Journal of Econometrics 256 (2026) 106266

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Estimation and inference in boundary discontinuity designs: Distance-based methods

Matias D. Cattaneo [a,*], Rocío Titiunik [b], Ruiqi (Rae) Yu [a]

[a] Department of Operations Research and Financial Engineering, Princeton University, United States
[b] Department of Politics, Princeton University, United States

ARTICLE INFO

Keywords:
Regression discontinuity
Treatment effects estimation
Causal inference
Isotropic nonparametric regression
Uniform inference
Minimax convergence rate

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.

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

* Corresponding author.
E-mail address: cattaneo@princeton.edu (M.D. Cattaneo).

https://doi.org/10.1016/j.jeconom.2026.106266
Received 4 November 2025; Received in revised form 26 March 2026; Accepted 21 May 2026

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):

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

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 $\mathbf{x} \in \mathscr{B}$, 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 $\mathbf{x} \in \mathscr{B}$. 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 $\mathbf{x} \in \mathscr{B}$. 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 $\mathscr{B}$. 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 $i = 1, \dots, n$,

$$D_i(\mathbf{x}) = \big(\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_1) - \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_0)\big) \cdot d(\mathbf{X}_i, \mathbf{x}), \qquad \mathbf{x} \in \mathscr{B},$$

where $d(\cdot,\cdot)$ denotes a distance function, $\mathscr{X} = \mathscr{A}_0 \cup \mathscr{A}_1$ with $\mathscr{A}_0$ and $\mathscr{A}_1$ denoting the disjoint (connected) control and treatment regions, respectively, and $\mathscr{B} = \mathtt{bd}(\mathscr{A}_0) \cap \mathtt{bd}(\mathscr{A}_1)$ with $\mathtt{bd}(\mathscr{A}_t)$ denoting the boundary of the set $\mathscr{A}_t$ for $t \in \{0, 1\}$. Without loss of generality, we assume that the assignment boundary belongs to the treatment region, that is, $\mathscr{B} \subset \mathscr{A}_1$ and $\mathscr{B} \cap \mathscr{A}_0 = \emptyset$. A canonical example is the Euclidean distance $d(\mathbf{X}_i, \mathbf{x}) = |\mathbf{X}_i - \mathbf{x}| = \sqrt{(X_{1i} - x_1)^2 + (X_{2i} - x_2)^2}$ for $\mathbf{x} = (x_1, x_2)^\top \in \mathscr{B}$, although alternative distance measures may be appropriate in spatial applications (Banerjee, 2005). For each $\mathbf{x} \in \mathscr{B}$, 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 $Y_i = \mathbb{1}(D_i(\mathbf{x}) \in \mathscr{T}_0) Y_i(0) + \mathbb{1}(D_i(\mathbf{x}) \in \mathscr{T}_1) Y_i(1)$, where $\mathscr{T}_0 = (-\infty, 0)$ and $\mathscr{T}_1 = [0, \infty)$. Following standard practice in the RD literature, the distance-based local polynomial treatment effect estimator for each $\mathbf{x} \in \mathscr{B}$ is

$$\widehat{\vartheta}(\mathbf{x}) = \mathbf{e}_0^\top \widehat{\boldsymbol{\gamma}}_1(\mathbf{x}) - \mathbf{e}_0^\top \widehat{\boldsymbol{\gamma}}_0(\mathbf{x}), \qquad \mathbf{x} \in \mathscr{B},$$

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

$$\widehat{\boldsymbol{\gamma}}_t(\mathbf{x}) = \arg\min_{\boldsymbol{\gamma}\in\mathbb{R}^{p+1}} \frac{1}{n}\sum_{i=1}^{n} \big(Y_i - \mathbf{r}_p(D_i(\mathbf{x})/h)^\top \boldsymbol{\gamma}\big)^2 K_h(D_i(\mathbf{x}))\mathbb{1}(D_i(\mathbf{x}) \in \mathscr{I}_t), \tag{1}$$

with $\mathbf{e}_0$ the conformable intercept selector, $\mathbf{r}_p(u) = (1, u, u^2, \ldots, u^p)^\top$ the usual polynomial basis, and $K_h(u) = K(u/h)/h^2$ a distance-based kernel with bandwidth h. The kernel down-weights observations as their distance to $\mathbf{x} \in \mathscr{B}$ increases, while the bandwidth controls the degree of localization. The normalization by h^2 reflects the bivariate score dimension; multiplying all weights by a common positive constant does not affect the weighted least squares estimator. In this setup, observations contribute isotropically according to their univariate distance $\mathscr{d}(\mathbf{X}_i, \mathbf{x})$ from the boundary point $\mathbf{x}$.

The estimator $\widehat{\vartheta}(\mathbf{x})$ is thus the difference of two univariate (isotropic) local polynomial regression estimators evaluated along the one-dimensional manifold $\mathscr{B}$. 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

$$\theta_{t,\mathbf{x}}(r) = \mathbb{E}\big[Y_i \big| D_i(\mathbf{x}) = r\big] = \mathbb{E}\big[Y_i(t) \big| \mathscr{d}(\mathbf{X}_i, \mathbf{x}) = |r|, \mathbf{X}_i \in \mathscr{A}_t\big], \qquad r \in \mathscr{I}_t, \tag{2}$$

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 $\widehat{\vartheta}(\mathbf{x})$ 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 $\mu_t(\mathbf{x}) = \mathbb{E}[Y_i(t) \mid \mathbf{X}_i = \mathbf{x}]$. Consequently, without restrictions on the data-generating process, the geometry of the assignment boundary, and the distance function, $\lim_{r\to 0}\theta_{t,\mathbf{x}}(r)$ need not coincide with $\mu_t(\mathbf{x})$. Theorem 1 provides sufficient conditions for these parameters to agree, including restrictions ensuring that $\mathscr{B}$ 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 $\mathscr{B}$ is a well-behaved one-dimensional manifold, the pth-order estimator $\widehat{\vartheta}(\mathbf{x})$ 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 $\mathscr{B}$ 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 $\widehat{\vartheta}(\mathbf{x})$. Theorem 4 establishes pointwise and uniform convergence rates, while Theorem 5 provides valid uncertainty quantification both pointwise and uniformly over $\mathscr{B}$. 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 $\mathscr{B}$ 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 $(\tau(\mathbf{x}) : \mathbf{x} \in \mathscr{B})$, 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 $\widehat{\vartheta}(\mathbf{x})$ 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 $\mathscr{B}$, 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 $\mathscr{B}$ 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.

The supplemental appendix further extends the distance-based framework to policy-relevant functionals of the BATEC: the *Weighted Boundary Average Treatment Effect* (WBATE) and *Largest Boundary Average Treatment Effect* (LBATE), as well as to fuzzy (imperfect treatment compliance) BD designs, covering fuzzy BATEC, fuzzy WBATE, and fuzzy LBATE estimands. Section 8 summarizes

these extensions and also discusses adjustments for pre-intervention covariates aimed at either improving efficiency or heterogeneity analysis.

Concrete guidance for empirical implementation based on the results developed in this paper and in our ongoing companion work (Cattaneo et al., 2026b,c) is given in Section 9. All our methodological results are implemented in our companion package `rd2d` available in `Python`, `R` and `Stata`; see https://rdpackages.github.io/rd2d and Cattaneo et al. (2025b) for further details.

From a broader theoretical perspective, this paper establishes a 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 foundational 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 nonparametric perspective, we develop 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 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 $perim(\mathcal{S}) = \sup_{g\in\mathscr{D}_2(\mathcal{X})} \int_{\mathbb{R}^2} \mathbb{1}(\mathbf{x} \in \mathcal{S}) \operatorname{div} g(\mathbf{x}) d\mathbf{x} / \|g\|_\infty$, where div is the divergence operator, and $\mathscr{D}_2(\mathcal{X})$ 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 $\mathtt{bd}(\mathcal{S})$ is a smooth simple closed curve, $perim(\mathcal{S})$ simplifies to the curve length of $\mathtt{bd}(\mathcal{S})$. A curve $\mathscr{B} \subseteq \mathbb{R}^2$ is a *rectifiable curve* if there exists a Lipschitz map $\gamma : [0,1] \mapsto \mathbb{R}^2$ such that $\mathscr{B} = \gamma([0,1])$. For a function $f : \mathbb{R}^2 \mapsto \mathbb{R}$, $supp(f)$ denotes the closure of the set $\{\mathbf{x} \in \mathbb{R}^2 : f(\mathbf{x}) \neq 0\}$. For real sequences, $a_n = o(b_n)$ means $\limsup_{n\to\infty} |a_n/b_n| = 0$, and $a_n \lesssim b_n$ means there exist constants $C < \infty$ and $N < \infty$ such that $|a_n| \leq C|b_n|$ for all $n > N$. For sequences of random variables, $a_n = o_{\mathbb{P}}(b_n)$ means $|a_n/b_n| \to_{\mathbb{P}} 0$, and $a_n \lesssim_{\mathbb{P}} b_n$ means $\limsup_{M\to\infty} \limsup_{n\to\infty} \mathbb{P}[|a_n/b_n| \geq M] = 0$. Throughout, $\mathbf{x}$ denotes a generic point in $\mathcal{X}$ and, when $\mathbf{x} \in \mathscr{B}$, a generic boundary point; $\mathbf{b}$ and $\mathbf{b}_j$ denote selected or discretized evaluation points on the boundary. 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 presents a minimax convergence rate result for isotropic (distance-based) regression estimation. Section 7 illustrates the finite-sample performance of the distance-based methods. Section 8 discusses extensions, Section 9 provides implementation guidance, 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 $t \in \{0,1\}$, $p \geq 0$, and $\upsilon \geq 2$.

(i) $(Y_1(t), \mathbf{X}_1^\top)^\top, \ldots, (Y_n(t), \mathbf{X}_n^\top)^\top$ 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 $\mathcal{X} = [\mathtt{L}, \mathtt{U}]^2$, for $-\infty < \mathtt{L} < \mathtt{U} < \infty$.
(iii) $\mu_t(\mathbf{x}) = \mathbb{E}[Y_i(t) \mid \mathbf{X}_i = \mathbf{x}]$ has a $(p+1)$ times continuously differentiable extension to an open neighborhood of $\mathcal{X}$.
(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) $\sup_{\mathbf{x}\in\mathcal{X}} \mathbb{E}\big[|Y_i(t)|^{2+\upsilon} \mid \mathbf{X}_i = \mathbf{x}\big] < \infty$.

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

$$\mathbf{\Psi}_{t,\mathbf{x}} = \mathbb{E}\Big[\mathbf{r}_p\Big(\frac{D_i(\mathbf{x})}{h}\Big)\mathbf{r}_p\Big(\frac{D_i(\mathbf{x})}{h}\Big)^\top K_h(D_i(\mathbf{x}))\mathbb{1}(D_i(\mathbf{x}) \in \mathcal{I}_t)\Big]$$

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) $\mathscr{B} \subset \text{int}(\mathscr{X})$ is a rectifiable curve with positive length.
(ii) The distance function $\mathscr{d} : \mathbb{R}^2 \times \mathbb{R}^2 \to [0,\infty)$ is a metric satisfying $\|\mathbf{x}_1 - \mathbf{x}_2\| \lesssim \mathscr{d}(\mathbf{x}_1,\mathbf{x}_2) \lesssim \|\mathbf{x}_1 - \mathbf{x}_2\|$ for all $\mathbf{x}_1,\mathbf{x}_2 \in \mathscr{X}$.
(iii) $K : \mathbb{R} \to [0,\infty)$ is either compactly supported and Lipschitz continuous, or $K(u) = \mathbb{1}(u \in [-1,1])$ and the distance balls $\{\mathbf{v} \in \mathscr{X} : \mathscr{d}(\mathbf{v},\mathbf{x}) \le r\}$, indexed by $(\mathbf{x},r) \in \mathscr{X} \times \mathbb{R}_+$, form a VC class.
(iv) $\liminf_{h\downarrow 0} \inf_{\mathbf{x}\in\mathscr{B}} \lambda_{\min}(\boldsymbol{\Psi}_{t,\mathbf{x}}) \gtrsim 1$.
(v) For each $\mathbf{x} \in \mathscr{B}$ and $t \in \{0,1\}$, the Hausdorff-measure denominator defining $\theta_{t,\mathbf{x}}(r)$ is positive and finite for all sufficiently small $|r| > 0$ with $r \in \mathscr{I}_t$.

Assumption 2(i) imposes minimal regularity on the assignment boundary $\mathscr{B}$, 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 a metric equivalent (up to constants) to the Euclidean distance. Assumption 2(iii) imposes standard conditions on the univariate kernel function; for the uniform kernel, the additional VC condition is automatic for Euclidean distance balls and other primitive cases covered by Lemma SA-1 in the supplemental appendix. Assumption 2(iv) further restricts (implicitly) the geometry of the boundary $\mathscr{B}$ relative to the kernel and distance functions, ruling out highly irregular shapes that would generate regions with too few observations. Assumption 2(v) ensures that the induced one-dimensional conditional expectations are well defined near the cutoff. Lemma SA-2 in the supplemental appendix provides primitive conditions for Assumption 2(iv) when $\mathscr{d}(\cdot)$ is the Euclidean norm. The conditions imposed by Assumption 2 are sufficient for uniform (over $\mathscr{B}$) estimation and inference, although some can be weakened for pointwise results; see the supplemental appendix.

3. Identification and interpretation

For each boundary point $\mathbf{x} \in \mathscr{B}$ and corresponding signed distance score $D_i(\mathbf{x})$, the univariate distance-based local polynomial estimator is $\widehat{\vartheta}(\mathbf{x}) = \widehat{\theta}_{1,\mathbf{x}}(0) - \widehat{\theta}_{0,\mathbf{x}}(0)$, where $\widehat{\theta}_{t,\mathbf{x}}(0) = \mathbf{e}_0^\top \widehat{\boldsymbol{\gamma}}_t(\mathbf{x}) = \mathbf{r}_p(0)^\top \widehat{\boldsymbol{\gamma}}_t(\mathbf{x})$. Conceptually, the estimator $\widehat{\theta}_{t,\mathbf{x}}(0)$ 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 $\mathbf{x} \in \mathscr{B}$.

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), and (v) 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 $\mathbf{x} \in \mathscr{B}$.

