

Estimation and Inference in Boundary Discontinuity Designs: Distance-Based Methods*

Matias D. Cattaneo[†] Rocio Titiunik[‡] Ruiqi (Rae) Yu[§]

March 26, 2026

Abstract

We study nonparametric distance-based (isotropic) local polynomial methods for estimating the boundary average treatment effect curve, a causal functional that captures treatment effect heterogeneity in boundary discontinuity designs. We establish identification, estimation, and inference results both pointwise and uniformly along the treatment assignment boundary. We show that the geometric regularity of the boundary, a one-dimensional manifold, plays a central role in determining feasible convergence rates and valid inference procedures. Our theoretical contributions are threefold. First, we derive uniform lower and upper bounds on the convergence rate of the misspecification bias of isotropic local polynomial estimators. Second, we obtain uniform distributional approximations that justify boundary-robust inference. Third, we establish minimax lower bounds for a broad class of nonparametric isotropic regression estimators. These results yield practical guidance for empirical implementation, including new bandwidth selection rules that adapt to local irregularities of the treatment-assignment boundary. We illustrate the proposed methods using simulation evidence and an empirical application, and provide companion general-purpose software.

Keywords: regression discontinuity, treatment effects estimation, causal inference, isotropic non-parametric regression, uniform inference, minimax convergence rate.

*We thank Alberto Abadie, Boris Hanin, Kosuke Imai, Xinwei Ma, Victor Panaretos, Jörg Stoye, and Jeff Wooldridge for comments and discussions. The co-Editor, Associate Editor, and two reviewers provided excellent comments and suggestions. Cattaneo and Titiunik gratefully acknowledge financial support from the National Science Foundation (SES-2019432, DMS-2210561, SES-2241575 and SES-2342226). Cattaneo gratefully acknowledges financial support from the National Institute for Food and Agriculture (NIFA) through grant 2024-67023-42704, and the Data-Driven Social Science initiative at Princeton University.

[†]Department of Operations Research and Financial Engineering, Princeton University.

[‡]Department of Politics, Princeton University.

[§]Department of Operations Research and Financial Engineering, Princeton University.

Contents

1	Introduction	1
1.1	Contributions	3
1.2	Notation	7
1.3	Organization	8
2	Setup and Assumptions	8
3	Identification and Interpretation	9
4	Approximation Bias	12
5	Estimation and Inference	15
5.1	Convergence Rates	15
5.2	Uncertainty Quantification	16
5.3	Discussion and Implementation	18
6	Distance-based Minimax Convergence Rate	21
7	Numerical Results	23
7.1	Simulation Study	25
7.2	Empirical Application	26
8	Extensions	29
8.1	Regression Kink Designs	29
8.2	Imperfect Compliance	29
8.3	Pre-treatment Covariates	30
9	Recommendations for Practice	31
10	Conclusion	33

1 Introduction

Discontinuities in treatment assignment are widely used in observational studies to investigate causal effects. In regression discontinuity (RD) designs, each unit receives a univariate score (or running variable), and treatment is assigned according to a threshold rule: units with scores equal to or above a known cutoff are assigned to treatment, while units with scores below the cutoff remain in the control condition. Local treatment effects at the cutoff are identifiable under continuity (Hahn et al., 2001) or local randomization (Cattaneo et al., 2015) assumptions, and estimation and inference are typically conducted using local polynomial least squares regression methods. See Cattaneo and Titiunik (2022) for an overview of the methodological RD literature and Cattaneo et al. (2020) for a practical introduction.

Boundary discontinuity (BD) designs generalize canonical RD settings to contexts in which units receive a bivariate score $\mathbf{X}_i = (X_{1i}, X_{2i})^\top$ with support $\mathcal{X} \subseteq \mathbb{R}^2$, and treatment assignment is determined by a threshold rule based on a one-dimensional boundary curve \mathcal{B} that partitions the support of the score. A prototypical example is the geographic RD design and its variants (Keele and Titiunik, 2015; Keele et al., 2015; Keele and Titiunik, 2016; Keele et al., 2017; Galiani et al., 2017; Rischard et al., 2021; Diaz and Zubizarreta, 2023), where a geographic boundary separates units into treatment and control areas according to their location. These designs are also referred to as Multi-Score RD Designs (Papay et al., 2011; Reardon and Robinson, 2012; Wong et al., 2013). A practical introduction to BD designs is provided in Cattaneo et al. (2024b, Section 5), while Cattaneo et al. (2026a) offers an overview of the empirical literature.

Using standard causal inference notation (Hernán and Robins, 2020), let $Y_i(0)$ and $Y_i(1)$ denote the potential outcomes for unit $i = 1, 2, \dots, n$ under control and treatment assignments, respectively. The key building block in BD designs is the *Boundary Average Treatment Effect Curve* (BATEC):

$$\tau(\mathbf{x}) = \mathbb{E}[Y_i(1) - Y_i(0) | \mathbf{X}_i = \mathbf{x}], \quad \mathbf{x} \in \mathcal{B},$$

a functional causal parameter that characterizes heterogeneous treatment effects along the assignment boundary. Identification, estimation, and inference for the BATEC can proceed through two

distinct methodological approaches:

- *Location-Based Methods.* This approach directly exploits the bivariate score \mathbf{X}_i , thereby extending canonical RD methods to multidimensional settings. Early identification and estimation results are provided by [Papay et al. \(2011\)](#), [Reardon and Robinson \(2012\)](#), [Wong et al. \(2013\)](#), and [Keele and Titiunik \(2015\)](#), while [Cattaneo et al. \(2026b\)](#) develop modern pointwise and uniform estimation and inference procedures.
- *Distance-Based Methods.* This approach reduces the bivariate score \mathbf{X}_i to a univariate distance-based running variable defined for each $\mathbf{x} \in \mathcal{B}$, thereby enabling the use of standard univariate RD methods. [Reardon and Robinson \(2012, Section 4\)](#) refer to this strategy as distance-based RD, and [Keele and Titiunik \(2015\)](#) discuss its use in geographic RD applications. [Velez and Newman \(2019\)](#) provide an empirical illustration employing distance-based polynomial regression to study treatment-effect heterogeneity in a geographic RD design.

The main goal of this paper is to study the identification, estimation, and inference properties of distance-based methods for the BATEC in BD designs, and to leverage the resulting theoretical insights to provide practical guidance for empirical implementation. The defining feature of these methods is the use of univariate local polynomial regression, where the running variable is the distance between \mathbf{X}_i and each point $\mathbf{x} \in \mathcal{B}$. From a nonparametric smoothing perspective, this approach is closely related to isotropic regression techniques studied in statistics and machine learning. By contrast, location-based methods employ local polynomial regressions that directly incorporate the bivariate score into the polynomial approximation.

In practice, researchers implementing BD designs frequently transform the multidimensional score into a univariate measure of distance to the assignment boundary and conduct estimation and inference using this distance as the running variable. Evidence compiled in [Cattaneo et al. \(2026a, Table 1\)](#) documents more than 80 recent empirical studies in economics and related quantitative disciplines adopting this strategy. When the goal is to study treatment-effect heterogeneity along the boundary, distance-based polynomial regression replaces the full bivariate location information in \mathbf{X}_i with a scalar distance defined relative to each boundary point $\mathbf{x} \in \mathcal{B}$. In other applications, the objective is instead to pool treatment effects along the boundary, in which case researchers often use the distance to the closest point on \mathcal{B} . This alternative strategy gives rise to pooling-

based methods, which are not the focus of this paper. Pooling-based methods discard location information and therefore identify a scalar (weighted) *Boundary Average Treatment Effect* (BATE) causal parameter. Cattaneo et al. (2026a) provide an introductory discussion, while ongoing work (Cattaneo et al., 2026c) develops formal identification, estimation, and inference results for pooling-based methods.

We focus on distance-based methods. Beyond their empirical prevalence, distance-based methods offer important practical advantages. In many applications, precise location information may be unavailable due to confidentiality restrictions, data aggregation, or measurement limitations, whereas distance-to-boundary measures can still be constructed or released. In such settings, distance-based procedures may provide the only feasible way to estimate the heterogeneous treatment effects summarized by the boundary average treatment effect curve $\tau(\mathbf{x})$. Moreover, by reducing the dimensionality of the running variable, these methods simplify implementation, facilitate interpretation, and enhance comparability with standard univariate RD analyses. These considerations underscore the importance of formally understanding the statistical properties of distance-based estimators in BD designs.

1.1 Contributions

To outline our main contributions, we first formalize the distance-based approach. The core ingredient is the scalar signed distance-based score for each unit $i = 1, \dots, n$,

$$D_i(\mathbf{x}) = (\mathbf{1}(\mathbf{X}_i \in \mathcal{A}_1) - \mathbf{1}(\mathbf{X}_i \in \mathcal{A}_0)) \cdot \mathcal{d}(\mathbf{X}_i, \mathbf{x}), \quad \mathbf{x} \in \mathcal{B},$$

where $\mathcal{d}(\cdot, \cdot)$ denotes a distance function, $\mathcal{X} = \mathcal{A}_0 \cup \mathcal{A}_1$ with \mathcal{A}_0 and \mathcal{A}_1 denoting the disjoint (connected) control and treatment regions, respectively, and $\mathcal{B} = \text{bd}(\mathcal{A}_0) \cap \text{bd}(\mathcal{A}_1)$ with $\text{bd}(\mathcal{A}_t)$ denoting the boundary of the set \mathcal{A}_t for $t \in \{0, 1\}$. Without loss of generality, we assume that the boundary belongs to the treatment region, that is, $\text{bd}(\mathcal{A}_1) \subset \mathcal{A}_1$ and $\mathcal{B} \cap \mathcal{A}_0 = \emptyset$. A canonical example is the Euclidean distance $\mathcal{d}(\mathbf{X}_i, \mathbf{x}) = |\mathbf{X}_i - \mathbf{x}| = \sqrt{(X_{1i} - x_1)^2 + (X_{2i} - x_2)^2}$ for $\mathbf{x} = (x_1, x_2)^\top \in \mathcal{B}$, although alternative distance measures may be appropriate in spatial applications (Banerjee, 2005). For each $\mathbf{x} \in \mathcal{B}$, the distance-based setup is analogous to a standard univariate RD design in which $D_i(\mathbf{x}) \in \mathbb{R}$ is the running variable, the cutoff is $c = 0$, and $D_i(\mathbf{x}) \geq 0$ (< 0)

corresponds to treatment (control) assignment.

The observed outcome is $Y_i = \mathbf{1}(D_i(\mathbf{x}) \in \mathcal{J}_0), Y_i(0) + \mathbf{1}(D_i(\mathbf{x}) \in \mathcal{J}_1), Y_i(1)$, where $\mathcal{J}_0 = (-\infty, 0)$ and $\mathcal{J}_1 = [0, \infty)$. Following standard practice in the RD literature, the distance-based local polynomial treatment effect estimator for each $\mathbf{x} \in \mathcal{B}$ is

$$\hat{\vartheta}(\mathbf{x}) = \mathbf{e}_1^\top \hat{\gamma}_1(\mathbf{x}) - \mathbf{e}_1^\top \hat{\gamma}_0(\mathbf{x}), \quad \mathbf{x} \in \mathcal{B},$$

where, for $t \in \{0, 1\}$,

$$\hat{\gamma}_t(\mathbf{x}) = \arg \min_{\gamma \in \mathbb{R}^{p+1}} \frac{1}{n} \sum_{i=1}^n (Y_i - \mathbf{r}_p(D_i(\mathbf{x}))^\top \gamma)^2 K_h(D_i(\mathbf{x})) \mathbf{1}(D_i(\mathbf{x}) \in \mathcal{J}_t), \quad (1)$$

with \mathbf{e}_1 the conformable first unit vector, $\mathbf{r}_p(u) = (1, u, u^2, \dots, u^p)^\top$ the usual polynomial basis, and $K_h(u) = K(u/h)/h$ a univariate kernel with bandwidth h . The kernel down-weights observations as their distance to $\mathbf{x} \in \mathcal{B}$ increases, while the bandwidth controls the degree of localization. In this setup, observations contribute isotropically according to their univariate distance $\mathcal{d}(\mathbf{X}_i, \mathbf{x})$ from the boundary point \mathbf{x} .

The estimator $\hat{\vartheta}(\mathbf{x})$ is thus the difference of two univariate (isotropic) local polynomial regression estimators evaluated along the one-dimensional manifold \mathcal{B} . Importantly, this estimator does not directly target $\tau(\mathbf{x})$. Rather, it targets the difference $\theta_{1,\mathbf{x}}(0) - \theta_{0,\mathbf{x}}(0)$, where

$$\theta_{t,\mathbf{x}}(r) = \mathbb{E}[Y_i | D_i(\mathbf{x}) = r, D_i(\mathbf{x}) \in \mathcal{J}_t] = \mathbb{E}[Y_i(t) | \mathcal{d}(\mathbf{X}_i, \mathbf{x}) = |r|], \quad (2)$$

are univariate conditional expectations induced by transforming the bivariate score through the distance function. Our objective is therefore to characterize the conditions under which valid pointwise and uniform identification, estimation, and inference for the BATEC $\tau(\mathbf{x})$ can be conducted using the distance-based estimator $\hat{\vartheta}(\mathbf{x})$ and its associated estimand $\theta_{1,\mathbf{x}}(0) - \theta_{0,\mathbf{x}}(0)$.

We begin by studying identification of $\tau(\mathbf{x})$ through distance-based approximations. The induced conditional expectations $\theta_{t,\mathbf{x}}(r)$ generally differ from the bivariate regression functions $\mu_t(\mathbf{x}) = \mathbb{E}[Y_i(t) | \mathbf{X}_i = \mathbf{x}]$. Consequently, without restrictions on the data-generating process, the geometry of the assignment boundary, and the distance function, $\lim_{r \rightarrow 0} \theta_{t,\mathbf{x}}(r)$ need not coincide with $\mu_t(\mathbf{x})$. Theorem 1 provides sufficient conditions for these parameters to agree, including restrictions ensur-

ing that \mathcal{B} is a rectifiable curve (Federer, 2014). This result contributes to the multidimensional RD literature by establishing a new identification theory for distance-based methods (cf., Hahn et al., 2001; Papay et al., 2011; Reardon and Robinson, 2012; Wong et al., 2013; Cattaneo et al., 2016).

The geometry of the boundary influences not only identification but also bias approximation. We establish two complementary results. First, Theorem 2 shows that near irregular points of the assignment boundary, such as kinks that may arise even when \mathcal{B} is a well-behaved one-dimensional manifold, the p th-order estimator $\hat{\vartheta}(\mathbf{x})$ exhibits an irreducible bias of order h , regardless of the polynomial order or the smoothness of the underlying regression functions $\mu_0(\mathbf{x})$ and $\mu_1(\mathbf{x})$. This phenomenon arises because the induced regression function $\theta_{t,\mathbf{x}}(r)$ is at most Lipschitz continuous uniformly in neighborhoods of such irregular points. Second, Theorem 3 shows that when \mathcal{B} is smooth, the usual approximation bias of order h^{p+1} from the nonparametric smoothing literature is recovered (Härdle et al., 2004). Hence, the bias of the distance-based estimator varies along the boundary according to its local geometric features.

These findings imply that standard univariate RD procedures can exhibit unexpected behavior when applied to BD designs through distance-based running variables: geometric irregularities of the assignment boundary may induce substantially larger misspecification bias. This issue is particularly relevant in geographic RD applications, where boundaries often contain kinks or other irregularities. For instance, the influential work of Mandelbrot (1967, 1983) has argued that geographic borders (and other shapes in nature) are fractals; see Avnir et al. (1998), and references therein, for a recent discussion on this ongoing debate among mathematicians and philosophers. From a practical standpoint, our geometric analysis yields concrete foundational guidance for bandwidth choice, point estimation, and statistical inference.

We next develop estimation and inference theory for $\tau(\mathbf{x})$ based on $\hat{\vartheta}(\mathbf{x})$. Theorem 4 establishes pointwise and uniform convergence rates, while Theorem 5 provides valid uncertainty quantification both pointwise and uniformly over \mathcal{B} . These results are formulated in terms of a generic misspecification bias component to accommodate the different interactions between the distance function and the boundary geometry. Under appropriate regularity conditions and when \mathcal{B} is sufficiently smooth, the estimator achieves the optimal nonparametric convergence rate pointwise and uniformly (Tsybakov, 2008). We also construct confidence intervals for $\tau(\mathbf{x})$ and confidence bands

for the entire curve $(\tau(\mathbf{x}) : \mathbf{x} \in \mathcal{B})$, leveraging a new strong approximation result for empirical processes developed in the supplemental appendix.

We also establish a minimax optimality result that highlights the fundamental limitations of distance-based methods when the assignment boundary is not smooth. Theorem 6 derives a minimax lower bound convergence rate for isotropic nonparametric estimators of a smooth bivariate regression function evaluated along a one-dimensional rectifiable manifold. We show that, irrespective of the smoothness of the underlying bivariate regression function, isotropic univariate estimators can achieve at most the uniform convergence rate $n^{-1/4}$ (up to polylogarithmic factors). This rate coincides with the optimal minimax rate for Lipschitz continuous bivariate regression functions over the same domain, thereby demonstrating that the convergence rate attained by $\hat{\vartheta}(\mathbf{x})$ is unimprovable without imposing stronger geometric restrictions on the boundary or exploiting additional structural features of the data-generating process.

Our theoretical results characterize identification, estimation, and inference for the BATEC, both pointwise and uniformly over \mathcal{B} , when this causal parameter is estimated using univariate distance-based (isotropic) local polynomial regression methods. As discussed in Section 5, these findings also inform empirical practice by motivating new “regularized” distance-based estimation and inference procedures when additional information about the geometry of the assignment boundary \mathcal{B} is available. Section 7 presents numerical evidence based on the dataset of [Londoño-Vélez et al. \(2020\)](#), illustrating the finite-sample performance of the proposed methods. These results are implemented in our companion R package `rd2d` (<https://rdpackages.github.io/rd2d>); see [Cattaneo et al. \(2025b\)](#) for further details.

Section 8 discusses several extensions of our theoretical framework, including regression kink designs, settings with imperfect treatment compliance, and adjustments for pre-intervention covariates aimed at either improving efficiency or heterogeneity analysis. In addition, Section 9 provides concrete guidance for empirical implementation based on the results developed in this paper as well as those in our ongoing companion work ([Cattaneo et al., 2026b,c](#)).

From a broader theoretical perspective, this paper establishes a new connection between the econometric methodology for BD designs and the statistical theory of nonparametric smoothing over submanifolds, thereby contributing to two distinct yet closely related literatures in econometrics, statistics, machine learning, and data science. Methodologically, we provide the first founda-

tional results for the analysis and interpretation of BD designs based on distance-based estimators, an approach widely used in applied work but lacking formal theoretical justification. From a non-parametric perspective, we develop new pointwise and uniform estimation and inference results for isotropic local polynomial regression on submanifolds, thereby advancing the theory of smoothing methods in low-dimensional geometric settings.

Our results are complementary to concurrent work by [Chen and Gao \(2026\)](#), who study estimation and inference for integral functionals on submanifolds using series (sieve) methods. In contrast, we focus on estimation based on one-dimensional distance transformations evaluated pointwise along the submanifold, analyze isotropic local polynomial smoothing procedures, and uncover new phenomena related to uniform misspecification bias. In addition, we establish a novel minimax lower bound convergence rate for isotropic nonparametric regression estimators in this setting.

