Causal Treatment Effect Aggregation in Boundary Discontinuity Designs*

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

July 4, 2025

Abstract

This paper analyzes two distinct approaches for estimating aggregated average treatment effects in boundary discontinuity designs: the widely-used *pre-aggregation* approach, and a novel *post-aggregation* approach. The pre-aggregation approach bundles all units near a treatment assignment boundary first, and then estimates a single, aggregated local average treatment effect. While commonly used in empirical work, its formal statistical properties remain underexplored. In contrast, the post-aggregation approach first estimates heterogeneous treatment effects along the assignment boundary, and then aggregates them, offering flexibility in weighting and robustness to boundary irregularities. We provide a precise definition of the aggregated causal parameters identified by each approach, and establish valid large-sample estimation and inference methods based on them. The theoretical results elucidate the relative merits of each aggregation approach, providing practical guidance for past and future applications of boundary discontinuity designs. General-purpose companion software is provided, and illustrated using an empirical application.

Keywords: regression discontinuity, treatment effects estimation, causal inference.

*We thank Alberto Abadie, Boris Hanin, Kosuke Imai, Xinwei Ma, and Jeff Wooldridge for comments and discussions. In particular, we thank Alberto Abadie for suggesting the terminology "pre-aggregation" and "post-aggregation", and Kosuke Imai for suggesting the procedure discussed in Remark 1. Cattaneo and Titiunik gratefully acknowledge financial support from the National Science Foundation (SES-2019432, DMS-2210561, and SES-2241575). Cattaneo gratefully acknowledge financial support from the Data-Driven Social Science initiative at Princeton University.

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

‡Department of Politics, Princeton University.

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

Contents

1 Introduction

Boundary discontinuity designs are used to estimate average treatment effects for units located near a boundary inducing a discontinuous change in treatment assignment. Early influential examples include Card and Krueger [1994], Black [1999], and Dell [2010]; see also Jardim et al. [2024] for a recent application and more references. Keele and Titiunik [2016] give an introduction to causal identification, estimation, and inference methods based on geography. This research design, sometimes called a Multi-dimensional or Geographic Regression Discontinuity (RD), is a generalization of the canonical RD design [Cattaneo and Titiunik, 2022].

A common approach in the literature is to first bundle all units that are near to the assignment boundary, according to some location metric, and then conduct the empirical analysis using a "local randomization" framework to estimate an aggregated, local to the boundary, average RD treatment effect. We call this popular empirical method the *pre-aggregation* approach because all units "close" to the boundary are pooled together first, and then a single average RD treatment effect is estimated. An alternative approach for causal treatment effect aggregation is to estimate the average treatment effect curve along the boundary first, which captures a collection of local RD heterogeneous treatment effects at each point on the assignment boundary, and bundle them after to recover an aggregated, local to the boundary, average treatment effect. We call this alternative empirical method the *post-aggregation* approach.

The pre-aggregation approach is widely used empirical work, but its formal statistical properties have not been unexplored before. The post-aggregation approach is new to the literature, thereby providing an alternative method for causal treatment effect aggregation in empirical work. This paper studies identification, estimation, and inference for both pre-aggregation and post-aggregation approaches to estimate the local (to the boundary) average treatment effect in boundary discontinuity designs. Our results describe precisely the relative merits of each approach, and thus help inform past and future applications.

The pre-aggregation approach does not require specific location information because aggregation is done directly for all units located within a local region covering the assignment boundary. As a result, approximation of the aggregated, local to the boundary, RD average treatment effect is the main and only goal, which is achieved by assuming that the covering region shrinks towards

the assignment boundary as the sample size increases. Heuristically, localization is achieve along level sets expanding the assignment boundary. The most common implementation is a simple difference-in-means estimator. If, in addition, location or distance-to-the-boundary information is available, then the estimator may include regression adjustments to account for location-fixed-effects or polynomial expansions based on distance. Our main results give a precise definition of the causal parameter uncovered by the pre-aggregation approach, and provide estimation and inference results. We also highlight potential issues related to the lack of smoothness of the assignment boundary.

The post-aggregation approach requires specific location information for each unit, and builds on the identification, estimation, and inference results for the average treatment effect curve along the boundary recently established by Cattaneo et al. [2025a]. It has several advantages for empirical practice. First, it builds off an explicit heterogeneity treatment effect analysis, thereby complementing and progressively expanding those causal findings. Second, it allows for different weighting schemes for aggregation, giving the researcher more options for causal estimation and inference. Third, it is robust to issues related to non-smooth or otherwise "too" complex assignment boundaries, which can affect the performance of the pre-aggregation approach. Fourth, due to its robustness and simplicity, optimal bandwidth selection and robust bias correction can be developed and implemented.

Our theoretical results build on precise concepts from geometric measure theory [Federer, 2014], which allow us to provide both high-level and primitive conditions on the geometry of the assignment boundary and the data generating process. For each estimator approach, we establish a mean square error (MSE) approximation and an asymptotic Gaussian distributional approximation. We employ these results to deduce consistency, convergence rates, and uncertainty quantification methods. These results elucidate the relative merits of each aggregation approach, providing practical guidance for past and future applications of boundary discontinuity designs. We also provide general-purpose companion software, and illustrate the feasibility of our methods by revising the empirical application of Londoño-Vélez et al. [2020].

This work contributes to a burgeoning methodological literature on causal treatment effect aggregation in boundary discontinuity and multi-dimensional RD designs, including in particular the popular geographic RD design. Previous work has largely focused on identification in specific settings [Papay et al., 2011, Reardon and Robinson, 2012, Keele and Titiunik, 2015], or on

methodology for geographic RD designs [Keele et al., 2015, 2017, Galiani et al., 2017, Diaz and Zubizarreta, 2023]. Most recently, Cattaneo et al. [2025a] provided foundational work on identification, estimation, and inference for the average treatment effect curve along the assignment boundary, which our post-aggregation approach builds upon and extends. This paper is the first to formally study identification, estimation and inference methods for treatment effect aggregation in boundary discontinuity designs.

The remainder of the paper is organized as follows. Section 2 presents the setup. Section 3 presents results for the pre-aggregation approach, while Section 4 discusses the post-aggregation approach. Section 5 reports the empirical application. Finally, Section 6 concludes. The supplemental appendix reports more general theoretical results, proofs, and other technical results.