This identification result is obtained by representing $\theta_{t,\mathbf{x}}(r)$ as a sequence of integrals over submanifolds near $\mathbf{x} \in \mathscr{B}$, that is, over level sets of shrinking spheres generated by the distance function $\mathscr{d}(\cdot,\mathbf{x})$. Without restrictions on the data generating process, the geometry of the assignment boundary, and the distance function, the limits $\lim_{r\to 0} \theta_{t,\mathbf{x}}(r)$ 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

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

for all $\mathbf{x} \in \mathscr{B}$. 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 $\mathscr{B}$. 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 $\widehat{\boldsymbol{\gamma}}_t(\mathbf{x})$ as the sample analogue of the coefficients from the best local mean-square approximation of the conditional expectation $\mathbb{E}[Y_i(t) \mid D_i(\mathbf{x})]$ using the polynomial basis $\mathbf{r}_p(D_i(\mathbf{x})/h)$:

$$\boldsymbol{\gamma}_t^*(\mathbf{x}) = \arg\min_{\boldsymbol{\gamma}\in\mathbb{R}^{p+1}} \mathbb{E}\Big[\big(Y_i - \mathbf{r}_p(D_i(\mathbf{x})/h)^\top \boldsymbol{\gamma}\big)^2 K_h(D_i(\mathbf{x}))\mathbb{1}(D_i(\mathbf{x}) \in \mathscr{I}_t)\Big].$$

Letting $\theta_{t,\mathbf{x}}^*(r) = \mathbf{r}_p(r/h)^\top \boldsymbol{\gamma}_t^*(\mathbf{x})$, so that $\theta_{t,\mathbf{x}}^*(0) = \mathbf{e}_0^\top \boldsymbol{\gamma}_t^*(\mathbf{x}) = \mathbf{r}_p(0)^\top \boldsymbol{\gamma}_t^*(\mathbf{x})$, we obtain the standard least-squares decomposition

$$\widehat{\theta}_{t,\mathbf{x}}(0) - \theta_{t,\mathbf{x}}(0) = \big[\theta_{t,\mathbf{x}}^*(0) - \theta_{t,\mathbf{x}}(0)\big] + \mathbf{e}_0^\top \boldsymbol{\Psi}_{t,\mathbf{x}}^{-1} \mathbf{O}_{t,\mathbf{x}} + \big[\mathbf{e}_0^\top (\widehat{\boldsymbol{\Psi}}_{t,\mathbf{x}}^{-1} - \boldsymbol{\Psi}_{t,\mathbf{x}}^{-1}) \mathbf{O}_{t,\mathbf{x}}\big], \tag{3}$$

where

$$\widehat{\boldsymbol{\Psi}}_{t,\mathbf{x}} = \mathbb{E}_n\Big[\mathbf{r}_p\Big(\frac{D_i(\mathbf{x})}{h}\Big)\mathbf{r}_p\Big(\frac{D_i(\mathbf{x})}{h}\Big)^\top K_h(D_i(\mathbf{x}))\mathbb{1}(D_i(\mathbf{x}) \in \mathscr{I}_t)\Big], \qquad \text{and}$$

$$\mathbf{O}_{t,\mathbf{x}} = \mathbb{E}_n\Big[\mathbf{r}_p\Big(\frac{D_i(\mathbf{x})}{h}\Big) K_h(D_i(\mathbf{x}))(Y_i - \theta_{t,\mathbf{x}}^*(D_i(\mathbf{x})))\mathbb{1}(D_i(\mathbf{x}) \in \mathscr{I}_t)\Big].$$

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

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

where

$$\mathfrak{B}(\mathbf{x}) = \theta^*_{1,\mathbf{x}}(0) - \theta^*_{0,\mathbf{x}}(0) - \tau(\mathbf{x}),$$
$$\mathfrak{L}(\mathbf{x}) = \mathbf{e}_0^\top \boldsymbol{\Psi}_{1,\mathbf{x}}^{-1}\mathbf{O}_{1,\mathbf{x}} - \mathbf{e}_0^\top \boldsymbol{\Psi}_{0,\mathbf{x}}^{-1}\mathbf{O}_{0,\mathbf{x}},$$
$$\mathfrak{Q}(\mathbf{x}) = \mathbf{e}_0^\top (\widehat{\boldsymbol{\Psi}}_{1,\mathbf{x}}^{-1} - \boldsymbol{\Psi}_{1,\mathbf{x}}^{-1})\mathbf{O}_{1,\mathbf{x}} - \mathbf{e}_0^\top (\widehat{\boldsymbol{\Psi}}_{0,\mathbf{x}}^{-1} - \boldsymbol{\Psi}_{0,\mathbf{x}}^{-1})\mathbf{O}_{0,\mathbf{x}}.$$

Here, $\mathfrak{B}(\mathbf{x})$ denotes the nonrandom approximation bias, $\mathfrak{L}(\mathbf{x})$ is an unconditional mean-zero linear statistic, and $\mathfrak{Q}(\mathbf{x})$ 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 $\mathscr{B}$ 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

$$\frac{\mathfrak{B}(\mathbf{x})}{\sqrt{\mathbb{V}[\mathfrak{L}(\mathbf{x})]}} \lesssim \sqrt{nh^2}\mathfrak{B}(\mathbf{x}) \to 0 \tag{4}$$

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 $\mathscr{d}(\cdot,\cdot)$, and the geometric regularity of the boundary $\mathscr{B}$. Consequently, the bias of the distance-based local polynomial estimator may be affected by the shape of the boundary $\mathscr{B}$, regardless of the polynomial order p imposed in Assumption 1.

Fig. 1 illustrates this issue graphically. For a point $\mathbf{x} \in \mathscr{B}$ 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 $\mathscr{d}(\cdot,\cdot)$, a sufficiently small value of r yields a complete arc $\{\mathbf{u} \in \mathscr{A}_1 : \mathscr{d}(\mathbf{u},\mathbf{x}) = r\}$, whereas for larger values of r this arc becomes truncated by the boundary. In the example depicted in Fig. 1, the function $\theta_{1,\mathbf{x}}(r)$ is smooth for $r < r_3$ 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 $\mathscr{B}$ 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 $p \geq 0$ 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 $L \geq 1$, let $\mathscr{P}$ be the class of data generating processes satisfying Assumption 1(i)-(iii) with $\mathscr{X} \subseteq [-L,L]^2$, Assumption 2, and the following additional conditions, with (ii) and (iii) imposed for each $t \in \{0,1\}$:*

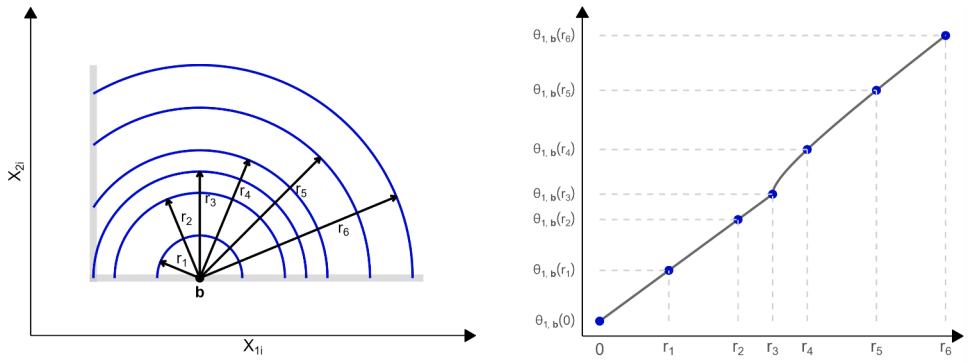
(i) $L^{-1} \leq \inf_{\mathbf{x}\in\mathscr{X}} f_X(\mathbf{x}) \leq \sup_{\mathbf{x}\in\mathscr{X}} f_X(\mathbf{x}) \leq L$,

(ii) $\max_{0\leq|\nu|\leq p} \sup_{\mathbf{x}\in\mathscr{X}} |\partial^\nu \mu_t(\mathbf{x})| + \max_{0\leq|\nu|\leq p} \sup_{\mathbf{x},\mathbf{y}\in\mathscr{X}} \frac{|\partial^\nu \mu_t(\mathbf{x}) - \partial^\nu \mu_t(\mathbf{y})|}{\|\mathbf{x}-\mathbf{y}\|} \leq L$, *and*

(iii) $\liminf_{h\downarrow 0} \inf_{\mathbf{x}\in\mathscr{B}} \int_{\mathscr{A}_t} K_h((2t-1)\mathscr{d}(\mathbf{u},\mathbf{x}))d\mathbf{u} \geq L^{-1}$.

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

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



(a) Distance to $\mathbf{b} \in \mathscr{B}$. (b) Distance-based Conditional Expectation.

Fig. 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 $D_i(\mathbf{b}) = \mathscr{d}(\mathbf{X}_i, \mathbf{b}) = \|\mathbf{X}_i - \mathbf{b}\|$ to point $\mathbf{b} \in \mathscr{B}$ 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 $r = r_3$.

This result precisely characterizes the uniform (over $\mathscr{B}$ 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. Let $R_K < \infty$ satisfy $supp(K) \subseteq [-R_K, R_K]$ and set $\bar{R}_K = \max\{1, R_K\}$. Suppose $\mathbf{X}_i \sim \mathsf{Uniform}([-2\bar{R}_K, 2\bar{R}_K]^2)$, $\mu_0(x_1, x_2) = 0$, $\mu_1(x_1, x_2) = x_2$ for all $(x_1, x_2) \in [-2\bar{R}_K, 2\bar{R}_K]^2$, and $Y_i(0)|\mathbf{X}_i \sim \mathsf{Normal}(\mu_0(\mathbf{X}_i), 1)$ and $Y_i(1)|\mathbf{X}_i \sim \mathsf{Normal}(\mu_1(\mathbf{X}_i), 1)$. Let $\mathscr{d}(\cdot,\cdot)$ be the Euclidean distance, and define the treatment and control regions as $\mathscr{A}_1 = \{(x,y) \in [-2\bar{R}_K, 2\bar{R}_K]^2 : x \geq 0, y \geq 0\}$ and $\mathscr{A}_0 = [-2\bar{R}_K, 2\bar{R}_K]^2 \setminus \mathscr{A}_1$. The resulting assignment boundary $\mathscr{B} = ([0, 2\bar{R}_K] \times \{0\}) \cup (\{0\} \times [0, 2\bar{R}_K])$ 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 $\mathbf{x} \in \mathscr{B}$ 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 $\mathscr{B}$ induce nondifferentiability in $\theta_{t,\mathbf{x}}(r)$, as illustrated in Fig. 1b.

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

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

(i) For $\mathbf{x} \in \mathscr{B}$, and for some $\delta, \varepsilon > 0$, suppose that $\mathscr{B} \cap \{\mathbf{y} : \|\mathbf{y} - \mathbf{x}\| \leq \varepsilon\} = \gamma([-\delta, \delta])$, where $\gamma : \mathbb{R} \to \mathbb{R}^2$ is a one-to-one function in $C^{\kappa+2}([-\delta, \delta], \mathbb{R}^2)$. Let $m_{\mathbf{x}} = \kappa \wedge (p+1)$. Then, for each $t \in \{0,1\}$, $r \mapsto \theta_{t,\mathbf{x}}(r)$ is $m_{\mathbf{x}}$-times continuously differentiable on $\mathscr{I}_t$ near zero. Therefore, there exists a positive constant $C_{\mathbf{x}}$ such that $|\mathfrak{B}(\mathbf{x})| \leq C_{\mathbf{x}} h^{m_{\mathbf{x}}}$ for all sufficiently small h.

(ii) Suppose $\mathscr{B} = \gamma([0, L])$ 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 $\mathbf{x} \in \mathscr{B}^o = \gamma([\delta, L - \delta])$, $r \in [0, \varepsilon]$, and $t = 0, 1$, the set $\{\mathbf{u} \in \mathbb{R}^2 : \mathscr{d}(\mathbf{u}, \mathbf{x}) = r\}$ intersects $\mathtt{bd}(\mathscr{A}_t)$ at only two points in $\mathscr{B}$. Let $m = \iota \wedge (p+1)$. Then, for each $t \in \{0,1\}$, $r \mapsto \theta_{t,\mathbf{x}}(r)$ is m-times continuously differentiable on $\mathscr{I}_t$ near zero with derivatives uniformly bounded over $\mathbf{x} \in \mathscr{B}^o$. In particular, the one-sided limits at zero of the derivatives of order $0 \leq \upsilon \leq m$ exist and are finite. Therefore, there exists a positive constant C such that $\sup_{\mathbf{x} \in \mathscr{B}^o} |\mathfrak{B}(\mathbf{x})| \leq C h^m$ for all sufficiently small h.

This theorem provides sufficient smoothness conditions on the boundary $\mathscr{B}$ 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 Fig. 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 $\mathscr{B}$, 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* $n^{\frac{1+v}{2+v}}h^2/\log(1/h) \to \infty$ *and* $h \to 0$, *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* $\mathbf{x} \in \mathscr{B}$, *and*

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

This theorem establishes pointwise (for each $\mathbf{x} \in \mathscr{B}$) and uniform (over $\mathscr{B}$) convergence rates for the distance-based treatment effect estimator. In particular, by Theorem 2, $\widehat{\vartheta}(\mathbf{x}) \to_{\mathbb{P}} \tau(\mathbf{x})$ both pointwise and uniformly under the stated bandwidth condition. 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 $\mathscr{B}$ 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, approximate pointwise and integrated (over $\mathscr{B}$) MSE criteria can be formulated from the same linearization, with leading variance and squared-bias terms matching the rates in Theorem 4. Details are provided in the supplemental appendix.

5.2. Uncertainty quantification

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

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

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

$$\widehat{\Xi}_{\mathbf{x}_1,\mathbf{x}_2} = \widehat{\Xi}_{0,\mathbf{x}_1,\mathbf{x}_2} + \widehat{\Xi}_{1,\mathbf{x}_1,\mathbf{x}_2}, \qquad \widehat{\Xi}_{t,\mathbf{x}_1,\mathbf{x}_2} = \frac{1}{nh^2}\mathbf{e}_0^\top \widehat{\mathbf{\Psi}}_{t,\mathbf{x}_1}^{-1} \widehat{\mathbf{\Upsilon}}_{t,\mathbf{x}_1,\mathbf{x}_2} \widehat{\mathbf{\Psi}}_{t,\mathbf{x}_2}^{-1} \mathbf{e}_0,$$

and

$$\begin{aligned}\widehat{\Upsilon}_{t,\mathbf{x}_1,\mathbf{x}_2} = h^2\mathbb{E}_n\Big[&\mathbf{r}_p\Big(\frac{D_i(\mathbf{x}_1)}{h}\Big)K_h(D_i(\mathbf{x}_1))\big(Y_i - \mathbf{r}_p(D_i(\mathbf{x}_1)/h)^\top \widehat{\boldsymbol{\gamma}}_t(\mathbf{x}_1)\big)\mathbb{1}(D_i(\mathbf{x}_1) \in \mathscr{I}_t)\\ &\times \mathbf{r}_p\Big(\frac{D_i(\mathbf{x}_2)}{h}\Big)^\top K_h(D_i(\mathbf{x}_2))\big(Y_i - \mathbf{r}_p(D_i(\mathbf{x}_2)/h)^\top \widehat{\boldsymbol{\gamma}}_t(\mathbf{x}_2)\big)\mathbb{1}(D_i(\mathbf{x}_2) \in \mathscr{I}_t)\Big].\end{aligned}$$

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

$$\widehat{I}_\alpha(\mathbf{x}) = \Big[\widehat{\vartheta}(\mathbf{x}) - \mathfrak{q}_\alpha\sqrt{\widehat{\Xi}_{\mathbf{x},\mathbf{x}}}\,,\ \widehat{\vartheta}(\mathbf{x}) + \mathfrak{q}_\alpha\sqrt{\widehat{\Xi}_{\mathbf{x},\mathbf{x}}}\Big], \qquad \mathbf{x} \in \mathscr{B},$$

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

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

Uniform inference over $\mathscr{B}$ is more challenging. The stochastic process $(\widehat{T}(\mathbf{x}) : \mathbf{x} \in \mathscr{B})$ is generally not asymptotically tight and therefore does not converge weakly in the space of uniformly bounded real-valued functions on $\mathscr{B}$ 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 $\mathscr{B}$ and the reliance on signed distance scores introduce additional technical complications. We address these issues through Gaussian approximations for the supremum of $(\widehat{T}(\mathbf{x}) : \mathbf{x} \in \mathscr{B})$, with a stronger process-level coupling provided in the supplemental appendix. These approximations allow uniform inference based on critical values $\mathfrak{q}_\alpha$ satisfying

$$\mathbb{P}\big[\tau(\mathbf{x}) \in \widehat{I}_\alpha(\mathbf{x}) \text{ , for all } \mathbf{x} \in \mathscr{B}\big] = \mathbb{P}\Big[\sup_{\mathbf{x}\in\mathscr{B}} \big|\widehat{T}(\mathbf{x})\big| \le \mathfrak{q}_\alpha\Big].$$

Our technical analysis combines a process-level strong approximation (Theorem SA-7 in the supplemental appendix), the corresponding confidence-band result (Theorem SA-8), and a direct supremum approximation (Theorem SA-9) 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 $\big((Y_i, (D_i(\mathbf{x}) : \mathbf{x} \in \mathscr{B})) : 1 \le i \le n\big)$.