1.2 Notation

We employ standard concepts and notation from empirical process theory ([van der Vaart and Wellner, 1996](#); [Giné and Nickl, 2016](#)) and geometric measure theory ([Simon et al., 1984](#); [Federer, 2014](#)). For a random variable \mathbf{V}_i , we write $\mathbb{E}_n[g(\mathbf{V}_i)] = n^{-1} \sum_{i=1}^n g(\mathbf{V}_i)$. For a vector $\mathbf{v} \in \mathbb{R}^k$, the Euclidean norm is $\|\mathbf{v}\| = (\sum_{j=1}^k \mathbf{v}_j^2)^{1/2}$. For a matrix \mathbf{A} , $\lambda_{\min}(\mathbf{A})$ denotes the smallest eigenvalue. $C^k(\mathcal{X}, \mathcal{Y})$ denotes the class of k -times continuously differentiable functions from \mathcal{X} to \mathcal{Y} , and $C^k(\mathcal{X})$ is a shorthand for $C^k(\mathcal{X}, \mathbb{R})$. For a Borel set $\mathcal{S} \subseteq \mathcal{X}$, the De Giorgi perimeter of \mathcal{S} is $\text{perim}(\mathcal{S}) = \sup_{g \in \mathcal{D}_2(\mathcal{X})} \int_{\mathbb{R}^2} \mathbf{1}(\mathbf{x} \in \mathcal{S}) \text{div } g(\mathbf{x}) d\mathbf{x} / \|g\|_{\infty}$, where div is the divergence operator, and $\mathcal{D}_2(\mathcal{X})$ denotes the space of C^∞ functions with values in \mathbb{R}^2 and with compact support included in \mathcal{X} . When \mathcal{S} is connected, and the boundary $\text{bd}(\mathcal{S})$ is a smooth simple closed curve, $\text{perim}(\mathcal{S})$ simplifies to the curve length of $\text{bd}(\mathcal{S})$. A curve $\mathcal{B} \subseteq \mathbb{R}^2$ is a *rectifiable curve* if there exists a Lipschitz map $\gamma : [0, 1] \mapsto \mathbb{R}^2$ such that $\mathcal{B} = \gamma([0, 1])$. For a function $f : \mathbb{R}^2 \mapsto \mathbb{R}$, $\text{Supp}(f)$ denotes the closure of the set $\{\mathbf{x} \in \mathbb{R}^2 : f(\mathbf{x}) \neq 0\}$. For real sequences $a_n = o(b_n)$ if $\limsup_{n \rightarrow \infty} \frac{|a_n|}{|b_n|} = 0$, $a_n \lesssim b_n$ if there exists some constant C and $N > 0$ such that $n > N$ implies $|a_n| \leq C|b_n|$. For sequences of random variables $a_n = o_{\mathbb{P}}(b_n)$ if $\text{plim}_{n \rightarrow \infty} \frac{|a_n|}{|b_n|} = 0$, $|a_n| \lesssim_{\mathbb{P}} |b_n|$ if $\limsup_{M \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}[|\frac{a_n}{b_n}| \geq M] = 0$. Finally, $\Phi(x)$ denotes the standard Gaussian cumulative distribution function.

1.3 Organization

Section 2 introduces the setup and assumptions used throughout the paper. Section 3 studies identification of $\tau(\mathbf{x})$ via distance-based approximations, while Section 4 analyzes the bias properties of the distance-based estimator $\widehat{\vartheta}(\mathbf{x})$. Section 5 develops estimation and inference procedures for $\tau(\mathbf{x})$ based on distance-based local polynomial regression. Section 6 examines the fundamental minimax estimation problem underlying Theorem 2. Section 7 illustrates the finite-sample performance of the distance-based methods using simulated data. Section 8 discusses extensions of our theoretical framework, Section 9 provides implementation guidance for empirical researchers, and Section 10 concludes.

The supplemental appendix contains generalizations of the main theoretical results, their proofs, and additional findings that may be of independent interest. In particular, it presents a new strong approximation theorem for empirical processes with multiplicative separable structure and bounded polynomial moments, extending recent work by Cattaneo and Yu (2025).

2 Setup and Assumptions

The following assumption collects the core conditions imposed on the underlying data generating process.

Assumption 1 (Data Generating Process). *Let $t \in \{0, 1\}$, $p \geq 1$, and $v \geq 2$.*

- (i) $(Y_1(t), \mathbf{X}_1^\top)^\top, \dots, (Y_n(t), \mathbf{X}_n^\top)^\top$ are independent and identically distributed random vectors.
- (ii) The distribution of \mathbf{X}_i admits a Lebesgue density $f_X(\mathbf{x})$ that is continuous and bounded away from zero on its support $\mathcal{X} = [\mathbf{L}, \mathbf{U}]^2$, for $-\infty < \mathbf{L} < \mathbf{U} < \infty$.
- (iii) $\mu_t(\mathbf{x}) = \mathbb{E}[Y_i(t) \mid \mathbf{X}_i = \mathbf{x}]$ is $(p + 1)$ times continuously differentiable on \mathcal{X} .
- (iv) $\sigma_t^2(\mathbf{x}) = \mathbb{V}[Y_i(t) \mid \mathbf{X}_i = \mathbf{x}]$ is continuous and bounded away from zero on \mathcal{X} .
- (v) $\sup_{\mathbf{x} \in \mathcal{X}} \mathbb{E}[|Y_i(t)|^{2+v} \mid \mathbf{X}_i = \mathbf{x}] < \infty$.

Assumption 1 generalizes standard regularity conditions used in univariate RD designs (see, e.g., Cattaneo and Titiunik, 2022, and references therein). Additional restrictions specific to distance-

based methods in BD designs are also required. Let

$$\Psi_{t,\mathbf{x}} = \mathbb{E} \left[\mathbf{r}_p \left(\frac{D_i(\mathbf{x})}{h} \right) \mathbf{r}_p \left(\frac{D_i(\mathbf{x})}{h} \right)^\top K_h(D_i(\mathbf{x})) \mathbf{1}(D_i(\mathbf{x}) \in \mathcal{J}_t) \right]$$

denote the fixed- h population Gram matrix associated with the distance-based estimator.

Assumption 2 (Kernel, Distance, and Boundary). *Let $t \in \{0, 1\}$.*

- (i) \mathcal{B} is a rectifiable curve.
- (ii) The distance function $\mathcal{d} : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow [0, \infty)$ satisfies $\|\mathbf{x}_1 - \mathbf{x}_2\| \lesssim \mathcal{d}(\mathbf{x}_1, \mathbf{x}_2) \lesssim \|\mathbf{x}_1 - \mathbf{x}_2\|$ for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}$.
- (iii) $K : \mathbb{R} \rightarrow [0, \infty)$ is compactly supported and Lipschitz continuous, or $K(u) = \mathbf{1}(u \in [-1, 1])$.
- (iv) $\liminf_{h \downarrow 0} \inf_{\mathbf{x} \in \mathcal{B}} \lambda_{\min}(\Psi_{t,\mathbf{x}}) \gtrsim 1$.

Assumption 2(i) imposes minimal regularity on the assignment boundary \mathcal{B} , which facilitates the evaluation of integrals and the derivation of uniform results over the one-dimensional submanifold. Assumption 2(ii) requires the distance function to be equivalent (up to constants) to the Euclidean distance. Assumption 2(iii) imposes standard conditions on the univariate kernel function. Assumption 2(iv) further restricts (implicitly) the geometry of the boundary \mathcal{B} relative to the kernel and distance functions, ruling out highly irregular shapes that would generate regions with too few observations. Lemma SA-1 in the supplemental appendix provides primitive conditions for Assumption 2(iv) when $\mathcal{d}(\cdot)$ is the Euclidean norm. The conditions imposed by Assumption 2 are sufficient for uniform (over \mathcal{B}) estimation and inference, although some can be weakened for pointwise results; see the supplemental appendix.

3 Identification and Interpretation

For each boundary point $\mathbf{x} \in \mathcal{B}$ and corresponding signed distance score $D_i(\mathbf{x})$, the univariate distance-based local polynomial estimator is $\hat{\vartheta}(\mathbf{x}) = \hat{\theta}_{1,\mathbf{x}}(0) - \hat{\theta}_{0,\mathbf{x}}(0)$, where $\hat{\theta}_{t,\mathbf{x}}(0) = \mathbf{e}_1^\top \hat{\gamma}_t(\mathbf{x}) = \mathbf{r}_p(0)^\top \hat{\gamma}_t(\mathbf{x})$. Conceptually, the estimator $\hat{\theta}_{t,\mathbf{x}}(0)$ targets the estimand $\theta_{t,\mathbf{x}}(0)$ defined in (2), which is the univariate conditional expectation induced by the distance transformation applied to the bivariate location variable \mathbf{X}_i for each point $\mathbf{x} \in \mathcal{B}$.

The following theorem establishes identification of the causal functional parameter $\tau(\mathbf{x})$ through the distance-based induced conditional expectations.

Theorem 1 (Identification). *Suppose Assumptions 1(i)–(iii) and 2(i)–(ii) hold. Then,*

$$\tau(\mathbf{x}) = \lim_{r \downarrow 0} \theta_{1,\mathbf{x}}(r) - \lim_{r \uparrow 0} \theta_{0,\mathbf{x}}(r)$$

for all $\mathbf{x} \in \mathcal{B}$.

This identification result is obtained by representing $\theta_{t,\mathbf{x}}(r)$ as a sequence of integrals over submanifolds near $\mathbf{x} \in \mathcal{B}$, that is, over level sets of shrinking spheres generated by the distance function $\mathcal{d}(\cdot, \mathbf{x})$. Without restrictions on the data generating process, the geometry of the assignment boundary, and the distance function, the limits $\lim_{r \rightarrow 0} \theta_{t,\mathbf{x}}(r)$ and $\mu_t(\mathbf{x})$ need not coincide.

The identification result in Theorem 1 is new to the literature. Related continuity-based identification arguments in RD designs include [Hahn et al. \(2001\)](#) for one-dimensional score settings; [Papay et al. \(2011\)](#), [Reardon and Robinson \(2012\)](#), and [Keele and Titiunik \(2015\)](#) for multi-score settings; and [Cattaneo et al. \(2016\)](#) for multi-cutoff designs. For example, identification of the BATEC using location-based methods takes the form

$$\tau(\mathbf{x}) = \mathbb{E}[Y_i(1) - Y_i(0) | \mathbf{X}_i = \mathbf{x}] = \lim_{\mathbf{u} \rightarrow \mathbf{x}, \mathbf{u} \in \mathcal{A}_1} \mathbb{E}[Y_i | \mathbf{X}_i = \mathbf{u}] - \lim_{\mathbf{u} \rightarrow \mathbf{x}, \mathbf{u} \in \mathcal{A}_0} \mathbb{E}[Y_i | \mathbf{X}_i = \mathbf{u}],$$

for all $\mathbf{x} \in \mathcal{B}$. Our result does not cover pooling-based methods, where identification, estimation, and inference rely on the distance to the closest boundary point and therefore aggregate treatment effects along \mathcal{B} . See [Cattaneo et al. \(2026a\)](#) for an overview; formal analysis of pooling-based methods is the subject of ongoing work ([Cattaneo et al., 2026c](#)).

The isotropic nonparametric procedures underlying distance-based methods naturally introduce approximation (misspecification) error when targeting $\tau(\mathbf{x})$. To study this issue, we interpret the estimator $\widehat{\gamma}_t(\mathbf{x})$ as the sample analogue of the coefficients from the best local mean-square approximation of the conditional expectation $\mathbb{E}[Y_i(t) | D_i(\mathbf{x})]$ using the polynomial basis $\mathbf{r}_p(D_i(\mathbf{x}))$:

$$\gamma_t^*(\mathbf{x}) = \arg \min_{\gamma \in \mathbb{R}^{p+1}} \mathbb{E} \left[(Y_i - \mathbf{r}_p(D_i(\mathbf{x}))^\top \gamma)^2 K_h(D_i(\mathbf{x})) \mathbf{1}(D_i(\mathbf{x}) \in \mathcal{J}_t) \right].$$

Letting $\theta_{t,\mathbf{x}}^*(0) = \mathbf{e}_1^\top \boldsymbol{\gamma}_t^*(\mathbf{x}) = \mathbf{r}_p(0)^\top \boldsymbol{\gamma}_t^*(\mathbf{x})$, we obtain the standard least-squares decomposition

$$\widehat{\theta}_{t,\mathbf{x}}(0) - \theta_{t,\mathbf{x}}(0) = [\theta_{t,\mathbf{x}}^*(0) - \theta_{t,\mathbf{x}}(0)] + \mathbf{e}_1^\top \boldsymbol{\Psi}_{t,\mathbf{x}}^{-1} \mathbf{O}_{t,\mathbf{x}} + [\mathbf{e}_1^\top (\widehat{\boldsymbol{\Psi}}_{t,\mathbf{x}}^{-1} - \boldsymbol{\Psi}_{t,\mathbf{x}}^{-1}) \mathbf{O}_{t,\mathbf{x}}], \quad (3)$$

where

$$\begin{aligned} \widehat{\boldsymbol{\Psi}}_{t,\mathbf{x}} &= \mathbb{E}_n \left[\mathbf{r}_p \left(\frac{D_i(\mathbf{x})}{h} \right) \mathbf{r}_p \left(\frac{D_i(\mathbf{x})}{h} \right)^\top K_h(D_i(\mathbf{x})) \mathbf{1}(D_i(\mathbf{x}) \in \mathcal{J}_t) \right], \quad \text{and} \\ \mathbf{O}_{t,\mathbf{x}} &= \mathbb{E}_n \left[\mathbf{r}_p \left(\frac{D_i(\mathbf{x})}{h} \right) K_h(D_i(\mathbf{x})) (Y_i - \theta_{t,\mathbf{x}}^*(D_i(\mathbf{x}))) \mathbf{1}(D_i(\mathbf{x}) \in \mathcal{J}_t) \right]. \end{aligned}$$

In decomposition (3), the first term represents the nonrandom approximation bias, the second term is a stochastic linear component (a sample average of mean-zero random variables), and the third term is a higher-order linearization error arising from estimation of the Gram matrix. Importantly, our analysis does not impose conditional mean-zero restrictions, allowing for general forms of local misspecification.

Theorem 1 together with decomposition (3) implies that

$$\widehat{\vartheta}(\mathbf{x}) - \tau(\mathbf{x}) = \mathfrak{B}(\mathbf{x}) + \mathfrak{L}(\mathbf{x}) + \mathfrak{Q}(\mathbf{x}), \quad \mathbf{x} \in \mathcal{B},$$

where $\mathfrak{B}(\mathbf{x}) = \theta_{1,\mathbf{x}}^*(0) - \theta_{0,\mathbf{x}}^*(0) - \tau(\mathbf{x})$ denotes the nonrandom approximation bias, $\mathfrak{L}(\mathbf{x}) = \mathbf{e}_1^\top \boldsymbol{\Psi}_{1,\mathbf{x}}^{-1} \mathbf{O}_{1,\mathbf{x}} - \mathbf{e}_1^\top \boldsymbol{\Psi}_{0,\mathbf{x}}^{-1} \mathbf{O}_{0,\mathbf{x}}$ is an unconditional mean-zero linear statistic, and $\mathfrak{Q}(\mathbf{x}) = \mathbf{e}_1^\top (\widehat{\boldsymbol{\Psi}}_{1,\mathbf{x}}^{-1} - \boldsymbol{\Psi}_{1,\mathbf{x}}^{-1}) \mathbf{O}_{1,\mathbf{x}} - \mathbf{e}_1^\top (\widehat{\boldsymbol{\Psi}}_{0,\mathbf{x}}^{-1} - \boldsymbol{\Psi}_{0,\mathbf{x}}^{-1}) \mathbf{O}_{0,\mathbf{x}}$ captures higher-order linearization effects.

In conventional local polynomial settings, $\mathfrak{B}(\mathbf{x})$ is of order h^{p+1} , $\mathfrak{L}(\mathbf{x})$ is approximately Gaussian, and $\mathfrak{Q}(\mathbf{x})$ is asymptotically negligible. In the present setting, however, these standard results need not hold because estimation is conducted along the one-dimensional boundary \mathcal{B} and relies on isotropic smoothing based on distance to each boundary point. We show that additional geometric and smoothness conditions are required to recover the usual approximation properties, and that these properties may fail in some empirically relevant configurations.

More specifically, the approximation bias $\mathfrak{B}(\mathbf{x})$ plays a central role in both pointwise and uniform inference for the BATEC. As emphasized in the RD literature (Calonico et al., 2014) and in kernel-based nonparametric inference more broadly (Calonico et al., 2018, 2022), valid large-sample

inference requires a standardized “small bias” condition of the form

$$\frac{\mathfrak{B}(\mathbf{x})}{\sqrt{\mathbb{V}[\mathfrak{L}(\mathbf{x})]}} \lesssim \sqrt{nh}\mathfrak{B}(\mathbf{x}) \rightarrow 0 \quad (4)$$

for the bandwidth sequence employed. Because bandwidths are typically chosen to minimize mean squared error (MSE), this condition may fail in general, implying that inference cannot ignore first-order misspecification bias.

Robust bias correction addresses this issue by explicitly estimating and removing the leading bias term and adjusting the variance accordingly. A common implementation proceeds in two steps: first, an MSE-optimal bandwidth and corresponding MSE-optimal point estimator are constructed using a polynomial of order p ; second, inference is conducted using a higher-order polynomial (typically $p+1$) while retaining the bandwidth from the first step. This approach yields valid confidence intervals when $\mathfrak{B}(\mathbf{x}) \lesssim h^{p+1}$. While such bias behavior is standard under smoothness conditions in canonical nonparametric settings, we show in the next section that there exist configurations in which $\mathfrak{B}(\mathbf{x}) \lesssim h$ regardless of the polynomial order. Consequently, approximation bias can substantially affect both estimation and robust bias-corrected inference when employing distance-based methods.

4 Approximation Bias

The smoothness of the induced distance-based conditional expectation function $r \mapsto \theta_{t,\mathbf{x}}(r) = \mathbb{E}[Y_i(t)|D_i(\mathbf{x}) = r]$ depends on the smoothness of the conditional expectation $\mu_t(\mathbf{x}) = \mathbb{E}[Y_i(t)|\mathbf{X}_i = \mathbf{x}]$, the distance function $\mathcal{d}(\cdot, \cdot)$, and the geometric regularity of the boundary \mathcal{B} . Consequently, the bias of the distance-based local polynomial estimator may be affected by the shape of the boundary \mathcal{B} , regardless of the polynomial order p imposed in Assumption 1.

Figure 1 illustrates this issue graphically. For a point $\mathbf{x} \in \mathcal{B}$ sufficiently close to a kink of the boundary, the conditional expectation $\theta_{1,\mathbf{x}}(r)$ may fail to be differentiable for some $r \geq 0$. The problem arises because, given the distance function $\mathcal{d}(\cdot, \cdot)$, a sufficiently small value of r yields a complete arc $\{\mathbf{x} \in \mathcal{A}_1 : \mathcal{d}(\mathbf{X}_i, \mathbf{x}) = r\}$, whereas for larger values of r this arc becomes truncated by the boundary. In the example depicted in Figure 1, the function $\theta_{1,\mathbf{x}}(r)$ is smooth for $r < r_3$

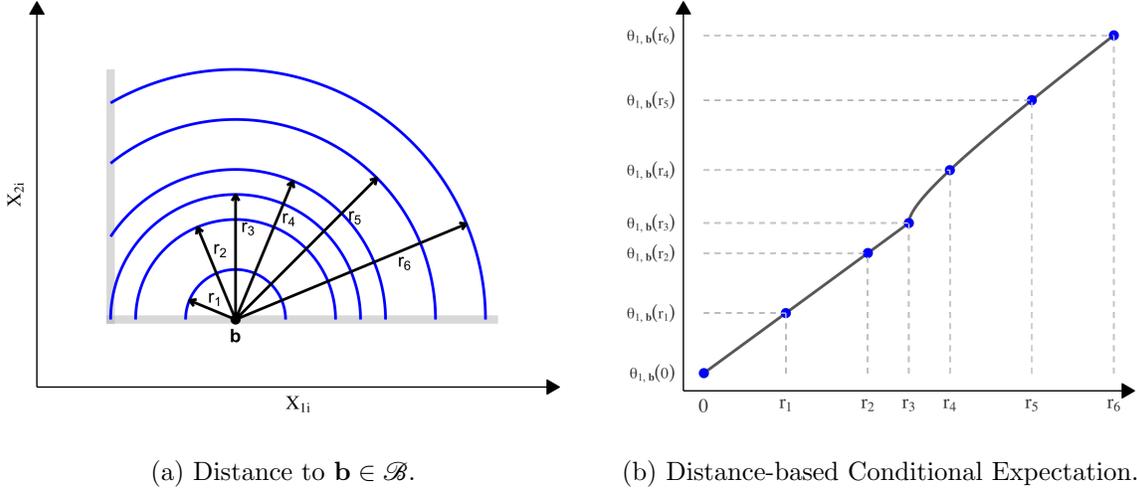


Figure 1: Lack of smoothness of distance-based conditional expectation near a kink.

