

# Estimation and Inference in Boundary Discontinuity Designs: Location-Based Methods\*

Matias D. Cattaneo<sup>†</sup>      Rocio Titiunik<sup>‡</sup>      Ruiqi (Rae) Yu<sup>§</sup>

May 13, 2026

## Abstract

Boundary discontinuity designs are used to learn about causal treatment effects along a continuous assignment boundary that splits units into control and treatment groups according to a bivariate location score. We analyze location-based local polynomial treatment effect estimators that directly employ the bivariate score of each unit. We develop pointwise and uniform estimation and inference methods for the *Boundary Average Treatment Effect Curve* (BATEC), as well as for two aggregated causal parameters: the *Weighted Boundary Average Treatment Effect* (WBATE) and the *Largest Boundary Average Treatment Effect* (LBATE). Our results cover both sharp and fuzzy (imperfect compliance) designs. We illustrate the methods with an empirical application, and provide companion general-purpose software. The supplemental appendix includes additional substantive theoretical results, methodological details, and simulation evidence.

*Keywords:* regression discontinuity, multi-score regression discontinuity, treatment effects estimation, causal inference, uniform inference, local polynomial methods.

---

\*We thank Alberto Abadie, Xiaohong Chen, Boris Hanin, Kosuke Imai, Oliver Linton, Xinwei Ma, Victor Panaretos, Jörg Stoye, and Jeff Wooldridge for comments and discussions. We also thank three anonymous referees for helpful 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, the Data-Driven Social Science initiative at Princeton University, and a 2026 Guggenheim Fellowship.

<sup>†</sup>Department of Operations Research and Financial Engineering, Princeton University.

<sup>‡</sup>Department of Politics, Princeton University.

<sup>§</sup>Department of Operations Research and Financial Engineering, Princeton University.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Organization and Related Literature . . . . .	4
1.2	Notation . . . . .	6
<b>2</b>	<b>Setup</b>	<b>6</b>
<b>3</b>	<b>Boundary Average Treatment Effect Curve</b>	<b>9</b>
3.1	Treatment Effect Estimation . . . . .	9
3.2	Uncertainty Quantification . . . . .	11
3.3	Implementation . . . . .	13
<b>4</b>	<b>Weighted Boundary Average Treatment Effect</b>	<b>14</b>
<b>5</b>	<b>Largest Boundary Average Treatment Effect</b>	<b>16</b>
<b>6</b>	<b>Imperfect Compliance</b>	<b>17</b>
6.1	Fuzzy BATEC . . . . .	19
6.2	Fuzzy WBATE . . . . .	21
6.3	Fuzzy LBATE . . . . .	22
<b>7</b>	<b>The Causal Effects of SPP on College Attendance</b>	<b>23</b>
<b>8</b>	<b>Extensions and Future Research</b>	<b>26</b>
8.1	Multidimensional Scores and General Assignment Boundaries . . . . .	26
8.2	Derivatives and Regression Kink Designs . . . . .	28
8.3	Covariate Adjustment . . . . .	28
8.4	Multidimensional RD Plots . . . . .	29
8.5	Clustered and Spatially Correlated Data . . . . .	29

# 1 Introduction

We study estimation and inference in boundary discontinuity (BD) designs, where treatment assignment changes discontinuously along a known boundary that separates treatment and control assignment regions according to a bivariate score. This setup is also known as a Multi-Score Regression Discontinuity (RD) design [48, 49, 56], with Geographic RD designs as an important special case [42, 43, 41, 44, 34, 50, 31]. As in the univariate RD design, the discontinuous assignment rule can be used to identify causal effects under continuity restrictions on potential outcomes, even when treatment is targeted to units with the greatest need. This makes BD designs a central tool in non-experimental policy evaluation. See [11] for an overview of the RD literature, [19, Section 5] for a practical introduction to Multi-Dimensional RD designs, and [23] for a review of empirical practice employing BD designs.

Our motivating application is the *Ser Pilo Paga* (SPP) program studied by [46]. SPP was a Colombian government subsidy that provided tuition support for post-secondary students to attend a government-certified higher education institution. Eligibility was based on both merit and economic need: students had to obtain a high grade in Colombia’s national standardized high school exit exam, SABER 11, and they also had to come from economically disadvantaged families, as measured by the survey-based wealth index SISBEN. Thus the assignment rule was deterministic but genuinely bivariate: to qualify for the program, students had to obtain a SABER 11 score in the top 9 percent of scores and their household’s SISBEN index had to be below a region-specific threshold. After recentering both scores at their corresponding thresholds, Figure 1a plots the bivariate score  $\mathbf{X}_i = (\text{SABER11}_i, \text{SISBEN}_i)^\top$  for the 2014 cohort ( $n = 363,096$  observations), together with the assignment boundary  $\mathcal{B}$  and 40 evenly-spaced boundary points  $\mathbf{b}_1, \dots, \mathbf{b}_{40}$ .

The main object of interest is the collection of treatment effects along the assignment boundary. Using standard potential outcomes notation [38, and references therein], we define the *Boundary Average Treatment Effect Curve* (BATEC) as

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

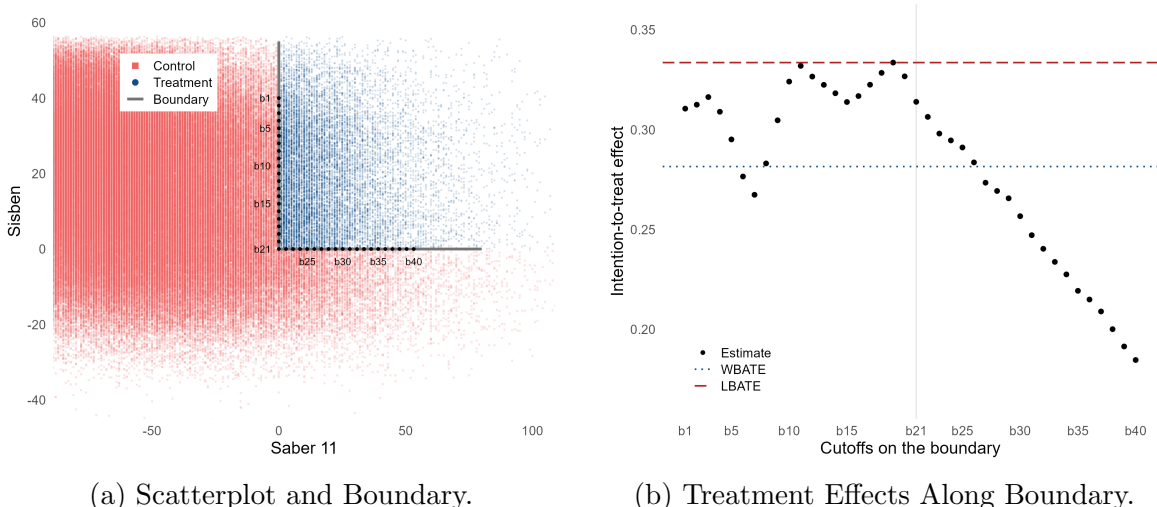


Figure 1: Scatterplot, Assignment Boundary, and Treatment Effects Using SPP data.

*Notes.* Panel (a) presents a scatterplot of the bivariate score  $\mathbf{X}_i$  using the SPP data, and also plots the treatment boundary  $\mathcal{B}$  with 40 marked grid points. Panel (b) presents causal treatment effect estimates over the 40 boundary grid points depicted in Panel (a). Specifically, the black solid dots correspond to  $\hat{\tau}(\mathbf{b}_j)$  (BATEC), the blue dotted line corresponds to  $\hat{\tau}_{\text{WBATE}}$  (WBATE), and the red dot-dash line corresponds to  $\hat{\tau}_{\text{LBATE}}$  (LBATE).

where  $Y_i(0)$  and  $Y_i(1)$  are the potential outcomes under control and treatment assignment,  $\mathbf{X}_i = (X_{1i}, X_{2i})^\top$  is the bivariate score, and  $\mathcal{B}$  is the assignment boundary. In the SPP application,  $Y_i$  is an indicator for college enrollment, so  $\tau(\mathbf{x})$  is the effect of SPP on the probability of college attendance for students at the margin of eligibility at boundary point  $\mathbf{x} \in \mathcal{B}$ . For example, in Figure 1a,  $\tau(\mathbf{b}_1)$  corresponds to students with relatively high SISBEN score and low SABER11 score, while  $\tau(\mathbf{b}_{40})$  corresponds to students with relatively low SISBEN score and high SABER11 score. Identification is the natural bivariate analogue of the continuity argument in univariate RD designs [36]: treatment assignment changes abruptly at  $\mathcal{B}$ , while the potential outcome regression functions are assumed to vary continuously through the boundary. For more discussion, see [48], [49], [56], [42], [13], and references therein.

We focus on a *location-based* approach to BD designs: local polynomial regressions are fit using the two-dimensional score itself near each boundary point. This distinguishes our approach from analyses that first transform the score into a scalar distance-to-boundary measure, or that use frontier, binned, or pooled one-dimensional RD analyses as approximations [23, and references therein]. Those alternatives can be useful benchmarks, but they

do not directly deliver a theory for location-based local polynomial estimation using the full bivariate score. Although pointwise versions of these estimators are natural extensions of familiar univariate local polynomial methods, their statistical properties in BD designs are different from those in univariate RD designs [22, 24].

The BATEC captures treatment effect heterogeneity along  $\mathcal{B}$ , but researchers may also want scalar summaries of that heterogeneity. We study two leading examples. The *Weighted Boundary Average Treatment Effect* (WBATE) is

$$\tau_{\text{WBATE}} = \frac{\int_{\mathcal{B}} \tau(\mathbf{x})w(\mathbf{x}) \, d\mathbf{x}}{\int_{\mathcal{B}} w(\mathbf{x}) \, d\mathbf{x}},$$

where  $w(\mathbf{x})$  is a weighting function. Motivated by the SPP application, the main paper assumes that  $\mathcal{B}$  is a piecewise-linear boundary with finitely many line segments, so the integrals above are line integrals; the supplemental appendix covers more general assignment boundaries, including irregular boundaries sometimes encountered in geographic RD applications. Intuitively,  $\tau_{\text{WBATE}}$  averages all boundary treatment effects; see [49] and [56] for related discussion. The *Largest Boundary Average Treatment Effect* (LBATE) is

$$\tau_{\text{LBATE}} = \sup_{\mathbf{x} \in \mathcal{B}} \tau(\mathbf{x}),$$

which captures the largest causal effect along the assignment boundary. See [1] for discussion of extreme treatment effects in policy decisions. In the SPP application, WBATE and LBATE summarize, respectively, the average and largest causal effect of receiving the college subsidy across students near the eligibility boundary.

We make four contributions. First, we develop pointwise and uniform estimation and inference methods for the location-based local polynomial estimator of the BATEC. The uniform results are new and allow researchers to form confidence bands for the entire treatment effect curve along  $\mathcal{B}$ . Second, we provide mean squared error (MSE) expansions, bandwidth selection rules, robust bias-corrected inference, and implementation guidance for the location-based approach. Third, we develop estimation and inference methods for

WBATE and LBATE, thereby covering both average and extreme summaries of treatment effect heterogeneity along the boundary. Fourth, we extend the results to BD designs with imperfect compliance, delivering pointwise and uniform inference for fuzzy BATEC, fuzzy WBATE, and fuzzy LBATE parameters. The supplemental appendix gives more general theory for higher-dimensional scores, more general boundaries, derivatives, and coordinate-specific bandwidths, while the companion R package `rd2d` implements the proposed methods.

Figure 1b previews the empirical value of these methods in the SPP application. It reports estimated treatment effects  $\hat{\tau}(\mathbf{b}_j)$  at the 40 boundary points in Figure 1a, together with point estimates  $\hat{\tau}_{\text{WBATE}}$  and  $\hat{\tau}_{\text{LBATE}}$  of WBATE and LBATE, respectively. The pattern suggests meaningful treatment effect heterogeneity: among students near the threshold of academic merit, the program appears to have a roughly constant effect across wealth levels; among students near the threshold of wealth, the effect appears to decline as academic performance improves. This finding suggests that students from underprivileged backgrounds benefit more from SPP than wealthier and high-achieving students. The application also features one-sided imperfect compliance: roughly 41% of eligible students opted not to take up the SPP subsidy. We use the fuzzy methods to account for imperfect compliance in the analysis.

## 1.1 Organization and Related Literature

Section 2 introduces the formal causal inference framework and location-based estimator. For a review of the methodological RD literature see [11], and for practical introductions see [14, 19]. For an overview of empirical practice employing BD designs see [23].

Section 3 presents pointwise, integrated mean squared, and uniform estimation and inference methods for the BATEC. The pointwise results generalize well-known results in the nonparametric literature [see 37, and references therein], and early BD design results in [48]. The uniform results are new and provide valid inference for the entire treatment effect curve and transformations thereof. We also discuss bandwidth selection based on MSE expansions, robust bias-corrected inference, and implementation details, building on and generalizing state-of-the-art univariate RD methods; see [7] for bandwidth selection, and

[3, 5, 8] for robust bias correction.

Sections 4 and 5 study the aggregated causal parameters  $\tau_{\text{WBATE}}$  and  $\tau_{\text{LBATE}}$ . WBATE is an integral functional over a one-dimensional manifold of a two-dimensional conditional expectation contrast. Related integral functionals on submanifolds have only been studied recently; in concurrent work, [25] developed series/sieve methods for such functionals, whereas our results use local polynomial regression as the nonparametric ingredient. See [49] and [56] for early methodological discussion in the context of BD designs. Our large-sample nonparametric results complement the design-based methods developed in [43] and [31], and the Bayesian methods developed in [50] for BD designs.

Section 6 considers imperfect compliance in BD designs. We present estimation and inference methods for fuzzy BATEC, fuzzy WBATE, and fuzzy LBATE, while noting that the causal interpretation of these parameters is an active research area; see [29], [51], [40], [13, 15] for fuzzy multidimensional RD designs, as well as [2] for univariate fuzzy RD designs.

Section 7 applies the methods to the SPP data, revising the main results reported in [46] and documenting sharp and fuzzy treatment effect heterogeneity along  $\mathcal{B}$ . Section 8 discusses derivative estimation, multidimensional scores, covariate adjustment for efficiency improvements and heterogeneity analysis, bivariate RD Plots, and clustered and spatially correlated data.