Theorem 5 (Statistical Inference). *Suppose Assumptions 1 and 2 hold, and let $h \to 0$.*

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

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

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

(ii) If $\liminf_{n\to\infty} \frac{\log h}{\log n} > -\infty$, $\frac{n^{\frac{v}{2+v}} h^2}{(\log n)^3} \to \infty$, $nh^2 \log n \sup_{\mathbf{x}\in\mathscr{B}} \mathfrak{B}^2(\mathbf{x}) \to 0$, and $\mathit{perim}(\{\mathbf{y} \in \mathscr{A}_t : (2t-1)\mathscr{d}(\mathbf{y},\mathbf{x})/h \in \mathit{supp}(K)\}) \lesssim h$ for all $\mathbf{x} \in \mathscr{B}$ and $t \in \{0,1\}$, then

$$\mathbb{P}\big[\tau(\mathbf{x}) \in \widehat{I}_\alpha(\mathbf{x}),\ \text{for all } \mathbf{x} \in \mathscr{B}\big] \to 1-\alpha,$$

for $\mathfrak{q}_\alpha = \inf\{c > 0 : \mathbb{P}[\sup_{\mathbf{x}\in\mathscr{B}} |\widehat{Z}_n(\mathbf{x})| \ge c|\mathcal{U}_n] \le \alpha\}$, where $(\widehat{Z}_n : \mathbf{x} \in \mathscr{B})$ is a Gaussian process conditional on $\mathcal{U}_n$ satisfying $\mathbb{E}[\widehat{Z}_n(\mathbf{x}_1)|\mathcal{U}_n] = 0$ and $\mathbb{E}[\widehat{Z}_n(\mathbf{x}_1)\widehat{Z}_n(\mathbf{x}_2)|\mathcal{U}_n] = \widehat{\Xi}_{\mathbf{x}_1,\mathbf{x}_2}/\sqrt{\widehat{\Xi}_{\mathbf{x}_1,\mathbf{x}_1}\widehat{\Xi}_{\mathbf{x}_2,\mathbf{x}_2}}$ for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathscr{B}$.

This theorem establishes asymptotically valid inference procedures based on the distance-based local polynomial treatment-effect estimator $\widehat{\vartheta}(\mathbf{x})$. For uniform inference, an additional geometric restriction on the assignment boundary $\mathscr{B}$ is required. Specifically, the De Giorgi perimeter condition can be verified when the boundary of the set $\{\mathbf{y} \in \mathscr{A}_t : (2t-1)\mathscr{d}(\mathbf{y},\mathbf{x})/h \in \mathit{supp}(K)\}$ has length of order h. Because this set is contained in a constant multiple of an h-ball centered at $\mathbf{x}$ when K is compactly supported, the condition holds provided that the local segment $\mathscr{B} \cap \{\mathbf{y} \in \mathbb{R}^2 : \mathscr{d}(\mathbf{y},\mathbf{x}) \lesssim h\}$ is not excessively irregular. Intuitively, the one-dimensional boundary $\mathscr{B}$ must not be locally "too long" or highly oscillatory.

5.3. Discussion and implementation

Although the distance-based estimator $\widehat{\vartheta}(\mathbf{x})$ 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 $\widehat{T}(\mathbf{x})$ 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 $\mathscr{B}$ (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 $\mathscr{B}$ 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 $h_{\mathtt{1d},\mathbf{x}} \asymp n^{-1/(3+2p)}$, whereas the *correct* MSE-optimal bandwidth is $h_{\mathtt{MSE},\mathbf{x}} \asymp n^{-1/(4+2p)}$. 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 $\mathscr{B}$ 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 $\mathscr{B}$, complicating automatic implementation. A simple global strategy is therefore to set $h = \widehat{\mathtt{C}} \cdot n^{-1/4}$ for all $\mathbf{x} \in \mathscr{B}$, where $\widehat{\mathtt{C}}$ 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

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

for all $\mathbf{x} \in \mathscr{B}$, where $\mathscr{B}_{\mathtt{kink}}$ 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 $\mathscr{B}$ 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 $\mathscr{B}$ 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 $\mathbf{x} \in \mathscr{B}$. 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 $\mathbf{x} \in \mathscr{B}$ when the smoothness of $\mathscr{B}$ 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 $\mathscr{B}$ 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 $\hat{h}_{\mathtt{kink},\mathbf{x}}(\mathscr{B})$ 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 $\mathbf{b}_1, \dots, \mathbf{b}_M \in \mathscr{B}$. The conditional Gaussian process $(\hat{Z}_n(\mathbf{x}) : \mathbf{x} \in \mathscr{B})$ 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 $\mathbb{E}[\hat{Z}_n(\mathbf{b}_1)\hat{Z}_n(\mathbf{b}_2)|\mathcal{U}_n]$. 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 $\mathfrak{q}_\alpha$ are then obtained by simulating the distribution of $\max_{1\le l\le M} |\hat{Z}_n(\mathbf{b}_l)|$.

6. Distance-based minimax convergence rate

Theorems 2 and 3 provide precise conditions characterizing the pointwise and uniform (over $\mathscr{B}$) bias and convergence rates of the distance-based local polynomial estimator $\hat{\vartheta}(\mathbf{x})$. 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 $\mathscr{B}$ 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 $q \ge 1$ and $L \ge 1$, let $\mathscr{P}_{\mathtt{NP}} = \mathscr{P}_{\mathtt{NP}}(L, q)$ 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) $((Y_i, \mathbf{X}_i) : 1 \le i \le n)$ *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 $L^{-1} \le \inf_{\mathbf{x}\in\mathcal{X}} f(\mathbf{x}) \le \sup_{\mathbf{x}\in\mathcal{X}} f(\mathbf{x}) \le L$, and $\mathscr{B} = \mathtt{bd}(\mathcal{X})$ is a rectifiable curve.*

(iii) $\mu(\mathbf{x}) = \mathbb{E}[Y_i|\mathbf{X}_i = \mathbf{x}]$ *belongs to a Hölder ball of smoothness q on $\mathcal{X}$ with*

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

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

In addition, let $\mathscr{T}$ be the class of all distance-based estimators $T_n(\mathbf{U}_n(\mathbf{x}))$ with $\mathbf{U}_n(\mathbf{x}) = ((Y_i, \|\mathbf{X}_i - \mathbf{x}\|)^\top : 1 \le i \le n)$ for each $\mathbf{x} \in \mathcal{X}$. Then,

$$\liminf_{n\to\infty} n^{1/4} \inf_{T_n\in\mathscr{T}} \sup_{\mathbb{P}\in\mathscr{P}_{\mathtt{NP}}} \mathbb{E}_{\mathbb{P}}\Big[\sup_{\mathbf{x}\in\mathscr{B}} \Big|T_n(\mathbf{U}_n(\mathbf{x})) - \mu(\mathbf{x})\Big|\Big] \gtrsim 1,$$

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

$$\liminf_{n\to\infty} \Big(\frac{n}{\log n}\Big)^{\frac{q}{2q+2}} \inf_{S_n\in\mathcal{S}} \sup_{\mathbb{P}\in\mathscr{P}_{\mathtt{NP}}} \mathbb{E}_{\mathbb{P}}\Big[\sup_{\mathbf{x}\in\mathscr{B}} \Big|S_n(\mathbf{x}; \mathbf{V}_n) - \mu(\mathbf{x})\Big|\Big] \gtrsim 1,$$

where $\mathcal{S}$ denotes the unrestricted class of estimators based on $\mathbf{V}_n = ((Y_i, \mathbf{X}_i^\top)^\top : 1 \le i \le n)$. 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 $\mathscr{B}$ 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 $\mathscr{T}$, 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 2 and 4, the distance-based pth order local polynomial estimator

$$\hat{\mu}(\mathbf{x}) = \mathbf{e}_0^\top \hat{\boldsymbol{\gamma}}(\mathbf{x}), \qquad \hat{\boldsymbol{\gamma}}(\mathbf{x}) = \arg\min_{\boldsymbol{\gamma}\in\mathbb{R}^{p+1}} \mathbb{E}_n\Big[\big(Y_i - \mathbf{r}_p(D_i(\mathbf{x})/h)^\top\boldsymbol{\gamma}\big)^2 K_h(D_i(\mathbf{x}))\Big], \qquad \mathbf{x} \in \mathscr{B},$$

satisfies

$$\limsup_{M\to\infty}\limsup_{n\to\infty}\sup_{\mathbb{P}\in\mathscr{P}}\mathbb{P}\Big[\Big(\frac{n}{\log n}\Big)^{1/4}\sup_{\mathbf{x}\in\mathscr{B}}\big|\widehat{\mu}(\mathbf{x})-\mu(\mathbf{x})\big|\geq M\mathscr{B}ig]=0.$$

Since $\widehat{\mu}(\mathbf{x})\in\mathscr{T}$, 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 $\mathscr{B}$ 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 $\mathbf{X}_i=(X_{1i},X_{2i})^\top=(\texttt{SABER11}_i,\texttt{SISBEN}_i)^\top$, 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 $\mathscr{B}=\{(\texttt{SABER11},\texttt{SISBEN}):(\texttt{SABER11}\geq 0,\ \texttt{SISBEN}=0)\}\ \cup\ \{(\texttt{SABER11}=0,\ \texttt{SISBEN}\geq 0)\}$. Because the support of $\mathscr{X}$ is compact, the boundary $\mathscr{B}$ contains three kinks: the interior point $(0,0)$, which is of primary interest in our analysis, and two additional points where $\mathscr{B}$ intersects the boundary of the support of $\mathscr{X}$. Our empirical analysis considers 21 evenly spaced cutoff points $\{\mathbf{b}_1,\dots,\mathbf{b}_{21}\}\subset\mathscr{B}$, located near the interior kink and away from the boundary of $\mathscr{X}$, in order to highlight the impact of a single kink on the bias of the distance-based estimator. Throughout, we employ the Euclidean distance, $\mathscr{d}(\mathbf{X}_i,\mathbf{x})=\|\mathbf{X}_i-\mathbf{x}\|$, to construct the signed score $D_i(\mathbf{x})$ for each $\mathbf{x}\in\mathscr{B}$.

We analyze both simulated data calibrated to match salient features of the SPP dataset and the 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/. Fig. 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)–(d) 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 $(100\cdot\mathsf{Beta}(3,4)-25,\ 100\cdot\mathsf{Beta}(3,4)-25)^\top$ 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 $\mathbf{b}_{11}$.

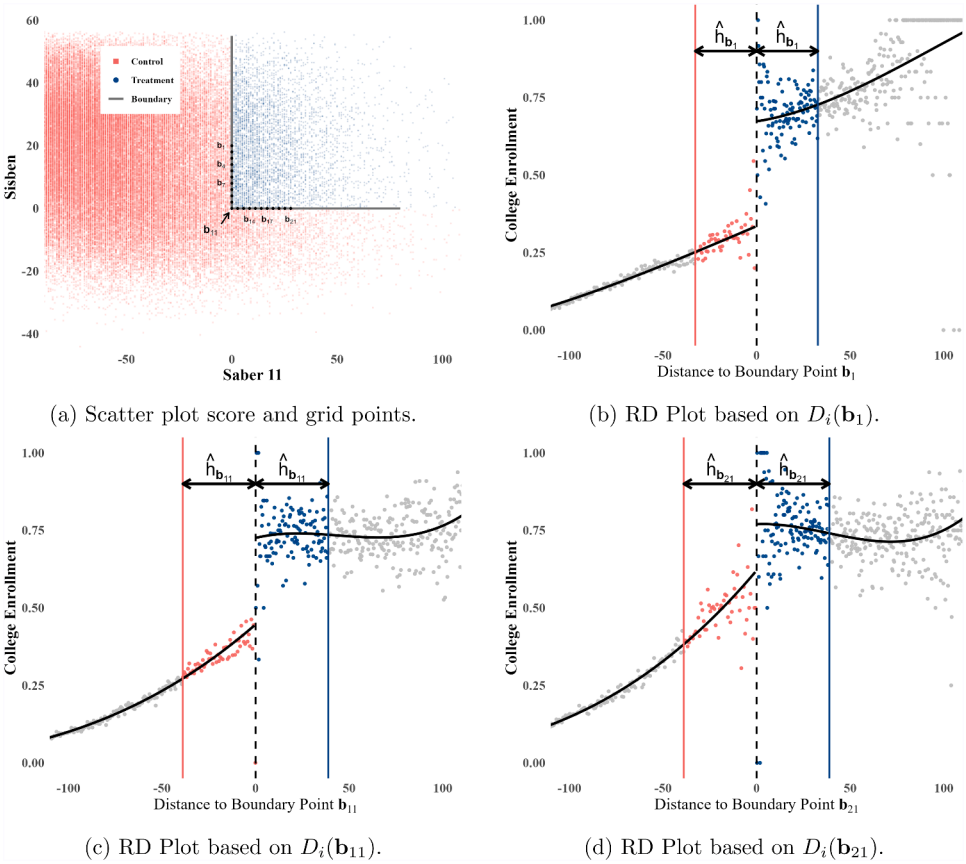
For each $t\in\{0,1\}$ and $i=1,\dots,n$, potential outcomes are generated according to the regression model $Y_i(t)=\mu_t(\mathbf{X}_i)+\sigma_t(\mathbf{X}_i)u_{t,i}$, where $\mathbf{X}_i$, $u_{0,i}$ and $u_{1,i}$ are mutually independent, and

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

with $u_{t,i}\sim\mathsf{Normal}(0,1)$. 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., $\delta_{t,11}=\delta_{t,12}=\delta_{t,21}=\delta_{t,22}=\delta_{t,23}=0$) or heteroskedastic disturbances. Table 1 reports the corresponding parameter values.

The simulation design sets $n=20,000$ and uses 5,000 Monte Carlo replications. We evaluate performance in terms of point estimation and uncertainty quantification, both pointwise and uniformly over the boundary $\mathscr{B}$. 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 use the selector $h=\widehat{h}_{\texttt{MSE},\mathbf{b}}$. Panels (b) report results based on the kink-adaptive selector $h=\widehat{h}_{\texttt{kink},\mathbf{b}}(\mathscr{B})$, which exploits knowledge of the kink location. Panels (c) use the rule $h=\widehat{\mathrm{C}}\cdot n^{-1/4}$, which remains valid regardless of whether kinks are present in $\mathscr{B}$. Panels (d) employ the bandwidth $h=h_{\texttt{1d},\mathbf{b}}$ 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 Fig. 3, which reports the average value of each point estimator across the 5,000 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 $\mathscr{B}$ 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



(a) Scatter plot score and grid points.

(b) RD Plot based on $D_i(\mathbf{b}_1)$.

(c) RD Plot based on $D_i(\mathbf{b}_{11})$.

(d) RD Plot based on $D_i(\mathbf{b}_{21})$.

Fig. 2. Scatter plot and selected distance-based RD plots (SPP data).

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 Fig. 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 $\mathbf{b}_{11}$–$\mathbf{b}_{21}$).

8. Extensions

The theoretical results developed in the supplemental appendix primarily cover multidimensional location scores with $d \geq 2$, recover the usual one-dimensional RD case for the results not tied to continuum boundary averaging, and also provide several theoretical and methodological extensions of our results. This section highlights three such extensions that are of practical interest.