2 Setup

Let $(Y_i(0), Y_i(1), \mathbf{X}_i^\top)^\top$, $i = 1, 2, \ldots, n$, be a random sample with $Y_i(0)$ and $Y_i(1)$ denoting the potential outcomes for unit i under control and treatment assignment, respectively, and $\mathbf{X}_i = (X_{1i}, X_{2i})^\top$ being a continuous score supported on $\mathscr{X} \subseteq \mathbb{R}^2$. Let $\mathscr{X} = \mathscr{A}_0 \cup \mathscr{A}_1$ with $\mathscr{A}_0$ and $\mathscr{A}_1$ be the control and treatment disjoint (connected) regions, respectively, and $\mathscr{B} = \mathtt{bd}(\mathscr{A}_0) \cap \mathtt{bd}(\mathscr{A}_1)$ be the assignment boundary, where $\mathtt{bd}(\mathscr{A}_t)$ denotes the boundary of the set $\mathscr{A}_t$. Without loss of generality, we assume the known one-dimensional boundary curve $\mathscr{B}$ belongs to the treatment assignment area, that is, $\mathtt{bd}(\mathscr{A}_1) \subset \mathscr{A}_1$ and $\mathscr{B} \cap \mathscr{A}_0 = \emptyset$. Units are assigned the control group if $\mathscr{X} \in \mathscr{A}_0$, or otherwise to the treatment group if $\mathscr{X} \in \mathscr{A}_1$. Thus, the observed data is (Y_i, T_i), $i = 1, \ldots, n$, where

$$Y_i = T_i Y_i(0) + (1 - T_i) Y_i(1), \qquad T_i = \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_1),$$

with $\mathbb{1}(\cdot)$ denoting the indicator function. The score variable $\mathbf{X}_i$ for each unit, or certain transformation thereof such as a scalar distance measure to the boundary, may or may not be directly observed depending on the setting considered. We discuss these different cases below.

We impose throughout the paper the following basic regularity conditions on the data generating process.

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

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

(ii) *The distribution of* $\mathbf{X}_i$ *has a Lebesgue density* $f(\mathbf{x})$ *that is continuous and bounded away from zero on its support compact support* $\mathscr{X}$*.*

(iii) $\mu_t(\mathbf{x}) = \mathbb{E}[Y_i(t)|\mathbf{X}_i = \mathbf{x}]$ *is* $(p+1)$*-times continuously differentiable on* $\mathscr{X}$*.*

(iv) $\sigma_t^2(\mathbf{x}) = \mathbb{V}[Y_i(t)|\mathbf{X}_i = \mathbf{x}]$ *is bounded away from zero and continuous on* $\mathscr{X}$*.*

(v) $\sup_{\mathbf{x}\in\mathscr{X}} \mathbb{E}[|Y_i(t)|^{2+v}|\mathbf{X}_i = \mathbf{x}] < \infty$*.*

These conditions are all standard in the literature. The *average treatment effect curve along the boundary* is

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

This functional parameter captures the possibly heterogeneous, average RD treatment effect at each point on the assignment boundary. See Cattaneo et al. [2025a] for a foundational analysis on identification, estimation, and inference for $\tau(\mathbf{x})$. While this functional parameter is certainly of practical interest, researchers often would like to also report a single causal treatment effect via some aggregation along (a region of) the assignment boundary $\mathscr{B}$. In many applications, most notably in geographic RD designs, the assignment boundary may exhibit kinks or other irregularities, which can make the aggregation process empirically and technically challenging.

Because the domain of aggregation $\mathscr{B}$ forms a possibly irregular one-dimensional manifold in $\mathbb{R}^2$, we need to introduce some concepts from geometric measure theory to define formally the class of parameters interest studied in this paper. The upcoming definitions will be essential in proving our main theoretical results: see the supplemental appendix for details, and Federer [2014] for a classical background reference.

2.1 Integration over 1-dimensional Manifolds

Integrating functions over one-dimensional manifolds in the plane extends the familiar concept of integration from intervals to curves. This extension is crucial in various fields, from physics (e.g., calculating work done by a force along a path) to engineering (e.g., fluid flow over a surface). In

our context, we seek to formally define expressions such as

$$\int_{\mathscr{B}} \tau(\mathbf{b}) w(\mathbf{b})\, d\mathbf{b},$$

where $w : \mathscr{B} \mapsto \mathbb{R}$ is a known weighting function. When the boundary $\mathscr{B}$ is "nice" enough, the above integral can be understood as a line integral. While a one-dimensional manifold on the plane can be intuitively thought of as a smooth curve, to establish a rigorous framework for such integrals for more general one-dimensional domains $\mathscr{B}$, we need to consider curves that might not be smooth everywhere, or easily parameterized by differentiable functions. We thus need to delve into the concepts of the *Hausdorff measure* and *rectifiable sets*.

(i) *Hausdorff Measure.* The foundation for measuring the "length" of a general one-dimensional curve in the plane is the 1-dimensional Hausdorff measure, denoted by $\mathfrak{H}^1$. Unlike the Lebesgue measure, which is primarily designed for integer dimensions, the Hausdorff measure can be defined for any non-negative real dimension. For a set $E \subset \mathbb{R}^2$ and $\delta > 0$, we define the δ-cover $\mathfrak{H}^1_\delta(E)$ as the infimum of sums $\sum_{i=1}^{\infty} \operatorname{diam}(A_i)$ over all countable covers $\{A_i\}$ of E such that $\operatorname{diam}(A_i) < \delta$. The 1-dimensional Hausdorff measure is then given by $\mathfrak{H}^1(E) = \lim_{\delta \to 0^+} \mathfrak{H}^1_\delta(E)$.

(ii) *Rectifiable Set.* A subset E of $\mathbb{R}^2$ is said to be 1-rectifiable if it is of Hausdorff dimension 1 and, roughly speaking, E can be covered, up to a set of 1-dimensional Hausdorff measure zero, by a countable union of images of Lipschitz functions from subsets of $\mathbb{R}$. More precisely, a Borel set $E \subset \mathbb{R}$ is 1-rectifiable if $\mathfrak{H}^1(E) < \infty$ and there exist a countable collection of Lipschitz maps $g_i : \mathbb{R} \mapsto \mathbb{R}$ such that $E \subset \bigcup_{i=1}^{\infty} g_i(\mathbb{R})$. Heuristically, this means that the curve can be "approximated" by pieces that are images of Lipschitz functions from intervals.

The definition of rectifiable curve is more general than that of a continuously differentiable curve and, crucially, it allows for curves with "corners" or points where a tangent might not exist in the classical sense, yet still possess a well-defined length. For a smooth curve, this definition precisely yields its arc length. For rectifiable curves, the 1-dimensional Hausdorff measure also corresponds to the intuitive notion of length, even when classical arc length formulas relying on differentiability might fail. The key property of rectifiable sets is that they admit an "approximate tangent space"

almost everywhere, which is essential for defining the integration process.