The supplemental appendix presents more general theoretical results, reports all proofs, and gives other technical results that may be of independent interest. In particular, it generalizes the main results to (i)  $d$ -dimensional scores, for  $d \geq 2$ ; (ii) more general assignment boundaries  $\mathcal{B} \subseteq \mathbb{R}^d$  with effective dimension  $d - 1$ ; (iii) derivative estimation and inference; and (iv) different bandwidths for each coordinate of the score. Those results leverage ideas from geometric measure theory [53, 33], thereby accommodating geographic and other BD settings where  $\mathcal{B}$  is irregular. From a technical perspective, Section SA-7 leverages ideas in [12] and presents new strong approximation theorems for empirical processes with polynomially bounded moments. In addition, the supplemental appendix gives practical implementation details in Section SA-9, empirical results in Section SA-10, and simulation evidence in Section SA-11.

Finally, a companion software article [20] discusses the general-purpose R software package `rd2d` (<https://rdpackages.github.io/rd2d>) implementing the methods developed in this paper. We also provide complete replication files for our numerical results.

## 1.2 Notation

We employ standard concepts and notations from empirical process theory [55, 35]. For a random variable  $Z_i$ , we write  $\mathbb{E}_n[g(Z_i)] = n^{-1} \sum_{i=1}^n g(Z_i)$ . For sequences, we write  $a_n = o(b_n)$  if  $\limsup_{n \rightarrow \infty} \frac{|a_n|}{|b_n|} = 0$ , and  $a_n \lesssim b_n$  if there exist constants  $C$  and  $N > 0$  such that  $n > N$  implies  $|a_n| \leq C|b_n|$ . For sequences of random variables, we write  $a_n = o_{\mathbb{P}}(b_n)$  if  $\text{plim}_{n \rightarrow \infty} \frac{|a_n|}{|b_n|} = 0$ , and  $|a_n| \lesssim_{\mathbb{P}} |b_n|$  if  $\limsup_{M \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}[|\frac{a_n}{b_n}| \geq M] = 0$ . For a set  $A$ ,  $\text{bd}(A)$  denotes the boundary of  $A$  and  $\text{int}(A)$  denotes the interior of  $A$ . Finally,  $\Phi(x)$  denotes the standard Gaussian cumulative distribution function.

## 2 Setup

We begin with sharp BD designs, where all units comply perfectly with their treatment assignment. Let  $\mathcal{A}_0$  and  $\mathcal{A}_1$  denote the control and treatment regions, respectively, and let  $\mathcal{B} = \text{bd}(\mathcal{A}_0) \cap \text{bd}(\mathcal{A}_1)$  denote their common assignment boundary. We use the convention that boundary points belong to the treatment region, so  $\mathcal{B} \subseteq \mathcal{A}_1$  and  $\mathcal{B} \cap \mathcal{A}_0 = \emptyset$ . Units are assigned to the control or treatment conditions according to the rule  $\mathbf{1}(\mathbf{X}_i \in \mathcal{A}_1)$ , and the observed outcome is  $Y_i = \mathbf{1}(\mathbf{X}_i \in \mathcal{A}_0)Y_i(0) + \mathbf{1}(\mathbf{X}_i \in \mathcal{A}_1)Y_i(1)$ . We start by assuming that all units assigned to the treatment condition do receive the treatment, and none of the units assigned to the control condition receive the treatment. Section 6 discusses fuzzy BD designs, where treatment status need not coincide with treatment assignment.

We impose the following basic conditions on the underlying data generating process.

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

- (i)  $(Y_1(t), \mathbf{X}_1^\top)^\top, \dots, (Y_n(t), \mathbf{X}_n^\top)^\top$  are independent and identically distributed random vectors.

- (ii) *The distribution of  $\mathbf{X}_i$  has compact support  $\mathcal{X} \subseteq \mathbb{R}^2$  with nonempty interior and a Lebesgue density  $f_X(\mathbf{x})$  that is continuous and bounded away from zero on  $\mathcal{X}$ .*
- (iii)  *$\mu_t(\mathbf{x}) = \mathbb{E}[Y_i(t)|\mathbf{X}_i = \mathbf{x}]$  has a  $(p+1)$ -times continuously differentiable extension to an open neighborhood of  $\mathcal{X}$ .*
- (iv)  *$\sigma_t^2(\mathbf{x}) = \mathbb{V}[Y_i(t)|\mathbf{X}_i = \mathbf{x}]$  is bounded away from zero and continuous on  $\mathcal{X}$ .*
- (v)  *$\sup_{\mathbf{x} \in \mathcal{X}} \mathbb{E}[|Y_i(t)|^{2+v}|\mathbf{X}_i = \mathbf{x}] < \infty$  for some  $v \geq 2$ .*

These conditions are the bivariate analogues of standard continuity-based RD assumptions. Part (i) imposes random sampling; this may be restrictive in some geographic applications with spatial dependence, an issue we discuss in Section 8. Part (ii) imposes compact score support and ensures that local neighborhoods around boundary points contain observations with positive probability. Because Assumption 2 places  $\mathcal{B}$  in the interior of  $\mathcal{X}$ , the boundary-local neighborhoods used by the estimators remain away from the edge of the score support for small bandwidths. Parts (iii)–(v) impose standard smoothness, variance, and moment conditions on the potential outcomes.

Under Assumption 1, the BATEC introduced in Section 1 is identified at every boundary point as  $\tau(\mathbf{x}) = \mu_1(\mathbf{x}) - \mu_0(\mathbf{x})$  for  $\mathbf{x} \in \mathcal{B}$ . Because treatment assignment changes discontinuously at  $\mathcal{B}$  while  $\mu_0(\cdot)$  and  $\mu_1(\cdot)$  are continuous, the conditional mean of the observed outcome on side  $\mathcal{A}_t$  identifies  $\mu_t(\mathbf{x})$  as one approaches  $\mathbf{x} \in \mathcal{B}$  from that side. This is the usual RD continuity argument adapted to a bivariate score; see [48], [49], [42], [13], and references therein. Identification of  $\tau_{\text{WBATE}}$  and  $\tau_{\text{LBATE}}$  follows from identification of  $\tau(\mathbf{x})$  over  $\mathcal{B}$ . Our focus is on the statistical properties of local polynomial methods for estimating these objects and conducting pointwise, uniform, and aggregate inference.

For each  $\mathbf{x} \in \mathcal{B}$ , the location-based estimator fits two local polynomial regressions, one using observations from the treatment side and one using observations from the control side. The resulting BATEC estimator is

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

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

$$\widehat{\boldsymbol{\beta}}_t(\mathbf{x}) = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^{p_p}} \mathbb{E}_n \left[ (Y_i - \mathbf{r}_p(\mathbf{X}_i - \mathbf{x})^\top \boldsymbol{\beta})^2 K_h(\mathbf{X}_i - \mathbf{x}) \mathbf{1}(\mathbf{X}_i \in \mathcal{A}_t) \right], \quad (1)$$

with  $p_p = (2+p)(1+p)/2$  and  $\mathbf{r}_p(\mathbf{u})$  collecting all monomials  $u_1^{a_1} u_2^{a_2}$  with  $a_1 + a_2 \leq p$ , starting with the constant term. Let  $\mathbf{e}_0$  denote the unit vector selecting this constant term. The kernel is  $K_h(\mathbf{u}) = K(u_1/h, u_2/h)/h^2$ , where  $K(\cdot)$  is a bivariate function and  $h$  is a bandwidth. We use a common bandwidth for expositional simplicity and because researchers can standardize the two score coordinates before implementation. The supplemental appendix allows for different bandwidths across score coordinates.

We impose the following assumption on the assignment boundary curve and bivariate kernel function.

**Assumption 2** (Boundary and Kernel). (i)  $\mathcal{B} \subset \text{int}(\mathcal{X})$ , and  $\mathcal{B}$  is piecewise linear with finitely many line segments and positive finite length.

(ii)  $K : \mathbb{R}^2 \rightarrow [0, \infty)$  is compactly supported and Lipschitz continuous, or  $K(\mathbf{u}) = \mathbf{1}(\mathbf{u} \in [-1, 1]^2)$ .

(iii)  $K(\mathbf{u}) > 0$  for all  $\mathbf{u}$  in a neighborhood of the origin.

Assumption 2(i) gives a simple and transparent boundary condition. It ensures that line integrals over  $\mathcal{B}$  are well defined, that the boundary can be discretized for uniform inference and aggregate estimands, and that the boundary is away from the edge of the support of the score. Assumption 2(ii) imposes standard regularity conditions on the kernel. Assumption 2(iii) imposes local positivity around the origin; together with the piecewise-linear boundary in Assumption 2(i), this condition implies that, after local rescaling, the kernel assigns nonnegligible mass to each side of the boundary and hence each side contains enough kernel-weighted variation to identify the local polynomial coefficients uniformly along  $\mathcal{B}$ . Assumption 2 may be restrictive in geographic or other applications where the boundary cannot be well approximated by piecewise linear functions. In the supplemental appendix, we give more primitive geometric conditions that allow for irregular boundaries.

Finally, to ensure that  $\tau_{\text{WBATE}}$  is well-defined, we impose the following conditions on the

weight function.

**Assumption 3** (Weight Function and Boundary). *Let  $w : \mathcal{B} \rightarrow (0, \infty)$  satisfy  $\sup_{\mathbf{x} \in \mathcal{B}} w(\mathbf{x}) < \infty$ ,  $\inf_{\mathbf{x} \in \mathcal{B}} w(\mathbf{x}) > 0$ , and  $\int_{\mathcal{B}} w(\mathbf{x}) d\mathbf{x} < \infty$ .*

Weights may be normalized without loss of generality. In particular,  $w(\mathbf{x}) = f_X(\mathbf{x})$  satisfies Assumption 3. In this case, the WBATE reduces to the *boundary average treatment effect* (BATE):  $\tau_{\text{BATE}} = \int_{\mathcal{B}} \tau(\mathbf{x}) f(\mathbf{x} | \mathbf{X}_i \in \mathcal{B}) d\mathbf{x}$ , with  $f(\mathbf{x} | \mathbf{X}_i \in \mathcal{B}) = \frac{f_X(\mathbf{x})}{\int_{\mathcal{B}} f_X(\mathbf{x}) d\mathbf{x}}$ . More formally, the expression  $f(\mathbf{x} | \mathbf{X}_i \in \mathcal{B})$  denotes the normalized boundary density with respect to arc length, with Radon–Nikodym derivative proportional to  $f_X(\mathbf{x})$  on  $\mathcal{B}$ . This parameter is the density-weighted average causal effect along the assignment boundary, and is discussed by [56]. See [22] for an alternative regression-based approach commonly used in practice to estimate the BATE.

### 3 Boundary Average Treatment Effect Curve

This section studies pointwise, integrated, and uniform estimation and inference for the BATEC ( $\tau(\mathbf{x}) : \mathbf{x} \in \mathcal{B}$ ) in sharp BD designs. We focus on a bivariate score, a scalar bandwidth  $h$ , and treatment-effect levels; in the supplemental appendix, we generalize this setup to treatment-effect derivatives, diagonal bandwidth matrices, and score of three or more dimensions. Throughout this section, integrals over  $\mathcal{B}$  are line integrals with respect to arc length; in the appendix these are written as integrals with respect to a Hausdorff measure [33]. Pointwise results follow from standard local polynomial arguments once the one-sided local design is nonsingular. The integrated MSE expansion and the uniform distribution approximation require additional empirical-process and geometric arguments, all of which are developed in the supplemental appendix.

#### 3.1 Treatment Effect Estimation

The first result gives pointwise and uniform convergence rates for the location-based estimator. The stochastic terms are the usual variance component and a polynomial-moment

empirical-process remainder, while the smoothing bias is of order  $h^{p+1}$ .

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

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

Theorem 1 implies consistency of the treatment effect estimator based on the bivariate location score under bandwidth sequences that make the displayed rate vanish; for uniform consistency, it suffices that  $n^{(1+v)/(2+v)}h^2/\log(1/h) \rightarrow \infty$  and  $h \rightarrow 0$ . The bias order is  $h^{p+1}$  because smoothing is performed in the two-dimensional score space on each side of the boundary; it does not require reparameterizing the boundary by a scalar running variable. This feature is important when  $\mathcal{B}$  has corners or other piecewise-linear changes in direction; see [24] for more discussion.

Next, we state the conditional MSE expansion. The pointwise expansion has the standard bias–variance form; the integrated expansion averages this local expansion along the one-dimensional assignment boundary. Recall that  $t \in \{0, 1\}$ . The population Gram matrix is

$$\mathbf{\Gamma}_{t,\mathbf{x}} = \mathbb{E} \left[ \mathbf{r}_p \left( \frac{\mathbf{X}_i - \mathbf{x}}{h} \right) \mathbf{r}_p \left( \frac{\mathbf{X}_i - \mathbf{x}}{h} \right)^\top K_h(\mathbf{X}_i - \mathbf{x}) \mathbf{1}(\mathbf{X}_i \in \mathcal{A}_t) \right].$$

The leading non-random pointwise variance is  $V_{\mathbf{x}} = V_{1,\mathbf{x}} + V_{0,\mathbf{x}}$ , where  $V_{t,\mathbf{x}} = \mathbf{e}_0^\top \mathbf{\Gamma}_{t,\mathbf{x}}^{-1} \boldsymbol{\Sigma}_{t,\mathbf{x},\mathbf{x}} \mathbf{\Gamma}_{t,\mathbf{x}}^{-1} \mathbf{e}_0$  with

$$\boldsymbol{\Sigma}_{t,\mathbf{x},\mathbf{x}} = h^2 \mathbb{E} \left[ \mathbf{r}_p \left( \frac{\mathbf{X}_i - \mathbf{x}}{h} \right) \mathbf{r}_p \left( \frac{\mathbf{X}_i - \mathbf{x}}{h} \right)^\top K_h(\mathbf{X}_i - \mathbf{x})^2 \sigma_t^2(\mathbf{X}_i) \mathbf{1}(\mathbf{X}_i \in \mathcal{A}_t) \right].$$

Using standard multi-index notation, the leading non-random pointwise bias is  $B_{\mathbf{x}} = B_{1,\mathbf{x}} - B_{0,\mathbf{x}}$ , where

$$B_{t,\mathbf{x}} = \mathbf{e}_0^\top \mathbf{\Gamma}_{t,\mathbf{x}}^{-1} \sum_{|\mathbf{k}|=p+1} \frac{\mu_t^{(\mathbf{k})}(\mathbf{x})}{\mathbf{k}!} \mathbb{E} \left[ \mathbf{r}_p \left( \frac{\mathbf{X}_i - \mathbf{x}}{h} \right) \left( \frac{\mathbf{X}_i - \mathbf{x}}{h} \right)^{\mathbf{k}} K_h(\mathbf{X}_i - \mathbf{x}) \mathbf{1}(\mathbf{X}_i \in \mathcal{A}_t) \right].$$

The following theorem gives the MSE expansions. Let  $\mathbf{X} = (\mathbf{X}_1^\top, \dots, \mathbf{X}_n^\top)$ .