Note: Analytic example of $\theta_{1,\mathbf{b}}(r) = \mathbb{E}[Y(1)|D_i(\mathbf{b}) = r]$, $r \geq 0$, for distance transformation $D_i(\mathbf{b}) = \mathcal{d}(\mathbf{X}_i, \mathbf{b}) = \|\mathbf{X}_i - \mathbf{b}\|$ to point $\mathbf{b} \in \mathcal{B}$ near a kink point on the boundary, based on location $\mathbf{X}_i = (X_{1i}, X_{2i})^\top$. The induced univariate conditional expectation $r \mapsto \theta_{1,\mathbf{b}}(r)$ is continuous but not differentiable at $r = r_3$.

and for $r > r_3$, but is not differentiable at $r = r_3$. In particular, the left derivative is finite while the right derivative diverges. Details of this analytic example are provided in the supplemental appendix.

Although the smoothness of the boundary \mathcal{B} influences the regularity of $\theta_{t,\mathbf{x}}(r)$, the locality of the distance-based estimator implies that the approximation error cannot be smaller than that of a local constant estimator, regardless of the polynomial order $p \geq 1$ or the smoothness assumptions imposed on $\mu_t(\mathbf{x})$. The following theorem formalizes this observation and shows that the resulting bias order cannot be improved by increasing p .

Theorem 2 (Approximation Bias: Uniform Guarantee). *For some $L > 0$, let \mathcal{P} be the class of data generating processes satisfying Assumptions 1(i)-(iii) with $\mathcal{X} \subseteq [-L, L]^2$, Assumption 2, and the following three additional conditions:*

- (i) $L^{-1} \leq \inf_{\mathbf{x} \in \mathcal{X}} f_X(\mathbf{x}) \leq \sup_{\mathbf{x} \in \mathcal{X}} f_X(\mathbf{x}) \leq L$,
- (ii) $\max_{0 \leq |\nu| \leq p} \sup_{\mathbf{x} \in \mathcal{X}} |\partial^\nu \mu_t(\mathbf{x})| + \max_{0 \leq |\nu| \leq p} \sup_{\mathbf{x}, \mathbf{y} \in \mathcal{X}} \frac{|\partial^\nu \mu_t(\mathbf{x}) - \partial^\nu \mu_t(\mathbf{y})|}{\|\mathbf{x} - \mathbf{y}\|} \leq L$, and
- (iii) $\liminf_{h \downarrow 0} \inf_{\mathbf{x} \in \mathcal{B}} \int_{\mathcal{A}_t} k_h(\mathcal{d}(\mathbf{u}, \mathbf{x})) d\mathbf{u} \geq L^{-1}$.

For any $p \geq 1$, if $nh^2 \rightarrow \infty$ and $h \rightarrow 0$, then

$$1 \lesssim \liminf_{n \rightarrow \infty} \sup_{\mathbb{P} \in \mathcal{P}} \sup_{\mathbf{x} \in \mathcal{B}} \frac{\mathfrak{B}(\mathbf{x})}{h} \leq \limsup_{n \rightarrow \infty} \sup_{\mathbb{P} \in \mathcal{P}} \sup_{\mathbf{x} \in \mathcal{B}} \frac{\mathfrak{B}(\mathbf{x})}{h} \lesssim 1.$$

This result precisely characterizes the uniform (over \mathcal{B} and the data generating process) approximation bias of the distance-based local polynomial estimator $\widehat{\vartheta}(\mathbf{x})$. The upper bound follows from the fact that, in general, $|\theta_{t,\mathbf{x}}(0) - \theta_{t,\mathbf{x}}(r)| \lesssim r$ for $t \in 0, 1$. The lower bound is established using the following example. Suppose $\mathbf{X}_i \sim \text{Uniform}([-2, 2]^2)$, $\mu_0(x_1, x_2) = 0$, $\mu_1(x_1, x_2) = x_2$ for all $(x_1, x_2) \in [-2, 2]^2$, and $Y_i(0)|\mathbf{X}_i \sim \text{Normal}(\mu_0(\mathbf{X}_i), 1)$ and $Y_i(1)|\mathbf{X}_i \sim \text{Normal}(\mu_1(\mathbf{X}_i), 1)$. Let $\mathcal{d}(\cdot, \cdot)$ be the Euclidean distance, and define the treatment and control regions as $\mathcal{A}_1 = \{(x, y) \in \mathbb{R}^2 : x \leq 0, y \geq 0\}$ and $\mathcal{A}_0 = \mathbb{R}^2 / \mathcal{A}_1$. The resulting assignment boundary $\mathcal{B} = \{(x, y) \in \mathbb{R}^2 : -2 \leq x \leq 0, y = 0 \text{ or } x = 0, 0 \leq y \leq 2\}$ is L-shaped, with a 90° kink at $\mathbf{x} = (0, 0)$, similar to the empirical application considered in Section 7. This data generating process satisfies the conditions of Theorem 2. The supplemental appendix establishes the lower bound through a detailed analysis of the induced approximation bias.

As a point of contrast, one might expect that $\mathfrak{B}(\mathbf{x}) \lesssim h^{p+1}$ pointwise in $\mathbf{x} \in \mathcal{B}$ for sufficiently small bandwidth h , provided that kinks on the boundary are sufficiently separated relative to the bandwidth and other regularity conditions hold. However, Theorem 2 shows that, regardless of sample size (and hence bandwidth), there always exists a neighborhood of a boundary kink where the misspecification bias of $\widehat{\vartheta}(\mathbf{x})$ is at best of order h , independently of the polynomial order p . This phenomenon arises because nonsmooth changes in the geometry of \mathcal{B} induce nondifferentiability in $\theta_{t,\mathbf{x}}(r)$, as illustrated in Figure 1b.

When the boundary \mathcal{B} is sufficiently smooth, sharper bias bounds can be obtained.

Theorem 3 (Approximation Bias: Smooth Boundary). *Suppose Assumption 1(i)-(iii) and Assumptions 2 hold, with $\mathcal{d}(\cdot, \cdot)$ the Euclidean distance. Let $h \rightarrow 0$.*

- (i) *For $\mathbf{x} \in \mathcal{B}$, and for some $\delta, \varepsilon > 0$, suppose that $\mathcal{B} \cap \{\mathbf{y} : \|\mathbf{y} - \mathbf{x}\| \leq \varepsilon\} = \gamma([- \delta, \delta])$, where $\gamma : \mathbb{R} \rightarrow \mathbb{R}^2$ is a one-to-one function in $C^{\kappa+2}([- \delta, \delta], \mathbb{R}^2)$. Then, $\theta_{0,\mathbf{x}}(\cdot)$ and $\theta_{1,\mathbf{x}}(\cdot)$ are $(\kappa \wedge (p+1))$ -times continuously differentiable on $[0, \varepsilon]$. Therefore, there exists a positive constant C such that $|\mathfrak{B}(\mathbf{x})| \leq Ch^{\kappa \wedge (p+1)}$ for all $h \in [0, \varepsilon]$ and $\mathbf{x} \in \mathcal{B}$.*
- (ii) *Suppose $\mathcal{B} = \gamma([0, L])$ where γ is a one-to-one function in $C^{\iota+2}([0, L], \mathbb{R}^2)$ for some $L > 0$. Suppose there exists $\delta, \varepsilon > 0$ such that for all $\mathbf{x} \in \mathcal{B}^\circ = \gamma([\delta, L - \delta])$, $r \in [0, \varepsilon]$, and $t = 0, 1$, and the set $\{\mathbf{u} \in \mathbb{R}^2 : \mathcal{d}(\mathbf{u}, \mathbf{x}) = r\}$ intersects with $\text{bd}(\mathcal{A}_t)$ at only two points in \mathcal{B} . Then, for all $\mathbf{x} \in \gamma([\delta, L - \delta])$, $\theta_{0,\mathbf{x}}(\cdot)$ and $\theta_{1,\mathbf{x}}(\cdot)$ are $(\iota \wedge (p+1))$ -times continuously differentiable on*

$[0, \varepsilon]$, and $\lim_{r \downarrow 0} \frac{d^v}{dr^v} \theta_{t, \mathbf{x}}(r)$ exists and is finite for all $0 \leq v \leq p+1$ and $t \in \{0, 1\}$. Therefore, there exists a positive constant C such that $\sup_{\mathbf{x} \in \mathcal{B}^\circ} |\mathfrak{B}(\mathbf{x})| \leq Ch^{\nu \wedge (p+1)}$ for all $h \in [0, r]$.

This theorem provides sufficient smoothness conditions on the boundary \mathcal{B} under which the distance-based estimator $\widehat{\vartheta}(\mathbf{x})$ attains the usual nonparametric smoothing bias rate, improving upon the minimal guarantee established in Theorem 2. Achieving a uniform bias bound requires that the boundary be uniformly smooth in the sense that it admits a global smooth parameterization. This condition is essential because, as demonstrated by Theorem 2, even a single kink in an otherwise smooth boundary can deteriorate the convergence rate of the misspecification bias; see Figure 1.

5 Estimation and Inference

The results in this section are established for a general misspecification bias $\mathfrak{B}(\mathbf{x})$ of the distance-based estimator $\widehat{\vartheta}(\mathbf{x})$, which depends on the properties of the kernel and distance functions as well as on the geometry (smoothness) of the assignment boundary \mathcal{B} , as demonstrated by Theorems 2 and 3. We also discuss the implications of these results for empirical implementation and compare them with standard methods for univariate RD designs.

5.1 Convergence Rates

Using technical results established in the supplemental appendix, we obtain the following convergence rates for the univariate distance-based local polynomial treatment effect estimator.

Theorem 4 (Convergence Rates). *Suppose Assumptions 1 and 2 hold. If $nh^2/\log(1/h) \rightarrow \infty$ and $h \rightarrow 0$, then*

- (i) $|\widehat{\vartheta}(\mathbf{x}) - \tau(\mathbf{x})| \lesssim_{\mathbb{P}} \frac{1}{\sqrt{nh^2}} + \frac{1}{n^{\frac{1+v}{2+v}} h^2} + |\mathfrak{B}(\mathbf{x})|$ for $\mathbf{x} \in \mathcal{B}$, and
- (ii) $\sup_{\mathbf{x} \in \mathcal{B}} |\widehat{\vartheta}(\mathbf{x}) - \tau(\mathbf{x})| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^2}} + \frac{\log(1/h)}{n^{\frac{1+v}{2+v}} h^2} + \sup_{\mathbf{x} \in \mathcal{B}} |\mathfrak{B}(\mathbf{x})|$.

This theorem establishes pointwise (for each $\mathbf{x} \in \mathcal{B}$) and uniform (over \mathcal{B}) convergence rates for the distance-based treatment effect estimator. In particular, by Theorem 2, $\widehat{\vartheta}(\mathbf{x}) \rightarrow_{\mathbb{P}} \tau(\mathbf{x})$ both pointwise and uniformly whenever $h \rightarrow 0$. Furthermore, by Theorem 3(i), we obtain the pointwise bound $|\mathfrak{B}(\mathbf{x})| \lesssim h^{p+1}$ for sufficiently small h when the boundary \mathcal{B} satisfies appropriate

smoothness conditions. Similarly, Theorem 3(ii) shows that the uniform bias rate of $\widehat{\vartheta}(\mathbf{x})$ improves as the assignment boundary becomes smoother.

The “variance” component of the distance-based estimator $\widehat{\vartheta}(\mathbf{x})$ is of order $(nh^2)^{-1}$, even though the procedure is formally a univariate local polynomial estimator. A naïve analogy with standard univariate nonparametric regression would instead suggest a rate of order $(nh)^{-1}$. This difference is expected because the target estimand involves differences of bivariate regression functions, and the isotropic distance-based approach cannot circumvent the curse of dimensionality. See also Section 6 for further discussion.

Combining Theorems 2 and 4(ii), and for an appropriate bandwidth choice, the estimator $\widehat{\vartheta}(\mathbf{x})$ can attain the uniform convergence rate $(n/\log n)^{-1/4}$. Section 6 shows that, over the class of rectifiable boundaries, no isotropic nonparametric estimator can achieve a uniform convergence rate faster than $n^{-1/4}$, which coincides with the minimax mean square rate for estimating Lipschitz continuous bivariate regression functions (see, e.g., [Tsybakov, 2008](#), and references therein). Consequently, the distance-based estimator $\widehat{\vartheta}(\mathbf{x})$ is minimax optimal (up to logarithmic factors) for suitable bandwidth sequences.

Finally, valid pointwise, or integrated (over \mathcal{B}), MSE convergence rates matching those in Theorem 4(i) can also be established. Details are provided in the supplemental appendix.

5.2 Uncertainty Quantification

To develop companion pointwise and uniform inference procedures along the treatment-assignment boundary \mathcal{B} , consider the feasible t -statistic for a given bandwidth choice and each boundary point $\mathbf{x} \in \mathcal{B}$:

$$\widehat{\mathbf{T}}(\mathbf{x}) = \frac{\widehat{\vartheta}(\mathbf{x}) - \tau(\mathbf{x})}{\sqrt{\widehat{\Xi}_{\mathbf{x},\mathbf{x}}}},$$

where, using standard least squares algebra, for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{B}$ and $t \in \{0, 1\}$,

$$\widehat{\Xi}_{\mathbf{x}_1, \mathbf{x}_2} = \widehat{\Xi}_{0, \mathbf{x}_1, \mathbf{x}_2} + \widehat{\Xi}_{1, \mathbf{x}_1, \mathbf{x}_2}, \quad \widehat{\Xi}_{t, \mathbf{x}_1, \mathbf{x}_2} = \frac{1}{nh^2} \mathbf{e}_1^\top \widehat{\Psi}_{t, \mathbf{x}_1}^{-1} \widehat{\Upsilon}_{t, \mathbf{x}_1, \mathbf{x}_2} \widehat{\Psi}_{t, \mathbf{x}_2}^{-1} \mathbf{e}_1,$$

and

$$\begin{aligned} \widehat{\mathbf{Y}}_{t, \mathbf{x}_1, \mathbf{x}_2} = & h^2 \mathbb{E}_n \left[\mathbf{r}_p \left(\frac{D_i(\mathbf{x}_1)}{h} \right) k_h(D_i(\mathbf{x}_1)) (Y_i - \mathbf{r}_p(D_i(\mathbf{x}_1))^\top \widehat{\boldsymbol{\gamma}}_t(\mathbf{x}_1)) \mathbf{1}(D_i(\mathbf{x}_1) \in \mathcal{J}_t) \right. \\ & \left. \times \mathbf{r}_p \left(\frac{D_i(\mathbf{x}_2)}{h} \right)^\top k_h(D_i(\mathbf{x}_2)) (Y_i - \mathbf{r}_p(D_i(\mathbf{x}_2))^\top \widehat{\boldsymbol{\gamma}}_t(\mathbf{x}_2)) \mathbf{1}(D_i(\mathbf{x}_2) \in \mathcal{J}_t) \right]. \end{aligned}$$

Accordingly, feasible confidence intervals and confidence bands over \mathcal{B} take the form

$$\widehat{\mathbf{I}}_\alpha(\mathbf{x}) = \left[\widehat{\boldsymbol{\vartheta}}(\mathbf{x}) - \boldsymbol{q}_\alpha \sqrt{\widehat{\boldsymbol{\Xi}}_{\mathbf{x}, \mathbf{x}}}, \widehat{\boldsymbol{\vartheta}}(\mathbf{x}) + \boldsymbol{q}_\alpha \sqrt{\widehat{\boldsymbol{\Xi}}_{\mathbf{x}, \mathbf{x}}} \right], \quad \mathbf{x} \in \mathcal{B},$$

for any $\alpha \in (0, 1)$, where \boldsymbol{q}_α denotes the appropriate critical value depending on the inference procedure.

For pointwise inference, we show in the supplemental appendix that $\sup_{t \in \mathbb{R}} |\mathbb{P}[\widehat{\mathbf{T}}(\mathbf{x}) \leq t] - \Phi(t)| \rightarrow 0$ for each $\mathbf{x} \in \mathcal{B}$, under standard regularity conditions and provided that $nh^2 \mathfrak{B}^2(\mathbf{x}) \rightarrow 0$. In this case, $\boldsymbol{q}_\alpha = \Phi^{-1}(1 - \alpha/2)$ is asymptotically valid.

Uniform inference over \mathcal{B} is more challenging. The stochastic process $(\widehat{\mathbf{T}}(\mathbf{x}) : \mathbf{x} \in \mathcal{B})$ is generally not asymptotically tight and therefore does not converge weakly in the space of uniformly bounded real-valued functions on \mathcal{B} equipped with the supremum norm (van der Vaart and Wellner, 1996; Giné and Nickl, 2016). Consequently, classical empirical process arguments based on weak convergence cannot be applied directly. Moreover, the geometry of the manifold \mathcal{B} and the reliance on signed distance scores introduce additional technical complications. We address these issues by first establishing a novel strong approximation result for the entire stochastic process $(\widehat{\mathbf{T}}(\mathbf{x}) : \mathbf{x} \in \mathcal{B})$ under the usual ‘‘small bias’’ condition $nh^2 \sup_{\mathbf{x} \in \mathcal{B}} \mathfrak{B}^2(\mathbf{x}) \rightarrow 0$. This approximation allows uniform inference based on critical values \boldsymbol{q}_α satisfying

$$\mathbb{P}[\tau(\mathbf{x}) \in \widehat{\mathbf{I}}_\alpha(\mathbf{x}), \text{ for all } \mathbf{x} \in \mathcal{B}] = \mathbb{P} \left[\sup_{\mathbf{x} \in \mathcal{B}} |\widehat{\mathbf{T}}(\mathbf{x})| \leq \boldsymbol{q}_\alpha \right].$$

Our technical analysis combines a new strong approximation theorem (Theorem SA-8 in the supplemental appendix) with recent results from Chernozhukov et al. (2014a,b) and Chernozhukov et al. (2022), thereby improving upon classical coupling approaches such as Yurinskii’s coupling (see also Cattaneo et al., 2025a). Let \mathcal{U}_n denote the σ -algebra generated by $((Y_i, (D_i(\mathbf{x}) : \mathbf{x} \in \mathcal{B})) :$

$1 \leq i \leq n$).

Theorem 5 (Statistical Inference). *Suppose Assumptions 1 and 2 hold.*

(i) *For all $\mathbf{x} \in \mathcal{B}$, if $n^{\frac{v}{2+v}}h^2 \rightarrow \infty$ and $nh^2\mathfrak{B}^2(\mathbf{x}) \rightarrow 0$, then*

$$\mathbb{P}[\tau(\mathbf{x}) \in \hat{\mathbb{I}}_\alpha(\mathbf{x})] \rightarrow 1 - \alpha,$$

for $q_\alpha = \Phi^{-1}(1 - \alpha/2)$.

(ii) *If $n^{\frac{v}{2+v}}h^2/\log n \rightarrow \infty$, $\liminf_{n \rightarrow \infty} \frac{\log h}{\log n} > -\infty$, $nh^2 \sup_{\mathbf{x} \in \mathcal{B}} \mathfrak{B}^2(\mathbf{x}) \rightarrow 0$, and $\text{perim}(\{\mathbf{y} \in \mathcal{A}_t : \mathcal{d}(\mathbf{y}, \mathbf{x})/h \in \text{Supp}(k)\}) \lesssim h$ for all $\mathbf{x} \in \mathcal{B}$ and $t \in \{0, 1\}$, then*

$$\mathbb{P}[\tau(\mathbf{x}) \in \hat{\mathbb{I}}_\alpha(\mathbf{x}), \text{ for all } \mathbf{x} \in \mathcal{B}] \rightarrow 1 - \alpha,$$

for $q_\alpha = \inf\{c > 0 : \mathbb{P}[\sup_{\mathbf{x} \in \mathcal{B}} |\hat{Z}_n(\mathbf{x})| \geq c|\mathcal{U}_n] \leq \alpha\}$, where $(\hat{Z}_n : \mathbf{x} \in \mathcal{B})$ is a Gaussian process conditional on \mathcal{U}_n , with $\mathbb{E}[\hat{Z}_n(\mathbf{x}_1)|\mathcal{U}_n] = 0$ and $\mathbb{E}[\hat{Z}_n(\mathbf{x}_1)\hat{Z}_n(\mathbf{x}_2)|\mathcal{U}_n] = \hat{\Xi}_{\mathbf{x}_1, \mathbf{x}_2} / \sqrt{\hat{\Xi}_{\mathbf{x}_1, \mathbf{x}_1}\hat{\Xi}_{\mathbf{x}_2, \mathbf{x}_2}}$, for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{B}$.