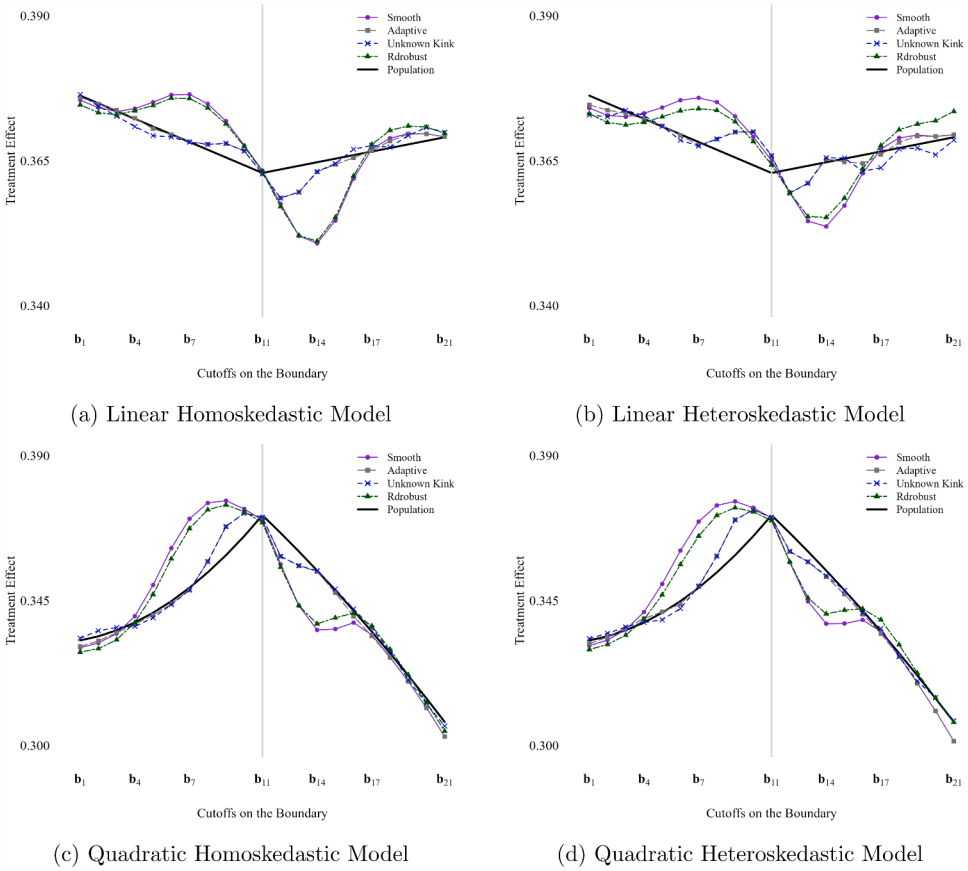
8.1. Aggregated treatment effects

As discussed in Cattaneo et al. (2026b), and references therein, two other interesting causal parameters are the *Weighted Boundary Average Treatment Effect* (WBATE)

$$\tau_{\texttt{WBATE}} = \frac{\int_{\mathscr{B}} \tau(\mathbf{x}) w(\mathbf{x})\, \mathrm{d}\mathfrak{H}^1(\mathbf{x})}{\int_{\mathscr{B}} w(\mathbf{x})\, \mathrm{d}\mathfrak{H}^1(\mathbf{x})}$$

Table 1
Parameters calibrated using the SPP dataset.

(a) Population parameters: $\mu_t(\mathbf{X}_i)$				
	Linear		Quadratic	
	$t=0$	$t=1$	$t=0$	$t=1$
$\beta_{Y,t,0}$	3.35×10^{-1}	6.98×10^{-1}	3.72×10^{-1}	7.44×10^{-1}
$\beta_{Y,t,11}$	2.52×10^{-3}	2.74×10^{-3}	4.23×10^{-3}	2.29×10^{-3}
$\beta_{Y,t,12}$	-1.27×10^{-3}	-6.05×10^{-4}	-2.45×10^{-3}	-5.84×10^{-3}
$\beta_{Y,t,21}$	0	0	1.25×10^{-5}	-1.33×10^{-7}
$\beta_{Y,t,22}$	0	0	-4.92×10^{-6}	2.14×10^{-5}
$\beta_{Y,t,23}$	0	0	3.12×10^{-5}	1.04×10^{-4}
(b) Population parameters: $\sigma^2_t(\mathbf{X}_i)$				
	Homoskedastic		Heteroskedastic	
	$t=0$	$t=1$	$t=0$	$t=1$
$\delta_{t,0}$	-2.20	-2.22	-1.57	-5.00
$\delta_{t,11}$	0	0	2.19×10^{-2}	-9.03×10^{-2}
$\delta_{t,12}$	0	0	-5.08×10^{-3}	1.11×10^{-1}
$\delta_{t,21}$	0	0	-1.15×10^{-4}	1.05×10^{-3}
$\delta_{t,22}$	0	0	5.23×10^{-4}	-7.46×10^{-4}
$\delta_{t,23}$	0	0	6.50×10^{-4}	-1.67×10^{-3}



(a) Linear Homoskedastic Model (b) Linear Heteroskedastic Model

(c) Quadratic Homoskedastic Model (d) Quadratic Heteroskedastic Model

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

where $w : \mathscr{B} \to (0, \infty)$ is a bounded weight function and integration is with respect to one-dimensional Hausdorff measure, and the *Largest Boundary Average Treatment Effect* (LBATE)

$$\tau_{\texttt{LBATE}} = \sup_{\mathbf{x} \in \mathscr{B}} \tau(\mathbf{x}).$$

Table 2
Linear homoskedastic model (Simulation results).

(a) $h = \widehat{h}_{\texttt{MSE},\mathbf{b}}$ (smooth $\mathscr{B}$)									
	h_0	h_1	$N_{\texttt{Co}}$	$N_{\texttt{Tr}}$	Bias	SD	RMSE	EC	IL
$\mathbf{b}_1$	19.531	19.531	2120	4259	−0.001	0.033	0.033	0.958	0.240
$\mathbf{b}_2$	19.133	19.133	2135	4107	−0.001	0.034	0.034	0.956	0.243
$\mathbf{b}_3$	18.879	18.879	2182	3953	−0.000	0.035	0.035	0.955	0.245
$\mathbf{b}_4$	18.784	18.784	2267	3798	0.002	0.036	0.036	0.952	0.246
$\mathbf{b}_5$	18.826	18.826	2383	3637	0.004	0.035	0.036	0.947	0.246
$\mathbf{b}_6$	18.978	18.978	2512	3469	0.007	0.035	0.035	0.953	0.245
$\mathbf{b}_7$	19.090	19.090	2617	3257	0.008	0.034	0.035	0.960	0.247
$\mathbf{b}_8$	19.093	19.093	2675	2989	0.008	0.035	0.036	0.959	0.252
$\mathbf{b}_9$	19.128	19.128	2711	2714	0.006	0.036	0.037	0.953	0.258
$\mathbf{b}_{10}$	19.376	19.376	2758	2482	0.003	0.038	0.038	0.955	0.268
$\mathbf{b}_{11}$	21.264	21.264	3050	2652	0.000	0.038	0.038	0.951	0.279
$\mathbf{b}_{12}$	19.240	19.240	2736	2567	−0.006	0.037	0.038	0.955	0.263
$\mathbf{b}_{13}$	19.204	19.204	2705	2964	−0.012	0.036	0.038	0.954	0.252
$\mathbf{b}_{14}$	19.259	19.259	2637	3357	−0.014	0.035	0.038	0.951	0.245
$\mathbf{b}_{15}$	19.026	19.026	2464	3617	−0.011	0.035	0.037	0.953	0.244
$\mathbf{b}_{16}$	18.641	18.641	2239	3751	−0.004	0.035	0.036	0.955	0.248
$\mathbf{b}_{17}$	18.403	18.403	2065	3814	0.000	0.035	0.035	0.950	0.251
$\mathbf{b}_{18}$	18.665	18.665	2013	3929	0.002	0.033	0.034	0.951	0.250
$\mathbf{b}_{19}$	19.455	19.455	2039	4153	0.002	0.034	0.034	0.953	0.245
$\mathbf{b}_{20}$	20.632	20.632	2094	4449	0.001	0.033	0.033	0.948	0.237
$\mathbf{b}_{21}$	21.791	21.791	2110	4669	0.000	0.033	0.033	0.950	0.232
`Uniform`								0.946	0.373
`WBATE`					−0.000	0.017	0.017	0.948	0.093
`LBATE`					0.043	0.027	0.050	0.984	0.390
(b) $h = \widehat{h}_{\texttt{kink},\mathbf{b}}(\mathscr{B})$ (kink adaptive)									
	h_0	h_1	$N_{\texttt{Co}}$	$N_{\texttt{Tr}}$	Bias	SD	RMSE	EC	IL
$\mathbf{b}_1$	18.845	18.845	2020	3988	−0.000	0.034	0.034	0.958	0.247
$\mathbf{b}_2$	17.539	17.539	1876	3510	−0.000	0.035	0.035	0.953	0.261
$\mathbf{b}_3$	15.909	15.909	1661	2913	0.000	0.038	0.038	0.948	0.285
$\mathbf{b}_4$	13.996	13.996	1382	2256	0.000	0.043	0.043	0.950	0.322
$\mathbf{b}_5$	12.000	12.000	1081	1641	−0.000	0.050	0.050	0.947	0.375
$\mathbf{b}_6$	10.000	10.000	788	1113	−0.000	0.060	0.060	0.947	0.451
$\mathbf{b}_7$	8.547	8.547	594	781	0.000	0.071	0.071	0.942	0.532
$\mathbf{b}_8$	8.365	8.365	597	678	0.001	0.075	0.075	0.939	0.559
$\mathbf{b}_9$	8.380	8.380	641	580	0.002	0.077	0.077	0.942	0.572
$\mathbf{b}_{10}$	8.489	8.489	701	479	0.002	0.080	0.080	0.942	0.586
$\mathbf{b}_{11}$	9.316	9.316	871	442	−0.000	0.086	0.086	0.931	0.636
$\mathbf{b}_{12}$	8.429	8.429	675	519	−0.005	0.078	0.078	0.939	0.580
$\mathbf{b}_{13}$	8.414	8.414	612	667	−0.005	0.075	0.075	0.940	0.557
$\mathbf{b}_{14}$	8.724	8.724	615	822	−0.002	0.070	0.070	0.942	0.522
$\mathbf{b}_{15}$	11.200	11.200	961	1419	−0.001	0.054	0.054	0.941	0.402
$\mathbf{b}_{16}$	13.996	13.996	1382	2258	−0.000	0.043	0.043	0.942	0.322
$\mathbf{b}_{17}$	16.487	16.487	1740	3124	0.000	0.037	0.037	0.946	0.275
$\mathbf{b}_{18}$	18.113	18.113	1931	3712	0.001	0.034	0.034	0.951	0.256
$\mathbf{b}_{19}$	19.403	19.403	2033	4132	0.002	0.034	0.034	0.953	0.245
$\mathbf{b}_{20}$	20.632	20.632	2094	4449	0.001	0.033	0.033	0.948	0.237
$\mathbf{b}_{21}$	21.791	21.791	2110	4669	0.000	0.033	0.033	0.950	0.232

Table 2
Continued

Uniform								0.934	0.616
WBATE					−0.000	0.021	0.021	0.952	0.120
LBATE					0.089	0.046	0.100	0.973	0.757
(c) $h = \widehat{C} \cdot n^{-1/4}$ (unknown kink location)									
	h_0	h_1	N_{Co}	N_{Tr}	Bias	SD	RMSE	EC	IL
$\mathbf{b}_1$	8.557	8.557	609	823	0.000	0.069	0.069	0.957	0.276
$\mathbf{b}_2$	8.382	8.382	597	801	−0.000	0.072	0.072	0.950	0.280
$\mathbf{b}_3$	8.271	8.271	588	786	−0.001	0.074	0.074	0.949	0.282
$\mathbf{b}_4$	8.229	8.229	582	777	−0.001	0.073	0.073	0.956	0.284
$\mathbf{b}_5$	8.248	8.248	579	772	−0.002	0.074	0.074	0.952	0.283
$\mathbf{b}_6$	8.314	8.314	576	770	−0.000	0.073	0.073	0.952	0.283
$\mathbf{b}_7$	8.364	8.364	574	750	−0.000	0.073	0.073	0.951	0.285
$\mathbf{b}_8$	8.365	8.365	597	678	0.001	0.075	0.075	0.951	0.291
$\mathbf{b}_9$	8.380	8.380	641	580	0.002	0.077	0.077	0.950	0.298
$\mathbf{b}_{10}$	8.489	8.489	701	479	0.002	0.080	0.080	0.951	0.309
$\mathbf{b}_{11}$	9.316	9.316	871	442	−0.000	0.086	0.086	0.947	0.330
$\mathbf{b}_{12}$	8.429	8.429	675	519	−0.005	0.078	0.078	0.950	0.303
$\mathbf{b}_{13}$	8.414	8.414	612	667	−0.005	0.075	0.075	0.953	0.291
$\mathbf{b}_{14}$	8.437	8.437	582	773	−0.002	0.074	0.074	0.949	0.282
$\mathbf{b}_{15}$	8.335	8.335	585	784	−0.001	0.074	0.074	0.945	0.281
$\mathbf{b}_{16}$	8.167	8.167	574	765	0.001	0.075	0.075	0.944	0.286
$\mathbf{b}_{17}$	8.063	8.063	561	746	0.001	0.075	0.075	0.947	0.289
$\mathbf{b}_{18}$	8.177	8.177	565	755	0.000	0.075	0.075	0.947	0.288
$\mathbf{b}_{19}$	8.523	8.523	588	795	0.002	0.073	0.073	0.947	0.282
$\mathbf{b}_{20}$	9.039	9.039	621	853	0.002	0.071	0.071	0.947	0.272
$\mathbf{b}_{21}$	9.547	9.547	640	896	0.001	0.069	0.069	0.945	0.266
Uniform								0.944	0.438
WBATE					−0.000	0.024	0.024	0.953	0.095
LBATE					0.123	0.049	0.132	0.981	0.457
(d) $h = \widehat{h}_{1d,\mathbf{b}}$ (rdrobust + rd2d.dist)									
	h_0	h_1	N_{Co}	N_{Tr}	Bias	SD	RMSE	EC	IL
$\mathbf{b}_1$	19.346	19.346	2103	4191	−0.002	0.035	0.035	0.960	0.244
$\mathbf{b}_2$	19.187	19.187	2155	4130	−0.002	0.036	0.036	0.957	0.245
$\mathbf{b}_3$	19.183	19.183	2247	4066	−0.001	0.037	0.037	0.958	0.244
$\mathbf{b}_4$	19.301	19.301	2372	3985	0.001	0.037	0.037	0.959	0.244
$\mathbf{b}_5$	19.499	19.499	2514	3876	0.004	0.037	0.038	0.953	0.242
$\mathbf{b}_6$	19.771	19.771	2662	3744	0.006	0.037	0.037	0.953	0.241
$\mathbf{b}_7$	20.043	20.043	2791	3576	0.008	0.036	0.037	0.960	0.241
$\mathbf{b}_8$	20.290	20.290	2888	3372	0.007	0.036	0.037	0.958	0.243
$\mathbf{b}_9$	20.611	20.611	2967	3165	0.006	0.037	0.037	0.955	0.246
$\mathbf{b}_{10}$	21.431	21.431	3097	3076	0.003	0.037	0.038	0.957	0.248
$\mathbf{b}_{11}$	23.256	23.256	3339	3247	0.000	0.038	0.038	0.956	0.260
$\mathbf{b}_{12}$	21.129	21.129	3057	3118	−0.006	0.037	0.038	0.957	0.246
$\mathbf{b}_{13}$	20.504	20.504	2936	3371	−0.012	0.037	0.039	0.956	0.242
$\mathbf{b}_{14}$	20.155	20.155	2800	3662	−0.014	0.037	0.039	0.953	0.240
$\mathbf{b}_{15}$	19.764	19.764	2605	3882	−0.010	0.038	0.039	0.956	0.241
$\mathbf{b}_{16}$	19.102	19.102	2335	3923	−0.004	0.038	0.038	0.957	0.247
$\mathbf{b}_{17}$	18.429	18.429	2083	3831	0.001	0.037	0.037	0.956	0.254
$\mathbf{b}_{18}$	18.250	18.250	1960	3778	0.003	0.036	0.036	0.954	0.258
$\mathbf{b}_{19}$	18.549	18.549	1927	3806	0.003	0.036	0.036	0.952	0.257
$\mathbf{b}_{20}$	19.074	19.074	1909	3854	0.002	0.037	0.038	0.949	0.256
$\mathbf{b}_{21}$	19.668	19.668	1873	3873	0.001	0.038	0.038	0.949	0.258
Uniform								0.949	0.371
WBATE					−0.000	0.017	0.017	0.945	0.092
LBATE					0.047	0.031	0.057	0.983	0.409

Because Theorem 1 identifies the BATEC using distance-based methods, these functionals are identified as well, which justifies the plug-in estimators

$$\widehat{\vartheta}_{\texttt{WBATE}} = \frac{\int_{\mathscr{B}} \widehat{\vartheta}(\mathbf{x}) w(\mathbf{x})\, \mathrm{d}\mathfrak{H}^1(\mathbf{x})}{\int_{\mathscr{B}} w(\mathbf{x})\, \mathrm{d}\mathfrak{H}^1(\mathbf{x})} \qquad \text{and} \qquad \widehat{\vartheta}_{\texttt{LBATE}} = \sup_{\mathbf{x}\in\mathscr{B}} \widehat{\vartheta}(\mathbf{x})$$

of $\tau_{\texttt{WBATE}}$ and $\tau_{\texttt{LBATE}}$, respectively. Sections SA-4 and SA-5 in the supplemental appendix give estimation and inference results for distance-based methods for WBATE and LBATE.