**Theorem 2** (MSE Expansions: Sharp BATEC). *Suppose Assumptions 1, 2, and 3 hold. If  $n^{v/(2+v)}h^2/\log(1/h) \rightarrow \infty$  and  $h \rightarrow 0$ , then*

- (i)  $\mathbb{E}[(\widehat{\tau}(\mathbf{x}) - \tau(\mathbf{x}))^2|\mathbf{X}] = h^{2(p+1)}B_{\mathbf{x}}^2 + \frac{1}{nh^2}V_{\mathbf{x}} + o_{\mathbb{P}}(\mathfrak{R})$ , and
- (ii)  $\int_{\mathcal{B}} \mathbb{E}[(\widehat{\tau}(\mathbf{x}) - \tau(\mathbf{x}))^2|\mathbf{X}]w(\mathbf{x}) d\mathbf{x} = h^{2(p+1)} \int_{\mathcal{B}} B_{\mathbf{x}}^2w(\mathbf{x}) d\mathbf{x} + \frac{1}{nh^2} \int_{\mathcal{B}} V_{\mathbf{x}}w(\mathbf{x}) d\mathbf{x} + o_{\mathbb{P}}(\mathfrak{R})$ ,

with  $\mathfrak{R} = h^{2p+2} + n^{-1}h^{-2}$ .

For a common scalar bandwidth  $h$ , ignoring higher-order terms gives the approximate MSE-optimal and IMSE-optimal bandwidth choices

$$h_{\text{MSE},\mathbf{x}} = \left( \frac{2V_{\mathbf{x}}}{(2p+2)B_{\mathbf{x}}^2} \frac{1}{n} \right)^{1/(2p+4)} \quad \text{and} \quad h_{\text{IMSE}} = \left( \frac{2 \int_{\mathcal{B}} V_{\mathbf{x}}w(\mathbf{x}) d\mathbf{x}}{(2p+2) \int_{\mathcal{B}} B_{\mathbf{x}}^2w(\mathbf{x}) d\mathbf{x}} \frac{1}{n} \right)^{1/(2p+4)},$$

provided that  $B_{\mathbf{x}} \neq 0$  and  $\int_{\mathcal{B}} B_{\mathbf{x}}^2w(\mathbf{x}) d\mathbf{x} \neq 0$ , respectively. These choices are infeasible because a preliminary bandwidth, as well as estimates of the conditional variances and higher-order derivatives of the conditional mean, are needed. We discuss feasible implementation in Section 3.3, and in the companion software article [20]. The same calculations extend in the appendix to diagonal bandwidth matrices, higher-dimensional scores, derivatives of the BATEC, and more general assignment boundaries.

The pointwise rates and MSE expansion mirror familiar nonparametric local polynomial results [see 54, for a textbook review]. The integrated MSE expansion in Theorem 2 requires additional care because the criterion integrates pointwise conditional MSE over the one-dimensional manifold  $\mathcal{B}$ . These point estimation results are also the inputs for the uniform and aggregation results below.

### 3.2 Uncertainty Quantification

Given a bandwidth choice, define the feasible  $t$ -statistic  $\widehat{T}(\mathbf{x}) = (\widehat{\tau}(\mathbf{x}) - \tau(\mathbf{x}))/\widehat{\Omega}_{\mathbf{x},\mathbf{x}}^{1/2}$  where, using standard least squares algebra, the covariance estimator is

$$\widehat{\Omega}_{\mathbf{x}_1,\mathbf{x}_2} = \frac{1}{nh^2} \mathbf{e}_0^\top \left[ \widehat{\Gamma}_{0,\mathbf{x}_1}^{-1} \widehat{\Sigma}_{0,\mathbf{x}_1,\mathbf{x}_2} \widehat{\Gamma}_{0,\mathbf{x}_2}^{-1} + \widehat{\Gamma}_{1,\mathbf{x}_1}^{-1} \widehat{\Sigma}_{1,\mathbf{x}_1,\mathbf{x}_2} \widehat{\Gamma}_{1,\mathbf{x}_2}^{-1} \right] \mathbf{e}_0$$

with  $\widehat{\Sigma}_{t,\mathbf{x}_1,\mathbf{x}_2} = h^2 \mathbb{E}_n [\mathbf{r}_p(\frac{\mathbf{X}_i - \mathbf{x}_1}{h}) \mathbf{r}_p(\frac{\mathbf{X}_i - \mathbf{x}_2}{h})^\top K_h(\mathbf{X}_i - \mathbf{x}_1) K_h(\mathbf{X}_i - \mathbf{x}_2) \widehat{\varepsilon}_i(\mathbf{x}_1) \widehat{\varepsilon}_i(\mathbf{x}_2) \mathbf{1}(\mathbf{X}_i \in \mathcal{A}_t)]$   
and  $\widehat{\varepsilon}_i(\mathbf{x}) = Y_i - \mathbf{1}(\mathbf{X}_i \in \mathcal{A}_0) \mathbf{r}_p(\mathbf{X}_i - \mathbf{x})^\top \widehat{\beta}_0(\mathbf{x}) - \mathbf{1}(\mathbf{X}_i \in \mathcal{A}_1) \mathbf{r}_p(\mathbf{X}_i - \mathbf{x})^\top \widehat{\beta}_1(\mathbf{x})$ , for all  $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{B}$  and  $t \in \{0, 1\}$ .

Wald-type feasible confidence intervals and confidence bands over  $\mathcal{B}$  take the form

$$\widehat{I}_\alpha(\mathbf{x}) = \left[ \widehat{\tau}(\mathbf{x}) - \varphi_\alpha \widehat{\Omega}_{\mathbf{x},\mathbf{x}}^{1/2}, \widehat{\tau}(\mathbf{x}) + \varphi_\alpha \widehat{\Omega}_{\mathbf{x},\mathbf{x}}^{1/2} \right], \quad \mathbf{x} \in \mathcal{B},$$

for any  $\alpha \in (0, 1)$ , where  $\varphi_\alpha$  denotes the appropriate quantile for the desired coverage objective. For pointwise inference, the usual normal approximation gives  $\sup_{u \in \mathbb{R}} |\mathbb{P}[\widehat{T}(\mathbf{x}) \leq u] - \Phi(u)| \rightarrow 0$  for each  $\mathbf{x} \in \mathcal{B}$ , provided that the ‘‘small bias’’ condition  $nh^{2p+4} \rightarrow 0$  holds. This result gives the usual pointwise confidence interval for  $\tau(\mathbf{x})$ .

For uniform inference over  $\mathcal{B}$ , two challenges arise: (i) the stochastic process  $(\widehat{T}(\mathbf{x}) : \mathbf{x} \in \mathcal{B})$  is not asymptotically tight, and thus it does not converge weakly in the space of uniformly bounded real functions supported on  $\mathcal{B}$  and equipped with the uniform norm [55, 35]; and (ii) the geometry of  $\mathcal{B}$  can affect the validity of the distributional approximation. We handle both issues by approximating the distribution of the supremum  $\sup_{\mathbf{x} \in \mathcal{B}} |\widehat{T}(\mathbf{x})|$  directly, rather than relying on weak convergence of the full process. This approach enables asymptotically valid confidence bands because  $\mathbb{P}[\tau(\mathbf{x}) \in \widehat{I}_\alpha(\mathbf{x}), \text{ for all } \mathbf{x} \in \mathcal{B}] = \mathbb{P}[\sup_{\mathbf{x} \in \mathcal{B}} |\widehat{T}(\mathbf{x})| \leq \varphi_\alpha]$ . The resulting Gaussian approximation for the supremum requires VC-type entropy and moment bounds. The supplemental appendix also gives a stronger process-level coupling under additional technical conditions. These approximation results leverage technical tools from [12], [16], [26, 27], [28], and [32]. Let  $\mathbf{V} = ((Y_1, \mathbf{X}_1^\top), \dots, (Y_n, \mathbf{X}_n^\top))^\top$ .

**Theorem 3** (Confidence Intervals and Bands: Sharp BATEC). *Suppose Assumptions 1 and 2 hold.*

- (i) *If  $nh^2 \rightarrow \infty$  and  $nh^{2p+4} \rightarrow 0$ , then  $\mathbb{P}[\tau(\mathbf{x}) \in \widehat{I}_\alpha(\mathbf{x})] \rightarrow 1 - \alpha$  with  $\varphi_\alpha = \Phi^{-1}(1 - \alpha/2)$ .*
- (ii) *If  $\liminf_{n \rightarrow \infty} \log h / \log n > -\infty$ ,  $n^{v/(2+v)} h^2 / (\log n)^9 \rightarrow \infty$ , and  $nh^{2p+4} \log n \rightarrow 0$ , then  $\mathbb{P}[\tau(\mathbf{x}) \in \widehat{I}_\alpha(\mathbf{x}), \text{ for all } \mathbf{x} \in \mathcal{B}] \rightarrow 1 - \alpha$ , with  $\varphi_\alpha = \inf \{c > 0 : \mathbb{P}[\sup_{\mathbf{x} \in \mathcal{B}} |\widehat{Z}(\mathbf{x})| \geq c | \mathbf{V}] \leq \alpha\}$ , where  $(\widehat{Z}(\mathbf{x}) : \mathbf{x} \in \mathcal{B})$  is a Gaussian process conditional on  $\mathbf{V}$  with  $\mathbb{E}[\widehat{Z}(\mathbf{x}) | \mathbf{V}] = 0$  and  $\text{Cov}[\widehat{Z}(\mathbf{x}_1), \widehat{Z}(\mathbf{x}_2) | \mathbf{V}] = \widehat{\Omega}_{\mathbf{x}_1, \mathbf{x}_2} / (\widehat{\Omega}_{\mathbf{x}_1, \mathbf{x}_1} \widehat{\Omega}_{\mathbf{x}_2, \mathbf{x}_2})^{1/2}$ , for all  $\mathbf{x}, \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{B}$ .*

$\mathcal{B}$ .

This theorem gives asymptotically valid pointwise and uniform uncertainty quantification for  $\tau(\mathbf{x})$  using the location-based estimator  $\hat{\tau}(\mathbf{x})$ . The appendix also shows how the result extends to higher-dimensional scores, derivatives, and more general boundaries under the corresponding geometric regularity conditions. As expected in a nonparametric smoothing setting, the pointwise undersmoothing condition  $nh^{2p+4} \rightarrow 0$  and its  $\log(n)$ -strengthened uniform analogue rule out the (I)MSE-optimal point estimator of  $\tau(\mathbf{x})$ . Thus, for implementation of both pointwise and uniform inference, we employ robust bias correction [3, 5, 8]; see Section 3.3 for details.

### 3.3 Implementation

It is straightforward to implement local and global bandwidth selectors based on Theorem 2. In particular, replacing the leading bias and variance quantities,  $B_{\mathbf{x}}$  and  $V_{\mathbf{x}}$ , with preliminary estimators, we obtain the feasible plug-in bandwidth selectors

$$\hat{h}_{\text{MSE},\mathbf{x}} = \left( \frac{2\hat{V}_{\mathbf{x}}}{(2p+2)\hat{B}_{\mathbf{x}}^2 n} \right)^{1/(2p+4)} \quad \text{and} \quad \hat{h}_{\text{IMSE}} = \left( \frac{2 \int_{\mathcal{B}} \hat{V}_{\mathbf{x}} w(\mathbf{x}) d\mathbf{x}}{(2p+2) \int_{\mathcal{B}} \hat{B}_{\mathbf{x}}^2 w(\mathbf{x}) d\mathbf{x} n} \right)^{1/(2p+4)},$$

where, for a preliminary bandwidth choice  $a \rightarrow 0$ ,  $\hat{B}_{\mathbf{x}} = \hat{B}_{1,\mathbf{x}} - \hat{B}_{0,\mathbf{x}}$  is constructed using  $\hat{B}_{t,\mathbf{x}} = \mathbf{e}_0^\top \hat{\Gamma}_{t,\mathbf{x}}^{-1} \sum_{|\mathbf{k}|=p+1} \frac{\hat{\mu}_t^{(\mathbf{k})}(\mathbf{x})}{\mathbf{k}!} \mathbb{E}_n \left[ \mathbf{r}_p \left( \frac{\mathbf{X}_i - \mathbf{x}}{a} \right) \left( \frac{\mathbf{X}_i - \mathbf{x}}{a} \right)^{\mathbf{k}} K_a(\mathbf{X}_i - \mathbf{x}) \mathbf{1}(\mathbf{X}_i \in \mathcal{A}_t) \right]$ , with  $\hat{\Gamma}_{t,\mathbf{x}}$  computed using the preliminary bandwidth  $a$ , and where  $\hat{\mu}_t^{(\mathbf{k})}(\mathbf{x})$  is a preliminary estimator of  $\mu_t^{(\mathbf{k})}(\mathbf{x})$ , and  $\hat{V}_{\mathbf{x}} = na^2 \hat{\Omega}_{\mathbf{x},\mathbf{x}}$  is constructed using the variance estimator with the preliminary bandwidth  $a$ . Omitted implementation details are discussed in the companion software article [20]. See also [7] for a review on modern bandwidth selection methods in RD designs with univariate score.