For a measurable function $g : \mathbb{R}^2 \to \mathbb{R}$ and a 1-rectifiable set $\mathscr{D} \subset \mathbb{R}^2$, the integral of g over $\mathscr{D}$ is defined as $\int_{\mathscr{D}} g \, d\mathfrak{H}^1$, where the integral is formally defined using standard measure theory. For sufficiently "nice" functions (e.g., continuous functions with compact support), this integral can be further understood using parameterizations and the *Area Formula*. Since $\mathscr{D}$ is a 1-rectifiable set, it can be (almost entirely) represented as the disjoint union of the images of a countable collection of Lipschitz maps $\phi_j : I_j \mapsto \mathbb{R}^2$, where $I_j \subseteq \mathbb{R}$ are intervals. Then,

$$\int_{\mathscr{D}} g \, d\mathfrak{H}^1 = \sum_{j=1}^{\infty} \int_{I_j} g(\phi_j(u)) J_1\phi_j(u) du, \tag{1}$$

where $J_1\phi_j(u)$ is the 1-dimensional Jacobian of the Lipschitz map ϕ_j at the point u; the Jacobian is defined for $\mathfrak{m}$-almost every $u \in I_j$, where $\mathfrak{m}$ denotes the Lebesgue measure on $\mathbb{R}$. In particular, if $\phi_j(u) = (\phi_{j1}(u), \phi_{j2}(u))^\top$ is differentiable, then the 1-dimensional Jacobian is simply the magnitude of the derivative vector: $J_1\phi_j(u) = \|\frac{d}{du}(\phi_{j1}(u), \phi_{j2}(u))^\top\| = \sqrt{(\frac{d}{du}\phi_{j1}(u))^2 + (\frac{d}{du}\phi_{j2}(u))^2}$, where $\|\cdot\|$ denotes the Euclidean norm.

The integral (1) is a natural generalization of the standard line integral, where $d\mathfrak{H}^1$ serves as the rigorous "length element," enabling a robust framework for integration over complex one-dimensional domains in the plane. More precisely, if $\mathscr{D}$ is a smooth curve parameterized by $\gamma : [a, b] \mapsto \mathbb{R}^2$, then we can set $\phi_1 = \gamma$ and $I_1 = [a, b]$, and $\phi_j = 0$ for all $j \geq 2$, to obtain

$$\int_{\mathscr{D}} g \, d\mathfrak{H}^1 = \int_{I_1} g(\gamma(u)) \Big\| \frac{d}{du}\gamma(u) \Big\| du,$$

because the 1-dimensional Jacobian $J_1\phi_j(u) = \big\|\frac{d}{du}\gamma(u)\big\|$ is precisely the speed of the parameterization. For notational simplicity, and whenever there is no confusion, we write

$$\int_{\mathscr{D}} g(\mathbf{u}) d\mathbf{u} = \int_{\mathscr{D}} g \, d\mathfrak{H}^1,$$

with the understanding that the left-hand-side integral is just notation for the rigorously defined right-hand-side integral.

2.2 Aggregated Causal Treatment Effects

We consider identification, estimation, and inference for a class of parameters of interest that aggregates the heterogeneous treatment effects, $(\tau(\mathbf{b}) : \mathbf{b} \in \mathscr{B})$, along the often non-smooth assignment boundary $\mathscr{B}$. Assuming that $\mathscr{B}$ is (at least) rectifiable, and for $w : \mathscr{X} \mapsto \mathbb{R}$ a known function, the *aggregated boundary treatment effect* is

$$\tau_w = \frac{\int_{\mathscr{B}} \tau(\mathbf{b}) w(\mathbf{b}) d\mathbf{b}}{\int_{\mathscr{B}} w(\mathbf{b}) d\mathbf{b}},$$

which defines a class of treatment effects indexed by the choice of weights.

The definition of τ_w depends on the geometry of the assignment boundary $\mathscr{B}$, and the choice of weights w. For example, if $\mathscr{B}$ is a smooth curve parameterized by $\gamma : [a, b] \mapsto \mathbb{R}^2$, then

$$\tau_w = \frac{\int_{[a,b]} \tau(\gamma(u)) w(\gamma(u)) \big\| \frac{d}{du}\gamma(u) \big\| du}{\int_{[a,b]} w(\gamma(u)) \big\| \frac{d}{du}\gamma(u) \big\| du}.$$

In particular, if $\mathscr{B}$ is a straight line with $\mathscr{B} = \{\mathbf{a} + s\mathbf{b} : 0 \leq s \leq L\}$. Then

$$\tau_w = \frac{\int_{[0,L]} \tau(\mathbf{a} + s\mathbf{b}) w(\mathbf{a} + s\mathbf{b}) ds}{\int_{[0,L]} w(\mathbf{a} + s\mathbf{b}) ds}.$$

3 Pre-Aggregation Approach

A common empirical approach to estimate an aggregated boundary treatment effect is to first pool all observations with score within a small region covering the assignment boundary $\mathscr{B}$, and then estimate the "local" average treatment effect using only those observations. This localization approach does not employ $\mathbf{X}_i$ directly, but rather only information about whether observations are "close" to some point on the boundary.

To formalize the pre-aggregation approach, let $h > 0$ be the bandwidth controlling the width of the region $\mathscr{R}(h)$ covering the assignment boundary $\mathscr{B}$. Then, the tubular localization region is

$$\mathscr{R}(h) = \mathscr{R}_0(h) \cup \mathscr{R}_1(h), \qquad \mathscr{R}_t(h) = \Big\{ \mathbf{x} \in \mathscr{A}_t : \inf_{\mathbf{b} \in \mathscr{B}} \mathscr{d}(\mathbf{b}, \mathbf{x}) \leq h \Big\},$$

for $t \in \{0,1\}$. The pre-aggregation boundary average treatment effect estimator is

$$\widehat{\tau}_{\text{pre}} = \widehat{\tau}_{\text{pre},1} - \widehat{\tau}_{\text{pre},0}, \tag{2}$$

where

$$\widehat{\tau}_{\text{pre},t} = \frac{1}{N_t(h)} \sum_{i=1}^{n} \mathbb{1}(\mathbf{X}_i \in \mathscr{R}_t(h)) Y_i, \qquad N_t(h) = \sum_{i=1}^{n} \mathbb{1}(\mathbf{X}_i \in \mathscr{R}_t(h))$$