Table 3
Linear heteroskedastic model (Simulation results).

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

	h_0	h_1	$N_{\texttt{Co}}$	$N_{\texttt{Tr}}$	Bias	SD	RMSE	EC	IL
$\mathbf{b}_1$	19.917	19.917	2174	4411	−0.002	0.044	0.044	0.956	0.312
$\mathbf{b}_2$	19.329	19.329	2166	4181	−0.002	0.044	0.045	0.958	0.315
$\mathbf{b}_3$	18.839	18.839	2175	3939	−0.001	0.045	0.045	0.957	0.318
$\mathbf{b}_4$	18.488	18.488	2211	3699	0.001	0.046	0.046	0.953	0.320
$\mathbf{b}_5$	18.264	18.264	2270	3459	0.003	0.046	0.046	0.954	0.321
$\mathbf{b}_6$	18.133	18.133	2338	3215	0.006	0.046	0.046	0.958	0.323
$\mathbf{b}_7$	18.010	18.010	2394	2947	0.008	0.046	0.046	0.960	0.325
$\mathbf{b}_8$	17.712	17.712	2394	2613	0.008	0.047	0.047	0.956	0.332
$\mathbf{b}_9$	17.533	17.533	2395	2303	0.007	0.048	0.048	0.949	0.339
$\mathbf{b}_{10}$	17.461	17.461	2393	2015	0.005	0.048	0.049	0.954	0.343
$\mathbf{b}_{11}$	18.552	18.552	2578	1984	0.002	0.048	0.048	0.949	0.340
$\mathbf{b}_{12}$	17.414	17.414	2382	2113	−0.004	0.050	0.050	0.953	0.351
$\mathbf{b}_{13}$	17.561	17.561	2372	2520	−0.010	0.050	0.051	0.959	0.351
$\mathbf{b}_{14}$	17.865	17.865	2349	2953	−0.011	0.051	0.052	0.953	0.349
$\mathbf{b}_{15}$	17.984	17.984	2251	3296	−0.008	0.052	0.053	0.957	0.354
$\mathbf{b}_{16}$	18.195	18.195	2153	3603	−0.003	0.053	0.053	0.956	0.361
$\mathbf{b}_{17}$	18.627	18.627	2099	3896	0.000	0.054	0.054	0.954	0.367
$\mathbf{b}_{18}$	19.308	19.308	2100	4180	0.002	0.054	0.054	0.952	0.372
$\mathbf{b}_{19}$	20.440	20.440	2160	4548	0.002	0.053	0.053	0.952	0.371
$\mathbf{b}_{20}$	21.899	21.899	2232	4967	0.001	0.053	0.053	0.951	0.370
$\mathbf{b}_{21}$	23.484	23.484	2276	5358	0.000	0.054	0.054	0.957	0.371
`Uniform`								0.950	0.516
`WBATE`					0.000	0.022	0.022	0.952	0.121
`LBATE`					0.066	0.038	0.076	0.982	0.553

(b) $h = \widehat{h}_{\texttt{kink},\mathbf{b}}(\mathscr{B})$ (kink adaptive)

	h_0	h_1	$N_{\texttt{Co}}$	$N_{\texttt{Tr}}$	Bias	SD	RMSE	EC	IL
$\mathbf{b}_1$	19.062	19.062	2049	4073	−0.002	0.045	0.045	0.954	0.324
$\mathbf{b}_2$	17.606	17.606	1887	3536	−0.001	0.046	0.046	0.951	0.342
$\mathbf{b}_3$	15.908	15.908	1661	2914	−0.000	0.050	0.050	0.948	0.371
$\mathbf{b}_4$	13.995	13.995	1382	2256	−0.000	0.055	0.055	0.945	0.414
$\mathbf{b}_5$	12.000	12.000	1082	1641	0.000	0.063	0.063	0.943	0.477
$\mathbf{b}_6$	10.000	10.000	789	1112	−0.000	0.075	0.075	0.944	0.571
$\mathbf{b}_7$	8.241	8.241	553	727	−0.001	0.090	0.090	0.937	0.688
$\mathbf{b}_8$	7.762	7.762	511	593	0.002	0.099	0.099	0.942	0.736
$\mathbf{b}_9$	7.681	7.681	537	496	0.004	0.103	0.103	0.934	0.751
$\mathbf{b}_{10}$	7.650	7.650	574	392	0.006	0.102	0.102	0.938	0.769
$\mathbf{b}_{11}$	8.128	8.128	681	328	0.003	0.101	0.101	0.938	0.759
$\mathbf{b}_{12}$	7.629	7.629	554	431	−0.004	0.105	0.105	0.937	0.776
$\mathbf{b}_{13}$	7.694	7.694	508	569	−0.003	0.104	0.105	0.942	0.775
$\mathbf{b}_{14}$	8.450	8.450	578	772	0.001	0.095	0.095	0.941	0.710
$\mathbf{b}_{15}$	11.200	11.200	962	1418	−0.001	0.074	0.074	0.946	0.551
$\mathbf{b}_{16}$	13.992	13.992	1382	2256	−0.001	0.062	0.062	0.943	0.459
$\mathbf{b}_{17}$	16.564	16.564	1751	3153	−0.001	0.056	0.056	0.949	0.407
$\mathbf{b}_{18}$	18.582	18.582	1993	3894	0.001	0.054	0.054	0.950	0.384
$\mathbf{b}_{19}$	20.310	20.310	2144	4494	0.001	0.053	0.053	0.952	0.373
$\mathbf{b}_{20}$	21.890	21.890	2231	4963	0.001	0.053	0.053	0.951	0.370
$\mathbf{b}_{21}$	23.483	23.483	2276	5358	0.000	0.054	0.054	0.957	0.371

Table 3
Continued

Uniform								0.927	0.829
WBATE					0.000	0.027	0.027	0.949	0.153
LBATE					0.127	0.061	0.141	0.967	1.012
(c) $h = \widehat{C} \cdot n^{-1/4}$ (unknown kink location)									
	h_0	h_1	N_{Co}	N_{Tr}	Bias	SD	RMSE	EC	IL
$\mathbf{b}_1$	8.726	8.726	630	856	−0.003	0.093	0.093	0.950	0.356
$\mathbf{b}_2$	8.468	8.468	608	818	−0.002	0.093	0.093	0.948	0.360
$\mathbf{b}_3$	8.253	8.253	585	781	0.000	0.094	0.094	0.949	0.365
$\mathbf{b}_4$	8.100	8.100	566	751	0.001	0.095	0.095	0.947	0.369
$\mathbf{b}_5$	8.002	8.002	549	725	0.000	0.096	0.096	0.949	0.371
$\mathbf{b}_6$	7.944	7.944	532	701	−0.001	0.096	0.096	0.954	0.374
$\mathbf{b}_7$	7.890	7.890	514	670	−0.001	0.096	0.096	0.957	0.377
$\mathbf{b}_8$	7.760	7.760	511	593	0.002	0.099	0.100	0.951	0.384
$\mathbf{b}_9$	7.681	7.681	537	496	0.004	0.103	0.103	0.941	0.393
$\mathbf{b}_{10}$	7.650	7.650	574	392	0.006	0.102	0.102	0.951	0.397
$\mathbf{b}_{11}$	8.128	8.128	681	328	0.003	0.101	0.101	0.954	0.395
$\mathbf{b}_{12}$	7.629	7.629	554	431	−0.004	0.105	0.105	0.949	0.406
$\mathbf{b}_{13}$	7.693	7.693	508	569	−0.003	0.105	0.105	0.955	0.405
$\mathbf{b}_{14}$	7.827	7.827	507	665	0.001	0.104	0.104	0.953	0.403
$\mathbf{b}_{15}$	7.879	7.879	530	698	0.000	0.105	0.105	0.952	0.406
$\mathbf{b}_{16}$	7.971	7.971	550	726	−0.003	0.106	0.106	0.947	0.411
$\mathbf{b}_{17}$	8.161	8.161	573	762	−0.003	0.107	0.107	0.951	0.416
$\mathbf{b}_{18}$	8.459	8.459	600	807	−0.000	0.109	0.109	0.945	0.419
$\mathbf{b}_{19}$	8.955	8.955	641	878	−0.001	0.109	0.109	0.947	0.416
$\mathbf{b}_{20}$	9.594	9.594	688	963	−0.002	0.107	0.107	0.946	0.412
$\mathbf{b}_{21}$	10.288	10.288	727	1044	−0.001	0.106	0.106	0.953	0.411
Uniform								0.942	0.600
WBATE					−0.000	0.032	0.032	0.952	0.124
LBATE					0.175	0.069	0.188	0.979	0.636
(d) $h = \widehat{h}_{1d,\mathbf{b}}$ (rdrobust + rd2d.dist)									
	h_0	h_1	N_{Co}	N_{Tr}	Bias	SD	RMSE	EC	IL
$\mathbf{b}_1$	19.141	19.141	2074	4119	−0.003	0.049	0.049	0.958	0.328
$\mathbf{b}_2$	18.748	18.748	2087	3973	−0.003	0.049	0.049	0.955	0.329
$\mathbf{b}_3$	18.546	18.546	2138	3846	−0.002	0.050	0.050	0.959	0.328
$\mathbf{b}_4$	18.409	18.409	2206	3689	−0.001	0.051	0.051	0.954	0.327
$\mathbf{b}_5$	18.352	18.352	2291	3510	0.002	0.051	0.051	0.954	0.326
$\mathbf{b}_6$	18.342	18.342	2380	3306	0.004	0.050	0.051	0.964	0.325
$\mathbf{b}_7$	17.995	17.995	2387	2973	0.006	0.050	0.050	0.963	0.331
$\mathbf{b}_8$	17.794	17.794	2404	2669	0.007	0.051	0.052	0.957	0.337
$\mathbf{b}_9$	17.707	17.707	2421	2387	0.006	0.053	0.053	0.954	0.343
$\mathbf{b}_{10}$	17.845	17.845	2456	2148	0.004	0.053	0.053	0.957	0.344
$\mathbf{b}_{11}$	18.868	18.868	2619	2107	0.001	0.052	0.052	0.957	0.341
$\mathbf{b}_{12}$	17.862	17.862	2459	2264	−0.004	0.054	0.054	0.956	0.350
$\mathbf{b}_{13}$	18.205	18.205	2495	2731	−0.009	0.055	0.055	0.960	0.347
$\mathbf{b}_{14}$	18.657	18.657	2506	3220	−0.010	0.055	0.056	0.954	0.343
$\mathbf{b}_{15}$	18.922	18.922	2442	3620	−0.007	0.057	0.057	0.960	0.347
$\mathbf{b}_{16}$	18.870	18.870	2297	3851	−0.002	0.058	0.058	0.960	0.359
$\mathbf{b}_{17}$	18.966	18.966	2179	4030	0.001	0.060	0.060	0.959	0.370
$\mathbf{b}_{18}$	19.187	19.187	2096	4149	0.003	0.060	0.061	0.958	0.381
$\mathbf{b}_{19}$	19.676	19.676	2065	4268	0.004	0.061	0.061	0.957	0.390
$\mathbf{b}_{20}$	20.508	20.508	2065	4446	0.003	0.063	0.063	0.953	0.399
$\mathbf{b}_{21}$	21.347	21.347	2041	4551	0.004	0.065	0.065	0.959	0.411
Uniform								0.949	0.528
WBATE					0.000	0.023	0.023	0.947	0.122
LBATE					0.080	0.052	0.095	0.980	0.616

8.2. Imperfect compliance

Our theoretical framework can also accommodate imperfect compliance (fuzzy designs), where treatment assignment and treatment receipt may differ for some units. Using standard potential-outcomes notation (Hernán and Robins, 2020), for any $\mathbf{x} \in \mathscr{B}$, let $W_i = \mathbb{1}(D_i(\mathbf{x}) \in \mathscr{T}_0) \cdot W_i(0) + \mathbb{1}(D_i(\mathbf{x}) \in \mathscr{T}_1) \cdot W_i(1)$ denote the observed treatment receipt, where $W_i(t)$ is the potential treatment receipt under assignment $t \in \{0, 1\}$. The observed outcome is then $Y_i = \mathbb{1}(D_i(\mathbf{x}) \in \mathscr{T}_0) \cdot Y_i(0, W_i(0)) + \mathbb{1}(D_i(\mathbf{x}) \in \mathscr{T}_1) \cdot Y_i(1, W_i(1))$, where

Table 4
Quadratic homoskedastic model (Simulation results).