The bandwidth choices  $\hat{h}_{\text{MSE},\mathbf{x}}$  and  $\hat{h}_{\text{IMSE}}$  can be used to implement (I)MSE-optimal  $\hat{\tau}(\mathbf{x})$  treatment effect estimators, both pointwise and uniformly over  $\mathcal{B}$ . Furthermore, leveraging the results in Theorem 3, a simple application of robust bias-corrected inference proceeds by employing the same (I)MSE-optimal bandwidth (for  $p$ th order point estimation), but then

constructing the t-statistic  $\widehat{T}(\mathbf{x})$  with a  $(p+1)$ th polynomial order instead of  $p$ th polynomial order. The core idea is to simultaneously (i) debias the (I)MSE-optimal point estimator  $\widehat{\tau}(\mathbf{x})$ , and (ii) adjust the variance estimator to incorporate the additional uncertainty introduced by the bias correction. This inference approach has several theoretical advantages [3, 5, 8], and has been validated empirically [39, 30].

Finally, regarding the computation of the Gaussian process conditional on  $\mathbf{V}$ ,  $(\widehat{Z}(\mathbf{x}) : \mathbf{x} \in \mathcal{B})$ , there are two methodological issues to consider. First, simulation is implemented over a grid of points forming a discretization of the index set of the continuous stochastic process  $\widehat{Z}$ ; as the number of points in the mesh increases, the approximation becomes more accurate. Second, the estimated (discretized) covariance function may fail to be positive definite in finite samples, but this issue can be easily fixed via regularization; see [18] for a discussion and related technical results.

Section SA-9 in the supplemental appendix provides additional implementation details; see also the companion software package `rd2d` and article [20].

## 4 Weighted Boundary Average Treatment Effect

Without loss of generality, we normalize the weight function so that  $\int_{\mathcal{B}} w(\mathbf{x}) d\mathbf{x} = 1$ , and consider  $\tau_{\text{WBATE}} = \int_{\mathcal{B}} \tau(\mathbf{x})w(\mathbf{x}) d\mathbf{x}$ , where  $w$  satisfies Assumption 3. The WBATE aggregates the heterogeneous treatment effects  $(\tau(\mathbf{x}) : \mathbf{x} \in \mathcal{B})$  according to a user-specified weighting scheme. The same construction can be applied to a subset of the boundary, as in [49], in which case the estimand is the weighted average over that target subset.

The plug-in estimator is  $\widehat{\tau}_{\text{WBATE}} = \int_{\mathcal{B}} \widehat{\tau}(\mathbf{x})w(\mathbf{x}) d\mathbf{x}$ . When analytic line integration is inconvenient, the integral can be approximated on a boundary grid. For grid points  $(\mathbf{b}_j : j = 1, \dots, J)$  and quadrature weights  $(\omega_j : j = 1, \dots, J)$ , normalized so that  $\sum_{j=1}^J \omega_j w(\mathbf{b}_j) = 1$ ,  $\widehat{\tau}_{\text{WBATE}} \approx \sum_{j=1}^J \omega_j \widehat{\tau}(\mathbf{b}_j)w(\mathbf{b}_j)$ . In some applications the weights may instead be chosen directly from local empirical information, for example  $\omega_j w(\mathbf{b}_j) = N_j / \sum_{\ell=1}^J N_{\ell}$  with  $N_j = \sum_{i=1}^n \mathbf{1}(\|\mathbf{X}_i - \mathbf{b}_j\| \leq h)$ . See [20] for implementation details.

Our first result establishes a conditional MSE expansion for  $\widehat{\tau}_{\text{WBATE}}$ . Let  $B_{\text{WBATE}} = B_{1, \text{WBATE}} -$

$B_{0,\text{WBATE}}$ , with  $B_{t,\text{WBATE}} = \int_{\mathcal{B}} B_{t,\mathbf{x}} w(\mathbf{x}) d\mathbf{x}$ , and let  $\Omega_{\text{WBATE}} = \Omega_{1,\text{WBATE}} + \Omega_{0,\text{WBATE}}$ , with  $\Omega_{t,\text{WBATE}} = \int_{\mathcal{B}} \int_{\mathcal{B}} \Omega_{t,\mathbf{x}_1,\mathbf{x}_2} w(\mathbf{x}_1) w(\mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2$ , for  $t \in \{0, 1\}$ . For the variance-rate and inference results below, we impose the scalar aggregate nondegeneracy condition  $\Omega_{\text{WBATE}} \gtrsim (nh)^{-1}$ , which rules out cancellation of the local covariance kernel after weighting and integrating over  $\mathcal{B}$ .

**Theorem 4** (MSE Expansion: Sharp WBATE). *Suppose Assumptions 1, 2 and 3 hold, and  $\Omega_{\text{WBATE}} \gtrsim (nh)^{-1}$ . If  $n^{v/(2+v)} h^2 / \log(1/h) \rightarrow \infty$  and  $h \rightarrow 0$ , then  $\mathbb{E}[(\hat{\tau}_{\text{WBATE}} - \tau_{\text{WBATE}})^2 | \mathbf{X}] = \Omega_{\text{WBATE}} + h^{2p+2} B_{\text{WBATE}}^2 + o_{\mathbb{P}}(\mathfrak{R})$ , where  $(nh)^{-1} \lesssim \Omega_{\text{WBATE}} \lesssim (nh)^{-1}$ , and  $\mathfrak{R} = (nh)^{-1} + h^{2p+2}$ .*

The theorem implies  $\hat{\tau}_{\text{WBATE}} = \tau_{\text{WBATE}} + o_{\mathbb{P}}(1)$ . It also shows how aggregation changes the effective variance order. The pointwise estimator  $\hat{\tau}(\mathbf{x})$  has variance of order  $(nh^2)^{-1}$ , whereas  $\hat{\tau}_{\text{WBATE}}$  has variance of order  $(nh)^{-1}$  because only boundary locations within  $O(h)$  of one another have non-negligible covariance. Thus WBATE behaves like a one-dimensional non-parametric estimator.

Writing the leading variance term as  $\Omega_{\text{WBATE}} = V_{\text{WBATE}}/(nh)$ , the infeasible MSE-optimal scalar bandwidth is  $h_{\text{WBATE}} = \left( \frac{V_{\text{WBATE}}}{(2p+2)B_{\text{WBATE}}^2 n} \right)^{1/(2p+3)}$ , provided that  $B_{\text{WBATE}} \neq 0$ . A feasible counterpart can be constructed using plug-in estimators of  $B_{\text{WBATE}}$  and  $V_{\text{WBATE}}$ , as in Section 3.3. The exponent differs from the pointwise bandwidth because the WBATE variance is of order  $(nh)^{-1}$  rather than  $(nh^2)^{-1}$ .

For inference, define the feasible  $t$ -statistic  $\hat{\mathbb{T}}_{\text{WBATE}} = (\hat{\tau}_{\text{WBATE}} - \tau_{\text{WBATE}}) / \hat{\Omega}_{\text{WBATE}}^{1/2}$ , with  $\hat{\Omega}_{\text{WBATE}} = \hat{\Omega}_{1,\text{WBATE}} + \hat{\Omega}_{0,\text{WBATE}}$ , where  $\hat{\Omega}_{t,\text{WBATE}} = \int_{\mathcal{B}} \int_{\mathcal{B}} \hat{\Omega}_{t,\mathbf{x}_1,\mathbf{x}_2} w(\mathbf{x}_1) w(\mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2$ , for  $t \in \{0, 1\}$ . The associated confidence interval estimator is

$$\hat{\mathbb{I}}_{\alpha,\text{WBATE}} = \left[ \hat{\tau}_{\text{WBATE}} - \varphi_{\alpha} \hat{\Omega}_{\text{WBATE}}^{1/2}, \hat{\tau}_{\text{WBATE}} + \varphi_{\alpha} \hat{\Omega}_{\text{WBATE}}^{1/2} \right],$$

where  $\varphi_{\alpha} = \Phi^{-1}(1 - \alpha/2)$ .

**Theorem 5** (Confidence Intervals: Sharp WBATE). *Suppose Assumptions 1, 2, and 3 hold, and  $\Omega_{\text{WBATE}} \gtrsim (nh)^{-1}$ . If  $n^{v/(2+v)} h^2 / \log(1/h) \rightarrow \infty$  and  $nh^{2p+3} \rightarrow 0$ , then  $\mathbb{P}[\tau_{\text{WBATE}} \in \hat{\mathbb{I}}_{\alpha,\text{WBATE}}] \rightarrow 1 - \alpha$ .*

For implementation, the MSE-optimal WBATE point estimator can be paired with robust

bias correction [3, 5, 8], following the same logic discussed after Theorem 3. See Section SA-9 in the supplemental appendix, and the companion software package `rd2d` and article [20], for more details.

## 5 Largest Boundary Average Treatment Effect

As an alternative to the WBATE, researchers may be interested in the largest treatment effect along the boundary,  $\tau_{\text{LBATE}} = \sup_{\mathbf{x} \in \mathcal{B}} \tau(\mathbf{x})$ , with plug-in estimator  $\hat{\tau}_{\text{LBATE}} = \sup_{\mathbf{x} \in \mathcal{B}} \hat{\tau}(\mathbf{x})$ . The smallest boundary treatment effect can be handled analogously by applying the same construction to  $-\tau(\mathbf{x})$ . In practice, the supremum is computed over a sufficiently fine boundary grid,  $\hat{\tau}_{\text{LBATE}} \approx \max_{1 \leq j \leq J} \hat{\tau}(\mathbf{b}_j)$ . If the target set is restricted to a subset of  $\mathcal{B}$ , then both the estimand and estimator should be interpreted relative to that subset.

Theorem 1 gives  $|\hat{\tau}_{\text{LBATE}} - \tau_{\text{LBATE}}| \leq \sup_{\mathbf{x} \in \mathcal{B}} |\hat{\tau}(\mathbf{x}) - \tau(\mathbf{x})|$ , and therefore gives the corresponding convergence rate. This rate implies consistency under the same side conditions discussed after Theorem 1. For uncertainty quantification, we project the uniform confidence band from Theorem 3 onto the supremum functional. Define

$$\hat{\mathbf{I}}_{\alpha, \text{LBATE}} = \left[ \sup_{\mathbf{x} \in \mathcal{B}} \left( \hat{\tau}(\mathbf{x}) - \varrho_{\alpha} \hat{\Omega}_{\mathbf{x}, \mathbf{x}}^{1/2} \right), \sup_{\mathbf{x} \in \mathcal{B}} \left( \hat{\tau}(\mathbf{x}) + \varrho_{\alpha} \hat{\Omega}_{\mathbf{x}, \mathbf{x}}^{1/2} \right) \right]$$

where  $\varrho_{\alpha} = \inf\{c > 0 : \mathbb{P}(\sup_{\mathbf{x} \in \mathcal{B}} |\hat{Z}(\mathbf{x})| \geq c | \mathbf{V}) \leq \alpha\}$ , and  $\hat{Z}$  is the conditionally mean-zero Gaussian process from Theorem 3.

**Theorem 6** (Confidence Intervals: Sharp LBATE). *Suppose the assumptions and conditions in Theorem 3 hold. If  $\liminf_{n \rightarrow \infty} \log h / \log n > -\infty$ ,  $n^{v/(2+v)} h^2 / (\log n)^9 \rightarrow \infty$  and  $nh^{2p+4} \log n \rightarrow 0$ , then  $\mathbb{P}[\tau_{\text{LBATE}} \in \hat{\mathbf{I}}_{\alpha, \text{LBATE}}] \geq 1 - \alpha + o(1)$ .*

The resulting interval is generally conservative because it is obtained from simultaneous coverage of the entire BATEC. Implementation follows the same grid and Gaussian-simulation steps outlined for  $\hat{\mathbf{I}}_{\alpha}(\mathbf{x})$  and  $\hat{\mathbf{I}}_{\alpha, \text{WBATE}}$ . In practice, finite-sample regularization may be useful to ensure that  $\inf_{\mathbf{x} \in \mathcal{B}} \hat{\Omega}_{\mathbf{x}, \mathbf{x}}$  is bounded away from zero. See Section SA-9 in the

supplemental appendix for more discussion, and the companion software package `rd2d` and article [20] for implementation details.

## 6 Imperfect Compliance

We extend the results to settings with imperfect treatment compliance, where treatment assignment and treatment status may not be equal for some units; see [38] for an overview in causal inference. We only give a brief discussion to conserve space, but all omitted details can be found in the supplemental appendix (Section SA-6).

To formalize the fuzzy BD design setup, we need to expand the potential outcomes notation. Let  $W_i = \mathbf{1}(\mathbf{X}_i \in \mathcal{A}_0) \cdot W_i(0) + \mathbf{1}(\mathbf{X}_i \in \mathcal{A}_1) \cdot W_i(1)$  be the observed treatment status, where  $W_i(t)$  denotes the potential treatment status under treatment assignment  $t \in \{0, 1\}$  for each unit. In addition, the observed outcome is now  $Y_i = \mathbf{1}(\mathbf{X}_i \in \mathcal{A}_0) \cdot Y_i(0, W_i(0)) + \mathbf{1}(\mathbf{X}_i \in \mathcal{A}_1) \cdot Y_i(1, W_i(1))$ , where the potential outcomes are now functions of two arguments:  $Y_i(t, w)$  denotes the potential outcome for unit  $i$  when this unit is assigned to treatment  $t \in \{0, 1\}$  and takes treatment status  $w \in \{0, 1\}$ .

To recycle the notation and assumptions from the sharp BD design setup, we redefine  $Y_i(t) = Y_i(t, W_i(t))$  for each  $t \in \{0, 1\}$ . In addition, we impose the following assumption on each reduced-form potential outcome/status pair.