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

The (signed) induced "distance" to the closest point on the boundary for each unit is

$$D_i = (2T_i - 1) d(\mathbf{X}_i, \mathscr{B}), \qquad d(\mathbf{x}, \mathscr{B}) = \inf_{\mathbf{b} \in \mathscr{B}} \mathscr{d}(\mathbf{x}, \mathbf{b}),$$

where $\mathscr{d} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ denotes a distance metric on $\mathbb{R}$. Thus, for each unit, $|D_i|$ is the random distance to the closest point on the boundary for each unit distance, and $T_i = \mathbb{1}(D_i \geq 0)$. It follows that the estimator $\widehat{\tau}_{\text{pre}}$ is a "localized" difference-in-means estimator, which does not explicitly requires information about the score $\mathbf{X}_i$ or the induced random signed distance D_i: it only employs information about whether each unit is "close" enough to the boundary, that is, it only leverages location information through the indicator variables T_i, $\mathbb{1}(\mathbf{X}_i \in \mathscr{R}_0(h)) = \mathbb{1}(-h \leq D_i < 0)$ and $\mathbb{1}(\mathbf{X}_i \in \mathscr{R}_1(h)) = \mathbb{1}(0 \leq D_i \leq h)$, for each unit $i = 1, 2, \cdots, n$.

Interestingly, the estimator $\widehat{\tau}_{\text{pre}}$ can be thought of employing a local constant regression approximation based on the "level sets" in the tubular $\mathscr{R}$, as moving away from the assignment boundary $\mathscr{B}$. This interpretation is equivalent to a one-dimensional RD estimator based on the univariate random signed distance score D_i, and with cutoff $c = 0$. Thus, if information on $D_1, \ldots, D_n$ is also available, then a local polynomial debiasing approach could be used, as the following remark explains.

Remark 1 (Local Polynomial Generalization)**.** The pre-aggregation estimator $\widehat{\tau}_{\text{pre}}$ is expected to exhibit a smoothing bias of order h, since it implicitly employs a local-constant regression approximation based on the score variable $D_i = \inf_{\mathbf{b} \in \mathscr{B}} \mathscr{d}(\mathbf{b}, \mathbf{X}_i)$ for its construction. More precisely, if the random variables D_i, $i = 1, 2, \ldots, n$ are observed, then the local polynomial pre-aggregation

estimators for control and treatment groups are, for $t \in \{0,1\}$,

$$\widehat{\tau}_{\text{pre},t} = \mathbf{e}_1^\top \widehat{\boldsymbol{\xi}}_t, \qquad \widehat{\boldsymbol{\xi}}_t = \underset{\boldsymbol{\xi}\in\mathbb{R}^{p+1}}{\arg\min} \sum_{i=1}^{n} \left(Y_i - \mathbf{r}_p(D_i)^\top \boldsymbol{\xi}\right)^2 k_h(D_i) \mathbb{1}(T_i = t),$$

where $\mathbf{e}_1$ is the conformable first unit vector, $\mathbf{r}_p(u) = (1, u, u^2, \cdots, u^p)^\top$ is the usual univariate polynomial basis, and $k_h(u) = k(u/h)/h$ with $k(\cdot)$ a univariate kernel function. It follows that setting $p = 0$ and $k(u) = \mathbb{1}(|u| \leq 1)$ gives back the pre-aggregation estimator studied in (2). Under regularity conditions, it is reasonable to expect that the local polynomial pre-aggregation estimator $\widehat{\tau}_{\text{pre}} = \mathbf{e}_1^\top \widehat{\boldsymbol{\xi}}_1 - \mathbf{e}_1^\top \widehat{\boldsymbol{\xi}}_0$ could exhibit an improved bias of order h^{p+1}. However, this requires additional, specific restrictions on the geometry of $\mathscr{B}$, the choice of distance function $\mathscr{d}(\cdot)$, and the smoothness of the underlying induced conditional expectations. See Cattaneo et al. [2025a] for more discussion of distance-based methods. Since the local polynomial estimation approach is not common in the literature, we relegate its analysis to the supplemental appendix (Section SA-3.5). ⌟

3.1 Regularity Conditions

Our analysis of the pre-aggregation treatment effect estimator $\widehat{\tau}_{\text{pre}}$ in (2) proceeds under the following assumption.

Assumption 2 (Pre-Aggregation Approach)**.** *Let $\|\cdot\|$ be the Euclidean norm, and $t \in \{0,1\}$.*

(i) *There exists positive constants C_u and C_l such that $C_l \|\mathbf{x}_1 - \mathbf{x}_2\| \leq \mathscr{d}(\mathbf{x}_1, \mathbf{x}_2) \leq C_u \|\mathbf{x}_1 - \mathbf{x}_2\|$ for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathscr{X}$.*

(ii) *$J_1 d(\cdot, \mathscr{B}) \neq 0$ $\mathfrak{m}$-almost everywhere on $\mathscr{X}$, and $\int_{\mathscr{X}} |\frac{f(\mathbf{x})}{J_1 d(\mathbf{x},\mathscr{B})}| d\mathbf{x} < \infty$.*

(iii) *$\lim_{\epsilon \downarrow 0} F(s) = F(0)$ is finite, where $F(s) = \int_{d(\mathbf{x},\mathscr{B})=s} \frac{f(\mathbf{x})}{J_1 d(\mathbf{x},\mathscr{B})} d\mathfrak{H}^1(\mathbf{x})$ for all $s \geq 0$.*

(iv) *$r \mapsto \theta_t(r) = \mathbb{E}[Y_i | D_i = r, T_i = t] = \mathbb{E}[Y_i(t) | \mathscr{d}(\mathbf{X}_i, \mathscr{B}) = |r|]$ is continuous at 0.*

Part (i) in Assumption 2 restricts the underlying distance function to be equivalent to the Euclidean distance, while Parts (ii)–(iv) concern the geometry of the assignment boundary $\mathscr{B}$, as well as the smoothness of f and the induced regression functions, as they all interact with specific distance function used. These restrictions are high-level, and difficult to verify in general; see Section 2.1 for notation and definitions. If $\mathscr{B}$ enjoys more regularity (e.g., assuming $\mathscr{B}$ is piecewise parametrization by a Lipchitz function), then parts (ii)–(iv) in Assumption 2 may be verified under

primitive regularity conditions. See the supplemental appendix for more details; in particular, Section SA-3.5 verifies conditions (ii)–(iv) for the special case when $\mathscr{B}$ is piecewise linear (with finite many pieces).