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

	h_0	h_1	N_{Co}	N_{Tr}	Bias	SD	RMSE	EC	IL
$\mathbf{b}_1$	17.917	17.917	1899	3638	−0.002	0.036	0.036	0.948	0.260
$\mathbf{b}_2$	17.386	17.386	1869	3461	−0.002	0.037	0.037	0.950	0.266
$\mathbf{b}_3$	17.186	17.186	1889	3364	−0.001	0.038	0.038	0.957	0.268
$\mathbf{b}_4$	17.352	17.352	1993	3331	0.002	0.038	0.039	0.957	0.265
$\mathbf{b}_5$	17.854	17.854	2186	3338	0.009	0.038	0.039	0.953	0.259
$\mathbf{b}_6$	18.432	18.432	2398	3310	0.016	0.037	0.041	0.950	0.253
$\mathbf{b}_7$	18.885	18.885	2573	3201	0.021	0.036	0.042	0.958	0.250
$\mathbf{b}_8$	19.002	19.002	2655	2966	0.022	0.035	0.041	0.952	0.253
$\mathbf{b}_9$	19.131	19.131	2712	2716	0.017	0.036	0.040	0.950	0.258
$\mathbf{b}_{10}$	19.388	19.388	2760	2485	0.008	0.038	0.039	0.951	0.267
$\mathbf{b}_{11}$	21.252	21.252	3048	2650	−0.002	0.038	0.038	0.947	0.279
$\mathbf{b}_{12}$	19.198	19.198	2728	2557	−0.010	0.037	0.039	0.947	0.263
$\mathbf{b}_{13}$	19.022	19.022	2668	2915	−0.017	0.036	0.040	0.954	0.254
$\mathbf{b}_{14}$	18.697	18.697	2520	3197	−0.018	0.037	0.041	0.951	0.252
$\mathbf{b}_{15}$	17.741	17.741	2201	3228	−0.012	0.039	0.041	0.948	0.261
$\mathbf{b}_{16}$	16.747	16.747	1880	3135	−0.004	0.040	0.040	0.956	0.274
$\mathbf{b}_{17}$	16.235	16.235	1717	3042	−0.001	0.039	0.039	0.947	0.282
$\mathbf{b}_{18}$	16.400	16.400	1699	3076	−0.001	0.038	0.038	0.945	0.282
$\mathbf{b}_{19}$	17.082	17.082	1739	3239	−0.002	0.037	0.037	0.946	0.276
$\mathbf{b}_{20}$	17.979	17.979	1779	3429	−0.003	0.037	0.037	0.947	0.269
$\mathbf{b}_{21}$	18.924	18.924	1793	3580	−0.005	0.036	0.037	0.952	0.265
`Uniform`								0.948	0.398
`WBATE`					0.001	0.017	0.017	0.950	0.095
`LBATE`					0.031	0.029	0.043	0.990	0.417

(b) $h = \widehat{h}_{\mathrm{kink},\mathbf{b}}(\mathscr{B})$ (kink adaptive)

	h_0	h_1	N_{Co}	N_{Tr}	Bias	SD	RMSE	EC	IL
$\mathbf{b}_1$	17.644	17.644	1860	3529	−0.002	0.036	0.036	0.948	0.263
$\mathbf{b}_2$	16.703	16.703	1759	3205	−0.001	0.038	0.038	0.948	0.274
$\mathbf{b}_3$	15.611	15.611	1617	2813	−0.001	0.039	0.039	0.951	0.290
$\mathbf{b}_4$	13.979	13.979	1378	2252	−0.001	0.043	0.043	0.949	0.322
$\mathbf{b}_5$	12.000	12.000	1080	1642	−0.001	0.050	0.050	0.945	0.375
$\mathbf{b}_6$	10.000	10.000	787	1114	−0.001	0.059	0.059	0.947	0.450
$\mathbf{b}_7$	8.509	8.509	589	775	−0.001	0.070	0.070	0.944	0.535
$\mathbf{b}_8$	8.325	8.325	592	674	0.003	0.075	0.075	0.942	0.559
$\mathbf{b}_9$	8.381	8.381	641	581	0.009	0.076	0.077	0.940	0.571
$\mathbf{b}_{10}$	8.494	8.494	702	480	0.007	0.080	0.080	0.942	0.585
$\mathbf{b}_{11}$	9.311	9.311	870	442	−0.001	0.086	0.086	0.933	0.636
$\mathbf{b}_{12}$	8.411	8.411	672	517	−0.007	0.079	0.079	0.942	0.580
$\mathbf{b}_{13}$	8.334	8.334	601	656	−0.004	0.076	0.076	0.941	0.562
$\mathbf{b}_{14}$	8.644	8.644	605	807	−0.000	0.070	0.070	0.943	0.524
$\mathbf{b}_{15}$	11.200	11.200	962	1419	−0.001	0.053	0.053	0.945	0.401
$\mathbf{b}_{16}$	13.955	13.955	1376	2245	−0.002	0.044	0.044	0.947	0.322
$\mathbf{b}_{17}$	15.651	15.651	1620	2831	−0.001	0.040	0.040	0.944	0.290
$\mathbf{b}_{18}$	16.293	16.293	1684	3034	−0.001	0.038	0.038	0.945	0.283
$\mathbf{b}_{19}$	17.074	17.074	1738	3236	−0.002	0.037	0.037	0.946	0.276
$\mathbf{b}_{20}$	17.979	17.979	1779	3429	−0.003	0.037	0.037	0.947	0.269
$\mathbf{b}_{21}$	18.924	18.924	1793	3580	−0.005	0.036	0.037	0.952	0.265

Table 4
Continued

Uniform								0.930	0.629
WBATE					−0.001	0.021	0.021	0.947	0.121
LBATE					0.085	0.054	0.100	0.981	0.792
(c) $h = \widehat{C} \cdot n^{-1/4}$ (unknown kink location)									
	h_0	h_1	N_{Co}	N_{Tr}	Bias	SD	RMSE	EC	IL
$\mathbf{b}_1$	7.849	7.849	524	693	0.001	0.078	0.078	0.945	0.300
$\mathbf{b}_2$	7.617	7.617	504	662	0.002	0.079	0.079	0.949	0.307
$\mathbf{b}_3$	7.529	7.529	498	651	0.001	0.079	0.079	0.948	0.310
$\mathbf{b}_4$	7.602	7.602	506	663	−0.001	0.080	0.080	0.946	0.307
$\mathbf{b}_5$	7.822	7.822	527	696	−0.002	0.078	0.078	0.945	0.300
$\mathbf{b}_6$	8.075	8.075	547	729	−0.001	0.075	0.075	0.954	0.292
$\mathbf{b}_7$	8.274	8.274	563	736	−0.001	0.073	0.073	0.954	0.288
$\mathbf{b}_8$	8.325	8.325	592	674	0.003	0.075	0.075	0.950	0.292
$\mathbf{b}_9$	8.381	8.381	641	581	0.009	0.076	0.077	0.951	0.298
$\mathbf{b}_{10}$	8.494	8.494	702	480	0.007	0.080	0.080	0.946	0.308
$\mathbf{b}_{11}$	9.311	9.311	870	442	−0.001	0.086	0.086	0.947	0.329
$\mathbf{b}_{12}$	8.411	8.411	672	517	−0.007	0.079	0.079	0.946	0.303
$\mathbf{b}_{13}$	8.333	8.333	601	656	−0.004	0.076	0.076	0.951	0.293
$\mathbf{b}_{14}$	8.191	8.191	553	729	−0.000	0.075	0.075	0.947	0.290
$\mathbf{b}_{15}$	7.772	7.772	519	682	0.000	0.078	0.078	0.949	0.302
$\mathbf{b}_{16}$	7.337	7.337	476	616	0.001	0.082	0.082	0.945	0.317
$\mathbf{b}_{17}$	7.113	7.113	449	577	0.001	0.086	0.086	0.944	0.327
$\mathbf{b}_{18}$	7.185	7.185	449	579	0.001	0.084	0.084	0.945	0.327
$\mathbf{b}_{19}$	7.484	7.484	468	609	−0.001	0.084	0.084	0.943	0.319
$\mathbf{b}_{20}$	7.877	7.877	488	645	−0.002	0.080	0.080	0.947	0.311
$\mathbf{b}_{21}$	8.291	8.291	502	673	−0.001	0.078	0.078	0.948	0.306
Uniform								0.937	0.467
WBATE					0.000	0.025	0.025	0.948	0.098
LBATE					0.114	0.052	0.126	0.988	0.488
(d) $h = \widehat{h}_{1d,\mathbf{b}}$ (rdrobust + rd2d.dist)									
	h_0	h_1	N_{Co}	N_{Tr}	Bias	SD	RMSE	EC	IL
$\mathbf{b}_1$	17.226	17.226	1805	3381	−0.004	0.038	0.038	0.951	0.271
$\mathbf{b}_2$	16.871	16.871	1794	3282	−0.004	0.039	0.039	0.955	0.275
$\mathbf{b}_3$	16.770	16.770	1829	3228	−0.003	0.040	0.040	0.963	0.277
$\mathbf{b}_4$	16.999	16.999	1940	3229	−0.000	0.041	0.041	0.961	0.275
$\mathbf{b}_5$	17.517	17.517	2124	3259	0.006	0.042	0.043	0.956	0.270
$\mathbf{b}_6$	18.251	18.251	2356	3293	0.013	0.042	0.044	0.955	0.262
$\mathbf{b}_7$	18.950	18.950	2571	3265	0.018	0.040	0.044	0.956	0.256
$\mathbf{b}_8$	19.610	19.610	2754	3184	0.019	0.038	0.043	0.953	0.252
$\mathbf{b}_9$	20.292	20.292	2907	3079	0.016	0.037	0.040	0.950	0.249
$\mathbf{b}_{10}$	21.262	21.262	3066	3032	0.008	0.038	0.039	0.954	0.250
$\mathbf{b}_{11}$	23.045	23.045	3306	3188	−0.002	0.039	0.039	0.953	0.262
$\mathbf{b}_{12}$	20.859	20.859	3010	3043	−0.011	0.038	0.039	0.956	0.248
$\mathbf{b}_{13}$	19.881	19.881	2820	3194	−0.017	0.038	0.041	0.954	0.248
$\mathbf{b}_{14}$	18.689	18.689	2509	3233	−0.017	0.041	0.044	0.951	0.258
$\mathbf{b}_{15}$	17.167	17.167	2091	3088	−0.009	0.045	0.045	0.956	0.276
$\mathbf{b}_{16}$	15.870	15.870	1740	2869	−0.001	0.044	0.044	0.957	0.294
$\mathbf{b}_{17}$	15.296	15.296	1578	2731	0.002	0.043	0.043	0.948	0.302
$\mathbf{b}_{18}$	15.414	15.414	1561	2731	0.001	0.041	0.041	0.947	0.300
$\mathbf{b}_{19}$	15.959	15.959	1591	2833	0.000	0.041	0.041	0.949	0.294
$\mathbf{b}_{20}$	16.585	16.585	1605	2925	−0.001	0.041	0.041	0.955	0.290
$\mathbf{b}_{21}$	17.091	17.091	1577	2934	−0.003	0.042	0.042	0.955	0.292
Uniform								0.947	0.409
WBATE					0.001	0.018	0.018	0.949	0.096
LBATE					0.037	0.034	0.050	0.990	0.447

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 receipt $w \in \{0,1\}$.

The usual fuzzy estimand and estimator are

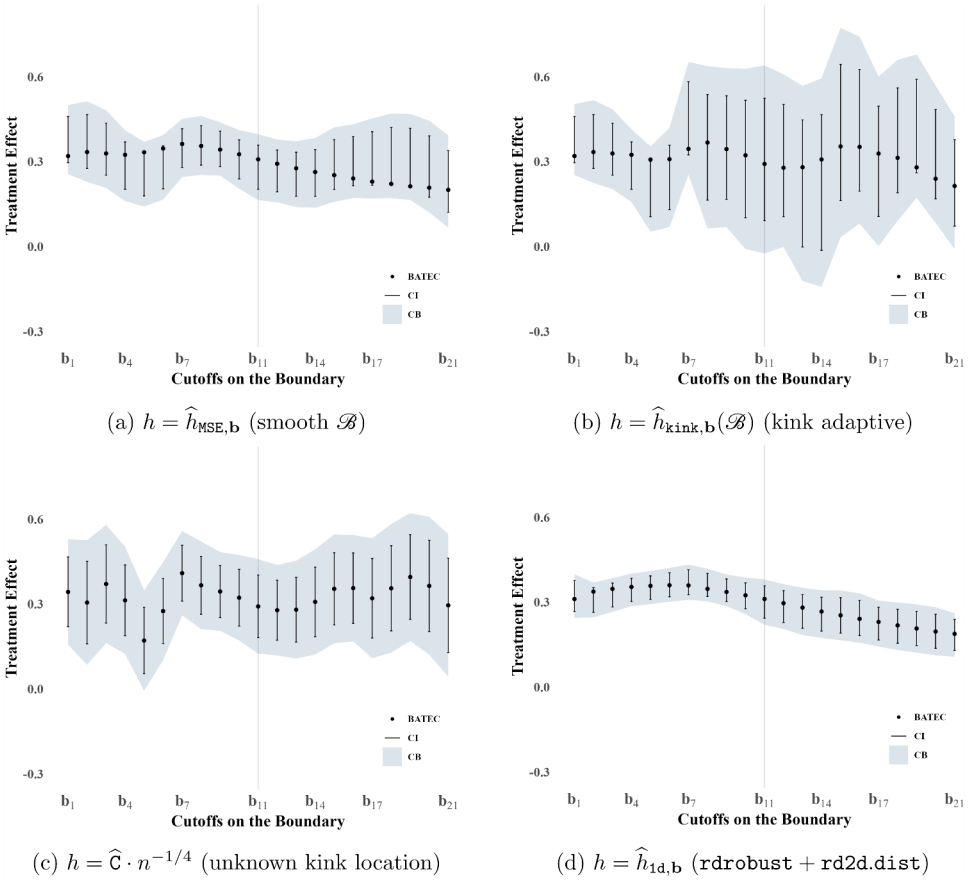
$$\zeta(\mathbf{x}) = \frac{\tau_Y(\mathbf{x})}{\tau_W(\mathbf{x})} \qquad \text{and} \qquad \widehat{\xi}(\mathbf{x}) = \frac{\widehat{\vartheta}_Y(\mathbf{x})}{\widehat{\vartheta}_W(\mathbf{x})},$$

Table 5
Quadratic heteroskedastic model (Simulation results).

(a) $h = \widehat{h}_{\mathrm{MSE},\mathbf{b}}$ (smooth $\mathscr{B}$)									
	h_0	h_1	N_{Co}	N_{Tr}	Bias	SD	RMSE	EC	IL
$\mathbf{b}_1$	18.550	18.550	1986	3881	−0.002	0.042	0.042	0.949	0.302
$\mathbf{b}_2$	17.907	17.907	1948	3654	−0.001	0.042	0.042	0.955	0.304
$\mathbf{b}_3$	17.562	17.562	1954	3494	0.000	0.042	0.042	0.960	0.303
$\mathbf{b}_4$	17.483	17.483	2019	3373	0.003	0.043	0.043	0.960	0.299
$\mathbf{b}_5$	17.596	17.596	2134	3256	0.009	0.042	0.043	0.958	0.293
$\mathbf{b}_6$	17.768	17.768	2262	3110	0.016	0.042	0.045	0.958	0.287
$\mathbf{b}_7$	17.881	17.881	2367	2912	0.021	0.041	0.046	0.951	0.283
$\mathbf{b}_8$	17.656	17.656	2382	2600	0.021	0.041	0.046	0.950	0.287
$\mathbf{b}_9$	17.529	17.529	2394	2304	0.017	0.041	0.045	0.949	0.291
$\mathbf{b}_{10}$	17.409	17.409	2382	2004	0.009	0.042	0.043	0.952	0.294
$\mathbf{b}_{11}$	18.492	18.492	2566	1971	−0.001	0.042	0.042	0.951	0.293
$\mathbf{b}_{12}$	17.353	17.353	2369	2100	−0.009	0.043	0.044	0.956	0.300
$\mathbf{b}_{13}$	17.438	17.438	2346	2490	−0.016	0.043	0.046	0.954	0.298
$\mathbf{b}_{14}$	17.577	17.577	2288	2874	−0.017	0.043	0.046	0.952	0.296
$\mathbf{b}_{15}$	17.244	17.244	2100	3076	−0.010	0.045	0.046	0.957	0.305
$\mathbf{b}_{16}$	16.769	16.769	1884	3142	−0.003	0.046	0.046	0.962	0.318
$\mathbf{b}_{17}$	16.620	16.620	1777	3181	−0.000	0.046	0.046	0.954	0.328
$\mathbf{b}_{18}$	17.013	17.013	1784	3306	−0.001	0.046	0.046	0.947	0.329
$\mathbf{b}_{19}$	17.833	17.833	1836	3526	−0.002	0.045	0.045	0.953	0.325
$\mathbf{b}_{20}$	18.869	18.869	1887	3768	−0.004	0.045	0.045	0.954	0.321
$\mathbf{b}_{21}$	20.049	20.049	1920	4002	−0.006	0.045	0.045	0.945	0.318
`Uniform`								0.952	0.458
`WBATE`					0.001	0.019	0.019	0.950	0.105
`LBATE`					0.042	0.033	0.053	0.988	0.481
(b) $h = \widehat{h}_{\mathtt{kink},\mathbf{b}}(\mathscr{B})$ (kink adaptive)									
	h_0	h_1	N_{Co}	N_{Tr}	Bias	SD	RMSE	EC	IL
$\mathbf{b}_1$	18.127	18.127	1925	3713	−0.001	0.043	0.043	0.950	0.307
$\mathbf{b}_2$	16.987	16.987	1800	3309	−0.000	0.044	0.043	0.950	0.317
$\mathbf{b}_3$	15.693	15.693	1630	2842	0.001	0.045	0.045	0.955	0.333
$\mathbf{b}_4$	13.973	13.973	1379	2251	0.001	0.048	0.048	0.948	0.366
$\mathbf{b}_5$	12.000	12.000	1082	1642	0.000	0.055	0.055	0.945	0.418
$\mathbf{b}_6$	10.000	10.000	788	1113	−0.001	0.065	0.065	0.942	0.494
$\mathbf{b}_7$	8.216	8.216	549	723	0.001	0.081	0.081	0.940	0.595
$\mathbf{b}_8$	7.739	7.739	508	590	0.005	0.088	0.088	0.934	0.634
$\mathbf{b}_9$	7.680	7.680	536	496	0.011	0.088	0.088	0.943	0.641
$\mathbf{b}_{10}$	7.627	7.627	570	390	0.008	0.087	0.087	0.946	0.651
$\mathbf{b}_{11}$	8.101	8.101	677	326	−0.001	0.090	0.090	0.937	0.648
$\mathbf{b}_{12}$	7.602	7.602	550	428	−0.006	0.091	0.091	0.937	0.659
$\mathbf{b}_{13}$	7.640	7.640	500	562	−0.003	0.089	0.089	0.944	0.651
$\mathbf{b}_{14}$	8.433	8.433	576	770	−0.002	0.080	0.080	0.943	0.590
$\mathbf{b}_{15}$	11.200	11.200	962	1419	−0.001	0.060	0.060	0.946	0.452
$\mathbf{b}_{16}$	13.943	13.943	1374	2241	−0.001	0.049	0.049	0.951	0.373
$\mathbf{b}_{17}$	15.831	15.831	1646	2896	−0.001	0.047	0.047	0.951	0.340
$\mathbf{b}_{18}$	16.848	16.848	1760	3241	−0.001	0.046	0.046	0.948	0.331
$\mathbf{b}_{19}$	17.819	17.819	1834	3520	−0.002	0.045	0.045	0.953	0.326
$\mathbf{b}_{20}$	18.869	18.869	1887	3768	−0.004	0.045	0.045	0.954	0.321
$\mathbf{b}_{21}$	20.049	20.049	1920	4002	−0.006	0.045	0.045	0.945	0.318