**Assumption 4** (Fuzzy DGP). *Let  $t \in \{0, 1\}$ .*

- (i)  $(Y_1(t), W_1(t), \mathbf{X}_1^\top)^\top, \dots, (Y_n(t), W_n(t), \mathbf{X}_n^\top)^\top$  are independent and identically distributed random vectors.
- (ii)  $\mu_{W,t}(\mathbf{x}) = \mathbb{E}[W_i(t) | \mathbf{X}_i = \mathbf{x}]$  has a  $(p+1)$ -times continuously differentiable extension to an open neighborhood of  $\mathcal{X}$ .
- (iii)  $\sigma_{W,t}^2(\mathbf{x}) = \mathbb{V}[W_i(t) | \mathbf{X}_i = \mathbf{x}]$  is continuous on  $\mathcal{X}$ .
- (iv)  $\boldsymbol{\Upsilon}_t(\mathbf{x}) = \text{Cov}[(Y_i(t), W_i(t))^\top | \mathbf{X}_i = \mathbf{x}]$  has entries that are continuous on  $\mathcal{X}$ .

The usual fuzzy estimators and estimand are, respectively,

$$\widehat{\zeta}(\mathbf{x}) = \frac{\widehat{\tau}_Y(\mathbf{x})}{\widehat{\tau}_W(\mathbf{x})} \quad \text{and} \quad \zeta(\mathbf{x}) = \frac{\tau_Y(\mathbf{x})}{\tau_W(\mathbf{x})},$$

for each  $\mathbf{x} \in \mathcal{B}$ , where  $\widehat{\tau}_Y(\mathbf{x}) = \mathbf{e}_0^\top \widehat{\boldsymbol{\beta}}_{Y,1}(\mathbf{x}) - \mathbf{e}_0^\top \widehat{\boldsymbol{\beta}}_{Y,0}(\mathbf{x})$  and  $\widehat{\tau}_W(\mathbf{x}) = \mathbf{e}_0^\top \widehat{\boldsymbol{\beta}}_{W,1}(\mathbf{x}) - \mathbf{e}_0^\top \widehat{\boldsymbol{\beta}}_{W,0}(\mathbf{x})$ , with  $\widehat{\boldsymbol{\beta}}_{A,t}(\mathbf{x})$  denoting the local polynomial fit (1) using the outcome variable  $A \in \{Y, W\}$  and side  $t \in \{0, 1\}$ , and  $\tau_Y(\mathbf{x}) = \mathbb{E}[Y_i(1, W_i(1)) - Y_i(0, W_i(0)) | \mathbf{X}_i = \mathbf{x}]$  and  $\tau_W(\mathbf{x}) = \mathbb{E}[W_i(1) - W_i(0) | \mathbf{X}_i = \mathbf{x}]$ .

The functional (causal) parameters  $\tau_Y(\mathbf{x})$  and  $\tau_W(\mathbf{x})$  correspond to the intention-to-treat effect and the first-stage effect, at the boundary point  $\mathbf{x}$ , respectively. The ratio  $\zeta(\mathbf{x})$  is the fuzzy BATEC parameter, which can be interpreted as a complier average treatment effect at each boundary point  $\mathbf{x}$  under additional assumptions; see [29], [51], [40], [13, 15] for more discussion of causal identification in fuzzy multidimensional RD designs. See also [2] and references therein for an overview of causal identification in univariate fuzzy RD designs.

Fuzzy analogues of WBATE and LBATE can be defined by replacing  $\tau(\mathbf{x})$  with  $\zeta(\mathbf{x})$  in the definitions of  $\tau_{\text{WBATE}}$  and  $\tau_{\text{LBATE}}$ , respectively. More precisely, after normalizing the weights so that  $\int_{\mathcal{B}} w(\mathbf{x}) d\mathbf{x} = 1$ , the fuzzy WBATE is  $\zeta_{\text{WBATE}} = \int_{\mathcal{B}} \zeta(\mathbf{x}) w(\mathbf{x}) d\mathbf{x}$ , and the fuzzy LBATE is  $\zeta_{\text{LBATE}} = \sup_{\mathbf{x} \in \mathcal{B}} \zeta(\mathbf{x})$ . Plug-in estimators of these parameters are  $\widehat{\zeta}_{\text{WBATE}} = \int_{\mathcal{B}} \widehat{\zeta}(\mathbf{x}) w(\mathbf{x}) d\mathbf{x}$  and  $\widehat{\zeta}_{\text{LBATE}} = \sup_{\mathbf{x} \in \mathcal{B}} \widehat{\zeta}(\mathbf{x})$ , respectively.

The properties of the fuzzy estimators can be established using the results for the sharp BD design, along with “continuous mapping” and “probability concentration” arguments. More precisely, under Assumptions 1, 2 and 4, and bandwidth sequences that make the sharp convergence rates vanish, our results imply that  $\widehat{\tau}_Y(\mathbf{x}) = \tau_Y(\mathbf{x}) + o_{\mathbb{P}}(1)$  and  $\widehat{\tau}_W(\mathbf{x}) = \tau_W(\mathbf{x}) + o_{\mathbb{P}}(1)$ , pointwise and uniformly over  $\mathbf{x} \in \mathcal{B}$ . For uniform consistency, it suffices that  $n^{(1+v)/(2+v)} h^2 / \log(1/h) \rightarrow \infty$  and  $h \rightarrow 0$ . Therefore, under the additional regularity condition  $\inf_{\mathbf{x} \in \mathcal{B}} |\tau_W(\mathbf{x})| > 0$ , and using the exact second-order “linearization”  $\widehat{\zeta}(\mathbf{x}) - \zeta(\mathbf{x}) = \mathbf{w}(\mathbf{x})^\top \boldsymbol{\mathfrak{T}}(\mathbf{x}) + \mathfrak{R}_{\mathbb{F}}(\mathbf{x})$  with

$$\mathbf{w}(\mathbf{x}) = \left[ \frac{1}{\tau_W(\mathbf{x})}, -\frac{\tau_Y(\mathbf{x})}{\tau_W(\mathbf{x})^2} \right]^\top \quad \text{and} \quad \boldsymbol{\mathfrak{T}}(\mathbf{x}) = \left[ \widehat{\tau}_Y(\mathbf{x}) - \tau_Y(\mathbf{x}), \widehat{\tau}_W(\mathbf{x}) - \tau_W(\mathbf{x}) \right]^\top,$$

and  $\sup_{\mathbf{x} \in \mathcal{B}} |\mathfrak{R}_F(\mathbf{x})| \lesssim_{\mathbb{P}} \sup_{\mathbf{x} \in \mathcal{B}} |\widehat{\tau}_W(\mathbf{x}) - \tau_W(\mathbf{x})|^2 + \sup_{\mathbf{x} \in \mathcal{B}} |\widehat{\tau}_Y(\mathbf{x}) - \tau_Y(\mathbf{x})| \sup_{\mathbf{x} \in \mathcal{B}} |\widehat{\tau}_W(\mathbf{x}) - \tau_W(\mathbf{x})|$ , which is negligible when the sharp convergence rates vanish. See the supplemental appendix (Section SA-6.1) for details.

Therefore, estimation and inference results for fuzzy BATEC, fuzzy WBATE, and fuzzy LBATE can be established using the linearization argument, and then applying the theoretical results in the supplemental appendix to the specific linear combination  $\mathfrak{w}(\mathbf{x})^\top \mathfrak{T}(\mathbf{x})$ . In the remainder of this section we briefly overview the main estimation and inference results for fuzzy BD designs. See the supplemental appendix (Section SA-6) for more results and omitted details.

## 6.1 Fuzzy BATEC

Pointwise and uniform convergence rates, AMSE/AIMSE expansions, and bandwidth selectors for the fuzzy BATEC estimator  $\widehat{\zeta}(\mathbf{x})$  can be established using the linearization argument outlined above, along with the results for the sharp BD design. To state the leading terms, let  $B_{A,\mathbf{x}} = B_{A,1,\mathbf{x}} - B_{A,0,\mathbf{x}}$  denote the leading bias constant from Section 3.1, with the outcome variable  $A \in \{Y, W\}$  in place of  $Y$ , and define  $B_{F,\mathbf{x}} = \mathfrak{w}(\mathbf{x})^\top (B_{Y,\mathbf{x}}, B_{W,\mathbf{x}})^\top$ . Similarly, let  $V_{A_1, A_2, \mathbf{x}}$  denote the leading covariance constant for the reduced-form local polynomial estimators using outcomes  $A_1, A_2 \in \{Y, W\}$ , and define

$$V_{F,\mathbf{x}} = \mathfrak{w}(\mathbf{x})^\top \begin{pmatrix} V_{Y,Y,\mathbf{x}} & V_{Y,W,\mathbf{x}} \\ V_{W,Y,\mathbf{x}} & V_{W,W,\mathbf{x}} \end{pmatrix} \mathfrak{w}(\mathbf{x}).$$

For inference below, we impose  $\inf_{\mathbf{x} \in \mathcal{B}} V_{F,\mathbf{x}} > 0$  directly. This is the scalar nondegeneracy condition for the linearized fuzzy Wald score and allows one-sided compliance because it does not require the treatment-status variances on both sides to be bounded away from zero.

**Theorem 7** (Convergence Rates: Fuzzy BATEC). *Suppose Assumptions 1, 2, and 4 hold, and  $\inf_{\mathbf{x} \in \mathcal{B}} |\tau_W(\mathbf{x})| > 0$ . If  $n^{(1+v)/(2+v)} h^2 / \log(1/h) \rightarrow \infty$  and  $h \rightarrow 0$ , then*

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

Using the linearization argument, we focus on the approximate MSE and approximate IMSE, respectively,  $\text{AMSE}_{\mathbb{F}}(\mathbf{x}) = \mathbb{E}[(\mathbf{w}(\mathbf{x})^\top \mathfrak{T}(\mathbf{x}))^2 | \mathbf{X}]$  and  $\text{AIMSE}_{\mathbb{F}} = \int_{\mathcal{B}} \text{AMSE}_{\mathbb{F}}(\mathbf{x}) w(\mathbf{x}) d\mathbf{x}$ .

**Theorem 8** (Approximate MSE Expansions: Fuzzy BATEC). *Suppose Assumptions 1, 2, 3, and 4 hold, and  $\inf_{\mathbf{x} \in \mathcal{B}} |\tau_W(\mathbf{x})| > 0$ . If  $n^{v/(2+v)} h^2 / \log(1/h) \rightarrow \infty$  and  $h \rightarrow 0$ , then*

- (i)  $\text{AMSE}_{\mathbb{F}}(\mathbf{x}) = h^{2p+2} B_{\mathbb{F},\mathbf{x}}^2 + \frac{1}{nh^2} V_{\mathbb{F},\mathbf{x}} + o_{\mathbb{P}}(\mathfrak{R})$ , and
- (ii)  $\text{AIMSE}_{\mathbb{F}} = h^{2p+2} \int_{\mathcal{B}} B_{\mathbb{F},\mathbf{x}}^2 w(\mathbf{x}) d\mathbf{x} + \frac{1}{nh^2} \int_{\mathcal{B}} V_{\mathbb{F},\mathbf{x}} w(\mathbf{x}) d\mathbf{x} + o_{\mathbb{P}}(\mathfrak{R})$ ,

with  $\mathfrak{R} = h^{2p+2} + n^{-1}h^{-2}$ .

Ignoring the higher-order terms, the approximate MSE-optimal and IMSE-optimal bandwidth selectors are obtained from the formulas in Section 3.1 by replacing  $B_{\mathbf{x}}, V_{\mathbf{x}}$  with  $B_{\mathbb{F},\mathbf{x}}, V_{\mathbb{F},\mathbf{x}}$ . For inference, we consider the usual non-Donsker t-statistic process, leading to the confidence intervals and bands

$$\widehat{\Gamma}_{\mathbb{F},\alpha}(\mathbf{x}) = \left[ \widehat{\zeta}(\mathbf{x}) - \varrho_\alpha \widehat{\Omega}_{\mathbb{F},\mathbf{x},\mathbf{x}}^{1/2}, \widehat{\zeta}(\mathbf{x}) + \varrho_\alpha \widehat{\Omega}_{\mathbb{F},\mathbf{x},\mathbf{x}}^{1/2} \right],$$

where  $\varrho_\alpha$  denotes an appropriate quantile depending on the desired confidence level  $\alpha \in (0, 1)$  and coverage objective (pointwise for  $\mathbf{x} \in \mathcal{B}$  versus uniform over  $\mathcal{B}$ ), and the variance estimator is derived from the above stochastic linearization using the feasible weights

$$\widehat{\mathbf{w}}(\mathbf{x}) = \left[ \frac{1}{\widehat{\tau}_W(\mathbf{x})}, -\frac{\widehat{\tau}_Y(\mathbf{x})}{\widehat{\tau}_W(\mathbf{x})^2} \right]^\top.$$

Specifically,

$$\widehat{\Omega}_{\mathbb{F},\mathbf{x}_1,\mathbf{x}_2} = \widehat{\mathbf{w}}(\mathbf{x}_1)^\top \begin{pmatrix} \widehat{\Omega}_{Y,Y,\mathbf{x}_1,\mathbf{x}_2} & \widehat{\Omega}_{Y,W,\mathbf{x}_1,\mathbf{x}_2} \\ \widehat{\Omega}_{W,Y,\mathbf{x}_1,\mathbf{x}_2} & \widehat{\Omega}_{W,W,\mathbf{x}_1,\mathbf{x}_2} \end{pmatrix} \widehat{\mathbf{w}}(\mathbf{x}_2), \quad \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{B},$$

where  $\widehat{\Omega}_{Y,W,\mathbf{x}_1,\mathbf{x}_2} = \frac{1}{nh^2} \mathbf{e}_0^\top \left[ \widehat{\Gamma}_{0,\mathbf{x}_1}^{-1} \widehat{\Sigma}_{Y,W,0,\mathbf{x}_1,\mathbf{x}_2} \widehat{\Gamma}_{0,\mathbf{x}_2}^{-1} + \widehat{\Gamma}_{1,\mathbf{x}_1}^{-1} \widehat{\Sigma}_{Y,W,1,\mathbf{x}_1,\mathbf{x}_2} \widehat{\Gamma}_{1,\mathbf{x}_2}^{-1} \right] \mathbf{e}_0$  with  $\widehat{\Sigma}_{Y,W,t,\mathbf{x}_1,\mathbf{x}_2} = h^2 \mathbb{E}_n \left[ \mathbf{r}_p \left( \frac{\mathbf{X}_i - \mathbf{x}_1}{h} \right) \mathbf{r}_p \left( \frac{\mathbf{X}_i - \mathbf{x}_2}{h} \right)^\top K_h(\mathbf{X}_i - \mathbf{x}_1) K_h(\mathbf{X}_i - \mathbf{x}_2) \widehat{\varepsilon}_{Y,i}(\mathbf{x}_1) \widehat{\varepsilon}_{W,i}(\mathbf{x}_2) \mathbf{1}(\mathbf{X}_i \in \mathcal{A}_t) \right]$  and  $\widehat{\varepsilon}_{A,i}(\mathbf{x}) = A_i - \mathbf{r}_p(\mathbf{X}_i - \mathbf{x})^\top [\mathbf{1}(\mathbf{X}_i \in \mathcal{A}_0) \widehat{\beta}_{A,0}(\mathbf{x}) + \mathbf{1}(\mathbf{X}_i \in \mathcal{A}_1) \widehat{\beta}_{A,1}(\mathbf{x})]$ , with  $A \in \{Y, W\}$ . The population covariance  $\Omega_{\mathbb{F},\mathbf{x}_1,\mathbf{x}_2}$  is defined analogously, replacing feasible weights and covariance

components by their population counterparts.