This theorem establishes asymptotically valid inference procedures based on the distance-based local polynomial treatment-effect estimator $\hat{\vartheta}(\mathbf{x})$. For uniform inference, an additional geometric restriction on the assignment boundary \mathcal{B} is required. Specifically, the De Giorgi perimeter condition can be verified when the boundary of the set $\mathbf{y} \in \mathcal{A}_t : \mathcal{d}(\mathbf{y}, \mathbf{x})/h \in \text{Supp}(k)$ has length of order h . Because this set is contained in an h -ball centered at \mathbf{x} , the condition holds provided that the local segment $\mathcal{B} \cap \{\mathbf{y} \in \mathbb{R}^2 : \mathcal{d}(\mathbf{y}, \mathbf{x}) \leq h\}$ is not excessively irregular. Intuitively, the one-dimensional boundary \mathcal{B} must not be locally “too long” or highly oscillatory.

5.3 Discussion and Implementation

Although the distance-based estimator $\hat{\vartheta}(\mathbf{x})$ resembles a univariate local polynomial procedure based on the scalar running variable $D_i(\mathbf{x})$, Theorem 4 shows that its pointwise and uniform variance convergence rates coincide with those of a bivariate nonparametric estimator and are therefore unimprovable. Theorem 5 further shows that inference procedures developed for univariate local polynomial regression can be directly applied in distance-based settings, provided the required

side conditions are satisfied. This result follows because $\widehat{\mathbf{T}}(\mathbf{x})$ is constructed as a self-normalizing statistic and is therefore *adaptive* to the fact that the scalar running variable $D_i(\mathbf{x})$ is induced by the underlying bivariate covariate \mathbf{X}_i . This finding highlights another advantage of pre-asymptotic variance estimation and self-normalization techniques for distributional approximation and inference (Calonico et al., 2018). Consequently, standard estimation and inference methods from the univariate RD design literature (Calonico et al., 2014) can be employed in BD designs based on distance-to-boundary running variables, provided that bandwidth choices ensure that the bias induced by the geometry of the assignment boundary \mathcal{B} (documented in Section 4) is sufficiently small.

For implementation, first consider the case in which $\mathfrak{B}(\mathbf{x}) \lesssim h^{p+1}$, that is, the assignment boundary \mathcal{B} is smooth (in the sense of Theorem 3) or the bias is otherwise negligible. Establishing an exact MSE expansion for $\widehat{\vartheta}(\mathbf{x})$ is cumbersome due to the additional complexity introduced by the distance transformation, but the relevant convergence rates follow from Theorem 4. The *incorrect* univariate MSE-optimal bandwidth is $h_{1d,\mathbf{x}} \asymp n^{-1/(3+2p)}$, whereas the *correct* MSE-optimal bandwidth is $h_{\text{MSE},\mathbf{x}} \asymp n^{-1/(4+2p)}$. Treating the distance variable as an intrinsically univariate score therefore leads to a smaller-than-optimal bandwidth choice because $n^{-1/(3+2p)} < n^{-1/(4+2p)}$, resulting in undersmoothing relative to the correct MSE-optimal bandwidth. As a consequence, the point estimator exhibits higher variance and lower bias, and associated inference procedures tend to be conservative. A simple remedy is to rescale the incorrect univariate bandwidth by the factor $n^{1/(3+2p)-1/(4+2p)}$, although this adjustment may be unnecessary when empirical bandwidth selection employs pre-asymptotic variance estimation, as implemented in the software package `rdrobust` (<https://rdpackages.github.io/rdrobust/>); see Calonico et al. (2020) and references therein.

When the assignment boundary \mathcal{B} exhibits kinks or other geometric irregularities, there always exists a neighborhood around each kink where the bias is irreducibly of order h , regardless of sample size or polynomial order, as established by Theorem 2. In this case, the correct local MSE-optimal bandwidth near each kink is $h \asymp n^{-1/4}$. Away from kink points, the MSE-optimal bandwidth behaves as described above. This lack of smoothness leads to spatially varying optimal bandwidths along \mathcal{B} , complicating automatic implementation. A simple global strategy is therefore to set $h = \widehat{\mathbf{C}}, n^{-1/4}$ for all $\mathbf{x} \in \mathcal{B}$, where $\widehat{\mathbf{C}}$ is a rule-of-thumb constant. Although generically suboptimal, this choice is never larger than the pointwise MSE-optimal bandwidth and therefore

yields a more variable (but conservative) estimator and associated inference procedure. A more adaptive alternative is to allow bandwidths to vary with proximity to kink points, for example by defining

$$\hat{h}_{\text{kink},\mathbf{x}}(\mathcal{B}) = \min \left\{ \hat{h}_{\text{MSE},\mathbf{x}}, \max \left\{ \hat{\mathbf{C}} \cdot n^{-1/4}, \min_{\mathbf{b} \in \mathcal{B}_{\text{kink}}} d(\mathbf{x}, \mathbf{b}) \right\} \right\}.$$

for all $\mathbf{x} \in \mathcal{B}$, where $\mathcal{B}_{\text{kink}}$ denotes the set of kink points on the boundary.

Given a bandwidth choice, valid statistical inference can be achieved by appropriately controlling the remaining misspecification bias. When the boundary \mathcal{B} is smooth, robust bias-correction methods developed for univariate RD designs remain applicable in the distance-based setting (Calonico et al., 2014, 2018, 2022). When \mathcal{B} contains kinks, however, undersmoothing relative to the MSE-optimal bandwidth for $p = 0$ becomes necessary because increasing the polynomial order does not guarantee a reduction in misspecification bias, rendering bias-correction techniques ineffective uniformly over $\mathbf{x} \in \mathcal{B}$. In such cases, bandwidth and polynomial order choices should account for proximity to boundary irregularities.

The companion software package `rd2d` implements a rule-of-thumb bandwidth selector targeting $h \asymp n^{-1/4}$ as the default choice for all $\mathbf{x} \in \mathcal{B}$ when the smoothness of \mathcal{B} is unknown. This choice is valid for point estimation regardless of whether the boundary is smooth. For inference, following Calonico et al. (2018), the package employs an undersmoothed bandwidth of order $n^{-1/3}$ to target coverage-error optimality. When \mathcal{B} is known to be smooth, the package instead uses a rule-of-thumb selector targeting $h \asymp n^{-1/(4+2p)}$ for point estimation and implements robust bias-corrected inference based on that bandwidth. If the locations of kink points are known, the adaptive selector $\hat{h}_{\text{kink},\mathbf{x}}(\mathcal{B})$ can also be employed. Section 7 illustrates the performance of these approaches and compares them with simple rule-of-thumb bandwidth choices obtained using the `rdrobust` package with $p = 0$. Further implementation details and simulation evidence are provided in Cattaneo et al. (2025b).

Finally, uniform inference based on Theorem 5(ii) is implemented by discretizing the boundary at evaluation points $\mathbf{b}_1, \dots, \mathbf{b}_M \in \mathcal{B}$. The conditional Gaussian process $(\hat{Z}_n(\mathbf{x}) : \mathbf{x} \in \mathcal{B})$ is then approximated by the M -dimensional (conditional) Gaussian vector $\hat{\mathbf{Z}}_n = (\hat{Z}_n(\mathbf{b}_1), \dots, \hat{Z}_n(\mathbf{b}_M))$ whose covariance matrix has generic element $\mathbb{E}[\hat{Z}_n(\mathbf{b}_1)\hat{Z}_n(\mathbf{b}_2)|\mathcal{U}]$. In finite samples this estimated

covariance matrix may fail to be positive definite, in which case simple regularization can be applied (see, e.g., Cattaneo et al., 2024a). Critical values q_α are then obtained by simulating the distribution of $\max_{1 \leq l \leq M} |\widehat{Z}n(\mathbf{b}l)|$.

6 Distance-based Minimax Convergence Rate

Theorems 2 and 3 provide precise conditions characterizing the pointwise and uniform (over \mathcal{B}) bias and convergence rates of the distance-based local polynomial estimator $\widehat{\vartheta}(\mathbf{x})$. In particular, the estimator can attain at best the uniform convergence rate $(n/\log n)^{-1/4}$ for an appropriate bandwidth sequence h . A natural follow-up theoretical (and potentially practical) question is whether alternative distance-based estimators could achieve faster rates. This section provides one answer: if the boundary \mathcal{B} is rectifiable, then no isotropic nonparametric estimator of a bivariate regression function can improve upon this rate, up to polylogarithmic factors, regardless of the degree of smoothness imposed on the underlying conditional expectation function.

This section is largely self-contained, as it concerns minimax theory for nonparametric estimation on curves (see, e.g., Tsybakov, 2008, and references therein). The following theorem establishes a minimax lower bound convergence rate for estimation of compactly supported bivariate regression functions using a class of isotropic nonparametric estimators.

Theorem 6 (Distance-based Minimax Convergence Rate). *For constants $q \geq 1$ and $L > 0$, let $\mathcal{P}_{\text{NP}} = \mathcal{P}_{\text{NP}}(L, q)$ be the class of (joint) probability laws \mathbb{P} of $(Y_1, \mathbf{X}_1), \dots, (Y_n, \mathbf{X}_n)$ satisfying the following:*

- (i) $((Y_i, \mathbf{X}_i) : 1 \leq i \leq n)$ are i.i.d. taking values in $\mathbb{R} \times \mathbb{R}^2$.
- (ii) \mathbf{X}_i admits a Lebesgue density f that is continuous on its compact support $\mathcal{X} \subseteq [-L, L]^2$, with $L^{-1} \leq \inf_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}) \leq \sup_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}) \leq L$, and $\mathcal{B} = \text{bd}(\mathcal{X})$ is a rectifiable curve.
- (iii) $\mu(\mathbf{x}) = \mathbb{E}[Y_i | \mathbf{X}_i = \mathbf{x}]$ is q -times continuously differentiable on \mathcal{X} with

$$\max_{0 \leq |\nu| \leq [q]} \sup_{\mathbf{x} \in \mathcal{X}} |\partial^\nu \mu(\mathbf{x})| + \max_{|\nu|=q} \sup_{\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}} \frac{|\partial^\nu \mu(\mathbf{x}_1) - \partial^\nu \mu(\mathbf{x}_2)|}{\|\mathbf{x}_1 - \mathbf{x}_2\|^{q-[\nu]}} \leq L.$$

- (iv) $\sigma^2(\mathbf{x}) = \mathbb{V}[Y_i | \mathbf{X}_i = \mathbf{x}]$ is continuous on \mathcal{X} with $L^{-1} \leq \inf_{\mathbf{x} \in \mathcal{X}} \sigma^2(\mathbf{x}) \leq \sup_{\mathbf{x} \in \mathcal{X}} \sigma^2(\mathbf{x}) \leq L$.

In addition, let \mathcal{T} be the class of all distance-based estimators $T_n(\mathbf{U}_n(\mathbf{x}))$ with $\mathbf{U}_n(\mathbf{x}) = ((Y_i, \|\mathbf{X}_i -$

$\mathbf{x}\|)^{\top} : 1 \leq i \leq n)$ for each $\mathbf{x} \in \mathcal{X}$. Then,

$$\liminf_{n \rightarrow \infty} n^{1/4} \inf_{T_n \in \mathcal{T}} \sup_{\mathbb{P} \in \mathcal{P}_{\text{NP}}} \mathbb{E}_{\mathbb{P}} \left[\sup_{\mathbf{x} \in \mathcal{B}} |T_n(\mathbf{U}_n(\mathbf{x})) - \mu(\mathbf{x})| \right] \gtrsim 1,$$

where $\mathbb{E}_{\mathbb{P}}[\cdot]$ denotes an expectation taken under the data generating process \mathbb{P} .

In sharp contrast, under the same assumptions the uniform minimax convergence rate over the unrestricted class of nonparametric estimators is

$$\liminf_{n \rightarrow \infty} \left(\frac{n}{\log n} \right)^{\frac{q}{2q+2}} \inf_{S_n \in \mathcal{S}} \sup_{\mathbb{P} \in \mathcal{P}_{\text{NP}}} \mathbb{E}_{\mathbb{P}} \left[\sup_{\mathbf{x} \in \mathcal{B}} |S_n(\mathbf{x}; \mathbf{W}_n) - \mu(\mathbf{x})| \right] \gtrsim 1,$$

where \mathcal{S} denotes the unrestricted class of estimators based on $\mathbf{W}_n = ((Y_i, \mathbf{X}_i^{\top})^{\top} : 1 \leq i \leq n)$. Therefore, Theorem 6 shows that restricting attention to estimators that use only the scalar distance $\|\mathbf{X}_i - \mathbf{x}\|$ (rather than the full covariate \mathbf{X}_i) necessarily limits the attainable uniform estimation accuracy along the boundary. In particular, for any such estimator there exists a data-generating process for which the maximal estimation error along \mathcal{B} is of order at least $n^{-1/4}$ when the boundary may contain countably many kinks. Notably, increasing the smoothness of the underlying regression function $\mu(\mathbf{x})$ does not improve the achievable uniform convergence rate within the class \mathcal{T} , because nonsmooth boundary geometry induces effective nondifferentiability in the induced regression problem.

Finally, these results can be connected to the analysis of BD designs. For expositional clarity, consider estimation of a single regression function rather than a treatment-effect difference. By Theorems 4 and 2, the distance-based p -th order local polynomial estimator

$$\hat{\mu}(\mathbf{x}) = \mathbf{e}_1^{\top} \hat{\boldsymbol{\gamma}}(\mathbf{x}), \quad \hat{\boldsymbol{\gamma}}(\mathbf{x}) = \arg \min_{\boldsymbol{\gamma} \in \mathbb{R}^{p+1}} \mathbb{E}_n \left[(Y_i - \mathbf{r}_p(D_i(\mathbf{x}))^{\top} \boldsymbol{\gamma})^2 k_h(D_i(\mathbf{x})) \right], \quad \mathbf{x} \in \mathcal{B},$$

satisfies

$$\limsup_{M \rightarrow \infty} \limsup_{n \rightarrow \infty} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left[\left(\frac{n}{\log n} \right)^{1/4} \sup_{\mathbf{x} \in \mathcal{B}} |\hat{\mu}(\mathbf{x}) - \mu(\mathbf{x})| \geq M \right] = 0.$$

Since $\hat{\mu}(\mathbf{x}) \in \mathcal{T}$, the distance-based local polynomial estimator is minimax optimal in the sense of Theorem 6, up to the logarithmic factor $\log^{1/4} n$, which we conjecture to be unimprovable. More

precisely, this factor arises from the use of the uniform norm, whereas the lower bound in Theorem 6 is derived using pointwise minimax arguments.

7 Numerical Results

We study the numerical performance of the distance-based methods using the real-world dataset analyzed by [Londoño-Vélez et al. \(2020\)](#), which evaluates the Colombian post-secondary education subsidy program *Ser Pilo Paga* (SPP). This anti-poverty policy provided tuition support to undergraduate students admitted to high-quality, government-certified higher education institutions. Eligibility for SPP was determined by a deterministic bivariate cutoff combining academic merit and economic need: students were required to obtain a SABER 11 high school exit exam score in the top 9 percent and to have a SISBEN wealth index below a region-specific threshold. The resulting treatment assignment rule induces an L-shaped boundary \mathcal{B} with a kink at the lowest joint levels of eligibility in both score dimensions (and two additional kinks at the limits of the score support), while remaining linear elsewhere.

The dataset contains $n = 363,096$ complete observations for the first program cohort (2014). Each observation $i = 1, \dots, n$ corresponds to an individual student with bivariate score $\mathbf{X}_i = (X_{1i}, X_{2i})^\top = (\text{SABER11}_i, \text{SISBEN}_i)^\top$, where the SABER11 test score ranges from -310 to 172 and the SISBEN wealth index ranges from -103.41 to 127.21 . Without loss of generality, both score components are recentered at their respective eligibility cutoffs and standardized to facilitate the use of a common distance function and scalar bandwidth parameter. Under this normalization, the treatment assignment boundary can be written as $\mathcal{B} = \{(\text{SABER11}, \text{SISBEN}) : (\text{SABER11} \geq 0, \text{SISBEN} = 0)\} \cup \{(\text{SABER11} = 0, \text{SISBEN} \geq 0)\}$. Because the support of \mathcal{X} is compact, the boundary \mathcal{B} contains three kinks: the interior point $(0, 0)$, which is of primary interest in our analysis, and two additional points where \mathcal{B} intersects the boundary of the support of \mathcal{X} . Our empirical analysis considers 21 evenly spaced cutoff points $\{\mathbf{b}_1, \dots, \mathbf{b}_{21}\} \subset \mathcal{B}$, located near the interior kink and away from the boundary of \mathcal{X} , in order to highlight the impact of a single kink on the bias of the distance-based estimator. Throughout, we employ the Euclidean distance, $D_i(\mathbf{x}) = \|\mathbf{X}_i - \mathbf{x}\|$ for each $\mathbf{x} \in \mathcal{B}$.

We analyze both simulated data calibrated to match salient features of the SPP dataset and the

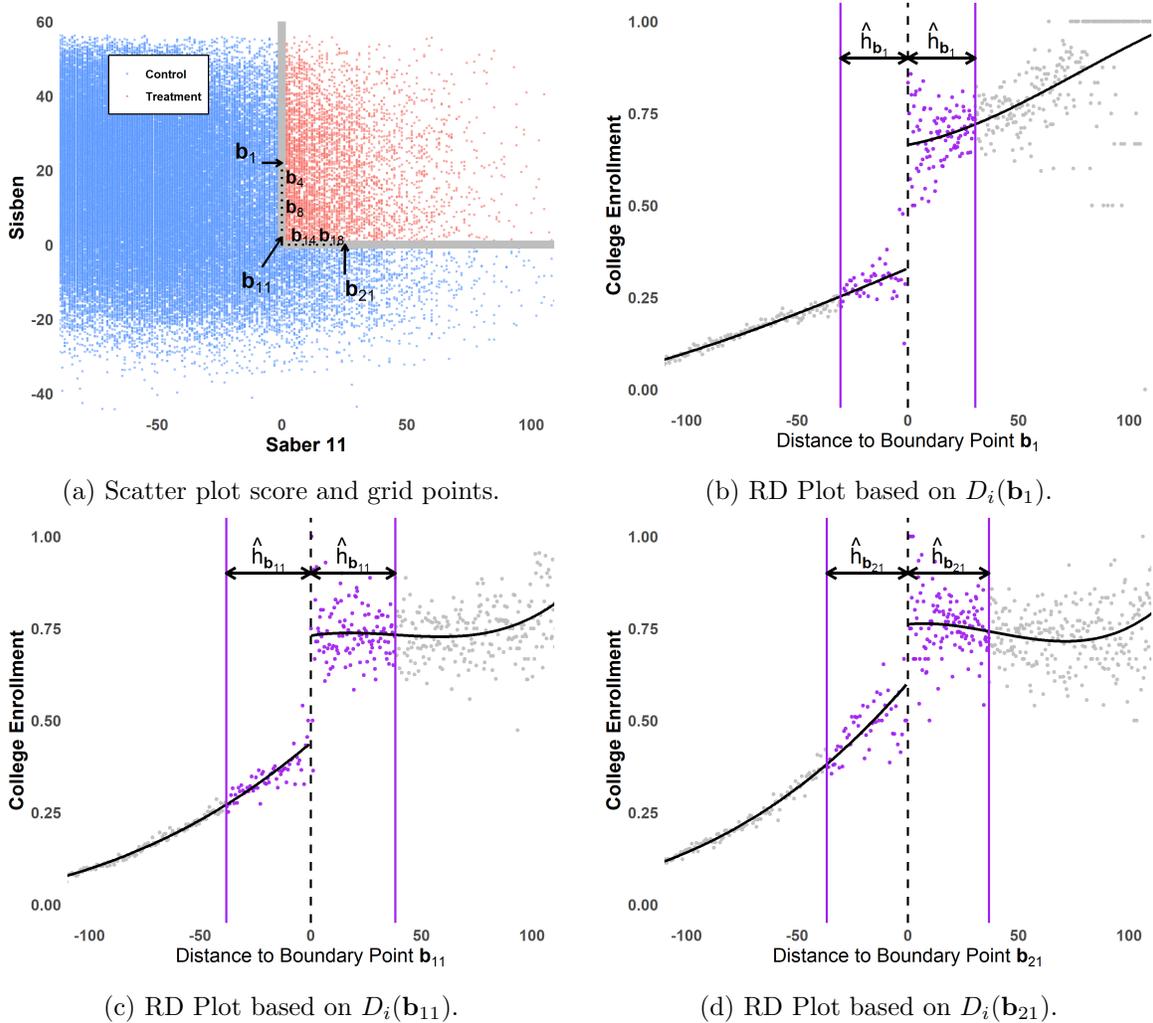


Figure 2: Scatter Plot and Selected Distance-Based RD Plots (SPP Data)

SPP dataset itself as an empirical application. All results reported in this section are implemented using our companion software package `rd2d`, together with the package `rdrobust` for RD designs with a univariate score. Additional implementation details are provided in the replication materials available at <https://rdpackages.github.io/>. Figure 2 summarizes key features of the SPP data. Panel (a) displays a scatterplot of the bivariate score, the treatment assignment boundary, and the 21 evaluation points. Panels (b)–(c) report representative RD plots based on the induced signed distance scores $D_i(\mathbf{b}_1)$, $D_i(\mathbf{b}_{11})$, and $D_i(\mathbf{b}_{21})$, respectively.