Table 5
Continued

Uniform								0.932	0.713
WBATE					−0.000	0.023	0.023	0.950	0.131
LBATE					0.102	0.058	0.117	0.979	0.888
(c) $h = \widehat{C} \cdot n^{-1/4}$ (unknown kink location)									
	h_0	h_1	N_{Co}	N_{Tr}	Bias	SD	RMSE	EC	IL
$\mathbf{b}_1$	8.127	8.127	558	743	0.001	0.090	0.090	0.948	0.347
$\mathbf{b}_2$	7.845	7.845	533	703	0.001	0.090	0.090	0.946	0.350
$\mathbf{b}_3$	7.694	7.694	518	680	0.001	0.088	0.088	0.953	0.349
$\mathbf{b}_4$	7.659	7.659	513	673	0.000	0.088	0.088	0.947	0.345
$\mathbf{b}_5$	7.709	7.709	514	674	−0.002	0.087	0.087	0.952	0.338
$\mathbf{b}_6$	7.785	7.785	513	674	−0.002	0.086	0.086	0.951	0.332
$\mathbf{b}_7$	7.834	7.834	507	661	0.000	0.086	0.086	0.947	0.328
$\mathbf{b}_8$	7.735	7.735	508	590	0.005	0.088	0.088	0.947	0.331
$\mathbf{b}_9$	7.680	7.680	536	496	0.011	0.088	0.088	0.948	0.335
$\mathbf{b}_{10}$	7.627	7.627	570	390	0.008	0.087	0.087	0.950	0.338
$\mathbf{b}_{11}$	8.101	8.101	677	326	−0.001	0.090	0.090	0.946	0.339
$\mathbf{b}_{12}$	7.602	7.602	550	428	−0.006	0.091	0.091	0.945	0.345
$\mathbf{b}_{13}$	7.640	7.640	500	562	−0.003	0.089	0.089	0.950	0.342
$\mathbf{b}_{14}$	7.701	7.701	493	645	−0.002	0.088	0.088	0.951	0.341
$\mathbf{b}_{15}$	7.555	7.555	492	643	−0.000	0.089	0.089	0.948	0.348
$\mathbf{b}_{16}$	7.347	7.347	476	618	−0.001	0.093	0.093	0.946	0.362
$\mathbf{b}_{17}$	7.281	7.281	468	606	0.001	0.096	0.096	0.945	0.372
$\mathbf{b}_{18}$	7.453	7.453	480	625	−0.001	0.097	0.097	0.941	0.372
$\mathbf{b}_{19}$	7.813	7.813	505	667	−0.002	0.094	0.094	0.950	0.367
$\mathbf{b}_{20}$	8.267	8.267	533	714	0.000	0.093	0.093	0.947	0.361
$\mathbf{b}_{21}$	8.784	8.784	556	759	0.000	0.094	0.094	0.941	0.357
Uniform								0.943	0.531
WBATE					0.000	0.028	0.028	0.947	0.107
LBATE					0.137	0.059	0.149	0.986	0.556
(d) $h = \widehat{h}_{1d,\mathbf{b}}$ (rdrobust + rd2d.dist)									
	h_0	h_1	N_{Co}	N_{Tr}	Bias	SD	RMSE	EC	IL
$\mathbf{b}_1$	17.834	17.834	1890	3614	−0.003	0.045	0.045	0.951	0.315
$\mathbf{b}_2$	17.392	17.392	1874	3473	−0.003	0.045	0.045	0.953	0.315
$\mathbf{b}_3$	17.123	17.123	1890	3353	−0.002	0.046	0.046	0.959	0.314
$\mathbf{b}_4$	17.070	17.070	1956	3254	0.001	0.047	0.047	0.964	0.311
$\mathbf{b}_5$	17.206	17.206	2067	3162	0.006	0.048	0.048	0.959	0.306
$\mathbf{b}_6$	17.450	17.450	2199	3053	0.011	0.048	0.049	0.958	0.299
$\mathbf{b}_7$	17.514	17.514	2288	2849	0.016	0.047	0.050	0.949	0.296
$\mathbf{b}_8$	17.563	17.563	2356	2616	0.018	0.047	0.050	0.948	0.295
$\mathbf{b}_9$	17.567	17.567	2393	2354	0.015	0.046	0.048	0.954	0.296
$\mathbf{b}_{10}$	17.679	17.679	2426	2107	0.008	0.045	0.046	0.956	0.296
$\mathbf{b}_{11}$	18.549	18.549	2564	2029	−0.002	0.046	0.046	0.958	0.298
$\mathbf{b}_{12}$	17.600	17.600	2410	2197	−0.009	0.047	0.048	0.957	0.302
$\mathbf{b}_{13}$	17.641	17.641	2383	2579	−0.014	0.047	0.050	0.956	0.301
$\mathbf{b}_{14}$	17.575	17.575	2288	2913	−0.013	0.049	0.051	0.953	0.303
$\mathbf{b}_{15}$	16.976	16.976	2057	3031	−0.006	0.051	0.052	0.962	0.318
$\mathbf{b}_{16}$	16.166	16.166	1798	2969	0.001	0.050	0.050	0.961	0.337
$\mathbf{b}_{17}$	15.779	15.779	1656	2908	0.004	0.051	0.051	0.960	0.349
$\mathbf{b}_{18}$	15.962	15.962	1637	2940	0.003	0.051	0.051	0.950	0.351
$\mathbf{b}_{19}$	16.531	16.531	1665	3054	0.001	0.050	0.050	0.955	0.350
$\mathbf{b}_{20}$	17.189	17.189	1679	3159	0.000	0.051	0.051	0.954	0.351
$\mathbf{b}_{21}$	17.845	17.845	1664	3220	−0.000	0.053	0.053	0.947	0.356
Uniform								0.954	0.480
WBATE					0.001	0.020	0.020	0.947	0.107
LBATE					0.052	0.042	0.067	0.991	0.540



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

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

Fig. 4. BATEC estimation and inference (SPP empirical application).

where, for each $\mathbf{x} \in \mathscr{B}$, $\tau_Y(\mathbf{x}) = \mathbb{E}[Y_i(1, W_i(1)) - Y_i(0, W_i(0))|\mathbf{X}_i = \mathbf{x}]$ and $\tau_W(\mathbf{x}) = \mathbb{E}[W_i(1) - W_i(0)|\mathbf{X}_i = \mathbf{x}]$, and $\widehat{\vartheta}_Y(\mathbf{x}) = \mathbf{e}_0^\top \widehat{\boldsymbol{\gamma}}_{Y,1}(\mathbf{x}) - \mathbf{e}_0^\top \widehat{\boldsymbol{\gamma}}_{Y,0}(\mathbf{x})$ and $\widehat{\vartheta}_W(\mathbf{x}) = \mathbf{e}_0^\top \widehat{\boldsymbol{\gamma}}_{W,1}(\mathbf{x}) - \mathbf{e}_0^\top \widehat{\boldsymbol{\gamma}}_{W,0}(\mathbf{x})$, where $\widehat{\boldsymbol{\gamma}}_{A,t}(\mathbf{x})$ denotes the local polynomial fit obtained using outcome variable $A \in \{Y, W\}$ and side $t \in \{0, 1\}$ in (1). A causal interpretation of $\zeta(\mathbf{x})$ can be obtained under additional assumptions; see Arai et al. (2022). Section SA-6 in the supplemental appendix gives estimation and inference results for fuzzy BATEC, fuzzy WBATE, and fuzzy LBATE when employing distance-based methods.

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 $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ denote covariates of dimension $d_Z \geq 1$.

For efficiency improvements, the covariate-adjusted estimator is

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

with

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

where $T_i = \mathbb{1}(D_i(\mathbf{x}) \geq 0)$, and $\widetilde{\mathbf{r}}_p(D_i(\mathbf{x}), T_i, \mathbf{Z}_i) = [\mathbf{r}_p(D_i(\mathbf{x})/h)^\top, T_i \cdot \mathbf{r}_p(D_i(\mathbf{x})/h)^\top, \mathbf{Z}_i^\top]^\top$ 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 $\mathfrak{p} = 2(p+1) + d_Z$.

For heterogeneity analysis, define

$$\check{\kappa}(\mathbf{x}, \mathbf{z}) = \mathbf{e}_{p+2}^\top \check{\boldsymbol{\gamma}}(\mathbf{x}) + \mathbf{z}^\top \Big[\mathbf{0}_{d_Z \times (2+d_Z)(p+1)} \quad \mathbf{I}_{d_Z} \quad \mathbf{0}_{d_Z \times p d_Z}\Big] \check{\boldsymbol{\gamma}}(\mathbf{x})$$

with

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

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 $\check{\mathbf{r}}_p(D_i(\mathbf{x}), T_i, \mathbf{Z}_i) = [\mathbf{r}_p(D_i(\mathbf{x})/h)^\top, T_i \cdot \mathbf{r}_p(D_i(\mathbf{x})/h)^\top, (\mathbf{r}_p(D_i(\mathbf{x})/h) \otimes \mathbf{Z}_i)^\top, T_i \cdot (\mathbf{r}_p(D_i(\mathbf{x})/h) \otimes \mathbf{Z}_i)^\top]^\top$ contains the full interaction between the polynomial basis function, the treatment assignment indicator, and the pre-intervention covariates.

Estimation and inference for covariate-adjusted distance-based methods can be established by leveraging the results in the supplemental appendix. A formal treatment is left for future work.

Table 6
BATEC estimation and inference (SPP empirical application).

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

$\mathbf{b} \in \mathscr{B}$	h	N_{Co}	N_{Tr}	$\tau_Y(\mathbf{b})$	p-value	95% RBC CI
$\mathbf{b}_1$	32.762	18,389	7264	0.320	0.000	(0.256, 0.500)
$\mathbf{b}_2$	27.994	12,482	5807	0.334	0.000	(0.229, 0.513)
$\mathbf{b}_3$	28.491	13,223	6043	0.329	0.000	(0.207, 0.481)
$\mathbf{b}_4$	31.169	16,559	7026	0.324	0.000	(0.161, 0.411)
$\mathbf{b}_5$	34.582	22,195	7670	0.333	0.000	(0.141, 0.370)
$\mathbf{b}_6$	33.580	21,212	6980	0.347	0.000	(0.167, 0.394)
$\mathbf{b}_7$	35.367	24,702	6882	0.363	0.000	(0.246, 0.451)
$\mathbf{b}_8$	36.083	26,282	6481	0.356	0.000	(0.253, 0.461)
$\mathbf{b}_9$	40.739	34,858	6790	0.343	0.000	(0.251, 0.439)
$\mathbf{b}_{10}$	38.273	30,105	5736	0.326	0.000	(0.205, 0.411)
$\mathbf{b}_{11}$	38.900	30,654	5244	0.308	0.000	(0.164, 0.397)
$\mathbf{b}_{12}$	38.456	27,144	5397	0.293	0.000	(0.157, 0.378)
$\mathbf{b}_{13}$	37.502	22,939	5419	0.277	0.000	(0.140, 0.372)
$\mathbf{b}_{14}$	36.557	19,479	5407	0.264	0.000	(0.137, 0.383)
$\mathbf{b}_{15}$	34.947	15,444	5273	0.252	0.000	(0.158, 0.421)
$\mathbf{b}_{16}$	36.527	15,081	5622	0.241	0.000	(0.172, 0.432)
$\mathbf{b}_{17}$	35.586	12,378	5524	0.230	0.000	(0.170, 0.453)
$\mathbf{b}_{18}$	35.712	10,813	5560	0.222	0.000	(0.171, 0.470)
$\mathbf{b}_{19}$	36.783	10,095	5743	0.213	0.000	(0.165, 0.469)
$\mathbf{b}_{20}$	36.701	8543	5667	0.208	0.000	(0.121, 0.445)
$\mathbf{b}_{21}$	38.896	8637	6018	0.201	0.000	(0.067, 0.393)
WBATE				0.290	0.000	(0.269, 0.340)
LBATE				0.363		(0.259, 0.510)

(b) $h = \widehat{h}_{\mathtt{kink},\mathbf{b}}(\mathscr{B})$ (kink adaptive)

$\mathbf{b} \in \mathscr{B}$	h	N_{Co}	N_{Tr}	$\tau_Y(\mathbf{b})$	p-value	95% RBC CI
$\mathbf{b}_1$	32.762	18,389	7264	0.320	0.000	(0.253, 0.503)
$\mathbf{b}_2$	27.994	12,482	5807	0.334	0.000	(0.225, 0.517)
$\mathbf{b}_3$	28.491	13,223	6043	0.329	0.000	(0.203, 0.485)
$\mathbf{b}_4$	31.169	16,559	7026	0.324	0.000	(0.158, 0.414)
$\mathbf{b}_5$	27.092	11,736	5753	0.307	0.000	(0.053, 0.354)
$\mathbf{b}_6$	22.577	7601	4394	0.309	0.000	(0.070, 0.419)
$\mathbf{b}_7$	18.061	4559	3160	0.345	0.000	(0.255, 0.652)
$\mathbf{b}_8$	13.546	2442	1982	0.368	0.000	(0.064, 0.637)
$\mathbf{b}_9$	14.134	2969	1828	0.345	0.000	(0.069, 0.631)
$\mathbf{b}_{10}$	13.278	2911	1329	0.323	0.003	(−0.009, 0.628)
$\mathbf{b}_{11}$	13.495	3364	936	0.292	0.005	(−0.024, 0.640)
$\mathbf{b}_{12}$	13.338	2776	1087	0.279	0.003	(−0.001, 0.608)
$\mathbf{b}_{13}$	13.007	2165	1194	0.280	0.052	(−0.121, 0.567)
$\mathbf{b}_{14}$	12.680	1668	1232	0.308	0.064	(−0.142, 0.594)
$\mathbf{b}_{15}$	12.121	1232	1169	0.354	0.001	(0.033, 0.773)
$\mathbf{b}_{16}$	14.000	1436	1402	0.352	0.000	(0.081, 0.741)
$\mathbf{b}_{17}$	16.800	1888	1836	0.329	0.002	(0.002, 0.600)
$\mathbf{b}_{18}$	19.600	2316	2252	0.314	0.000	(0.091, 0.659)
$\mathbf{b}_{19}$	22.400	2760	2713	0.280	0.000	(0.173, 0.679)
$\mathbf{b}_{20}$	25.200	3145	3178	0.240	0.000	(0.084, 0.569)
$\mathbf{b}_{21}$	28.000	3556	3543	0.214	0.004	(−0.009, 0.459)