The following result establishes the validity of these confidence intervals and bands. Let  $\mathbf{V} = ((Y_1, W_1, \mathbf{X}_1^\top)^\top, \dots, (Y_n, W_n, \mathbf{X}_n^\top)^\top)$  be the observed data.

**Theorem 9** (Confidence Intervals and Bands: Fuzzy BATEC). *Suppose Assumptions 1, 2 and 4 hold,  $\inf_{\mathbf{x} \in \mathcal{B}} |\tau_W(\mathbf{x})| > 0$ , and  $\inf_{\mathbf{x} \in \mathcal{B}} V_{F,\mathbf{x}} > 0$ .*

- (i) *If  $nh^2 \rightarrow \infty$  and  $nh^{2p+4} \rightarrow 0$ , then  $\mathbb{P}[\zeta(\mathbf{x}) \in \widehat{I}_{F,\alpha}(\mathbf{x})] \rightarrow 1 - \alpha$  with  $\mathcal{q}_\alpha = \Phi^{-1}(1 - \alpha/2)$ .*
- (ii) *If  $\liminf_{n \rightarrow \infty} \log h / \log n > -\infty$ ,  $n^{v/(2+v)} h^2 / (\log n)^9 \rightarrow \infty$ , and  $nh^{2p+4} \log n \rightarrow 0$ , then  $\mathbb{P}[\zeta(\mathbf{x}) \in \widehat{I}_{F,\alpha}(\mathbf{x}), \text{ for all } \mathbf{x} \in \mathcal{B}] \rightarrow 1 - \alpha$  with  $\mathcal{q}_\alpha = \inf\{c > 0 : \mathbb{P}[\sup_{\mathbf{x} \in \mathcal{B}} |\widehat{Z}_F(\mathbf{x})| \geq c|\mathbf{V}] \leq \alpha\}$ , where  $(\widehat{Z}_F(\mathbf{x}) : \mathbf{x} \in \mathcal{B})$  is a Gaussian process conditional on  $\mathbf{V}$  with  $\mathbb{E}[\widehat{Z}_F(\mathbf{x}_1)|\mathbf{V}] = 0$  and  $\text{Cov}[\widehat{Z}_F(\mathbf{x}_1), \widehat{Z}_F(\mathbf{x}_2)|\mathbf{V}] = \widehat{\Omega}_{F,\mathbf{x}_1,\mathbf{x}_2} / (\widehat{\Omega}_{F,\mathbf{x}_1,\mathbf{x}_1} \widehat{\Omega}_{F,\mathbf{x}_2,\mathbf{x}_2})^{1/2}$ , for all  $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{B}$ .*

See the supplemental appendix (Section SA-6.1) for more results.

## 6.2 Fuzzy WBATE

Without loss of generality, we set  $\int_{\mathcal{B}} w(\mathbf{x}) d\mathbf{x} = 1$ . The fuzzy WBATE and its plug-in estimator are

$$\check{\zeta}_{\text{WBATE}} = \int_{\mathcal{B}} \zeta(\mathbf{x}) w(\mathbf{x}) d\mathbf{x} \quad \text{and} \quad \widehat{\zeta}_{\text{WBATE}} = \int_{\mathcal{B}} \widehat{\zeta}(\mathbf{x}) w(\mathbf{x}) d\mathbf{x}.$$

Let  $B_{F,\text{WBATE}} = \int_{\mathcal{B}} B_{F,\mathbf{x}} w(\mathbf{x}) d\mathbf{x}$  and  $\Omega_{F,\text{WBATE}} = \int_{\mathcal{B}} \int_{\mathcal{B}} \Omega_{F,\mathbf{x}_1,\mathbf{x}_2} w(\mathbf{x}_1) w(\mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2$ , and define  $\widehat{\Omega}_{F,\text{WBATE}}$  analogously by replacing  $\Omega_{F,\mathbf{x}_1,\mathbf{x}_2}$  with  $\widehat{\Omega}_{F,\mathbf{x}_1,\mathbf{x}_2}$ . For the fuzzy WBATE AMSE and inference results, we impose the aggregate nondegeneracy condition  $\Omega_{F,\text{WBATE}} \gtrsim (nh)^{-1}$ , which rules out cancellation after applying the fuzzy Wald weights and integrating over  $\mathcal{B}$ . The approximate MSE is for the linearized aggregate

$$\check{\zeta}_{\text{WBATE}} = \int_{\mathcal{B}} \mathbf{w}(\mathbf{x})^\top \mathfrak{T}(\mathbf{x}) w(\mathbf{x}) d\mathbf{x}, \quad \text{AMSE}_{F,\text{WBATE}} = \mathbb{E}[\check{\zeta}_{\text{WBATE}}^2 | \mathbf{X}].$$

**Theorem 10** (Approximate MSE Expansion: Fuzzy WBATE). *Suppose Assumptions 1, 2, 3, and 4 hold,  $\inf_{\mathbf{x} \in \mathcal{B}} |\tau_W(\mathbf{x})| > 0$ ,  $\inf_{\mathbf{x} \in \mathcal{B}} V_{F,\mathbf{x}} > 0$ , and  $\Omega_{F,\text{WBATE}} \gtrsim (nh)^{-1}$ . If  $n^{v/(2+v)}h^2/\log(1/h) \rightarrow \infty$  and  $h \rightarrow 0$ , then  $\text{AMSE}_{F,\text{WBATE}} = \Omega_{F,\text{WBATE}} + h^{2p+2}B_{F,\text{WBATE}}^2 + o_{\mathbb{P}}(\mathfrak{R})$ , where  $(nh)^{-1} \lesssim \Omega_{F,\text{WBATE}} \lesssim (nh)^{-1}$  and  $\mathfrak{R} = (nh)^{-1} + h^{2p+2}$ .*

For inference, define  $\widehat{\mathbb{T}}_{F,\text{WBATE}} = (\widehat{\zeta}_{\text{WBATE}} - \zeta_{\text{WBATE}})/\widehat{\Omega}_{F,\text{WBATE}}^{1/2}$  with associated confidence interval

$$\widehat{\mathbb{I}}_{F,\alpha,\text{WBATE}} = \left[ \widehat{\zeta}_{\text{WBATE}} - \varrho_{\alpha} \widehat{\Omega}_{F,\text{WBATE}}^{1/2}, \widehat{\zeta}_{\text{WBATE}} + \varrho_{\alpha} \widehat{\Omega}_{F,\text{WBATE}}^{1/2} \right],$$

where  $\varrho_{\alpha} = \Phi^{-1}(1 - \alpha/2)$ .

**Theorem 11** (Confidence Intervals: Fuzzy WBATE). *Suppose Assumptions 1, 2, 3, and 4 hold,  $\inf_{\mathbf{x} \in \mathcal{B}} |\tau_W(\mathbf{x})| > 0$ ,  $\inf_{\mathbf{x} \in \mathcal{B}} V_{F,\mathbf{x}} > 0$ , and  $\Omega_{F,\text{WBATE}} \gtrsim (nh)^{-1}$ . If  $n^{v/(2+v)}h^2/\log(1/h) \rightarrow \infty$ ,  $nh^{2p+3} \rightarrow 0$ ,  $\sqrt{nh}h^{3/2}/\log(1/h) \rightarrow \infty$ , and  $n^{2v/(2+v)-1/2}h^{7/2}/\log^2(1/h) \rightarrow \infty$ , then  $\mathbb{P}[\zeta_{\text{WBATE}} \in \widehat{\mathbb{I}}_{F,\alpha,\text{WBATE}}] \rightarrow 1 - \alpha$ .*

See the supplemental appendix (Section SA-6.2) for more results.

### 6.3 Fuzzy LBATE

The fuzzy LBATE and its plug-in estimator are

$$\zeta_{\text{LBATE}} = \sup_{\mathbf{x} \in \mathcal{B}} \zeta(\mathbf{x}) \quad \text{and} \quad \widehat{\zeta}_{\text{LBATE}} = \sup_{\mathbf{x} \in \mathcal{B}} \widehat{\zeta}(\mathbf{x}).$$

Theorem 7 gives consistency and convergence rates for  $\widehat{\zeta}_{\text{LBATE}}$  by taking suprema over  $\mathcal{B}$ . For inference, use the Gaussian process from Theorem 9 and define

$$\widehat{\mathbb{I}}_{F,\alpha,\text{LBATE}} = \left[ \sup_{\mathbf{x} \in \mathcal{B}} \left( \widehat{\zeta}(\mathbf{x}) - \varrho_{\alpha} \widehat{\Omega}_{F,\mathbf{x},\mathbf{x}}^{1/2} \right), \sup_{\mathbf{x} \in \mathcal{B}} \left( \widehat{\zeta}(\mathbf{x}) + \varrho_{\alpha} \widehat{\Omega}_{F,\mathbf{x},\mathbf{x}}^{1/2} \right) \right],$$

where  $\varrho_{\alpha} = \inf\{c > 0 : \mathbb{P}[\sup_{\mathbf{x} \in \mathcal{B}} |\widehat{Z}_F(\mathbf{x})| \geq c | \mathbf{V}] \leq \alpha\}$ .

**Theorem 12** (Confidence Intervals: Fuzzy LBATE). *Suppose Assumptions 1, 2 and 4 hold,  $\inf_{\mathbf{x} \in \mathcal{B}} |\tau_W(\mathbf{x})| > 0$ , and  $\inf_{\mathbf{x} \in \mathcal{B}} V_{F,\mathbf{x}} > 0$ . If  $\liminf_{n \rightarrow \infty} \log h / \log n > -\infty$ ,  $n^{v/(2+v)}h^2/(\log n)^9 \rightarrow$*

$\infty$ , and  $nh^{2p+4} \log n \rightarrow 0$ , then  $\mathbb{P}[\hat{\zeta}_{\text{LBATE}} \in \hat{\mathbb{I}}_{\text{F},\alpha,\text{LBATE}}] \geq 1 - \alpha + o(1)$ .

See the supplemental appendix (Section SA-6.3) for additional results.

## 7 The Causal Effects of SPP on College Attendance

We illustrate our proposed causal inference methodology for BD designs with the SPP application introduced in Section 1. Recall that the dataset has  $n = 363,096$  complete observations for the first cohort of the program (2014), where each observation corresponds to one student, and the bivariate score is  $\mathbf{X}_i = (X_{1i}, X_{2i})^\top = (\text{SABER11}_i, \text{SISBEN}_i)^\top$ , where the first dimension is the student’s SABER11 test score (ranging from  $-310$  to  $172$ ) and the second dimension is the student’s SISBEN wealth index (ranging from  $-103.41$  to  $127.21$ ). Without loss of generality, each dimension of the score is recentered at its corresponding cutoff for program eligibility, so that the treatment assignment boundary is as shown in Figure 1a. All the results in this section are implemented using our companion R software package `rd2d`, and omitted details are given in the supplemental appendix (Sections SA-9 and SA-10), replication files, and companion software article [20].

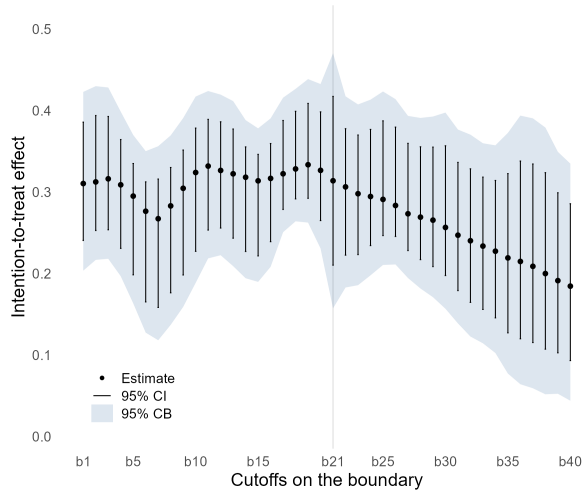
The outcome variable of interest is *college enrollment*, with  $Y_i = 1$  if the student enrolled in college and  $Y_i = 0$  otherwise. In this application, 15,423 students were eligible for the SPP program, but only 9,165 took up the benefit. We therefore report both sharp reduced-form effects of eligibility assignment and fuzzy effects that adjust for imperfect compliance. Specifically, we estimate the reduced-form BATEC  $\tau_Y(\mathbf{x})$ , the first-stage BATEC  $\tau_W(\mathbf{x})$ , and the fuzzy BATEC  $\zeta(\mathbf{x}) = \tau_Y(\mathbf{x})/\tau_W(\mathbf{x})$  at the 40 evenly spaced grid points  $\mathbf{b}_j \in \mathcal{B}$ ,  $j = 1, \dots, 40$ , depicted in Figure 1a. We also report the corresponding WBATE and LBATE parameters. The empirical results use the ITT bandwidth selector; in the supplemental appendix, we also report results obtained with fuzzy bandwidths. All local polynomial estimators use data-driven MSE-optimal bandwidths as discussed in Section 3.3; the two score coordinates are standardized for bandwidth selection and estimation, while the bandwidths reported in the tables are transformed back to the original SABER11 and SISBEN scales. Uncertainty quantification uses robust bias-corrected inference for pointwise confidence intervals and uniform

confidence bands.