7.1 Simulation Study

The score $\mathbf{X}_i = (X_{1i}, X_{2i})^\top$ is generated from the distribution $(100 \cdot \text{Beta}(3, 4) - 25, 100 \cdot \text{Beta}(3, 4) - 25)^\top$ with independent components, which approximately reproduces salient features of the SPP dataset. We employ the same L-shaped treatment assignment boundary as in the empirical application, featuring a kink at $\mathbf{x} = (0, 0)^\top$, corresponding to the evaluation grid point \mathbf{b}_{11} .

For each $t \in \{0, 1\}$ and $i = 1, \dots, n$, potential outcomes are generated according to the regression model $Y_i(t) = \mu_t(\mathbf{X}_i) + \sigma_t(\mathbf{X}_i)\varepsilon_{t,i}$, where \mathbf{X}_i , $\varepsilon_{0,i}$ and $\varepsilon_{1,i}$ mutually independent, and

$$\begin{aligned}\mu_t(\mathbf{X}_i) &= \beta_{t,0} + X_{1i}\beta_{t,11} + X_{2i}\beta_{t,12} + X_{1i}^2\beta_{t,21} + X_{1i}X_{2i}\beta_{t,22} + X_{2i}^2\beta_{t,23}, \\ \sigma_t^2(\mathbf{X}_i) &= \exp(\gamma_{t,0} + X_{1i}\gamma_{t,11} + X_{2i}\gamma_{t,12} + X_{1i}^2\gamma_{t,21} + X_{1i}X_{2i}\gamma_{t,22} + X_{2i}^2\gamma_{t,23}),\end{aligned}$$

with $\varepsilon_{t,i} \sim \text{Normal}(0, 1)$. The coefficient values are calibrated using estimates obtained from the SPP dataset. We consider four data-generating processes combining linear (i.e., $\beta_{t,21} = \beta_{t,22} = \beta_{t,23} = 0$) or quadratic specifications with homoskedastic (i.e., $\gamma_{t,11} = \gamma_{t,12} = \gamma_{t,21} = \gamma_{t,22} = \gamma_{t,23} = 0$) or heteroskedastic disturbances. Table 1 reports the corresponding parameter values.

	Linear		Quadratic		Homoskedastic		Heteroskedastic		
	$t = 0$	$t = 1$	$t = 0$	$t = 1$	$t = 0$	$t = 1$	$t = 0$	$t = 1$	
$\beta_{t,0}$	3.35×10^{-1}	6.98×10^{-1}	3.72×10^{-1}	7.44×10^{-1}	$\gamma_{t,0}$	-2.20	-1.66	-1.57	-2.37
$\beta_{t,11}$	2.52×10^{-3}	2.74×10^{-3}	4.23×10^{-3}	2.29×10^{-3}	$\gamma_{t,11}$	0	0	2.19×10^{-2}	9.92×10^{-4}
$\beta_{t,12}$	-1.27×10^{-3}	-6.05×10^{-4}	-2.45×10^{-3}	-5.84×10^{-3}	$\gamma_{t,12}$	0	0	-5.08×10^{-3}	4.96×10^{-2}
$\beta_{t,21}$	0	0	1.25×10^{-5}	-1.33×10^{-7}	$\gamma_{t,21}$	0	0	-1.15×10^{-4}	-3.36×10^{-4}
$\beta_{t,22}$	0	0	3.12×10^{-5}	1.04×10^{-4}	$\gamma_{t,22}$	0	0	6.50×10^{-4}	-8.78×10^{-4}
$\beta_{t,23}$	0	0	-4.92×10^{-6}	2.14×10^{-5}	$\gamma_{t,23}$	0	0	5.23×10^{-4}	-3.12×10^{-4}

(a) Population parameters: $\mu_t(\mathbf{X}_i)$

(b) Population parameters: $\sigma_t^2(\mathbf{X}_i)$

Table 1: Parameters calibrated using the SPP dataset.

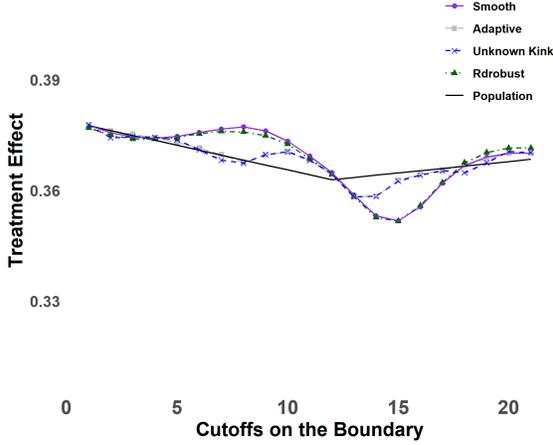
The simulation design sets $n = 20,000$ and uses 2,000 Monte Carlo replications. We evaluate performance in terms of point estimation and uncertainty quantification, both pointwise and uniformly over the boundary \mathcal{B} . Numerical results for the distance-based local polynomial procedures are reported in Tables 2–5, corresponding to the linear homoskedastic, linear heteroskedastic, quadratic homoskedastic, and quadratic heteroskedastic specifications, respectively. Each table contains four panels reflecting the bandwidth selection method employed. Panels (a) consider the case of a smooth boundary and uses the selector $h = \widehat{h}_{\text{MSE},\mathbf{b}}$. Panels (b) report results based on the kink-adaptive selector $h = \widehat{h}_{\text{MSE},\mathbf{b}}(\mathcal{B})$, which exploits knowledge of the kink location. Panels (c) use

the rule $h = \widehat{\mathcal{C}} \cdot n^{-1/4}$, which remains valid regardless of whether kinks are present in \mathcal{B} . Panels (d) employ the bandwidth $h = h_{1d,b}$ from the software package `rdrobust`, which is designed for univariate RD settings and therefore induces undersmoothing in the present BD design.

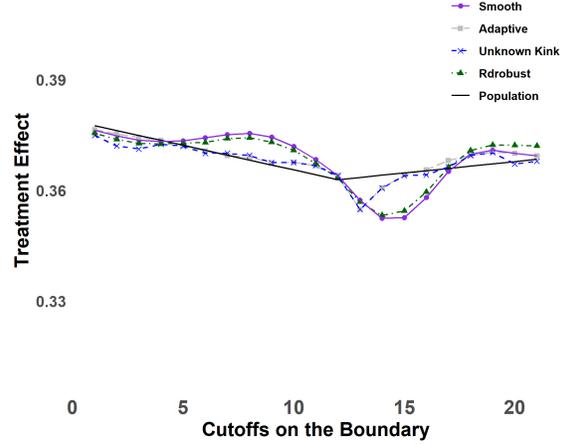
The main findings for bias are summarized in Figure 3, which reports the average value of each point estimator across the 2,000 replications and the target population regression function for each of the four data generating processes. The results show that distance-based methods that ignore the presence of the kink in \mathcal{B} exhibit a larger bias than procedures that explicitly account for kinks or other irregularities in the assignment boundary. This numerical evidence is consistent with the theoretical bias characterization established in Theorem 2. In this calibrated simulation study, however, the relatively large bias has limited consequences for inference because the signal-to-noise ratio is modest. As a result, coverage rates of confidence intervals and confidence bands remain close to their nominal levels across all specifications. These findings suggest that the bias induced by ignoring boundary irregularities can materially affect point estimation even when its impact on uncertainty quantification is muted. Naturally, in settings with stronger signals, failure to account for kinks or other non-smooth features of the assignment boundary may also lead to meaningful distortions in inference.

7.2 Empirical Application

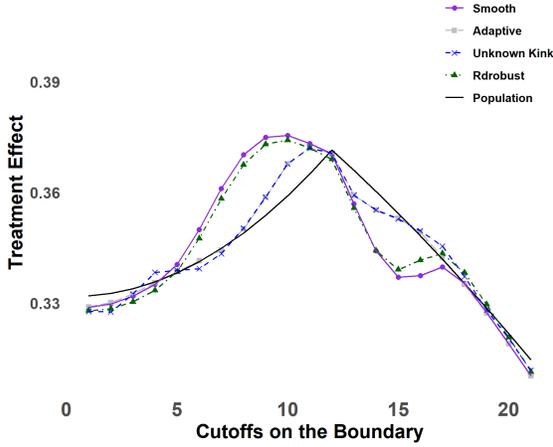
Table 6 and Figure 4 report pointwise and uniform estimation and inference for the BATEC using the distance-based procedures and the SPP dataset. The empirical findings are broadly consistent across the alternative methods proposed in this paper and align closely with the results obtained using location-based approaches in Cattaneo et al. (2026b). Overall, the evidence reveals heterogeneity in treatment effects along the assignment boundary: estimated effects are larger for students who are poorer and have lower academic achievement (evaluation points \mathbf{b}_1 – \mathbf{b}_{11}), and smaller for students who are relatively wealthier and have higher academic achievement (evaluation points \mathbf{b}_{11} – \mathbf{b}_{21}).



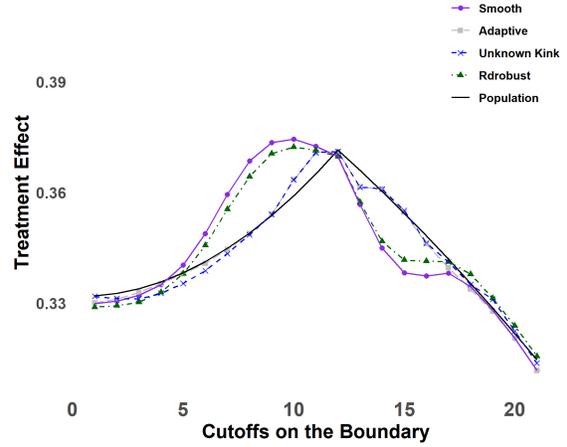
(a) Linear Homoskedastic Model



(b) Linear Heteroskedastic Model



(c) Quadratic Homoskedastic Model



(d) Quadratic Heteroskedastic Model

Figure 3: Average of BATEC Estimators (Simulation Results). Notes: (i) *Smooth* denotes using bandwidths $h = \hat{h}_{\text{MSE},\mathbf{x}}$ under the assumption that the boundary is smooth; (ii) *Adaptive* denotes using bandwidths $\hat{h}_{\text{kink},\mathbf{x}}(\mathcal{B})$ that adapts to the kink given information about the kink location; (iii) *Unknown Kink* denotes using the bandwidth $h = \hat{C} \cdot n^{-1/4}$ under kink rates; (iv) *Rdrobust* denotes the (incorrect) univariate MSE optimal bandwidth $h_{1d,\mathbf{x}}$ from *rdrobust*; and (v) *Population* denotes the population BATEC causal parameter, $\tau(\mathbf{x})$, calibrated using the SPP dataset (see Table 1).

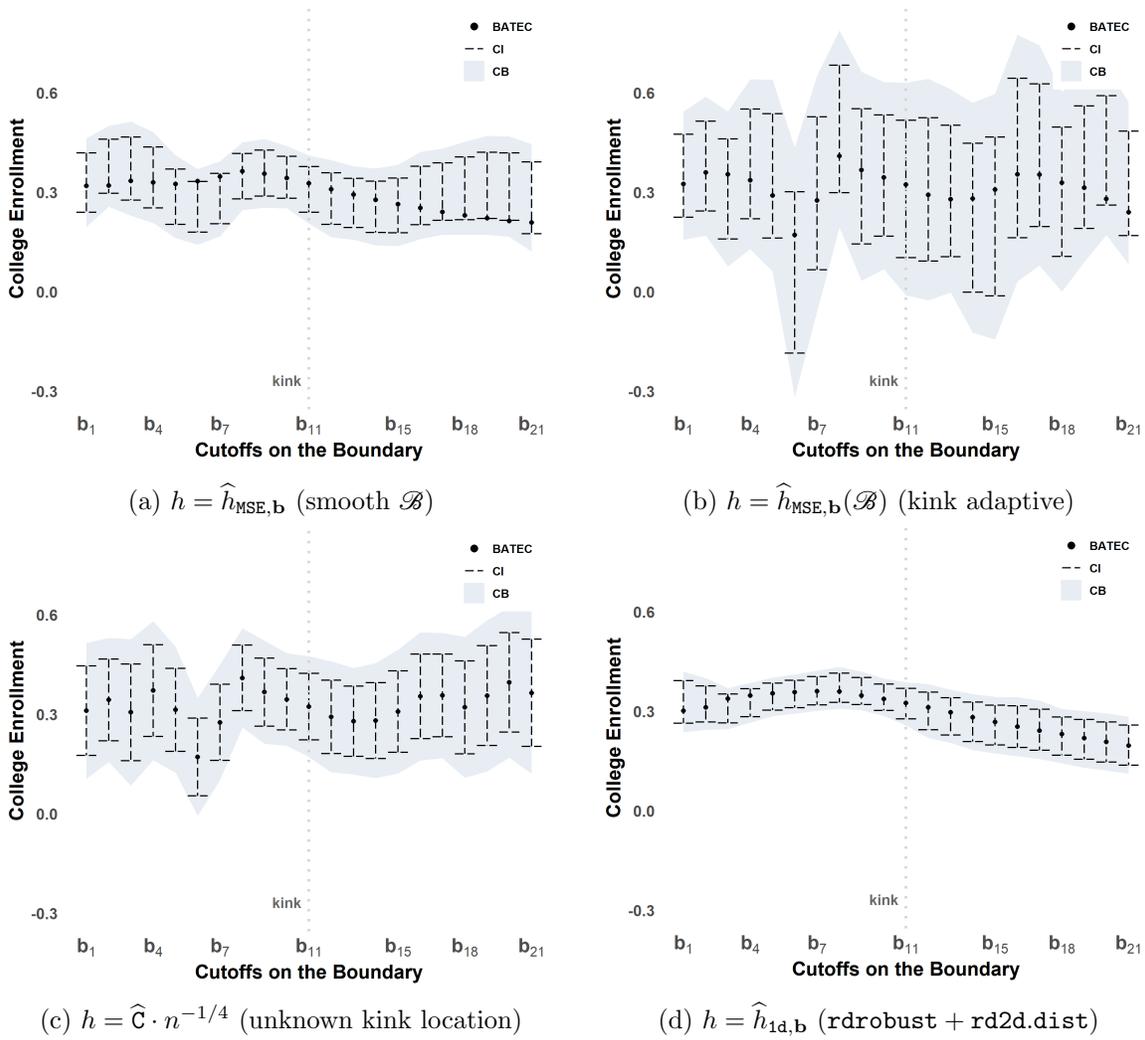


Figure 4: BATEC Estimation and Inference (SPP Empirical Application)

8 Extensions

The theoretical results developed in the supplemental appendix not only cover multidimensional location scores \mathbf{X}_i , but also provide the foundational tools needed to address several empirically relevant settings. This section discusses three natural extensions and generalizations.

8.1 Regression Kink Designs

All results presented in the paper extend immediately to derivative estimation, which is useful in the context of regression kink designs (Card et al., 2015). Specifically, the distance-based local polynomial regression estimator can be modified as

$$\widehat{\vartheta}^{(\nu)}(\mathbf{x}) = \mathbf{e}_{1+\nu}^\top \widehat{\gamma}_1(\mathbf{x}) - \mathbf{e}_{1+\nu}^\top \widehat{\gamma}_0(\mathbf{x}), \quad \mathbf{x} \in \mathcal{B},$$

where $\nu \in \{0, 1, \dots, p\}$ denotes the derivative order. All theoretical results follow directly by replacing the selector vector \mathbf{e}_1 with $\mathbf{e}_{1+\nu}$.

8.2 Imperfect Compliance

Our theoretical framework can also accommodate imperfect compliance (fuzzy RD designs), where treatment assignment and treatment receipt may differ for some units. Using standard potential-outcomes notation (Hernán and Robins, 2020), let $T_i = \mathbf{1}(D_i(\mathbf{x}) < 0) \cdot T_i(0) + \mathbf{1}(D_i(\mathbf{x}) \geq 0) \cdot T_i(1)$ denote the observed treatment status, where $T_i(t)$ is the potential treatment status under assignment $t \in \{0, 1\}$. The observed outcome is then $Y_i = \mathbf{1}(D_i(\mathbf{x}) < 0) \cdot Y_i(0, T_i(0)) + \mathbf{1}(D_i(\mathbf{x}) \geq 0) \cdot Y_i(1, T_i(1))$, where the potential outcome depends on both assignment and treatment receipt, that is, $Y_i(t, w)$ denotes the potential outcome for unit i when this unit is assigned to treatment $t \in \{0, 1\}$ and takes treatment status $w \in \{0, 1\}$.

The usual fuzzy estimand and estimator are

$$\zeta(\mathbf{x}) = \frac{\tau_Y(\mathbf{x})}{\tau_T(\mathbf{x})} \quad \text{and} \quad \widehat{\xi}(\mathbf{x}) = \frac{\widehat{\vartheta}_Y(\mathbf{x})}{\widehat{\vartheta}_T(\mathbf{x})},$$

where, for each $\mathbf{x} \in \mathcal{B}$, $\tau_Y(\mathbf{x}) = \mathbb{E}[Y_i(1, T_i(1)) - Y_i(0, T_i(0)) | \mathbf{X}_i = \mathbf{x}]$ and $\tau_T(\mathbf{x}) = \mathbb{E}[T_i(1) -$

$T_i(0)|\mathbf{X}_i = \mathbf{x}]$, and $\widehat{\vartheta}_Y(\mathbf{x}) = \mathbf{e}_1^\top \widehat{\gamma}_{Y,1}(\mathbf{x}) - \mathbf{e}_1^\top \widehat{\gamma}_{Y,0}(\mathbf{x})$ and $\widehat{\vartheta}_T(\mathbf{x}) = \mathbf{e}_1^\top \widehat{\gamma}_{D,1}(\mathbf{x}) - \mathbf{e}_1^\top \widehat{\gamma}_{T,0}(\mathbf{x})$, with $\widehat{\gamma}_{A,1}(\mathbf{x})$ denotes the local polynomial fit obtained using outcome variable $A \in \{Y, T\}$ in (1).

Under standard regularity conditions, our results imply that $\widehat{\vartheta}_Y(\mathbf{x}) = \tau_Y(\mathbf{x}) + o_{\mathbb{P}}(1)$ and $\widehat{\vartheta}_T(\mathbf{x}) = \tau_T(\mathbf{x}) + o_{\mathbb{P}}(1)$, both pointwise and uniformly over $\mathbf{x} \in \mathcal{B}$. Using the exact second-order linearization,

$$\widehat{\xi}(\mathbf{x}) - \zeta(\mathbf{x}) = \frac{1}{\tau_T(\mathbf{x})} (\widehat{\vartheta}_Y(\mathbf{x}) - \tau_Y(\mathbf{x})) - \frac{\tau_Y(\mathbf{x})}{\tau_T(\mathbf{x})^2} (\widehat{\vartheta}_T(\mathbf{x}) - \tau_T(\mathbf{x})) + \mathfrak{R}_n(\mathbf{x})$$

where

$$\mathfrak{R}_n(\mathbf{x}) = \frac{\tau_Y(\mathbf{x})}{\tau_T(\mathbf{x})^2 \widehat{\vartheta}_T(\mathbf{x})} (\widehat{\vartheta}_T(\mathbf{x}) - \tau_T(\mathbf{x}))^2 - \frac{1}{\tau_T(\mathbf{x}) \widehat{\vartheta}_T(\mathbf{x})} (\widehat{\vartheta}_Y(\mathbf{x}) - \tau_Y(\mathbf{x})) (\widehat{\vartheta}_T(\mathbf{x}) - \tau_T(\mathbf{x}))$$

it follows that $\widehat{\xi}(\mathbf{x}) - \zeta(\mathbf{x})$ is approximated by a linear combination of $\widehat{\vartheta}_Y(\mathbf{x}) - \tau_Y(\mathbf{x})$ and $\widehat{\vartheta}_T(\mathbf{x}) - \tau_T(\mathbf{x})$, yielding valid pointwise and uniform estimation and inference using the results in the supplemental appendix. Specifically, the remainder term $\mathfrak{R}_n(\mathbf{x})$ is of higher order and uniformly negligible provided $\inf_{\mathbf{x} \in \mathcal{B}} \tau_T(\mathbf{x}) > 0$ because $\sup_{\mathbf{x} \in \mathcal{B}} |\mathfrak{R}_n(\mathbf{x})| \lesssim_{\mathbb{P}} \sup_{\mathbf{x} \in \mathcal{B}} |\widehat{\vartheta}_T(\mathbf{x}) - \tau_T(\mathbf{x})|^2 + \sup_{\mathbf{x} \in \mathcal{B}} |\widehat{\vartheta}_Y(\mathbf{x}) - \tau_Y(\mathbf{x})| \sup_{\mathbf{x} \in \mathcal{B}} |\widehat{\vartheta}_T(\mathbf{x}) - \tau_T(\mathbf{x})|$, and hence Theorem 4 applies separately to the outcome and treatment estimators. A causal interpretation of $\zeta(\mathbf{x})$ can be obtained under additional assumptions; see [Arai et al. \(2022\)](#).

