $-L$ at the cutoff
- Smooth g_ϵ with $|g_\epsilon''| \le L$ interpolates the jump on $(-\epsilon, \epsilon)$ (shaded)
- At any fixed n its bias approaches the worst case as $\epsilon \to 0$, yet pointwise it is $o_p(h^2)$ Back

Stratified RD Estimator

Step 1 — Base RD within each group, with $h_j = h$ for all groups

$$\hat{\tau}_j(h_j) = e_1^\top \arg\min_{\beta \in \mathbb{R}^{2q+2}} \sum_{i=1}^{n} k(X_i/h_j)(Y_i - S_q(X_i)^\top \beta)^2 \mathbb{1}_{\{Z_i = j\}}$$

Step 2 — Group share at the cutoff, local polynomial regression for an interior point

$$\hat{g}(h_z) = e_1^\top \arg\min_{\beta \in \mathbb{R}^{q+1}} \sum_{i=1}^{n} k(X_i/h_z)(Z_i - \tilde{S}_q(X_i)^\top \beta)^2$$

with $\tilde{S}_q(X_i) = (1, X_i, \ldots, X_i^q)^\top$ and proportional bandwidths $h_z = \kappa h$

Step 3 — Stratified RD estimator

$$\hat{\tau}_S(h, h_z) = \hat{\tau}_1(h)\,\hat{g}(h_z) + \hat{\tau}_0(h)\,(1 - \hat{g}(h_z))$$

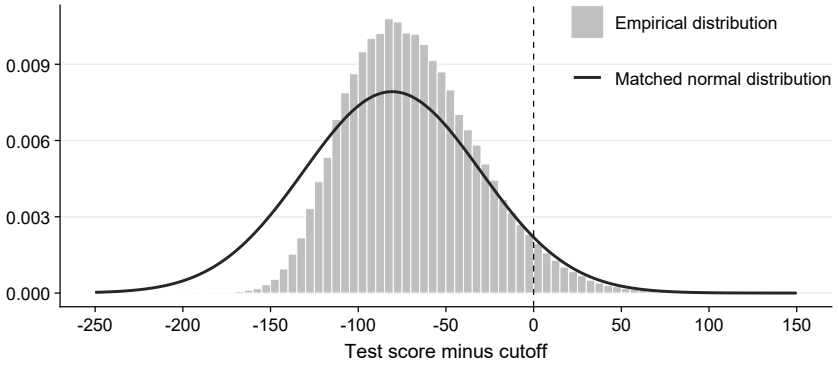
Back

Calibrated Simulation - Full Results

Bandwidth selection		Finite-sample bias ×100			SD ×100			CI length
n	*h*	Base	Strat.	Ratio	Base	Strat.	Ratio	Ratio
Panel A. CCT rdrobust								
10,000	58.9	1.330	1.135	0.85	3.63	3.63	1.00	0.69
80,000	38.9	0.516	0.445	0.86	1.59	1.60	1.00	0.69
320,000	29.5	0.336	0.283	0.84	0.91	0.91	1.00	0.69
Panel B. AK optimal bandwidth								
10,000	18.6	0.193	0.152	0.79	6.62	6.59	1.00	0.94
80,000	12.3	0.112	0.087	0.78	2.80	2.81	1.00	0.94
320,000	9.3	0.074	0.057	0.77	1.60	1.60	1.00	0.94
Panel C. AK rule of thumb								
10,000	30.2	0.349	0.290	0.83	5.04	5.03	1.00	0.82
80,000	19.9	0.209	0.169	0.81	2.20	2.19	1.00	0.81
320,000	15.1	0.148	0.117	0.79	1.27	1.27	1.00	0.81