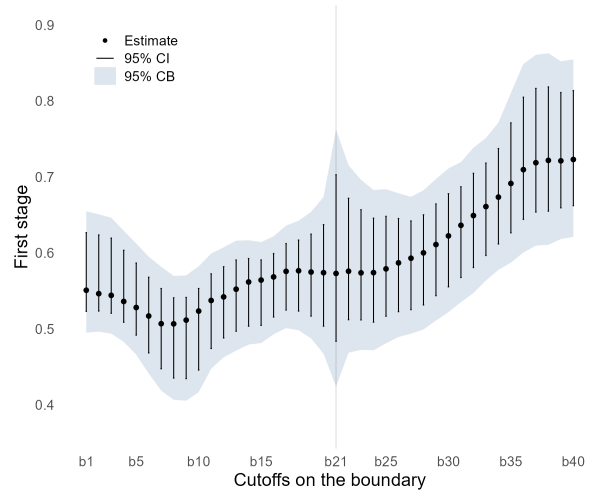
Figure 2 summarizes the main results for the BATEC graphically. Panel (a) reports reduced-form intention-to-treat effects of eligibility on college enrollment, Panel (b) reports the first-stage effects on SPP take-up, Panel (c) reports the fuzzy effects on college enrollment, and Panel (d) reports covariate balance using mother’s education as the outcome. Mother’s education is predetermined, so a credible BD design should deliver treatment effects on this covariate that are statistically indistinguishable from zero; see [14, Section 5] for related discussion of falsification tests in RD designs. The supplemental appendix reports additional heat maps for point estimates and robust bias-corrected  $p$ -values along the assignment boundary, as well as the complete numerical tables for all 40 boundary grid points.

The intention-to-treat effects in Panel (a) are statistically significant along the boundary. The pattern indicates roughly homogeneous effects along the portion of the boundary where students are marginal in terms of academic performance but satisfy the minimal poverty requirement ( $\mathbf{b}_1$  through approximately  $\mathbf{b}_{21}$ ), and more pronounced heterogeneity along the portion where students are marginal in terms of wealth ( $\mathbf{b}_{21}$  through  $\mathbf{b}_{40}$ ). In particular, the estimated reduced-form effect on college enrollment remains close to 30 percentage points along the first portion of the boundary and declines for the wealthiest eligible students as their SABER11 scores increase. This pattern is consistent with the program having larger enrollment effects among students for whom the subsidy is more likely to relax a binding financial constraint. The first-stage estimates in Panel (b) are strongly positive along the boundary, confirming substantial but imperfect take-up of the subsidy. Consequently, the fuzzy estimates in Panel (c) are larger than the corresponding intention-to-treat effects. The placebo estimates in Panel (d) are centered near zero and statistically insignificant, which supports the continuity-based identifying assumptions in this application.

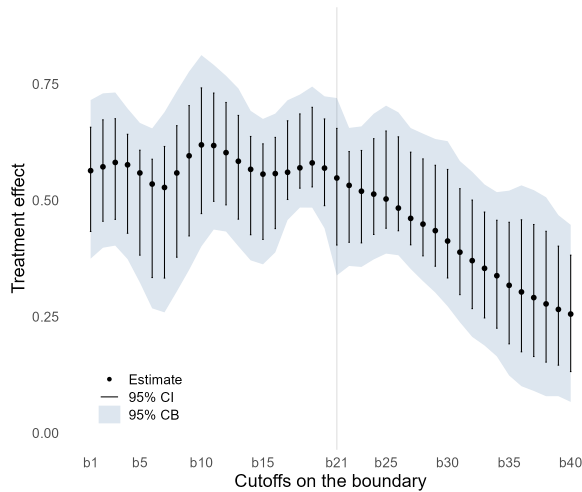
The magnitudes are also large relative to the control-side baseline enrollment rates. Section SA-10 of the supplemental appendix reports the estimates of  $\mu_0(\mathbf{x})$  and the corresponding ratios  $\hat{\tau}_Y(\mathbf{x})/\hat{\mu}_0(\mathbf{x})$  and  $\hat{\zeta}(\mathbf{x})/\hat{\mu}_0(\mathbf{x})$ . The reduced-form WBATE estimate is about 63 percent of the corresponding control-side baseline, while the fuzzy WBATE estimate is about 109



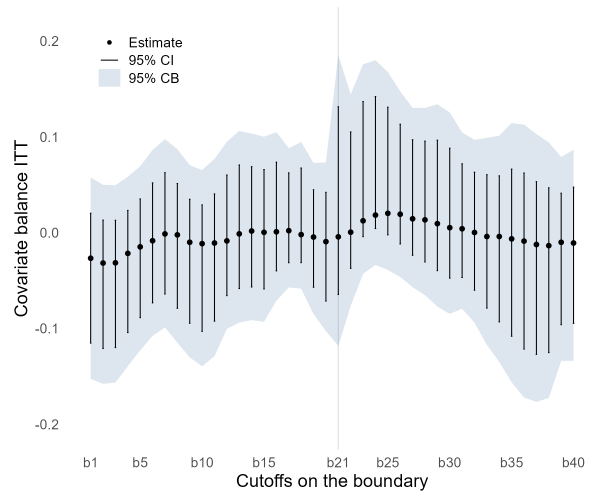
(a) Intention-to-Treat Effects ( $\tau_Y(\mathbf{x})$ ).



(b) First-Stage Effects ( $\tau_W(\mathbf{x})$ ).



(c) Fuzzy Effects ( $\zeta(\mathbf{x})$ ).



(d) Intention-to-Treat Covariate Balance.

Figure 2: Empirical Results for BATEC (SPP Application).

*Notes.* Panel (a) reports reduced-form effects of eligibility on college enrollment. Panel (b) reports first-stage effects on SPP take-up. Panel (c) reports fuzzy effects on college enrollment. Panel (d) reports a covariate analysis using mother's education as the outcome. Point estimates, robust bias-corrected confidence intervals (CI), and robust bias-corrected confidence bands (CB) using the 40 boundary grid points shown in Figure 1a. Bandwidths are MSE-optimal, selected separately at each grid point, and based on the ITT bandwidth selector.

percent of that baseline. Across the full boundary grid, the reduced-form effects range from about 28 to 95 percent of the local baseline enrollment rate, and the fuzzy effects range from about 38 to 182 percent. The results in Table SA-5 show that, in the piece of the boundary where students are near the academic threshold (points  $\mathbf{b}_1$ - $\mathbf{b}_{20}$ ) the control mean is roughly constant: between 37 and 40 percent of control students enroll in college. In contrast, in the region of the boundary where students are near the wealth threshold (points  $\mathbf{b}_{21}$ - $\mathbf{b}_{40}$ ), the share of control students who attend college increases steadily as SABER 11 increases; the treatment effect is there reduced in this region of the boundary because more students are already enrolling in college in the absence of the program. Near the marginal level of wealth, students who are highly achieving academically are therefore less constrained by wealth than their less academically achieving peers.

Table 1 presents a summary of the numerical results underlying Figure 2. Due to space limitations, each panel reports selected boundary points,  $\mathbf{b}_j$  with  $j = 1, 5, 10, 15, 20, 25, 30, 35, 40$ , together with the corresponding WBATE and LBATE estimates. The WBATE estimator uses equal weights over the 40 boundary grid points. The table also reports MSE-optimal bandwidths in the original score units, point estimates, robust bias-corrected test statistics,  $p$ -values, and confidence intervals. See Section SA-10 of the supplemental appendix for omitted details.

## 8 Extensions and Future Research

This section discusses several generalizations of our results that are useful for applied researchers. The theoretical results in the supplemental appendix already accommodate some of these extensions, while in other cases they serve as a foundation for future theoretical and methodological research.

### 8.1 Multidimensional Scores and General Assignment Boundaries

We considered bivariate scores and piecewise linear assignment boundaries with finitely many segments, a setting that covers many empirical BD designs while keeping the assumptions

(a) Intention-to-Treat Effects:  $\tau_Y(\mathbf{x})$ .

	$h_1$	$h_2$	$N_{Co}$	$N_{Tr}$	Estimate	$p$ -value	95% CI
$\mathbf{b}_1$	27.6	12.2	8921	3532	0.311	0.000	(0.204, 0.423)
$\mathbf{b}_5$	23.1	10.2	9044	3997	0.295	0.000	(0.164, 0.370)
$\mathbf{b}_{10}$	19.3	8.6	6925	3563	0.324	0.000	(0.189, 0.417)
$\mathbf{b}_{15}$	22.2	9.8	10063	4977	0.314	0.000	(0.190, 0.378)
$\mathbf{b}_{20}$	19.4	8.6	8945	2680	0.327	0.000	(0.231, 0.433)
$\mathbf{b}_{25}$	21.0	9.3	6188	3114	0.291	0.000	(0.211, 0.423)
$\mathbf{b}_{30}$	26.3	11.7	4698	4560	0.257	0.000	(0.157, 0.398)
$\mathbf{b}_{35}$	28.1	12.5	2875	2950	0.219	0.000	(0.077, 0.373)
$\mathbf{b}_{40}$	36.3	16.1	2916	2993	0.185	0.000	(0.044, 0.335)
WBATE					0.282	0.000	(0.248, 0.316)
LBATE					0.334		(0.263, 0.472)

(b) First-Stage Effects:  $\tau_W(\mathbf{x})$ .

	$h_1$	$h_2$	$N_{Co}$	$N_{Tr}$	Estimate	$p$ -value	95% CI
$\mathbf{b}_1$	27.6	12.2	8921	3532	0.551	0.000	(0.495, 0.655)
$\mathbf{b}_5$	23.1	10.2	9044	3997	0.528	0.000	(0.466, 0.612)
$\mathbf{b}_{10}$	19.3	8.6	6925	3563	0.524	0.000	(0.417, 0.582)
$\mathbf{b}_{15}$	22.2	9.8	10063	4977	0.564	0.000	(0.481, 0.614)
$\mathbf{b}_{20}$	19.4	8.6	8945	2680	0.574	0.000	(0.467, 0.674)
$\mathbf{b}_{25}$	21.0	9.3	6188	3114	0.579	0.000	(0.481, 0.684)
$\mathbf{b}_{30}$	26.3	11.7	4698	4560	0.623	0.000	(0.522, 0.711)
$\mathbf{b}_{35}$	28.1	12.5	2875	2950	0.691	0.000	(0.587, 0.811)
$\mathbf{b}_{40}$	36.3	16.1	2916	2993	0.723	0.000	(0.621, 0.855)
WBATE					0.592	0.000	(0.567, 0.620)
LBATE					0.723		(0.624, 0.860)

(c) Fuzzy Effects:  $\zeta(\mathbf{x})$ .

	$h_1$	$h_2$	$N_{Co}$	$N_{Tr}$	Estimate	$p$ -value	95% CI
$\mathbf{b}_1$	27.6	12.2	8921	3532	0.564	0.000	(0.374, 0.715)
$\mathbf{b}_5$	23.1	10.2	9044	3997	0.559	0.000	(0.323, 0.666)
$\mathbf{b}_{10}$	19.3	8.6	6925	3563	0.619	0.000	(0.401, 0.812)
$\mathbf{b}_{15}$	22.2	9.8	10063	4977	0.556	0.000	(0.362, 0.675)
$\mathbf{b}_{20}$	19.4	8.6	8945	2680	0.569	0.000	(0.440, 0.723)
$\mathbf{b}_{25}$	21.0	9.3	6188	3114	0.503	0.000	(0.385, 0.703)
$\mathbf{b}_{30}$	26.3	11.7	4698	4560	0.412	0.000	(0.272, 0.627)
$\mathbf{b}_{35}$	28.1	12.5	2875	2950	0.317	0.000	(0.123, 0.521)
$\mathbf{b}_{40}$	36.3	16.1	2916	2993	0.255	0.000	(0.067, 0.447)
WBATE					0.487	0.000	(0.435, 0.535)
LBATE					0.619		(0.487, 0.808)

(d) Intention-to-Treat Covariate Balance.

	$h_1$	$h_2$	$N_{Co}$	$N_{Tr}$	Estimate	$p$ -value	95% CI
$\mathbf{b}_1$	27.4	12.2	8750	3469	-0.027	0.170	(-0.153, 0.058)
$\mathbf{b}_5$	24.2	10.7	9767	4156	-0.015	0.398	(-0.123, 0.069)
$\mathbf{b}_{10}$	21.0	9.3	8085	3981	-0.011	0.272	(-0.139, 0.065)
$\mathbf{b}_{15}$	20.3	9.0	8051	4329	0.000	0.909	(-0.093, 0.100)
$\mathbf{b}_{20}$	20.2	8.9	9828	2825	-0.009	0.613	(-0.103, 0.073)
$\mathbf{b}_{25}$	18.7	8.3	4697	2598	0.020	0.058	(-0.039, 0.168)
$\mathbf{b}_{30}$	24.5	10.9	3842	3945	0.005	0.556	(-0.085, 0.126)
$\mathbf{b}_{35}$	23.1	10.3	1825	1842	-0.006	0.639	(-0.156, 0.114)
$\mathbf{b}_{40}$	27.8	12.3	1560	1511	-0.011	0.516	(-0.134, 0.087)
WBATE					-0.004	0.938	(-0.030, 0.028)
LBATE					0.020		(-0.031, 0.182)

Table 1: Selected Empirical Results for BATEC (SPP Application).

*Notes.* Panel (a) reports reduced-form effects of eligibility on college enrollment. Panel (b) reports take-up. Panel (c) reports fuzzy effects on college enrollment. Panel (d) reports a covariate balance using mother's education as the outcome. Point estimates and robust bias-corrected confidence intervals are computed using the 40 boundary grid points shown in Figure 1a. MSE-optimal bandwidths for each score coordinate are computed at each grid point using the ITT bandwidth selector. A subset of boundary grid points are reported for exposition. See the supplemental appendix for details and omitted results.

transparent. The supplemental appendix gives pointwise and uniform estimation and inference results when  $\mathbf{X}_i$  is a  $d$ -dimensional score and  $\mathcal{B}$  has effective dimension  $d - 1$ , with  $d \geq 2$ , allowing both diagonal bandwidth matrices and assignment boundaries that are not necessarily piecewise linear. The more general theory requires regularity conditions on the geometry of  $\mathcal{B}$ , formulated using ideas from geometric measure theory [33]. Practically, these conditions rule out boundary regions that are too irregular or too poorly supported by observations on one side for stable local polynomial estimation. Higher-dimensional scores are possible in principle, but the usual nonparametric curse of dimensionality remains; empirical work with  $d > 2$  should therefore include careful bandwidth, support, and sensitivity checks.