8.3 Pre-treatment Covariates

We next discuss how to incorporate pre-treatment covariates either to improve efficiency ([Calonico et al., 2019](#)) or to study treatment-effect heterogeneity ([Calonico et al., 2026](#)). Let $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ denote covariates of dimension $d_Z \geq 1$.

For efficiency improvements, the covariate-adjusted estimator is

$$\widetilde{\vartheta}(\mathbf{x}) = \mathbf{e}_{p+2}^\top \widetilde{\gamma}(\mathbf{x})$$

with

$$\widetilde{\gamma}(\mathbf{x}) = \arg \min_{\gamma \in \mathbb{R}^{2(p+1)+d_Z}} \mathbb{E}_n \left[(Y_i - \widetilde{\mathbf{r}}_p(D_i(\mathbf{x}), T_i, \mathbf{Z}_i)^\top \gamma)^2 K_h(D_i(\mathbf{x})) \mathbf{1}(D_i(\mathbf{x}) \in \mathcal{J}_t) \right],$$

where $T_i = \mathbf{1}(D_i(\mathbf{x}) \geq 0)$, and $\tilde{\mathbf{r}}_p(D_i(\mathbf{x}), T_i, \mathbf{Z}_i) = [\mathbf{r}_p(D_i(\mathbf{x}))^\top, T_i \cdot \mathbf{r}_p(D_i(\mathbf{x}))^\top, \mathbf{Z}_i^\top]^\top$ contains the full polynomial basis function for each treatment group but the pre-intervention covariates are not interacted with the treatment indicator; hence its dimension is $\mathbf{p} = 2(p + 1) + d_Z$.

For heterogeneity analysis, define

$$\check{\kappa}(\mathbf{x}, \mathbf{z}) = \mathbf{e}_{p+2}^\top \check{\gamma}(\mathbf{x}) + \check{\gamma}(\mathbf{x})^\top \begin{bmatrix} \mathbf{0}_{(2+d_Z)(p+1) \times d_Z} \\ \mathbf{I}_{d_Z} \\ \mathbf{0}_{pd_Z \times d_Z} \end{bmatrix} \mathbf{z}$$

with

$$\check{\gamma}(\mathbf{x}) = \arg \min_{\gamma \in \mathbb{R}^{2(p+1)(d_Z+1)}} \mathbb{E}_n \left[(Y_i - \tilde{\mathbf{r}}_p(D_i(\mathbf{x}), T_i, \mathbf{Z}_i)^\top \gamma)^2 K_h(D_i(\mathbf{x})) \mathbf{1}(D_i(\mathbf{x}) \in \mathcal{J}_t) \right],$$

where \mathbf{I}_d denotes the $(d \times d)$ identity matrix, $\mathbf{0}_{d_1 \times d_2}$ denotes the $(d_1 \times d_2)$ matrix of zeros, \mathbf{z} takes values on the support of \mathbf{Z}_i , and $\tilde{\mathbf{r}}_p(D_i(\mathbf{x}), T_i, \mathbf{Z}_i) = [\mathbf{r}_p(D_i(\mathbf{x}))^\top, T_i \cdot \mathbf{r}_p(D_i(\mathbf{x}))^\top, (\mathbf{r}_p(D_i(\mathbf{x})) \otimes \mathbf{Z}_i)^\top, T_i \cdot (\mathbf{r}_p(D_i(\mathbf{x})) \otimes \mathbf{Z}_i)^\top]^\top$ contains the full interaction between the polynomial basis function, the treatment assignment indicator, and the pre-intervention covariates.

9 Recommendations for Practice

[Cattaneo et al. \(2026a\)](#) review empirical applications of boundary discontinuity (BD) designs and document that researchers employ different empirical strategies depending on the objectives of the analysis, the assignment mechanism, and data availability. A useful taxonomy classifies methods according to how the multidimensional score \mathbf{X}_i is utilized. In practice, three levels of aggregation lead to three different types of methods.

- *Location-based methods.* These approaches exploit the full multidimensional score \mathbf{X}_i and therefore extend canonical regression discontinuity techniques to multidimensional assignment settings. By preserving local variation in all score dimensions, they enable the most flexible analysis of treatment-effect heterogeneity along the assignment boundary.
- *Distance-based methods.* These approaches reduce the multidimensional score \mathbf{X}_i to a uni-

variate distance measure defined relative to each boundary point $\mathbf{x} \in \mathcal{B}$, denoted by $D_i(\mathbf{x}) = \mathcal{d}(\mathbf{X}_i, \mathbf{x})$. This transformation allows researchers to implement standard univariate RD procedures locally to each point on the boundary while retaining the ability to study heterogeneous treatment effects summarized by the BATEC, denoted by $\tau(\mathbf{x}) = \mathbb{E}[Y_i(1) - Y_i(0) | \mathbf{X}_i = \mathbf{x}]$ for each $\mathbf{x} \in \mathcal{B}$.

- *Pooling-based methods.* These approaches further aggregate the multidimensional score by using the distance between each observation’s location and the closest boundary point, $D_i = \inf_{\mathbf{x} \in \mathcal{B}} \mathcal{d}(\mathbf{X}_i, \mathbf{x})$, thereby collapsing all boundary locations into a single scalar running variable. As a result, pooling-based methods identify an aggregate, scalar causal parameter corresponding to a weighted Boundary Average Treatment Effect (WBATE), given by $\tau_{\text{WBATE}} = \int_{\mathcal{B}} w(\mathbf{x})\tau(\mathbf{x}) d\mathbf{x}$.

Location-based and distance-based methods both target the BATEC as the fundamental causal functional parameter. This function can subsequently be summarized through policy-relevant functionals, such as the Weighted Boundary Average Treatment Effect, or the Largest Boundary Average Treatment Effect, given by $\tau_{\text{LBATE}} = \sup_{\mathbf{x} \in \mathcal{B}} \tau(\mathbf{x})$. In contrast, pooling-based methods directly estimate an aggregated WBATE parameter by construction. Location-based procedures are studied in our companion paper [Cattaneo et al. \(2026b\)](#), while formal analysis of pooling-based methods is the subject of ongoing research ([Cattaneo et al., 2026c](#)).

From a practical perspective, location-based methods are generally preferred whenever reliable information on the multidimensional score \mathbf{X}_i is available, as they offer the richest framework for identification, estimation and inference of the key building block functional parameter, $(\tau(\mathbf{x}) : \mathbf{x} \in \mathcal{B})$, and functionals thereof. When such information is limited or when dimensionality considerations make location-based approaches difficult to implement, distance-based methods provide a tractable and robust alternative. In particular, they enable estimation and inference for the BATEC, and functionals thereof, using well-understood univariate RD techniques.

At the same time, practitioners should recognize that distance-based estimators may exhibit different bias behavior near geometrically irregular regions of the treatment-assignment boundary, such as kink points. Empirical implementation should therefore account for boundary geometry when selecting bandwidths or interpreting results, following the results presented in this paper.

For example, researchers may restrict attention to locally smooth boundary segments or employ adaptive bandwidth strategies that vary with proximity to irregular points (e.g., our proposed bandwidth choice $\hat{h}_{\text{kink},x}(\mathcal{B})$). Such choices may affect the causal estimand and should be transparently reported. Nevertheless, the numerical evidence presented in Section 7 suggests that the proposed distance-based procedures can remain reasonably robust in empirically relevant configurations involving moderate boundary irregularities.

Finally, pooling-based methods remain popular in applied work because of their simplicity and ease of implementation. However, they provide the least informative causal analysis, as they aggregate potentially heterogeneous treatment effects into a single summary parameter. For this reason, we recommend that empirical researchers complement pooled estimates with analysis of the BATEC or related functionals whenever feasible, thereby obtaining a more comprehensive understanding of treatment-effect heterogeneity along the assignment boundary.

10 Conclusion

We studied the statistical properties of distance-based (isotropic) local polynomial estimation in boundary discontinuity designs. We established necessary and sufficient conditions for identification, estimation, and inference, both pointwise and uniformly along the treatment-assignment boundary. Our theoretical results highlight the central role played by the geometric regularity of the boundary—a one-dimensional manifold—along which estimation and inference are conducted. Building on these results, we provided concrete guidance for empirical implementation and illustrated the performance of the proposed methods using both simulated and real-world data. The companion general-purpose software package `rd2d` implements the main procedures developed in this paper (Cattaneo et al., 2025b).

Motivated by Theorem 3, which shows that distance-based methods can exhibit improved misspecification bias when the assignment boundary is sufficiently smooth, we conjecture that new distance-based (isotropic) nonparametric smoothing procedures that (i) explicitly incorporate geometric information about the boundary (e.g., knowledge of the location of kink points in \mathcal{B}), or (ii) rely on regularization of the boundary (e.g., via smooth approximations of \mathcal{B}), could achieve smaller misspecification bias. For example, one simple approach is to restrict estimation to locally

smooth segments of \mathcal{B} , thereby avoiding the bias issues highlighted by Theorem 2, at the cost of altering the target estimand through an additional regularization bias.

$\mathbf{b} \in \mathcal{B}$	h	Bias	SD	RMSE	EC	IL	$\mathbf{b} \in \mathcal{B}$	h	Bias	SD	RMSE	EC	IL
\mathbf{b}_1	19.905	0.000	0.033	0.033	0.948	0.238	\mathbf{b}_1	19.781	0.000	0.033	0.033	0.947	0.240
\mathbf{b}_2	19.466	-0.001	0.033	0.033	0.950	0.241	\mathbf{b}_2	18.810	-0.000	0.034	0.034	0.948	0.247
\mathbf{b}_3	19.087	-0.000	0.033	0.033	0.946	0.243	\mathbf{b}_3	17.512	0.000	0.035	0.035	0.943	0.262
\mathbf{b}_4	18.802	0.001	0.034	0.034	0.960	0.245	\mathbf{b}_4	15.901	0.001	0.037	0.037	0.948	0.285
\mathbf{b}_5	18.781	0.002	0.034	0.034	0.962	0.246	\mathbf{b}_5	13.996	0.001	0.042	0.042	0.957	0.322
\mathbf{b}_6	18.820	0.005	0.034	0.035	0.963	0.246	\mathbf{b}_6	12.000	0.001	0.049	0.049	0.950	0.375
\mathbf{b}_7	18.930	0.007	0.035	0.035	0.960	0.246	\mathbf{b}_7	10.000	0.000	0.059	0.059	0.948	0.451
\mathbf{b}_8	19.037	0.009	0.035	0.036	0.961	0.248	\mathbf{b}_8	8.532	-0.001	0.071	0.071	0.941	0.534
\mathbf{b}_9	19.054	0.009	0.036	0.037	0.951	0.253	\mathbf{b}_9	8.348	0.003	0.076	0.076	0.938	0.560
\mathbf{b}_{10}	19.148	0.008	0.037	0.038	0.954	0.258	\mathbf{b}_{10}	8.389	0.005	0.077	0.077	0.940	0.571
\mathbf{b}_{11}	19.387	0.005	0.037	0.038	0.951	0.267	\mathbf{b}_{11}	8.494	0.004	0.080	0.080	0.952	0.585
\mathbf{b}_{12}	21.239	0.002	0.038	0.038	0.945	0.279	\mathbf{b}_{12}	9.305	0.002	0.087	0.087	0.931	0.637
\mathbf{b}_{13}	19.195	-0.005	0.037	0.037	0.955	0.264	\mathbf{b}_{13}	8.410	-0.005	0.079	0.079	0.940	0.581
\mathbf{b}_{14}	19.151	-0.011	0.036	0.037	0.950	0.253	\mathbf{b}_{14}	8.390	-0.006	0.076	0.077	0.936	0.559
\mathbf{b}_{15}	19.215	-0.013	0.035	0.038	0.942	0.246	\mathbf{b}_{15}	8.724	-0.002	0.070	0.070	0.945	0.523
\mathbf{b}_{16}	18.953	-0.010	0.036	0.037	0.948	0.245	\mathbf{b}_{16}	11.200	-0.001	0.055	0.055	0.936	0.402
\mathbf{b}_{17}	18.628	-0.004	0.036	0.037	0.955	0.247	\mathbf{b}_{17}	13.996	-0.001	0.043	0.043	0.938	0.322
\mathbf{b}_{18}	18.436	0.000	0.035	0.035	0.956	0.251	\mathbf{b}_{18}	16.478	0.000	0.038	0.038	0.950	0.276
\mathbf{b}_{19}	18.663	0.002	0.035	0.035	0.949	0.250	\mathbf{b}_{19}	18.111	0.001	0.035	0.035	0.951	0.256
\mathbf{b}_{20}	19.409	0.002	0.034	0.034	0.954	0.245	\mathbf{b}_{20}	19.364	0.002	0.034	0.034	0.954	0.245
\mathbf{b}_{21}	20.644	0.002	0.034	0.034	0.950	0.237	\mathbf{b}_{21}	20.644	0.002	0.034	0.034	0.950	0.237
Uniform					0.945	0.374	Uniform					0.932	0.615

(a) $h = \widehat{h}_{\text{MSE},\mathbf{b}}$ (smooth \mathcal{B})(b) $h = \widehat{h}_{\text{MSE},\mathbf{b}}(\mathcal{B})$ (kink adaptive)

$\mathbf{b} \in \mathcal{B}$	h	Bias	SD	RMSE	EC	IL	$\mathbf{b} \in \mathcal{B}$	h	Bias	SD	RMSE	EC	IL
\mathbf{b}_1	8.720	0.000	0.071	0.071	0.946	0.274	\mathbf{b}_1	19.606	-0.001	0.035	0.035	0.950	0.244
\mathbf{b}_2	8.528	-0.002	0.072	0.072	0.945	0.277	\mathbf{b}_2	19.262	-0.001	0.035	0.035	0.950	0.246
\mathbf{b}_3	8.362	-0.000	0.074	0.074	0.942	0.280	\mathbf{b}_3	19.119	-0.001	0.035	0.035	0.950	0.246
\mathbf{b}_4	8.238	0.001	0.073	0.073	0.948	0.283	\mathbf{b}_4	19.116	0.000	0.036	0.036	0.960	0.245
\mathbf{b}_5	8.228	0.001	0.072	0.072	0.951	0.283	\mathbf{b}_5	19.240	0.002	0.036	0.036	0.963	0.244
\mathbf{b}_6	8.245	0.000	0.073	0.073	0.947	0.284	\mathbf{b}_6	19.455	0.004	0.036	0.036	0.965	0.243
\mathbf{b}_7	8.293	-0.001	0.073	0.073	0.957	0.284	\mathbf{b}_7	19.666	0.006	0.036	0.036	0.959	0.242
\mathbf{b}_8	8.340	-0.001	0.073	0.073	0.948	0.286	\mathbf{b}_8	19.819	0.008	0.037	0.038	0.962	0.244
\mathbf{b}_9	8.348	0.003	0.076	0.076	0.945	0.292	\mathbf{b}_9	20.168	0.008	0.038	0.039	0.950	0.245
\mathbf{b}_{10}	8.389	0.005	0.077	0.077	0.955	0.298	\mathbf{b}_{10}	20.661	0.007	0.038	0.039	0.955	0.245
\mathbf{b}_{11}	8.494	0.004	0.080	0.080	0.950	0.308	\mathbf{b}_{11}	21.467	0.004	0.038	0.038	0.948	0.248
\mathbf{b}_{12}	9.305	0.002	0.087	0.087	0.945	0.329	\mathbf{b}_{12}	23.273	0.002	0.038	0.038	0.952	0.260
\mathbf{b}_{13}	8.410	-0.005	0.079	0.079	0.948	0.304	\mathbf{b}_{13}	21.200	-0.005	0.036	0.037	0.954	0.245
\mathbf{b}_{14}	8.390	-0.006	0.076	0.077	0.950	0.292	\mathbf{b}_{14}	20.497	-0.011	0.037	0.038	0.953	0.242
\mathbf{b}_{15}	8.418	-0.002	0.074	0.074	0.948	0.283	\mathbf{b}_{15}	20.088	-0.013	0.037	0.039	0.946	0.240
\mathbf{b}_{16}	8.303	-0.001	0.075	0.075	0.944	0.282	\mathbf{b}_{16}	19.726	-0.009	0.038	0.040	0.949	0.241
\mathbf{b}_{17}	8.161	-0.001	0.076	0.076	0.940	0.285	\mathbf{b}_{17}	19.137	-0.004	0.039	0.039	0.963	0.247
\mathbf{b}_{18}	8.077	-0.002	0.074	0.074	0.953	0.289	\mathbf{b}_{18}	18.515	0.001	0.037	0.037	0.958	0.254
\mathbf{b}_{19}	8.176	0.000	0.074	0.074	0.946	0.288	\mathbf{b}_{19}	18.275	0.003	0.037	0.037	0.950	0.258
\mathbf{b}_{20}	8.503	0.003	0.072	0.072	0.951	0.282	\mathbf{b}_{20}	18.526	0.004	0.037	0.037	0.949	0.257
\mathbf{b}_{21}	9.044	0.002	0.070	0.070	0.951	0.272	\mathbf{b}_{21}	19.094	0.003	0.038	0.038	0.949	0.257
Uniform					0.936	0.438	Uniform					0.948	0.370

(c) $h = \widehat{\mathcal{C}} \cdot n^{-1/4}$ (unknown kink location)(d) $h = \widehat{h}_{\text{id},\mathbf{b}}$ (`rdrobust` + `rd2d.dist`)

Table 2: Linear Homoskedastic Model (Simulation Results)