3.2 MSE Approximation

The pre-aggregation estimator $\widehat{\tau}_{\text{pre}} = \widehat{\tau}_{\text{pre},1} - \widehat{\tau}_{\text{pre},0}$ in (2) admits the usual best linear decomposition from least squares algebra. For $t \in \{0, 1\}$,

$$\widehat{\tau}_{\text{pre},t} - \mu_{f,t} = \mathfrak{B}_{n,t} + L_{n,t} + Q_{n,t}, \qquad \mu_{f,t} = \frac{\int_{\mathscr{B}} \mu_t(\mathbf{b}) f(\mathbf{b})\, d\mathbf{b}}{\int_{\mathscr{B}} f(\mathbf{b})\, d\mathbf{b}}, \tag{3}$$

where

$$\begin{aligned}
\mathfrak{B}_{n,t} &= \theta_t^*(0) - \mu_{f,t}, \qquad & \theta_t^*(0) &= \frac{\mathbb{E}[\mathbb{1}(\mathbf{X}_i \in \mathscr{R}_t(h)) Y_i]}{\mathbb{E}[\mathbb{1}(\mathbf{X}_i \in \mathscr{R}_t(h))]}, \\
L_{n,t} &= \frac{1}{\mathbb{E}[\mathbb{1}(\mathbf{X}_i \in \mathscr{R}_t(h))]} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(\mathbf{X}_i \in \mathscr{R}_t(h))(Y_i - \theta_t^*(0)), \\
Q_{n,t} &= \Big(\frac{1}{\mathbb{E}[\mathbb{1}(\mathbf{X}_i \in \mathscr{R}_t(h))]} - \frac{n}{N_t(h)} \Big) \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(\mathbf{X}_i \in \mathscr{R}_t(h))(Y_i - \theta_t^*(0)).
\end{aligned}$$

The supplemental appendix studies the generalized version of this decomposition based on the estimator discussed in Remark 1. Employing the notation and definitions in Section 2.1, and leveraging the technical results in the supplemental appendix, we show that

$$\lim_{h \downarrow 0} \frac{1}{h} \mathbb{E}[\mathbb{1}(\mathbf{X}_i \in \mathscr{R}_t(h))] = \int_{\mathscr{B}} f(\mathbf{b}) d\mathbf{b},$$

and

$$\lim_{h \downarrow 0} \frac{1}{h} \mathbb{E}[\mathbb{1}(\mathbf{X}_i \in \mathscr{R}_t(h)) Y_i] = \int_{\mathscr{B}} \mu_t(\mathbf{b}) f(\mathbf{b}) d\mathbf{b},$$

for $t \in \{0, 1\}$, under Assumptions 1 and 2. Therefore, the non-random term $\mathfrak{B}_{n,t}$ captures the "smoothing bias" of the estimator $\widehat{\tau}_{\text{pre},t}$, which is difficult to characterized precisely in our current general setting. The term $L_{n,t}$ is a mean-zero average of independent random variables, and admits a Gaussian distributional approximation, thereby capturing the leading variability of the estimator.

Finally, because $L_{n,t} = O_{\mathbb{P}}(\frac{1}{\sqrt{nh}})$ and $Q_{n,t} = O_{\mathbb{P}}(\frac{1}{nh})$, it follows that the term $Q_{n,t}$ is negligible as $nh \to \infty$.

Putting all together, we obtain our first result. Let

$$\mathsf{MSE}[\widehat{\tau}_{\text{pre}}] = \mathbb{E}[(\mathfrak{B}_n + L_n)^2],$$

be the approximate mean square error based on the decomposition (3), where $\mathfrak{B}_n = \mathfrak{B}_{n,1} - \mathfrak{B}_{n,0}$, $L_n = L_{n,1} - L_{n,0}$. The higher-order term $Q_n = Q_{n,1} - Q_{n,0}$ has been explicitly ignored because it is asymptotically negligible under our assumptions, that is, $\mathsf{MSE}[\widehat{\tau}_{\text{pre}}]$ can be interpreted as the result of a Nagar-type expansion.

Theorem 1 (Pre-Aggregation: Convergence Rate and MSE Approximation)**.** *Suppose Assumptions 1 and 2 hold. If $nh \to \infty$ and $h \to 0$, then $Q_n = O_{\mathbb{P}}(\frac{1}{nh})$, and*

$$\mathsf{MSE}[\widehat{\tau}_{\text{pre}}] = \mathfrak{B}_n^2 + \frac{1}{nh}\mathfrak{V}_{\text{pre}}$$

where $\mathfrak{V}_{\text{pre}} = \mathfrak{V}_{\text{pre},1} + \mathfrak{V}_{\text{pre},0}$ with

$$\begin{aligned}
\mathfrak{V}_{\text{pre},t} &= \frac{h\mathbb{E}[\mathbb{1}(\mathbf{X}_i \in \mathscr{R}_t(h))(Y_i - \theta_t^*(0))^2]}{(\mathbb{E}[\mathbb{1}(\mathbf{X}_i \in \mathscr{R}_t(h))])^2} \\
&= \frac{\int_{\mathscr{B}} \sigma_t^2(\mathbf{b}) \frac{f(\mathbf{b})}{J_1 d(\mathbf{b},\mathscr{B})} d\mathbf{b}}{\left(\int_{\mathscr{B}} \frac{f(\mathbf{b})}{J_1 d(\mathbf{b},\mathscr{B})} d\mathbf{b}\right)^2} + \frac{\int_{\mathscr{B}} \left(\mu_t(\mathbf{b}) - \frac{\int_{\mathscr{B}} \mu_t(\mathbf{x}) \frac{f(\mathbf{x})}{J_1 d(\mathbf{x},\mathscr{B})} d\mathbf{x}}{\int_{\mathscr{B}} \frac{f(\mathbf{u})}{J_1 d(\mathbf{u},\mathscr{B})} d\mathbf{u}}\right)^2 \frac{f(\mathbf{b})}{J_1 d(\mathbf{b},\mathscr{B})} d\mathbf{b}}{\left(\int_{\mathscr{B}} \frac{f(\mathbf{b})}{J_1 d(\mathbf{b},\mathscr{B})} d\mathbf{b}\right)^2} + o(1),
\end{aligned}$$

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

This theorem establishes the convergence rates for each component in the decomposition of the pre-aggregation estimator. Since $\mathfrak{B}_n = o(1)$, it follows that $\widehat{\tau}_{\text{pre}}$ is consistent for τ_f, where

$$\tau_f = \mu_{f,1} - \mu_{f,0} = \frac{\int_{\mathscr{B}} \tau(\mathbf{b}) f(\mathbf{b})\, d\mathbf{b}}{\int_{\mathscr{B}} f(\mathbf{b})\, d\mathbf{b}}.$$

Unfortunately, characterizing the convergence rate of $\mathfrak{B}_n$, and hence of $\widehat{\tau}_{\text{pre}}$, is difficult at the current level of generality. Under additional regularity conditions on $\mathscr{B}$ and the data generating process, it is natural to expect that $|\mathfrak{B}_n| = O(h)$, but see Cattaneo et al. [2025a] more discussion

regarding possible challenges in establishing such a result. The asymptotic variance decomposes in the usual two terms: (i) variability for the correctly specified fit, and (ii) misspecification error. If $\mu_t(\mathbf{b}) = \mu$ for all $\mathbf{b} \in \mathscr{B}$, then the second term in the limit expression for $\mathfrak{V}_{\text{pre},t}$ is equal to zero.