## 8.2 Derivatives and Regression Kink Designs

The supplemental appendix also studies pointwise and uniform estimation and inference for derivatives of the BATEC, transformations of those derivatives, and their fuzzy counterparts. Derivative estimands are useful when the object of interest is not only the level of the treatment effect at the boundary, but also how that effect changes along  $\mathcal{B}$ . For example, derivatives can be used to assess monotonicity or other local patterns of treatment effect heterogeneity. The same theory covers regression kink designs [10] in BD settings, where the identifying variation comes from a change in slope rather than a change in level at the assignment boundary.

## 8.3 Covariate Adjustment

In RD designs, pre-treatment covariates can be used as placebo outcomes or balance checks [14, Section 5], incorporated in the local polynomial fit to improve precision [6], or used to study treatment effect heterogeneity [9]. These uses carry over naturally to BD designs. The empirical application above reports a placebo analysis using mother’s education as the outcome variable, and the supplemental appendix gives a basic formulation for incorporating pre-treatment covariates into location-based estimators. To conserve space, we present the

sharp design formulation in the supplemental appendix (Section SA-3.3), while extensions to the fuzzy estimands in Section 6 follow the same logic.

## 8.4 Multidimensional RD Plots

Visualization is an important part of univariate RD analysis [4, 45], but a single univariate RD plot is generally not enough in BD designs because treatment effects may vary along  $\mathcal{B}$ . A useful starting point is to display the assignment scores, boundary, and estimation grid, as in Figure 1a, followed by plots of the estimated BATEC and associated uncertainty measures, as in Figure 2. More formal bivariate RD plots can be constructed by extending binscatter methods: univariate RD plots are a particular instance of binscatter methods [17, 21], and analogous BD plots can be based on bins in the assignment space or on binned summaries local to the boundary. Such plots can help assess outcome discontinuities, visualize heterogeneity, and compare the location-based analysis with simpler binned or distance-to-boundary summaries. We plan to develop these visualization tools in future work.

## 8.5 Clustered and Spatially Correlated Data

Many BD applications involve observations that are clustered or spatially correlated, especially when the score contains geographic coordinates. The formal results in this paper are developed under random sampling, and therefore they should be applied directly only when that approximation is credible. When dependence is present and its structure is known or can be consistently approximated, the location-based estimators can in principle be paired with cluster-robust or spatially robust covariance estimators [47]. Pointwise results for nonparametric local-constant regression under spatial dependence have been recently developed [52]. Extending those ideas to general local polynomial BD estimators, especially for uniform inference over  $\mathcal{B}$ , requires additional empirical process and covariance-estimation results. We therefore view dependence-robust pointwise and uniform inference for BD designs as an important direction for future work.

## References

- [1] Isaiah Andrews, Toru Kitagawa, and Adam McCloskey. Inference on winners. *Quarterly Journal of Economics*, 139(1):305–358, 2024.
- [2] Yoichi Arai, Yu-Chin Hsu, Toru Kitagawa, Ismael Mourifié, and Yuanyuan Wan. Testing identifying assumptions in fuzzy regression discontinuity designs. *Quantitative Economics*, 13(1):1–28, 2022.
- [3] Sebastian Calonico, Matias D. Cattaneo, and Rocio Titiunik. Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326, 2014.
- [4] Sebastian Calonico, Matias D. Cattaneo, and Rocio Titiunik. Optimal data-driven regression discontinuity plots. *Journal of the American Statistical Association*, 110(512):1753–1769, 2015.
- [5] Sebastian Calonico, Matias D. Cattaneo, and Max H. Farrell. On the effect of bias estimation on coverage accuracy in nonparametric inference. *Journal of the American Statistical Association*, 113(522):767–779, 2018.
- [6] Sebastian Calonico, Matias D. Cattaneo, Max H. Farrell, and Rocio Titiunik. Regression discontinuity designs using covariates. *Review of Economics and Statistics*, 101(3):442–451, 2019.
- [7] Sebastian Calonico, Matias D. Cattaneo, and Max H. Farrell. Optimal bandwidth choice for robust bias corrected inference in regression discontinuity designs. *Econometrics Journal*, 23(2):192–210, 2020.
- [8] Sebastian Calonico, Matias D. Cattaneo, and Max H. Farrell. Coverage error optimal confidence intervals for local polynomial regression. *Bernoulli*, 28(4):2998–3022, 2022.
- [9] Sebastian Calonico, Matias D. Cattaneo, Max H. Farrell, Filippo Palomba, and Rocio Titiunik. Treatment effect heterogeneity in regression discontinuity designs. *arXiv preprint arXiv:2503.13696*, 2026.
- [10] David Card, David S. Lee, Zhuan Pei, and Andrea Weber. Inference on causal effects in a generalized regression kink design. *Econometrica*, 83(6):2453–2483, 2015.
- [11] Matias D. Cattaneo and Rocio Titiunik. Regression discontinuity designs. *Annual Review of Economics*, 14:821–851, 2022.
- [12] Matias D. Cattaneo and Ruiqi (Rae) Yu. Strong approximations for empirical processes indexed by lipschitz functions. *Annals of Statistics*, 53(3):1203–1229, 2025.

- [13] Matias D. Cattaneo, Luke Keele, Rocio Titiunik, and Gonzalo Vazquez-Bare. Interpreting regression discontinuity designs with multiple cutoffs. *Journal of Politics*, 78(4):1229–1248, 2016.
- [14] Matias D. Cattaneo, Nicolás Idrobo, and Rocio Titiunik. *A Practical Introduction to Regression Discontinuity Designs: Foundations*. Cambridge University Press, 2020.
- [15] Matias D. Cattaneo, Luke Keele, Rocio Titiunik, and Gonzalo Vazquez-Bare. Extrapolating treatment effects in multi-cutoff regression discontinuity designs. *Journal of the American Statistical Association*, 116(536):1941–1952, 2021.
- [16] Matias D. Cattaneo, Rajita Chandak, Michael Jansson, and Xinwei Ma. Boundary adaptive local polynomial conditional density estimators. *Bernoulli*, 30(4):3193–3223, 2024.
- [17] Matias D. Cattaneo, Richard K. Crump, Max H. Farrell, and Yingjie Feng. On binscatter. *American Economic Review*, 114(5):1488–1514, 2024.
- [18] Matias D. Cattaneo, Yingjie Feng, and William G. Underwood. Uniform inference for kernel density estimators with dyadic data. *Journal of the American Statistical Association*, 119(524):2695–2708, 2024.
- [19] Matias D. Cattaneo, Nicolás Idrobo, and Rocio Titiunik. *A Practical Introduction to Regression Discontinuity Designs: Extensions*. Cambridge University Press, 2024.
- [20] Matias D. Cattaneo, Rocio Titiunik, and Ruiqi (Rae) Yu. `rd2d`: Causal inference in boundary discontinuity designs. *Working paper*, 2025.
- [21] Matias D. Cattaneo, Richard K. Crump, Max H. Farrell, and Yingjie Feng. Nonlinear binscatter methods. *arXiv:2407.15276*, 2026.
- [22] Matias D. Cattaneo, Rocio Titiunik, and Ruiqi (Rae) Yu. Estimation and inference in boundary discontinuity designs: Pooling-based methods. *Working paper*, 2026.
- [23] Matias D. Cattaneo, Rocio Titiunik, and Ruiqi (Rae) Yu. Boundary discontinuity designs: Theory and practice. In *Invited book chapter for the 2025 Econometric Society World Congress*, volume 1, chapter 2. Cambridge University Press, 2026.
- [24] Matias D. Cattaneo, Rocio Titiunik, and Ruiqi (Rae) Yu. Estimation and inference in boundary discontinuity designs: Distance-based methods. *Working paper*, 2026.
- [25] Xiaohong Chen and Wayne Yuan Gao. Thin sets are not equally thin: Minimax learning of submanifold integrals. *arXiv preprint arXiv:2507.12673*, 2026.

- [26] Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Anti-concentration and honest, adaptive confidence bands. *Annals of Statistics*, 42(5):1787–1818, 2014.
- [27] Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Gaussian approximation of suprema of empirical processes. *Annals of Statistics*, 42(4):1564–1597, 2014.
- [28] Victor Chernozhukov, Denis Chetverikov, Kengo Kato, and Yuta Koike. Improved central limit theorem and bootstrap approximations in high dimensions. *Annals of Statistics*, 50(5):2562–2586, 2022.
- [29] Jin-young Choi and Myoung-jae Lee. Complier and monotonicity for fuzzy multi-score regression discontinuity with partial effects. *Economics Letters*, 228:111169, 2023.
- [30] Leandro De Magalhães, Dominik Hangartner, Salomo Hirvonen, Jaakko Meriläinen, Nelson A Ruiz, and Janne Tukiainen. When can we trust regression discontinuity design estimates from close elections? evidence from experimental benchmarks. *Political Analysis*, 2025.
- [31] Juan D Diaz and Jose R Zubizarreta. Complex discontinuity designs using covariates for policy impact evaluation. *Annals of Applied Statistics*, 17(1):67–88, 2023.
- [32] Richard M Dudley. *Uniform central limit theorems*, volume 142. Cambridge university press, 2014.
- [33] Herbert Federer. *Geometric measure theory*. Springer, 2014.
- [34] Sebastian Galiani, Patrick J. McEwan, and Brian Quistorff. External and internal validity of a geographic quasi-experiment embedded in a cluster-randomized experiment. In Matias D. Cattaneo and Juan Carlos Escanciano, editors, *Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics, volume 38)*, pages 195–236. Emerald Group Publishing, 2017.
- [35] Evarist Giné and Richard Nickl. *Mathematical Foundations of Infinite-dimensional Statistical Models*. Cambridge University Press, New York, 2016.
- [36] Jinyong Hahn, Petra Todd, and Wilbert van der Klaauw. Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1):201–209, 2001.
- [37] Wolfgang Härdle, Marlene Müller, Stefan Sperlich, and Axel Werwatz. *Nonparametric and Semiparametric Models*. Springer, Heidelberg, 2004.
- [38] Miguel A. Hernán and James M. Robins. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC, 2020.

- [39] Ari Hyytinen, Jaakko Meriläinen, Tuukka Saarimaa, Otto Toivanen, and Janne Tukiainen. When does regression discontinuity design work? evidence from random election outcomes. *Quantitative Economics*, 9(2):1019–1051, 2018.
- [40] Weiwei Jiang and Rong JB Zhu. Extrapolating treatment effects in multiple-score regression discontinuity designs. *Journal of Educational and Behavioral Statistics*, page 10769986261426399, 2026.
- [41] Luke Keele and Rocio Titiunik. Natural experiments based on geography. *Political Science Research and Methods*, 4(1):65–95, 2016.
- [42] Luke J. Keele and Rocio Titiunik. Geographic boundaries as regression discontinuities. *Political Analysis*, 23(1):127–155, 2015.
- [43] Luke J. Keele, Rocio Titiunik, and Jose Zubizarreta. Enhancing a geographic regression discontinuity design through matching to estimate the effect of ballot initiatives on voter turnout. *Journal of the Royal Statistical Society: Series A*, 178(1):223–239, 2015.
- [44] Luke J. Keele, Scott Lorch, Molly Passarella, Dylan Small, and Rocio Titiunik. An overview of geographically discontinuous treatment assignments with an application to children’s health insurance. In Matias D. Cattaneo and Juan Carlos Escanciano, editors, *Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics, volume 38)*, pages 147–194. Emerald Group Publishing, 2017.
- [45] Christina Korting, Carl Lieberman, Jordan Matsudaira, Zhuan Pei, and Yi Shen. Visual inference and graphical representation in regression discontinuity designs. *Quarterly Journal of Economics*, 138(3):1977–2019, 2023.
- [46] Juliana Londoño-Vélez, Catherine Rodríguez, and Fabio Sánchez. Upstream and downstream impacts of college merit-based financial aid for low-income students: Ser pilo paga in colombia. *American Economic Journal: Economic Policy*, 12(2):193–227, 2020.
- [47] James G MacKinnon, Morten Ørregaard Nielsen, and Matthew D Webb. Cluster-robust inference: A guide to empirical practice. *Journal of Econometrics*, 232(2):272–299, 2023.
- [48] John P Papay, John B Willett, and Richard J Murnane. Extending the regression-discontinuity approach to multiple assignment variables. *Journal of Econometrics*, 161(2):203–207, 2011.
- [49] Sean F Reardon and Joseph P Robinson. Regression discontinuity designs with multiple rating-score variables. *Journal of Research on Educational Effectiveness*, 5(1):83–104, 2012.

- [50] Maxime Rischard, Zach Branson, Luke Miratrix, and Luke Bornn. Do school districts affect nyc house prices? identifying border differences using a bayesian nonparametric approach to geographic regression discontinuity designs. *Journal of the American Statistical Association*, 116(534):619–631, 2021.
- [51] Philipp Alexander Schwarz, Oliver Schacht, Sven Klaassen, Johannes Oberpriller, and Martin Spindler. Effect identification and unit categorization in the multi-score regression discontinuity design with application to led manufacturing. *arXiv preprint arXiv:2508.15692*, 2025.
- [52] Yuya Shimizu. Nonparametric regression under cluster sampling. *Journal of Econometrics*, 252:106102, 2025.
- [53] Leon Simon et al. *Lectures on geometric measure theory*. Centre for Mathematical Analysis, Australian National University Canberra, 1984.
- [54] A.B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer, 2008.
- [55] Aad W. van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes*. Springer, 1996. doi: 10.1007/978-1-4757-2545-2.
- [56] Vivian C Wong, Peter M Steiner, and Thomas D Cook. Analyzing regression-discontinuity designs with multiple assignment variables: A comparative study of four estimation methods. *Journal of Educational and Behavioral Statistics*, 38(2):107–141, 2013.