$\mathbf{b} \in \mathcal{B}$	h	Bias	SD	RMSE	EC	IL	$\mathbf{b} \in \mathcal{B}$	h	Bias	SD	RMSE	EC	IL
\mathbf{b}_1	20.392	-0.001	0.040	0.040	0.952	0.284	\mathbf{b}_1	20.175	-0.001	0.040	0.040	0.952	0.286
\mathbf{b}_2	19.730	-0.001	0.041	0.041	0.949	0.284	\mathbf{b}_2	18.972	-0.001	0.041	0.041	0.948	0.293
\mathbf{b}_3	19.172	-0.001	0.041	0.041	0.948	0.284	\mathbf{b}_3	17.574	-0.001	0.043	0.043	0.950	0.306
\mathbf{b}_4	18.728	-0.000	0.041	0.041	0.955	0.284	\mathbf{b}_4	15.904	0.000	0.045	0.045	0.950	0.329
\mathbf{b}_5	18.433	0.001	0.041	0.041	0.960	0.283	\mathbf{b}_5	13.994	-0.000	0.049	0.049	0.951	0.364
\mathbf{b}_6	18.301	0.003	0.040	0.040	0.953	0.281	\mathbf{b}_6	12.000	-0.000	0.055	0.055	0.944	0.417
\mathbf{b}_7	18.209	0.006	0.040	0.040	0.958	0.279	\mathbf{b}_7	10.000	-0.000	0.065	0.065	0.948	0.494
\mathbf{b}_8	18.029	0.007	0.040	0.040	0.957	0.280	\mathbf{b}_8	8.238	0.001	0.079	0.079	0.944	0.591
\mathbf{b}_9	17.695	0.008	0.041	0.042	0.961	0.286	\mathbf{b}_9	7.756	0.001	0.085	0.085	0.956	0.630
\mathbf{b}_{10}	17.589	0.006	0.042	0.042	0.956	0.289	\mathbf{b}_{10}	7.706	0.002	0.089	0.089	0.928	0.637
\mathbf{b}_{11}	17.462	0.004	0.042	0.042	0.953	0.294	\mathbf{b}_{11}	7.650	0.003	0.086	0.086	0.941	0.651
\mathbf{b}_{12}	18.532	0.001	0.041	0.041	0.954	0.292	\mathbf{b}_{12}	8.119	0.001	0.087	0.087	0.932	0.645
\mathbf{b}_{13}	17.349	-0.006	0.042	0.043	0.957	0.299	\mathbf{b}_{13}	7.601	-0.009	0.090	0.091	0.930	0.657
\mathbf{b}_{14}	17.452	-0.012	0.042	0.043	0.957	0.298	\mathbf{b}_{14}	7.646	-0.003	0.089	0.089	0.929	0.650
\mathbf{b}_{15}	17.786	-0.012	0.041	0.043	0.956	0.292	\mathbf{b}_{15}	8.437	-0.001	0.079	0.079	0.942	0.589
\mathbf{b}_{16}	17.870	-0.007	0.042	0.042	0.957	0.294	\mathbf{b}_{16}	11.200	0.000	0.060	0.060	0.951	0.452
\mathbf{b}_{17}	17.874	-0.001	0.043	0.043	0.961	0.298	\mathbf{b}_{17}	13.986	0.002	0.049	0.049	0.953	0.371
\mathbf{b}_{18}	18.033	0.003	0.043	0.043	0.952	0.303	\mathbf{b}_{18}	16.418	0.003	0.045	0.045	0.948	0.327
\mathbf{b}_{19}	18.652	0.004	0.042	0.043	0.949	0.303	\mathbf{b}_{19}	18.185	0.003	0.043	0.043	0.948	0.309
\mathbf{b}_{20}	19.681	0.002	0.042	0.042	0.949	0.298	\mathbf{b}_{20}	19.629	0.002	0.042	0.042	0.949	0.298
\mathbf{b}_{21}	21.067	0.001	0.041	0.041	0.956	0.291	\mathbf{b}_{21}	21.067	0.001	0.041	0.041	0.956	0.291
Uniform					0.946	0.436	Uniform					0.922	0.697

(a) $h = \widehat{h}_{\text{MSE},\mathbf{b}}$ (smooth \mathcal{B})

(b) $h = \widehat{h}_{\text{MSE},\mathbf{b}}(\mathcal{B})$ (kink adaptive)

$\mathbf{b} \in \mathcal{B}$	h	Bias	SD	RMSE	EC	IL	$\mathbf{b} \in \mathcal{B}$	h	Bias	SD	RMSE	EC	IL
\mathbf{b}_1	8.934	-0.003	0.084	0.084	0.949	0.324	\mathbf{b}_1	19.768	-0.002	0.044	0.044	0.955	0.295
\mathbf{b}_2	8.644	-0.004	0.086	0.086	0.949	0.325	\mathbf{b}_2	19.279	-0.002	0.044	0.044	0.952	0.294
\mathbf{b}_3	8.399	-0.004	0.087	0.087	0.953	0.325	\mathbf{b}_3	18.916	-0.002	0.045	0.045	0.953	0.292
\mathbf{b}_4	8.205	-0.001	0.084	0.084	0.954	0.326	\mathbf{b}_4	18.722	-0.001	0.045	0.045	0.962	0.289
\mathbf{b}_5	8.076	-0.000	0.083	0.083	0.952	0.326	\mathbf{b}_5	18.546	0.000	0.045	0.045	0.962	0.286
\mathbf{b}_6	8.018	-0.001	0.084	0.084	0.952	0.323	\mathbf{b}_6	18.476	0.002	0.043	0.043	0.961	0.283
\mathbf{b}_7	7.977	0.001	0.082	0.082	0.954	0.323	\mathbf{b}_7	18.502	0.004	0.042	0.043	0.963	0.280
\mathbf{b}_8	7.899	0.001	0.083	0.083	0.955	0.324	\mathbf{b}_8	18.219	0.006	0.043	0.043	0.953	0.283
\mathbf{b}_9	7.752	0.001	0.085	0.085	0.951	0.329	\mathbf{b}_9	17.946	0.006	0.044	0.044	0.956	0.288
\mathbf{b}_{10}	7.706	0.002	0.089	0.089	0.945	0.334	\mathbf{b}_{10}	17.703	0.005	0.045	0.045	0.956	0.294
\mathbf{b}_{11}	7.650	0.003	0.086	0.086	0.957	0.338	\mathbf{b}_{11}	17.842	0.003	0.046	0.046	0.961	0.294
\mathbf{b}_{12}	8.119	0.001	0.087	0.087	0.945	0.338	\mathbf{b}_{12}	18.773	0.000	0.045	0.045	0.960	0.295
\mathbf{b}_{13}	7.601	-0.009	0.090	0.091	0.942	0.344	\mathbf{b}_{13}	17.709	-0.007	0.047	0.047	0.953	0.300
\mathbf{b}_{14}	7.646	-0.004	0.089	0.089	0.946	0.341	\mathbf{b}_{14}	17.820	-0.011	0.045	0.046	0.963	0.298
\mathbf{b}_{15}	7.792	-0.001	0.087	0.086	0.950	0.336	\mathbf{b}_{15}	18.141	-0.010	0.045	0.046	0.962	0.293
\mathbf{b}_{16}	7.829	-0.001	0.085	0.085	0.954	0.335	\mathbf{b}_{16}	18.370	-0.006	0.046	0.047	0.967	0.294
\mathbf{b}_{17}	7.831	0.000	0.087	0.087	0.960	0.338	\mathbf{b}_{17}	18.086	0.000	0.047	0.047	0.967	0.302
\mathbf{b}_{18}	7.901	0.003	0.087	0.087	0.958	0.342	\mathbf{b}_{18}	17.927	0.004	0.047	0.047	0.957	0.310
\mathbf{b}_{19}	8.172	0.003	0.089	0.089	0.948	0.341	\mathbf{b}_{19}	18.144	0.005	0.048	0.048	0.955	0.315
\mathbf{b}_{20}	8.622	-0.001	0.085	0.085	0.957	0.334	\mathbf{b}_{20}	18.610	0.004	0.047	0.047	0.956	0.316
\mathbf{b}_{21}	9.229	-0.001	0.083	0.083	0.948	0.325	\mathbf{b}_{21}	19.280	0.004	0.047	0.048	0.959	0.319
Uniform					0.945	0.506	Uniform					0.955	0.445

(c) $h = \widehat{\mathbf{C}} \cdot n^{-1/4}$ (unknown kink location)

(d) $h = \widehat{h}_{\text{id},\mathbf{b}}$ (`rdrobust` + `rd2d.dist`)

Table 3: Linear Heteroskedastic Model (Simulation Results)

$\mathbf{b} \in \mathcal{B}$	h	Bias	SD	RMSE	EC	IL	$\mathbf{b} \in \mathcal{B}$	h	Bias	SD	RMSE	EC	IL
\mathbf{b}_1	18.653	-0.003	0.035	0.035	0.948	0.254	\mathbf{b}_1	18.592	-0.003	0.035	0.035	0.948	0.254
\mathbf{b}_2	17.890	-0.003	0.035	0.035	0.945	0.261	\mathbf{b}_2	17.628	-0.002	0.035	0.035	0.945	0.263
\mathbf{b}_3	17.405	-0.002	0.035	0.035	0.951	0.266	\mathbf{b}_3	16.713	-0.001	0.037	0.037	0.952	0.274
\mathbf{b}_4	17.203	-0.001	0.037	0.037	0.946	0.268	\mathbf{b}_4	15.626	-0.000	0.039	0.039	0.946	0.290
\mathbf{b}_5	17.336	0.002	0.039	0.039	0.954	0.266	\mathbf{b}_5	13.971	0.000	0.044	0.044	0.948	0.323
\mathbf{b}_6	17.832	0.009	0.039	0.040	0.960	0.260	\mathbf{b}_6	12.000	0.000	0.050	0.050	0.954	0.376
\mathbf{b}_7	18.404	0.016	0.037	0.040	0.955	0.254	\mathbf{b}_7	10.000	-0.000	0.060	0.060	0.944	0.452
\mathbf{b}_8	18.879	0.021	0.035	0.041	0.953	0.251	\mathbf{b}_8	8.499	0.001	0.071	0.071	0.945	0.536
\mathbf{b}_9	18.981	0.021	0.035	0.041	0.950	0.254	\mathbf{b}_9	8.316	0.005	0.076	0.076	0.942	0.563
\mathbf{b}_{10}	19.089	0.016	0.036	0.040	0.948	0.259	\mathbf{b}_{10}	8.363	0.009	0.078	0.079	0.943	0.572
\mathbf{b}_{11}	19.336	0.008	0.038	0.038	0.945	0.269	\mathbf{b}_{11}	8.471	0.007	0.081	0.081	0.953	0.590
\mathbf{b}_{12}	21.238	-0.001	0.038	0.038	0.948	0.280	\mathbf{b}_{12}	9.304	-0.001	0.087	0.087	0.933	0.639
\mathbf{b}_{13}	19.214	-0.009	0.037	0.038	0.952	0.263	\mathbf{b}_{13}	8.418	-0.007	0.077	0.078	0.955	0.580
\mathbf{b}_{14}	19.046	-0.016	0.036	0.039	0.955	0.254	\mathbf{b}_{14}	8.344	-0.005	0.075	0.075	0.945	0.562
\mathbf{b}_{15}	18.746	-0.017	0.037	0.041	0.953	0.251	\mathbf{b}_{15}	8.648	-0.001	0.069	0.069	0.945	0.524
\mathbf{b}_{16}	17.792	-0.011	0.040	0.041	0.950	0.261	\mathbf{b}_{16}	11.200	0.001	0.054	0.054	0.932	0.403
\mathbf{b}_{17}	16.730	-0.002	0.040	0.040	0.942	0.276	\mathbf{b}_{17}	13.956	0.000	0.044	0.044	0.940	0.324
\mathbf{b}_{18}	16.232	-0.000	0.039	0.039	0.950	0.283	\mathbf{b}_{18}	15.636	-0.000	0.040	0.040	0.950	0.291
\mathbf{b}_{19}	16.437	-0.001	0.038	0.038	0.954	0.283	\mathbf{b}_{19}	16.324	-0.001	0.038	0.038	0.954	0.284
\mathbf{b}_{20}	17.181	-0.003	0.037	0.037	0.950	0.276	\mathbf{b}_{20}	17.174	-0.003	0.037	0.037	0.951	0.276
\mathbf{b}_{21}	18.119	-0.004	0.037	0.037	0.948	0.269	\mathbf{b}_{21}	18.119	-0.004	0.037	0.037	0.948	0.269
Uniform					0.949	0.397	Uniform					0.931	0.629

(a) $h = \hat{h}_{\text{MSE},\mathbf{b}}$ (smooth \mathcal{B})(b) $h = \hat{h}_{\text{MSE},\mathbf{b}}(\mathcal{B})$ (kink adaptive)

$\mathbf{b} \in \mathcal{B}$	h	Bias	SD	RMSE	EC	IL	$\mathbf{b} \in \mathcal{B}$	h	Bias	SD	RMSE	EC	IL
\mathbf{b}_1	8.172	-0.004	0.076	0.076	0.947	0.293	\mathbf{b}_1	17.750	-0.004	0.038	0.038	0.950	0.266
\mathbf{b}_2	7.838	-0.005	0.078	0.078	0.938	0.301	\mathbf{b}_2	17.219	-0.004	0.037	0.037	0.945	0.271
\mathbf{b}_3	7.625	-0.001	0.078	0.078	0.949	0.308	\mathbf{b}_3	16.908	-0.004	0.038	0.038	0.953	0.275
\mathbf{b}_4	7.537	0.003	0.082	0.082	0.944	0.310	\mathbf{b}_4	16.841	-0.002	0.039	0.039	0.957	0.276
\mathbf{b}_5	7.595	0.001	0.080	0.080	0.947	0.307	\mathbf{b}_5	17.075	0.000	0.041	0.041	0.954	0.274
\mathbf{b}_6	7.812	-0.002	0.077	0.077	0.951	0.301	\mathbf{b}_6	17.662	0.006	0.042	0.043	0.960	0.268
\mathbf{b}_7	8.063	-0.001	0.076	0.076	0.945	0.293	\mathbf{b}_7	18.350	0.013	0.042	0.044	0.954	0.261
\mathbf{b}_8	8.271	0.001	0.074	0.074	0.954	0.289	\mathbf{b}_8	18.995	0.019	0.040	0.044	0.954	0.256
\mathbf{b}_9	8.316	0.005	0.076	0.076	0.947	0.293	\mathbf{b}_9	19.565	0.019	0.038	0.043	0.949	0.253
\mathbf{b}_{10}	8.363	0.009	0.078	0.079	0.947	0.299	\mathbf{b}_{10}	20.067	0.015	0.038	0.041	0.949	0.253
\mathbf{b}_{11}	8.471	0.007	0.081	0.081	0.944	0.310	\mathbf{b}_{11}	21.059	0.007	0.039	0.039	0.952	0.254
\mathbf{b}_{12}	9.304	-0.001	0.087	0.087	0.944	0.331	\mathbf{b}_{12}	22.997	-0.003	0.039	0.039	0.953	0.264
\mathbf{b}_{13}	8.418	-0.007	0.077	0.078	0.955	0.304	\mathbf{b}_{13}	20.868	-0.010	0.038	0.039	0.959	0.248
\mathbf{b}_{14}	8.344	-0.005	0.075	0.075	0.955	0.293	\mathbf{b}_{14}	19.975	-0.016	0.037	0.041	0.951	0.247
\mathbf{b}_{15}	8.213	-0.002	0.074	0.074	0.954	0.290	\mathbf{b}_{15}	18.778	-0.015	0.041	0.044	0.957	0.258
\mathbf{b}_{16}	7.795	0.001	0.081	0.081	0.933	0.302	\mathbf{b}_{16}	17.261	-0.007	0.045	0.046	0.955	0.277
\mathbf{b}_{17}	7.330	0.004	0.085	0.085	0.941	0.320	\mathbf{b}_{17}	15.877	0.002	0.045	0.045	0.949	0.296
\mathbf{b}_{18}	7.112	0.002	0.085	0.085	0.944	0.327	\mathbf{b}_{18}	15.257	0.003	0.043	0.043	0.951	0.303
\mathbf{b}_{19}	7.201	-0.000	0.083	0.083	0.954	0.327	\mathbf{b}_{19}	15.423	0.001	0.041	0.041	0.955	0.301
\mathbf{b}_{20}	7.527	-0.001	0.083	0.083	0.947	0.320	\mathbf{b}_{20}	16.010	-0.001	0.041	0.041	0.953	0.295
\mathbf{b}_{21}	7.938	-0.003	0.081	0.081	0.944	0.311	\mathbf{b}_{21}	16.626	-0.003	0.042	0.042	0.953	0.291
Uniform					0.932	0.466	Uniform					0.949	0.407

(c) $h = \hat{\mathbf{C}} \cdot n^{-1/4}$ (unknown kink location)(d) $h = \hat{h}_{\text{id},\mathbf{b}}$ (`rdrobust` + `rd2d.dist`)

Table 4: Quadratic Homoskedastic Model (Simulation Results)

$\mathbf{b} \in \mathcal{B}$	h	Bias	SD	RMSE	EC	IL	$\mathbf{b} \in \mathcal{B}$	h	Bias	SD	RMSE	EC	IL
\mathbf{b}_1	19.460	-0.002	0.041	0.041	0.952	0.297	\mathbf{b}_1	19.323	-0.002	0.041	0.041	0.951	0.299
\mathbf{b}_2	18.555	-0.002	0.042	0.042	0.948	0.302	\mathbf{b}_2	18.151	-0.002	0.042	0.042	0.948	0.307
\mathbf{b}_3	17.879	-0.002	0.042	0.042	0.952	0.304	\mathbf{b}_3	16.985	-0.001	0.043	0.043	0.949	0.316
\mathbf{b}_4	17.581	-0.001	0.043	0.043	0.956	0.302	\mathbf{b}_4	15.681	-0.000	0.045	0.045	0.948	0.333
\mathbf{b}_5	17.457	0.002	0.044	0.044	0.960	0.299	\mathbf{b}_5	13.969	-0.000	0.049	0.049	0.955	0.366
\mathbf{b}_6	17.560	0.008	0.044	0.045	0.960	0.293	\mathbf{b}_6	12.000	-0.000	0.056	0.056	0.949	0.418
\mathbf{b}_7	17.789	0.015	0.043	0.045	0.951	0.286	\mathbf{b}_7	10.000	-0.000	0.066	0.066	0.949	0.493
\mathbf{b}_8	17.872	0.020	0.041	0.045	0.957	0.282	\mathbf{b}_8	8.206	-0.000	0.079	0.079	0.944	0.593
\mathbf{b}_9	17.632	0.020	0.040	0.045	0.950	0.287	\mathbf{b}_9	7.727	0.000	0.085	0.085	0.942	0.633
\mathbf{b}_{10}	17.496	0.015	0.041	0.044	0.953	0.290	\mathbf{b}_{10}	7.665	0.004	0.088	0.088	0.933	0.639
\mathbf{b}_{11}	17.393	0.007	0.042	0.042	0.958	0.295	\mathbf{b}_{11}	7.620	0.006	0.090	0.090	0.924	0.652
\mathbf{b}_{12}	18.550	-0.002	0.041	0.041	0.960	0.293	\mathbf{b}_{12}	8.127	-0.000	0.087	0.087	0.944	0.646
\mathbf{b}_{13}	17.358	-0.009	0.041	0.042	0.956	0.300	\mathbf{b}_{13}	7.604	-0.005	0.087	0.087	0.942	0.659
\mathbf{b}_{14}	17.415	-0.015	0.043	0.046	0.949	0.299	\mathbf{b}_{14}	7.630	0.001	0.090	0.090	0.935	0.651
\mathbf{b}_{15}	17.533	-0.016	0.045	0.048	0.950	0.297	\mathbf{b}_{15}	8.434	0.000	0.081	0.081	0.940	0.590
\mathbf{b}_{16}	17.216	-0.011	0.047	0.048	0.954	0.304	\mathbf{b}_{16}	11.200	-0.002	0.062	0.062	0.946	0.451
\mathbf{b}_{17}	16.813	-0.004	0.047	0.047	0.961	0.316	\mathbf{b}_{17}	13.937	-0.002	0.051	0.051	0.959	0.372
\mathbf{b}_{18}	16.660	-0.001	0.046	0.046	0.951	0.327	\mathbf{b}_{18}	15.819	-0.002	0.047	0.047	0.946	0.340
\mathbf{b}_{19}	17.005	-0.001	0.045	0.045	0.952	0.329	\mathbf{b}_{19}	16.847	-0.001	0.046	0.046	0.953	0.331
\mathbf{b}_{20}	17.834	-0.001	0.046	0.046	0.947	0.326	\mathbf{b}_{20}	17.822	-0.001	0.046	0.046	0.947	0.326
\mathbf{b}_{21}	18.871	-0.003	0.046	0.046	0.952	0.321	\mathbf{b}_{21}	18.871	-0.003	0.046	0.046	0.952	0.321
Uniform					0.959	0.455	Uniform					0.921	0.708

(a) $h = \widehat{h}_{\text{MSE},\mathbf{b}}$ (smooth \mathcal{B})

(b) $h = \widehat{h}_{\text{MSE},\mathbf{b}}(\mathcal{B})$ (kink adaptive)