3.3 Asymptotic Distribution and Inference

To conduct statistical inference we establish a central limit theorem for the usual t-statistic based on the pre-aggregation approach. Define

$$\widehat{\mathrm{T}}_{\text{pre}} = \frac{\widehat{\tau}_{\text{pre}} - \tau_f}{\widehat{\vartheta}_{\text{pre}}},$$

where

$$\widehat{\vartheta}^2_{\text{pre}} = \frac{1}{N_t(h)(N_t(h)-1)} \sum_{i=1}^{n} \mathbb{1}(\mathbf{X}_i \in \mathscr{R}_t(h))(Y_i - \widehat{\tau}_{\text{pre}})^2$$

is the "localized" sample variance estimator. Let $\rightsquigarrow$ denote weak convergence as $n \to \infty$, and let $\mathsf{N}(0,1)$ be the standard Gaussian distribution.

Theorem 2 (Pre-Aggregation: Asymptotic Distribution)**.** *Suppose Assumptions 1 and 2 hold. If* $nh \to \infty$ *and* $nh\mathfrak{B}_n^2 \to 0$*, then* $\widehat{\mathrm{T}}_{\text{pre}} \rightsquigarrow \mathsf{N}(0,1)$.

Theorem 2 gives large-sample validity of the usual confidence intervals for τ_f. For $\alpha \in (0,1)$, it follows that $\lim_{n\to\infty} \mathbb{P}[\tau_f \in \widehat{\mathrm{I}}_{\text{pre}}(\alpha)] = 1 - \alpha$, where

$$\widehat{\mathrm{I}}_{\text{pre}}(\alpha) = \left[\widehat{\tau}_{\text{pre}} - \mathfrak{c}_\alpha \widehat{\vartheta}_{\text{pre}} \ , \ \widehat{\tau}_{\text{pre}} + \mathfrak{c}_\alpha \widehat{\vartheta}_{\text{pre}}\right]$$

is the standard $(1-\alpha)$ confidence interval, with $\mathfrak{c}_\alpha = \Phi^{-1}(1-\alpha/2)$ and $\Phi(u) = \mathbb{P}[\mathsf{N}(0,1) \leq u]$.

Paralleling the discussion in Remark 1, our results justify estimation and inference based on the pre-aggregation approach, provided the small bias condition $nh\mathfrak{B}_n^2 \to 0$ holds. In particular, this approach can readily be implement using standard least squares regression:

$$\begin{bmatrix} \widehat{\zeta}_{\text{pre}} \\ \widehat{\tau}_{\text{pre}} \end{bmatrix} = \underset{\zeta,\tau\in\mathbb{R}^2}{\arg\min} \sum_{i=1}^{n} \big(Y_i - \zeta - T_i\tau\big)^2 \mathbb{1}(\mathbf{X}_i \in \mathscr{R}(h)),$$

where recall that $\mathbb{1}(\mathbf{X}_i \in \mathscr{R}(h)) = \mathbb{1}(|D_i| \leq h)$. Feasible inference for τ_f follows directly from standard least squares software, and is justified in large samples by Theorem 2.

4 Post-Aggregation Approach

This approach begins with a heterogeneous treatment effect estimator along the boundary, and then aggregates it to obtain a single boundary average treatment effect. Specifically, we consider the location-based local polynomial estimator proposed by Cattaneo et al. [2025a] for the treatment effect curve estimator of $\tau(\mathbf{x})$:

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

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

$$\widehat{\boldsymbol{\beta}}_t(\mathbf{x}) = \underset{\boldsymbol{\beta} \in \mathbb{R}^{\mathfrak{p}_p+1}}{\arg\min} \sum_{i=1}^{n} \big(Y_i - \mathbf{R}_p(\mathbf{X}_i - \mathbf{x})^\top \boldsymbol{\beta}\big)^2 K_h(\mathbf{X}_i - \mathbf{x}) \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t),$$

with $\mathfrak{p}_p = (2+p)(1+p)/2 - 1$, $\mathbf{R}_p(\mathbf{u}) = (1, u_1, u_2, u_1^2, u_2^2, u_1u_2, \cdots, u_1^p, u_2^p)^\top$ denotes the pth order polynomial expansion of the bivariate vector $\mathbf{u} = (u_1, u_2)^\top$, $K_h(\mathbf{u}) = K(u_1/h, u_2/h)/h^2$ for a bivariate kernel function $K(\cdot)$ and a bandwidth parameter h.

We impose the following conditions on the kernel function and assignment boundary.

Assumption 3 (Kernel Function)**.** *Let $t \in \{0,1\}$.*

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

(ii) $\liminf_{h \downarrow 0} \inf_{\mathbf{x} \in \mathscr{B}} \int_{\mathscr{A}_t} K_h(\mathbf{u} - \mathbf{x}) d\mathbf{u} \gtrsim 1$.

Part (i) in Assumption 3 imposes minimal regularity on the kernel function, and aligns with prior literature. Part (ii) in Assumption 3 ensures sufficient data availability for each point $\mathbf{x} \in \mathscr{B}$ in large samples, thereby making $\widehat{\tau}(\mathbf{x})$ well-defined. This is because $\mathbb{E}[K_h(\mathbf{X}_i - \mathbf{x})\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)] \gtrsim \int_{\mathscr{A}_t} K_h(\mathbf{u} - \mathbf{x}) d\mathbf{u}$ under Assumption 1(ii). The second part of the assumption rules out exotic boundaries, but is otherwise a minimal regularity condition on $\mathscr{B}$.

For a weighting scheme $w : \mathscr{X} \mapsto \mathbb{R}$, the post-aggregation boundary treatment effect estimator is

$$\widehat{\tau}_w = \frac{\int_{\mathscr{B}} \widehat{\tau}(\mathbf{b}) w(\mathbf{b})\, d\mathbf{b}}{\int_{\mathscr{B}} w(\mathbf{b})\, d\mathbf{b}},$$

where, without loss of generality, we assume $\int_{\mathscr{B}} w(\mathbf{b})\, d\mathbf{b} = 1$ to save notation.