Back

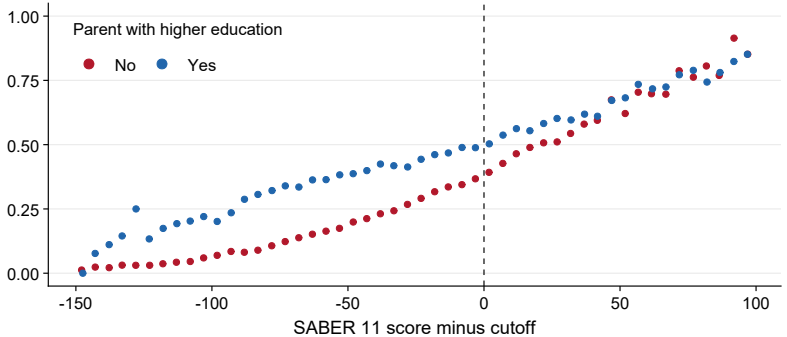
Calibrated Simulation - Running Variable



- Normal distribution matched to the data within 10 points on each side of the cutoff

Back

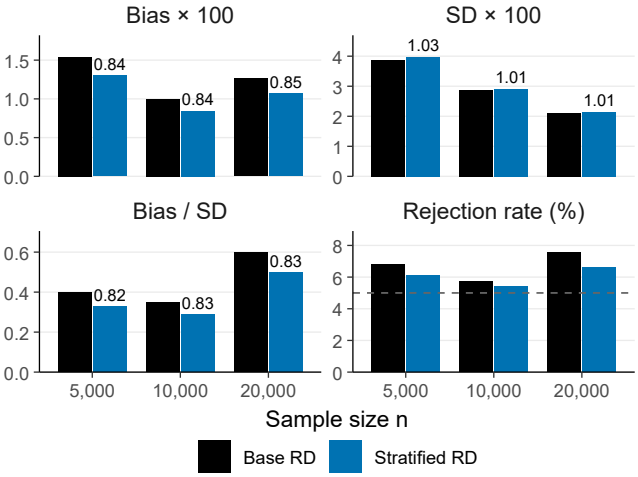
Placebo Study - Enrollment of SISBEN-Ineligible Students



- Monte Carlo placebo study following Noack et al. (2026)
- Ineligible for the treatment, so the true jump is zero ($N = 248{,}415$)
- 10,000 subsamples of size n without replacement
- Bandwidth chosen once in the full population and rescaled by $(N/n)^{1/5}$

Results Back

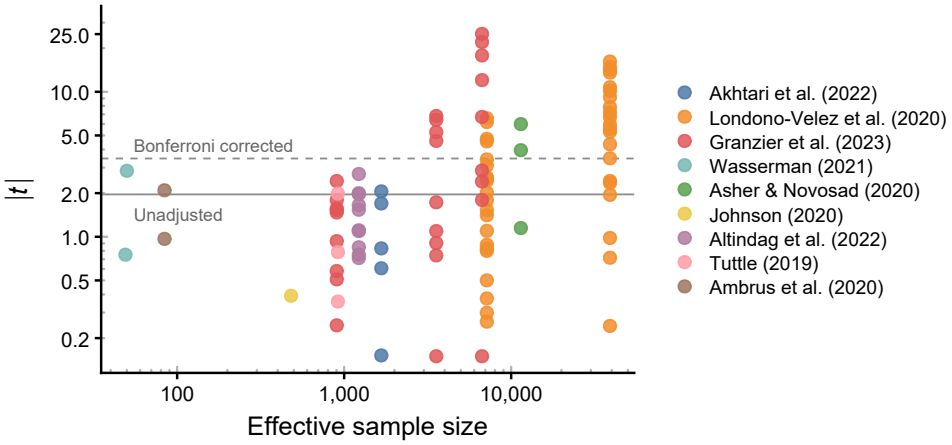
Placebo Study - CCT rdrobust



- rdrobust bandwidths: $h = 67.3, 58.6, 51.0$
- Same h for both estimators, $h_z = \sqrt{1/2}\, h$
- Bias 15-16% lower
- Slightly increased SD
- Conventional tests over-reject less for Stratified RD

Back

Covariate Balance in Replication Datasets



Back to This Paper
Back to Conclusion