Table 6
Continued

WBATE				0.312	0.000	(0.274, 0.376)
LBATE				0.368		(0.255, 0.773)
(c) $h = \widehat{C} \cdot n^{-1/4}$ (unknown kink location)						
$\mathbf{b} \in \mathscr{B}$	h	N_{Co}	N_{Tr}	$\tau_Y(\mathbf{b})$	p-value	95% RBC CI
$\mathbf{b}_1$	11.369	1536	1182	0.343	0.000	(0.157, 0.529)
$\mathbf{b}_2$	9.714	1071	893	0.306	0.000	(0.086, 0.525)
$\mathbf{b}_3$	9.887	1107	952	0.371	0.000	(0.163, 0.579)
$\mathbf{b}_4$	10.816	1373	1174	0.313	0.000	(0.124, 0.502)
$\mathbf{b}_5$	11.999	1776	1484	0.171	0.004	(−0.006, 0.348)
$\mathbf{b}_6$	11.651	1704	1507	0.275	0.000	(0.102, 0.448)
$\mathbf{b}_7$	12.270	1966	1691	0.409	0.000	(0.261, 0.558)
$\mathbf{b}_8$	12.519	2053	1707	0.367	0.000	(0.212, 0.521)
$\mathbf{b}_9$	14.134	2969	1828	0.345	0.000	(0.206, 0.484)
$\mathbf{b}_{10}$	13.278	2911	1329	0.323	0.000	(0.172, 0.474)
$\mathbf{b}_{11}$	13.495	3364	936	0.292	0.000	(0.125, 0.459)
$\mathbf{b}_{12}$	13.338	2776	1087	0.279	0.000	(0.119, 0.438)
$\mathbf{b}_{13}$	13.007	2165	1194	0.280	0.000	(0.108, 0.453)
$\mathbf{b}_{14}$	12.680	1668	1232	0.308	0.000	(0.122, 0.493)
$\mathbf{b}_{15}$	12.121	1232	1169	0.354	0.000	(0.162, 0.546)
$\mathbf{b}_{16}$	12.671	1191	1176	0.356	0.000	(0.168, 0.544)
$\mathbf{b}_{17}$	12.343	1037	995	0.321	0.000	(0.109, 0.532)
$\mathbf{b}_{18}$	12.386	904	889	0.356	0.000	(0.129, 0.583)
$\mathbf{b}_{19}$	12.758	862	847	0.396	0.000	(0.170, 0.622)
$\mathbf{b}_{20}$	12.729	754	750	0.364	0.000	(0.121, 0.608)
$\mathbf{b}_{21}$	13.492	769	755	0.296	0.000	(0.044, 0.547)
WBATE				0.325	0.000	(0.283, 0.367)
LBATE				0.409		(0.258, 0.626)
(d) $h = \widehat{h}_{1d,\mathbf{b}}$ (`rdrobust` + `rd2d.dist`)						
$\mathbf{b} \in \mathscr{B}$	h	N_{Co}	N_{Tr}	$\tau_Y(\mathbf{b})$	p-value	95% RBC CI
$\mathbf{b}_1$	49.759	52,499	12,202	0.311	0.000	(0.245, 0.399)
$\mathbf{b}_2$	64.201	102,394	13,859	0.338	0.000	(0.247, 0.369)
$\mathbf{b}_3$	65.363	108,053	13,749	0.347	0.000	(0.266, 0.385)
$\mathbf{b}_4$	67.149	115,026	13,663	0.354	0.000	(0.286, 0.401)
$\mathbf{b}_5$	65.313	107,722	13,130	0.358	0.000	(0.293, 0.410)
$\mathbf{b}_6$	62.364	95,870	12,384	0.360	0.000	(0.302, 0.421)
$\mathbf{b}_7$	57.821	79,327	11,142	0.359	0.000	(0.308, 0.433)
$\mathbf{b}_8$	64.047	97,346	11,579	0.347	0.000	(0.304, 0.418)
$\mathbf{b}_9$	67.489	105,280	11,545	0.336	0.000	(0.287, 0.398)
$\mathbf{b}_{10}$	60.813	81,239	9776	0.324	0.000	(0.258, 0.387)
$\mathbf{b}_{11}$	55.154	63,578	8088	0.312	0.000	(0.221, 0.381)
$\mathbf{b}_{12}$	52.754	53,179	7878	0.297	0.000	(0.206, 0.363)
$\mathbf{b}_{13}$	49.963	43,517	7558	0.281	0.000	(0.184, 0.351)
$\mathbf{b}_{14}$	50.847	41,209	7862	0.267	0.000	(0.174, 0.342)
$\mathbf{b}_{15}$	49.997	36,005	7847	0.254	0.000	(0.166, 0.342)
$\mathbf{b}_{16}$	52.637	36,840	8361	0.241	0.000	(0.157, 0.331)
$\mathbf{b}_{17}$	59.768	45,430	9574	0.231	0.000	(0.143, 0.306)
$\mathbf{b}_{18}$	60.682	42,938	9765	0.218	0.000	(0.131, 0.299)
$\mathbf{b}_{19}$	62.809	42,670	10,094	0.207	0.000	(0.122, 0.292)
$\mathbf{b}_{20}$	66.173	44,295	10,642	0.197	0.000	(0.113, 0.282)
$\mathbf{b}_{21}$	76.865	61,417	12,245	0.188	0.000	(0.108, 0.261)
WBATE				0.292	0.000	(0.260, 0.316)
LBATE				0.360		(0.308, 0.433)

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 univariate signed distance score, constructed from a distance measure relative to each boundary point $\mathbf{x} \in \mathscr{B}$. 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.
- *Pooling-based methods*. These approaches further aggregate the multidimensional score by using the distance between each observation's location and the closest boundary point, $D_i = \inf_{\mathbf{x}\in\mathscr{B}} \mathscr{d}(\mathbf{X}_i, \mathbf{x})$, 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 WBATE.

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 $\tau_{\texttt{LBATE}} = \sup_{\mathbf{x}\in\mathscr{B}} \tau(\mathbf{x})$. 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, $(\tau(\mathbf{x}) : \mathbf{x} \in \mathscr{B})$, 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 $\widehat{h}_{\texttt{kink},\mathbf{x}}(\mathscr{B})$). 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 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 $\mathscr{B}$), or (ii) rely on regularization of the boundary (e.g., via smooth approximations of $\mathscr{B}$), could achieve smaller misspecification bias. For example, one simple approach is to restrict estimation to locally smooth segments of $\mathscr{B}$, thereby avoiding the bias issues highlighted by Theorem 2, at the cost of altering the target estimand through an additional regularization bias.

Declaration of competing interest

The first author has a professional conflict of interest with Managing Editor Michael Jansson due to co-authorship. All three authors have a professional conflict of interest with Co-Editor Bo Honore due to university affiliation. The first author has a personal conflict of interest with Associated Editors Timothy Armstrong and Christoph Rothe.

Acknowledgement

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.

Supplementary material

Supplementary material associated with this article can be found in the online version at 10.1016/j.jeconom.2026.106266.

References

Arai, Y., Hsu, Y.-C., Kitagawa, T., Mourifié, I., Wan, Y., 2022. Testing identifying assumptions in fuzzy regression discontinuity designs. Quant. Econ. 13 (1), 1–28.
Avnir, D., Biham, O., Lidar, D., Malcai, O., 1998. Is the geometry of nature fractal? Science 279 (5347), 39–40.
Banerjee, S., 2005. On geodetic distance computations in spatial modeling. Biometrics 61 (2), 617–625.
Calonico, S., Cattaneo, M.D., Farrell, M.H., 2018. On the effect of bias estimation on coverage accuracy in nonparametric inference. J. Am. Stat. Assoc. 113 (522), 767–779.
Calonico, S., Cattaneo, M.D., Farrell, M.H., 2020. Optimal bandwidth choice for robust bias corrected inference in regression discontinuity designs. Econom. J. 23 (2), 192–210.
Calonico, S., Cattaneo, M.D., Farrell, M.H., 2022. Coverage error optimal confidence intervals for local polynomial regression. Bernoulli 28 (4), 2998–3022.
Calonico, S., Cattaneo, M.D., Farrell, M.H., Palomba, F., Titiunik, R., 2026. Treatment effect heterogeneity in regression discontinuity designs. arXiv:2503.13696.
Calonico, S., Cattaneo, M.D., Farrell, M.H., Titiunik, R., 2019. Regression discontinuity designs using covariates. Rev. Econ. Stat. 101 (3), 442–451.
Calonico, S., Cattaneo, M.D., Titiunik, R., 2014. Robust nonparametric confidence intervals for regression-discontinuity designs. Econometrica 82 (6), 2295–2326.
Cattaneo, M.D., Feng, Y., Underwood, W.G., 2024a. Uniform inference for kernel density estimators with dyadic data. J. Am. Stat. Assoc. 119 (524), 2695–2708.
Cattaneo, M.D., Frandsen, B., Titiunik, R., 2015. Randomization inference in the regression discontinuity design: an application to party advantages in the U.S. senate. J. Causal Inference 3 (1), 1–24.
Cattaneo, M.D., Idrobo, N., Titiunik, R., 2020. A Practical Introduction to Regression Discontinuity Designs: Foundations. Cambridge University Press.
Cattaneo, M.D., Idrobo, N., Titiunik, R., 2024b. A Practical Introduction to Regression Discontinuity Designs: Extensions. Cambridge University Press.
Cattaneo, M.D., Keele, L., Titiunik, R., Vazquez-Bare, G., 2016. Interpreting regression discontinuity designs with multiple cutoffs. J. Polit. 78 (4), 1229–1248.
Cattaneo, M.D., Masini, R., Underwood, W.G., 2025a. Yurinskii's coupling for martingales. Ann. Stat. 53 (5), 2179–2203.
Cattaneo, M.D., Titiunik, R., 2022. Regression discontinuity designs. Annu. Rev. Econ. 14, 821–851.
Cattaneo, M.D., Titiunik, R., Yu, R.R., 2025b. RD2D: causal inference in boundary discontinuity designs. arXiv:2505.07989.
Cattaneo, M.D., Titiunik, R., Yu, R.R., 2026a. Boundary discontinuity designs: theory and practice. In: Griffith, R., Gorodnichenko, Y., Kandori, M., Molinari, F. (Eds.), Advances in Economics and Econometrics: Thirteenth World Congress. Cambridge University Press. Vol. 1. chapter 2, p. forthcoming.
Cattaneo, M.D., Titiunik, R., Yu, R.R., 2026b. Estimation and inference in boundary discontinuity designs: location-based methods. arXiv:2505.05670.
Cattaneo, M.D., Titiunik, R., Yu, R.R., 2026c. Estimation and inference in boundary discontinuity designs: pooling-based methods. Working paper .
Cattaneo, M.D., Yu, R.R., 2025. Strong approximations for empirical processes indexed by lipschitz functions. Ann. Stat. 53 (3), 1203–1229.
Chen, X., Gao, W.Y., 2026. Thin sets are not equally thin: minimax learning of submanifold integrals. arXiv:2507.12673.
Chernozhukov, V., Chetverikov, D., Kato, K., 2014a. Anti-concentration and honest, adaptive confidence bands. Ann. Stat. 42 (5), 1787–1818.
Chernozhukov, V., Chetverikov, D., Kato, K., 2014b. Gaussian approximation of suprema of empirical processes. Ann. Stat. 42 (4), 1564–1597.
Chernozhukov, V., Chetverikov, D., Kato, K., Koike, Y., 2022. Improved central limit theorem and bootstrap approximations in high dimensions. Ann. Stat. 50 (5), 2562–2586.
Diaz, J.D., Zubizarreta, J.R., 2023. Complex discontinuity designs using covariates for policy impact evaluation. Ann. Appl. Stat. 17 (1), 67–88.
Federer, H., 2014. Geometric Measure Theory. Springer.
Galiani, S., McEwan, P.J., Quistorff, B., 2017. External and internal validity of a geographic quasi-experiment embedded in a cluster-randomized experiment. In: Cattaneo, M.D., Escanciano, J.C. (Eds.), Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics, vol. 38. Emerald Group Publishing, pp. 195–236.
Giné, E., Nickl, R., 2016. Mathematical Foundations of Infinite-dimensional Statistical Models. Cambridge University Press, New York.
Hahn, J., Todd, P., van der Klaauw, W., 2001. Identification and estimation of treatment effects with a regression-discontinuity design. Econometrica 69 (1), 201–209.
Härdle, W., Müller, M., Sperlich, S., Werwatz, A., 2004. Nonparametric and Semiparametric Models. Springer, Heidelberg.
Hernán, M.A., Robins, J.M., 2020. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC.
Keele, L., Titiunik, R., 2016. Natural experiments based on geography. Polit. Sci. Res. Methods 4 (1), 65–95.
Keele, L.J., Lorch, S., Passarella, M., Small, D., Titiunik, R., 2017. An overview of geographically discontinuous treatment assignments with an application to children's health insurance. In: Cattaneo, M.D., Escanciano, J.C. (Eds.), Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics, vol. 38. Emerald Group Publishing, pp. 147–194.
Keele, L.J., Titiunik, R., 2015. Geographic boundaries as regression discontinuities. Polit. Anal. 23 (1), 127–155.
Keele, L.J., Titiunik, R., Zubizarreta, J., 2015. Enhancing a geographic regression discontinuity design through matching to estimate the effect of ballot initiatives on voter turnout. J. R. Stat. Soc. A 178 (1), 223–239.
Londoño-Vélez, J., Rodríguez, C., Sánchez, F., 2020. Upstream and downstream impacts of college merit-based financial aid for low-income students: ser pilo paga in Colombia. Am. Econ. J. Econ. Policy 12 (2), 193–227.
Mandelbrot, B., 1967. How long is the coast of britain? Statistical self-similarity and fractional dimension. Science 156 (3775), 636–638.
Mandelbrot, B.B., 1983. The Fractal Geometry of Nature. New York .
Papay, J.P., Willett, J.B., Murnane, R.J., 2011. Extending the regression-discontinuity approach to multiple assignment variables. J. Econom. 161 (2), 203–207.
Reardon, S.F., Robinson, J.P., 2012. Regression discontinuity designs with multiple rating-score variables. J. Res. Educ. Eff. 5 (1), 83–104.
Rischard, M., Branson, Z., Miratrix, L., Bornn, L., 2021. Do school districts affect NYC house prices? Identifying border differences using a Bayesian nonparametric approach to geographic regression discontinuity designs. J. Am. Stat. Assoc. 116 (534), 619–631.
Simon, L., et al., 1984. Lectures on Geometric Measure Theory. Centre for Mathematical Analysis, Australian National University Canberra.
Tsybakov, A.B., 2008. Introduction to Nonparametric Estimation. Springer.
van der Vaart, A.W., Wellner, J.A., 1996. Weak Convergence and Empirical Processes. Springer.
Velez, Y.R., Newman, B.J., 2019. Tuning in, not turning out: evaluating the impact of ethnic television on political participation. Am. J. Polit. Sci. 63 (4), 808–823.
Wong, V.C., Steiner, P.M., Cook, T.D., 2013. Analyzing regression-discontinuity designs with multiple assignment variables: a comparative study of four estimation methods. J. Educ. Behav. Stat. 38 (2), 107–141.