4.1 MSE Expansion

To establish a valid MSE expansion, we need to introduce some notation. First, the leading pointwise conditional bias of the post-aggregation estimator $\widehat{\tau}(\mathbf{x})$ is $\bar{B}_{\mathbf{x}} = \bar{B}_{1,\mathbf{x}} - \bar{B}_{0,\mathbf{x}}$, where

$$\bar{B}_{t,\mathbf{x}} = \mathbf{e}_1^\top \widehat{\mathbf{\Gamma}}_{t,\mathbf{x}}^{-1} \sum_{|\mathbf{k}|=p+1} \frac{\mu_t^{(\mathbf{k})}(\mathbf{x})}{\mathbf{k}!} \frac{1}{n} \sum_{i=1}^{n} \mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{x}}{h}\Big)\Big(\frac{\mathbf{X}_i - \mathbf{x}}{h}\Big)^{\mathbf{k}} K_h(\mathbf{X}_i - \mathbf{x}) \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t),$$

using standard multi-index notation, and

$$\widehat{\mathbf{\Gamma}}_{t,\mathbf{x}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{r}_p\big(\frac{\mathbf{X}_i - \mathbf{x}}{h}\big) \mathbf{r}_p\big(\frac{\mathbf{X}_i - \mathbf{x}}{h}\big)^\top K_h(\mathbf{X}_i - \mathbf{x}) \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t),$$

for $t \in \{0, 1\}$ and $\mathbf{x} \in \mathscr{B}$. The pointwise conditional covariance of the post-aggregation estimator $\widehat{\tau}(\mathbf{x})$ is

$$\bar{V}_{\mathbf{x}_1,\mathbf{x}_2} = \bar{V}_{0,\mathbf{x}_1,\mathbf{x}_2} + \bar{V}_{1,\mathbf{x}_1,\mathbf{x}_2}, \qquad \bar{V}_{t,\mathbf{x}_1,\mathbf{x}_2} = \frac{1}{h} \mathbf{e}_1^\top \widehat{\mathbf{\Gamma}}_{t,\mathbf{x}_1}^{-1} \bar{\mathbf{\Sigma}}_{t,\mathbf{x}_1,\mathbf{x}_2} \widehat{\mathbf{\Gamma}}_{t,\mathbf{x}_2}^{-1} \mathbf{e}_1$$

with

$$\bar{\mathbf{\Sigma}}_{t,\mathbf{x}_1,\mathbf{x}_2} = \frac{h^2}{n} \sum_{i=1}^{n} \mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{x}_1}{h}\Big) \mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{x}_2}{h}\Big)^\top K_h(\mathbf{X}_i - \mathbf{x}_1) K_h(\mathbf{X}_i - \mathbf{x}_2) \varepsilon_i(\mathbf{x}_1) \varepsilon_i(\mathbf{x}_2) \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)$$

and $\varepsilon_i(\mathbf{x}) = Y_i - \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_0)\mu_0(\mathbf{x}) - \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_1)\mu_1(\mathbf{x})$, for $t \in \{0, 1\}$, for $t \in \{0, 1\}$ and $\mathbf{x}_1, \mathbf{x}_2 \in \mathscr{B}$.

The following theorem gives the MSE expansion.

Theorem 3 (Post-Aggregation: MSE Expansion)**.** *Suppose Assumptions 1 and 3 hold. If* $\frac{n^{\frac{v}{2+v}} h^2}{\log(1/h)} \to \infty$ *and* $h \to 0$*, then*

$$\mathbb{E}[(\widehat{\tau}_w - \tau_w)^2 | \mathbf{X}] = h^{2p+2} \bar{B}_{\mathscr{B}}^2 + \frac{1}{nh} \bar{V}_{\mathscr{B}} + o_{\mathbb{P}}\Big(h^{2p+2} + \frac{1}{nh}\Big).$$

where

$$\bar{B}_{\mathscr{B}} = \int_{\mathscr{B}} \bar{B}_{t,\mathbf{b}} w(\mathbf{b}) d\mathbf{b} = O_{\mathbb{P}}(1), \qquad \bar{V}_{\mathscr{B}} = \int_{\mathscr{B}} \int_{\mathscr{B}} \bar{V}_{\mathbf{b}_1,\mathbf{b}_2} w(\mathbf{b}_1) w(\mathbf{b}_2) d\mathbf{b}_1 d\mathbf{b}_2 = O_{\mathbb{P}}(1),$$

and $1/\bar{V}_{\mathscr{B}} = O_{\mathbb{P}}(1)$.

This theorem shows that the post-aggregation estimator is consistent for τ_w, with the one-dimensional convergence rate

$$|\widehat{\tau}_w - \tau_w| = O_{\mathbb{P}}\Big(h^{2p+2} + \frac{1}{nh}\Big).$$

Furthermore, the MSE-optimal bandwidth selector is

$$h_{\mathscr{B}} = \Big(\frac{\bar{V}_{\mathscr{B}}}{(2p+2)\bar{B}^2_{\mathscr{B}}} \frac{1}{n}\Big)^{1/(2p+3)},$$

provided that $\bar{B}_{\mathscr{B}} \neq 0$ with probability approaching one.

For implementation, it is easy to construct plug-in estimators of $\bar{B}_{\mathscr{B}}$ and $\bar{V}_{\mathscr{B}}$ using a preliminary bandwidth, and after replacing unknown quantities with estimates thereof. More precisely, for $\bar{B}_{\mathscr{B}}$, the unknowns $(\mu_t^{(\mathbf{k})}(\mathbf{x}) : |\mathbf{k}| = p+1)$ can be estimated using a higher-order local polynomial estimator, while for $\bar{V}_{\mathscr{B}}$, the unknown quantities $\bar{\mathbf{\Sigma}}_{0,\mathbf{x}_1,\mathbf{x}_2}$ and $\bar{\mathbf{\Sigma}}_{1,\mathbf{x}_1,\mathbf{x}_2}$ are replaced with

$$\widehat{\mathbf{\Sigma}}_{t,\mathbf{x}_1,\mathbf{x}_2} = \frac{h^2}{n} \sum_{i=1}^{n} \mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{x}_1}{h}\Big) \mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{x}_2}{h}\Big)^{\top} K_h(\mathbf{X}_i - \mathbf{x}_1) K_h(\mathbf{X}_i - \mathbf{x}_2) \widehat{\varepsilon}_i(\mathbf{x}_1) \widehat{\varepsilon}_i(\mathbf{x}_2) \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)$$

and $\widehat{\varepsilon}_i(\mathbf{x}) = Y_i - \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_0) \mathbf{R}_p(\mathbf{X}_i - \mathbf{x})^{\top} \widehat{\boldsymbol{\beta}}_0(\mathbf{x}) - \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_1) \mathbf{R}_p(\mathbf{X}_i - \mathbf{x})^{\top} \widehat{\boldsymbol{\beta}}_1(\mathbf{x})$, for $t \in \{0,1\}$ and $\mathbf{x}, \mathbf{x}_1, \mathbf{x}_2 \in \mathscr{B}$.