$\mathbf{b} \in \mathcal{B}$	h	Bias	SD	RMSE	EC	IL	$\mathbf{b} \in \mathcal{B}$	h	Bias	SD	RMSE	EC	IL
\mathbf{b}_1	8.525	-0.000	0.090	0.090	0.946	0.341	\mathbf{b}_1	18.492	-0.003	0.045	0.045	0.950	0.313
\mathbf{b}_2	8.129	-0.001	0.089	0.089	0.950	0.347	\mathbf{b}_2	17.885	-0.003	0.045	0.045	0.954	0.314
\mathbf{b}_3	7.833	-0.003	0.088	0.088	0.953	0.349	\mathbf{b}_3	17.365	-0.004	0.045	0.046	0.958	0.315
\mathbf{b}_4	7.702	-0.003	0.089	0.089	0.955	0.348	\mathbf{b}_4	17.094	-0.003	0.047	0.047	0.962	0.314
\mathbf{b}_5	7.648	-0.003	0.088	0.088	0.950	0.346	\mathbf{b}_5	17.083	-0.000	0.048	0.048	0.961	0.311
\mathbf{b}_6	7.693	-0.002	0.087	0.087	0.950	0.338	\mathbf{b}_6	17.228	0.004	0.050	0.050	0.964	0.306
\mathbf{b}_7	7.794	-0.001	0.087	0.087	0.951	0.330	\mathbf{b}_7	17.527	0.011	0.050	0.051	0.956	0.298
\mathbf{b}_8	7.830	-0.000	0.084	0.084	0.949	0.327	\mathbf{b}_8	17.508	0.015	0.047	0.049	0.954	0.295
\mathbf{b}_9	7.725	0.000	0.085	0.085	0.952	0.331	\mathbf{b}_9	17.545	0.017	0.045	0.048	0.946	0.294
\mathbf{b}_{10}	7.665	0.004	0.088	0.088	0.949	0.335	\mathbf{b}_{10}	17.551	0.013	0.045	0.047	0.955	0.296
\mathbf{b}_{11}	7.620	0.006	0.090	0.090	0.947	0.339	\mathbf{b}_{11}	17.736	0.006	0.045	0.045	0.956	0.295
\mathbf{b}_{12}	8.127	-0.000	0.087	0.087	0.957	0.338	\mathbf{b}_{12}	18.702	-0.002	0.044	0.044	0.964	0.296
\mathbf{b}_{13}	7.604	-0.005	0.087	0.087	0.950	0.345	\mathbf{b}_{13}	17.669	-0.009	0.045	0.046	0.961	0.300
\mathbf{b}_{14}	7.630	0.001	0.090	0.090	0.948	0.343	\mathbf{b}_{14}	17.572	-0.014	0.048	0.050	0.952	0.303
\mathbf{b}_{15}	7.681	0.001	0.090	0.089	0.952	0.342	\mathbf{b}_{15}	17.436	-0.013	0.051	0.052	0.954	0.306
\mathbf{b}_{16}	7.542	-0.002	0.090	0.090	0.951	0.347	\mathbf{b}_{16}	16.969	-0.007	0.052	0.053	0.967	0.317
\mathbf{b}_{17}	7.366	-0.001	0.092	0.092	0.954	0.360	\mathbf{b}_{17}	16.228	-0.001	0.052	0.052	0.964	0.335
\mathbf{b}_{18}	7.299	-0.000	0.095	0.095	0.950	0.371	\mathbf{b}_{18}	15.825	0.002	0.051	0.051	0.955	0.347
\mathbf{b}_{19}	7.450	0.002	0.097	0.097	0.949	0.372	\mathbf{b}_{19}	15.977	0.003	0.050	0.050	0.955	0.351
\mathbf{b}_{20}	7.813	0.001	0.094	0.094	0.948	0.368	\mathbf{b}_{20}	16.536	0.002	0.050	0.051	0.951	0.350
\mathbf{b}_{21}	8.267	-0.001	0.093	0.093	0.948	0.361	\mathbf{b}_{21}	17.148	0.001	0.051	0.051	0.955	0.352
Uniform					0.942	0.529	Uniform					0.952	0.475

(c) $h = \widehat{\mathbf{C}} \cdot n^{-1/4}$ (unknown kink location)

(d) $h = \widehat{h}_{\text{id},\mathbf{b}}$ (`rdrobust` + `rd2d.dist`)

Table 5: Quadratic Heteroskedastic Model (Simulation Results)

$\mathbf{b} \in \mathcal{B}$	h	$\tau(\mathbf{b})$	p-value	95% RBC CI
\mathbf{b}_1	30.458	0.319	0.000	(0.239, 0.419)
\mathbf{b}_2	32.762	0.320	0.000	(0.296, 0.459)
\mathbf{b}_3	27.994	0.334	0.000	(0.276, 0.466)
\mathbf{b}_4	28.491	0.329	0.000	(0.252, 0.436)
\mathbf{b}_5	31.169	0.324	0.000	(0.202, 0.370)
\mathbf{b}_6	34.582	0.333	0.000	(0.179, 0.332)
\mathbf{b}_7	33.580	0.347	0.000	(0.204, 0.356)
\mathbf{b}_8	35.367	0.363	0.000	(0.280, 0.417)
\mathbf{b}_9	36.083	0.356	0.000	(0.288, 0.427)
\mathbf{b}_{10}	40.739	0.343	0.000	(0.282, 0.408)
\mathbf{b}_{11}	38.273	0.326	0.000	(0.239, 0.377)
\mathbf{b}_{12}	38.900	0.308	0.000	(0.203, 0.358)
\mathbf{b}_{13}	38.456	0.293	0.000	(0.193, 0.342)
\mathbf{b}_{14}	37.502	0.277	0.000	(0.178, 0.333)
\mathbf{b}_{15}	36.557	0.264	0.000	(0.178, 0.343)
\mathbf{b}_{16}	34.947	0.252	0.000	(0.202, 0.378)
\mathbf{b}_{17}	36.527	0.241	0.000	(0.215, 0.389)
\mathbf{b}_{18}	35.586	0.230	0.000	(0.217, 0.406)
\mathbf{b}_{19}	35.712	0.222	0.000	(0.221, 0.421)
\mathbf{b}_{20}	36.783	0.213	0.000	(0.215, 0.418)
\mathbf{b}_{21}	36.701	0.208	0.000	(0.175, 0.391)

(a) $h = \hat{h}_{\text{MSE}, \mathbf{b}}$ (smooth \mathcal{B})

$\mathbf{b} \in \mathcal{B}$	h	$\tau(\mathbf{b})$	p-value	95% RBC CI
\mathbf{b}_1	22.000	0.325	0.000	(0.225, 0.475)
\mathbf{b}_2	20.000	0.359	0.000	(0.243, 0.514)
\mathbf{b}_3	18.000	0.354	0.000	(0.158, 0.461)
\mathbf{b}_4	16.000	0.336	0.000	(0.219, 0.550)
\mathbf{b}_5	14.000	0.290	0.000	(0.162, 0.536)
\mathbf{b}_6	12.000	0.171	0.640	(-0.185, 0.301)
\mathbf{b}_7	11.651	0.275	0.012	(0.066, 0.527)
\mathbf{b}_8	12.271	0.409	0.000	(0.299, 0.683)
\mathbf{b}_9	12.519	0.367	0.001	(0.144, 0.552)
\mathbf{b}_{10}	14.134	0.345	0.000	(0.167, 0.533)
\mathbf{b}_{11}	13.278	0.323	0.003	(0.102, 0.517)
\mathbf{b}_{12}	13.495	0.292	0.005	(0.092, 0.524)
\mathbf{b}_{13}	13.338	0.279	0.003	(0.105, 0.502)
\mathbf{b}_{14}	13.007	0.280	0.052	(-0.001, 0.447)
\mathbf{b}_{15}	12.680	0.308	0.064	(-0.013, 0.466)
\mathbf{b}_{16}	12.121	0.354	0.001	(0.162, 0.644)
\mathbf{b}_{17}	14.000	0.352	0.000	(0.196, 0.626)
\mathbf{b}_{18}	16.800	0.329	0.002	(0.106, 0.496)
\mathbf{b}_{19}	19.600	0.314	0.000	(0.190, 0.560)
\mathbf{b}_{20}	22.400	0.280	0.000	(0.261, 0.591)
\mathbf{b}_{21}	25.200	0.240	0.000	(0.169, 0.484)

(b) $h = \hat{h}_{\text{MSE}, \mathbf{b}}(\mathcal{B})$ (kink adaptive)

$\mathbf{b} \in \mathcal{B}$	h	$\tau(\mathbf{b})$	p-value	95% RBC CI
\mathbf{b}_1	10.569	0.310	0.000	(0.175, 0.446)
\mathbf{b}_2	11.369	0.343	0.000	(0.220, 0.467)
\mathbf{b}_3	9.714	0.306	0.000	(0.160, 0.451)
\mathbf{b}_4	9.887	0.371	0.000	(0.233, 0.509)
\mathbf{b}_5	10.816	0.313	0.000	(0.188, 0.438)
\mathbf{b}_6	11.999	0.171	0.004	(0.054, 0.288)
\mathbf{b}_7	11.651	0.275	0.000	(0.161, 0.390)
\mathbf{b}_8	12.271	0.409	0.000	(0.311, 0.508)
\mathbf{b}_9	12.519	0.367	0.000	(0.264, 0.469)
\mathbf{b}_{10}	14.134	0.345	0.000	(0.252, 0.437)
\mathbf{b}_{11}	13.278	0.323	0.000	(0.222, 0.423)
\mathbf{b}_{12}	13.495	0.292	0.000	(0.181, 0.403)
\mathbf{b}_{13}	13.338	0.279	0.000	(0.173, 0.384)
\mathbf{b}_{14}	13.007	0.280	0.000	(0.166, 0.395)
\mathbf{b}_{15}	12.680	0.308	0.000	(0.185, 0.431)
\mathbf{b}_{16}	12.121	0.354	0.000	(0.227, 0.481)
\mathbf{b}_{17}	12.671	0.356	0.000	(0.232, 0.481)
\mathbf{b}_{18}	12.343	0.321	0.000	(0.180, 0.461)
\mathbf{b}_{19}	12.386	0.356	0.000	(0.205, 0.506)
\mathbf{b}_{20}	12.758	0.396	0.000	(0.246, 0.545)
\mathbf{b}_{21}	12.729	0.364	0.000	(0.203, 0.526)

(c) $h = \hat{\mathcal{C}} \cdot n^{-1/4}$ (unknown kink location)

$\mathbf{b} \in \mathcal{B}$	h	$\tau(\mathbf{b})$	p-value	95% RBC CI
\mathbf{b}_1	42.728	0.301	0.000	(0.264, 0.392)
\mathbf{b}_2	49.759	0.311	0.000	(0.267, 0.376)
\mathbf{b}_3	64.201	0.338	0.000	(0.264, 0.352)
\mathbf{b}_4	65.363	0.347	0.000	(0.283, 0.368)
\mathbf{b}_5	67.149	0.354	0.000	(0.303, 0.385)
\mathbf{b}_6	65.312	0.358	0.000	(0.310, 0.393)
\mathbf{b}_7	62.363	0.360	0.000	(0.319, 0.404)
\mathbf{b}_8	57.820	0.359	0.000	(0.326, 0.415)
\mathbf{b}_9	64.046	0.347	0.000	(0.320, 0.402)
\mathbf{b}_{10}	67.489	0.336	0.000	(0.303, 0.382)
\mathbf{b}_{11}	60.813	0.324	0.000	(0.277, 0.368)
\mathbf{b}_{12}	55.154	0.312	0.000	(0.244, 0.358)
\mathbf{b}_{13}	52.754	0.297	0.000	(0.228, 0.341)
\mathbf{b}_{14}	49.963	0.281	0.000	(0.208, 0.327)
\mathbf{b}_{15}	50.847	0.267	0.000	(0.198, 0.318)
\mathbf{b}_{16}	49.997	0.254	0.000	(0.191, 0.316)
\mathbf{b}_{17}	52.637	0.241	0.000	(0.182, 0.306)
\mathbf{b}_{18}	59.768	0.231	0.000	(0.167, 0.282)
\mathbf{b}_{19}	60.682	0.218	0.000	(0.155, 0.275)
\mathbf{b}_{20}	62.809	0.207	0.000	(0.147, 0.268)
\mathbf{b}_{21}	66.173	0.197	0.000	(0.137, 0.258)

(d) $h = \hat{h}_{1d, \mathbf{b}}(\text{rdrobust} + \text{rd2d.dist})$

Table 6: BATEC Estimation and Inference (SPP Empirical Application)

References

- ARAI, Y., Y.-C. HSU, T. KITAGAWA, I. MOURIFIÉ, AND Y. WAN (2022): “Testing Identifying Assumptions in Fuzzy Regression Discontinuity Designs,” *Quantitative Economics*, 13, 1–28.
- AVNIR, D., O. BIHAM, D. LIDAR, AND O. MALCAI (1998): “Is the geometry of nature fractal?” *Science*, 279, 39–40.
- BANERJEE, S. (2005): “On Geodetic Distance Computations in Spatial Modeling,” *Biometrics*, 61, 617–625.
- CALONICO, S., M. D. CATTANEO, AND M. H. FARRELL (2018): “On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference,” *Journal of the American Statistical Association*, 113, 767–779.
- (2020): “Optimal Bandwidth Choice for Robust Bias Corrected Inference in Regression Discontinuity Designs,” *Econometrics Journal*, 23, 192–210.
- (2022): “Coverage Error Optimal Confidence Intervals for Local Polynomial Regression,” *Bernoulli*, 28, 2998–3022.
- CALONICO, S., M. D. CATTANEO, M. H. FARRELL, F. PALOMBA, AND R. TITIUNIK (2026): “Treatment Effect Heterogeneity in Regression Discontinuity Designs,” *Working paper*.
- CALONICO, S., M. D. CATTANEO, M. H. FARRELL, AND R. TITIUNIK (2019): “Regression Discontinuity Designs using Covariates,” *Review of Economics and Statistics*, 101, 442–451.
- CALONICO, S., M. D. CATTANEO, AND R. TITIUNIK (2014): “Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs,” *Econometrica*, 82, 2295–2326.
- CARD, D., D. S. LEE, Z. PEI, AND A. WEBER (2015): “Inference on Causal Effects in a Generalized Regression Kink Design,” *Econometrica*, 83, 2453–2483.
- CATTANEO, M. D., Y. FENG, AND W. G. UNDERWOOD (2024a): “Uniform Inference for Kernel Density Estimators with Dyadic Data,” *Journal of the American Statistical Association*, 119, 2695–2708.

- CATTANEO, M. D., B. FRANDBEN, AND R. TITIUNIK (2015): “Randomization Inference in the Regression Discontinuity Design: An Application to Party Advantages in the U.S. Senate,” *Journal of Causal Inference*, 3, 1–24.
- CATTANEO, M. D., N. IDROBO, AND R. TITIUNIK (2020): *A Practical Introduction to Regression Discontinuity Designs: Foundations*, Cambridge University Press.
- (2024b): *A Practical Introduction to Regression Discontinuity Designs: Extensions*, Cambridge University Press.
- CATTANEO, M. D., L. KEELE, R. TITIUNIK, AND G. VAZQUEZ-BARE (2016): “Interpreting Regression Discontinuity Designs with Multiple Cutoffs,” *Journal of Politics*, 78, 1229–1248.
- CATTANEO, M. D., R. MASINI, AND W. G. UNDERWOOD (2025a): “Yurinskii’s Coupling for Martingales,” *Annals of Statistics*, forthcoming.
- CATTANEO, M. D. AND R. TITIUNIK (2022): “Regression Discontinuity Designs,” *Annual Review of Economics*, 14, 821–851.
- CATTANEO, M. D., R. TITIUNIK, AND R. R. YU (2025b): “rd2d: Causal Inference in Boundary Discontinuity Designs,” *arXiv:2505.07989*.
- (2026a): “Boundary Discontinuity Designs: Theory and Practice,” in *Advances in Economics and Econometrics: Thirteenth World Congress*, ed. by R. Griffith, Y. Gorodnichenko, M. Kandori, and F. Molinari, Cambridge University Press, vol. 1, chap. 2, forthcoming.
- (2026b): “Estimation and Inference in Boundary Discontinuity Designs: Location-Based Methods,” *arXiv:2505.05670*.
- (2026c): “Estimation and Inference in Boundary Discontinuity Designs: Pooling-Based Methods,” *Working paper*.
- CATTANEO, M. D. AND R. R. YU (2025): “Strong Approximations for Empirical Processes Indexed by Lipschitz Functions,” *Annals of Statistics*, 53, 1203–1229.
- CHEN, X. AND W. Y. GAO (2026): “Thin Sets Are Not Equally Thin: Minimax Learning of Submanifold Integrals,” *arXiv:2507.12673*.

- CHERNOZHUKOV, V., D. CHETVERIKOV, AND K. KATO (2014a): “Anti-Concentration and Honest, Adaptive Confidence Bands,” *Annals of Statistics*, 42, 1787–1818.
- (2014b): “Gaussian Approximation of Suprema of Empirical Processes,” *Annals of Statistics*, 42, 1564–1597.
- CHERNOZHUKOV, V., D. CHETVERIKOV, K. KATO, AND Y. KOIKE (2022): “Improved central limit theorem and bootstrap approximations in high dimensions,” *Annals of Statistics*, 50, 2562–2586.
- DIAZ, J. D. AND J. R. ZUBIZARRETA (2023): “Complex Discontinuity Designs Using Covariates for Policy Impact Evaluation,” *Annals of Applied Statistics*, 17, 67–88.
- FEDERER, H. (2014): *Geometric measure theory*, Springer.
- GALIANI, S., P. J. MCEWAN, AND B. QUISTORFF (2017): “External and Internal Validity of a Geographic Quasi-Experiment Embedded in a Cluster-Randomized Experiment,” in *Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics, volume 38)*, ed. by M. D. Cattaneo and J. C. Escanciano, Emerald Group Publishing, 195–236.
- GINÉ, E. AND R. NICKL (2016): *Mathematical Foundations of Infinite-dimensional Statistical Models*, New York: Cambridge University Press.
- HAHN, J., P. TODD, AND W. VAN DER KLAUW (2001): “Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design,” *Econometrica*, 69, 201–209.
- HÄRDLE, W., M. MÜLLER, S. SPERLICH, AND A. WERWATZ (2004): *Nonparametric and Semiparametric Models*, Heidelberg: Springer.
- HERNÁN, M. A. AND J. M. ROBINS (2020): *Causal Inference: What If*, Boca Raton: Chapman & Hall/CRC.
- KEELE, L. AND R. TITIUNIK (2016): “Natural experiments based on geography,” *Political Science Research and Methods*, 4, 65–95.
- KEELE, L. J., S. LORCH, M. PASSARELLA, D. SMALL, AND R. TITIUNIK (2017): “An Overview of Geographically Discontinuous Treatment Assignments with an Application to Children’s Health

- Insurance,” in *Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics, volume 38)*, ed. by M. D. Cattaneo and J. C. Escanciano, Emerald Group Publishing, 147–194.
- KEELE, L. J. AND R. TITIUNIK (2015): “Geographic Boundaries as Regression Discontinuities,” *Political Analysis*, 23, 127–155.
- KEELE, L. J., R. TITIUNIK, AND J. ZUBIZARRETA (2015): “Enhancing a Geographic Regression Discontinuity Design Through Matching to Estimate the Effect of Ballot Initiatives on Voter Turnout,” *Journal of the Royal Statistical Society: Series A*, 178, 223–239.
- LONDOÑO-VÉLEZ, J., C. RODRÍGUEZ, AND F. SÁNCHEZ (2020): “Upstream and downstream impacts of college merit-based financial aid for low-income students: Ser Pilo Paga in Colombia,” *American Economic Journal: Economic Policy*, 12, 193–227.
- MANDELBROT, B. (1967): “How long is the coast of Britain? Statistical self-similarity and fractional dimension,” *science*, 156, 636–638.
- MANDELBROT, B. B. (1983): “The fractal geometry of nature,” *New York*.
- PAPAY, J. P., J. B. WILLETT, AND R. J. MURNANE (2011): “Extending the regression-discontinuity approach to multiple assignment variables,” *Journal of Econometrics*, 161, 203–207.
- REARDON, S. F. AND J. P. ROBINSON (2012): “Regression Discontinuity Designs with Multiple Rating-Score Variables,” *Journal of Research on Educational Effectiveness*, 5, 83–104.
- RISCHARD, M., Z. BRANSON, L. MIRATRIX, AND L. BORNN (2021): “Do School Districts Affect NYC House Prices? Identifying Border Differences using a Bayesian Nonparametric Approach to Geographic Regression Discontinuity Designs,” *Journal of the American Statistical Association*, 116, 619–631.
- SIMON, L. ET AL. (1984): *Lectures on geometric measure theory*, Centre for Mathematical Analysis, Australian National University Canberra.
- TSYBAKOV, A. (2008): *Introduction to Nonparametric Estimation*, Springer.

VAN DER VAART, A. W. AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes*, Springer.

VELEZ, Y. R. AND B. J. NEWMAN (2019): “Tuning in, not turning out: Evaluating the impact of ethnic television on political participation,” *American Journal of Political Science*, 63, 808–823.

WONG, V. C., P. M. STEINER, AND T. D. COOK (2013): “Analyzing regression-discontinuity designs with multiple assignment variables: A comparative study of four estimation methods,” *Journal of Educational and Behavioral Statistics*, 38, 107–141.