4.2 Asymptotic Distribution and Inference

We study the usual t-test based on the post-aggregation estimator $\widehat{\tau}_w$. That is, we consider the feasible statistic

$$\widehat{\mathrm{T}} = \frac{\widehat{\tau}_w - \tau_w}{\sqrt{\frac{1}{nh}\widehat{V}_{\mathscr{B}}}}, \qquad \widehat{V}_{\mathscr{B}} = \int_{\mathscr{B}} \int_{\mathscr{B}} (\widehat{V}_{0,\mathbf{b}_1,\mathbf{b}_2} + \widehat{V}_{1,\mathbf{b}_1,\mathbf{b}_2}) w(\mathbf{b}_1) w(\mathbf{b}_2) d\mathbf{b}_1 d\mathbf{b}_2,$$

where

$$\widehat{V}_{t,\mathbf{b}_1,\mathbf{b}_2} = \frac{1}{h}\mathbf{e}_1^\top \widehat{\mathbf{\Gamma}}_{t,\mathbf{b}_1}^{-1} \widehat{\mathbf{\Sigma}}_{t,\mathbf{b}_1,\mathbf{b}_2} \widehat{\mathbf{\Gamma}}_{t,\mathbf{b}_2}^{-1} \mathbf{e}_1$$

for $t \in \{0,1\}$ and $\mathbf{b}_1, \mathbf{b}_2 \in \mathscr{B}$.

The following theorem shows that $\widehat{\mathrm{T}}$ has a standard Gaussian asymptotic distribution.

Theorem 4 (Post-Aggregation: Asymptotic Distribution)**.** *Suppose Assumptions 1 and 3 hold. If* $\frac{n^{\frac{v}{2+v}}h^2}{\log(1/h)} \to \infty$ *and* $nh^{2p+3} \to 0$*, then* $\widehat{\mathrm{T}} \rightsquigarrow \mathsf{N}(0,1)$.

Provided that the bias condition $nh^{2p+3} \to 0$ holds, it follows from Theorem 4 that the confidence interval estimator

$$\widehat{\mathrm{I}}_{\mathrm{pos}}(\alpha) = \left[\widehat{\tau}_w - \mathfrak{c}_\alpha\sqrt{\frac{1}{nh}\widehat{V}_{\mathscr{B}}}\ ,\ \widehat{\tau}_w + \mathfrak{c}_\alpha\sqrt{\frac{1}{nh}\widehat{V}_{\mathscr{B}}}\right], \qquad \mathfrak{c}_\alpha = \Phi^{-1}(1-\alpha/2),$$

satisfies

$$\lim_{n\to\infty} \mathbb{P}[\tau_f \in \widehat{\mathrm{I}}_{\mathrm{pos}}(\alpha)] = 1-\alpha,$$

for any $\alpha \in (0,1)$. As it is well-known in the nonparametric smoothing literature, the bias condition $nh^{2p+3} \to 0$ requires an under-smoothed bandwidth choice (relative to the MSE-optimal $h_{\mathscr{B}}$). A theory-based approach to circumvent this problem is to employ robust bias-corrected inference [Calonico et al., 2014]. Specifically, the inference approach proceeds as follows:

1. **Step 1**. Construct the pth order local polynomial point estimator $\widehat{\tau}_w$, using the associated MSE-optimal bandwidth $h_{\mathscr{B}}$ (or a rate-consistent bandwidth estimate thereof). This gives an MSE-optimal treatment effect estimator $\widehat{\tau}_w$.

2. **Step 2**. Construct the test statistic $\widehat{\mathrm{T}}$ using a qth order local polynomial point estimator (both numerator and denominator), with $q \geq p+1$, and using the same bandwidth choice as in Step 1. Then, Theorem 4 holds provided Assumption 1 holds with p replaced by q. This gives a valid confidence interval estimator $\widehat{\mathrm{I}}(\alpha)$, constructed with a MSE-optimal point estimator and a robust bias-corrected t-statistic.

The theoretical properties of robust bias-corrected inference are discussed in Calonico et al. [2018, 2022], while its excellent empirical performance have been documented in Hyytinen et al. [2018] and De Magalhães et al. [2025]. See also Calonico et al. [2020] for more discussion on bandwidth selection in classical RD design settings. Our companion `R` software package `rd2d` [Cattaneo et al., 2025b] implements MSE-optimal point estimation and robust bias-corrected inference.

5 Empirical Illustration

We apply the pre-aggregation and post-aggregation approaches to the empirical study conducted by Londoño-Vélez et al. [2020]. They examined the impact of "Ser Pilo Paga" (SPP), a Colombian government subsidy designed to support post-secondary education. This anti-poverty initiative offered tuition assistance to undergraduate college students pursuing four- or five-year degrees at government-certified, high-quality higher education institutions. SPP eligibility combined merit and economic need: students were required a high score on Colombia's national standardized high school exit exam, SABER 11, and must originate from economically disadvantaged families, as indicated by the survey-based SISBEN wealth index. A deterministic rule with a fixed bivariate cutoff governed eligibility: students needed a SABER 11 score in the top 9 percent or higher and a household SISBEN index below a region-specific threshold. Formally, each student was assigned a bivariate score $\mathbf{X}_i = (\texttt{SABER11}_i, \texttt{SISBEN}_i)^\top$, where $X_{1i} = \texttt{SABER11}_i$ recorded the SABER11 score and $X_{2i} = \texttt{SISBEN}_i$ recorded the SISBEN wealth score, and the treatment assignment boundary is $\mathscr{B} = \{(\texttt{SABER11}, \texttt{SISBEN}) : (\texttt{SABER11}, \texttt{SISBEN}) \in \{\texttt{SABER11} \geq 0 \text{ and } \texttt{SISBEN} = 0\} \cup \{\texttt{SABER11} = 0 \text{ and } \texttt{SISBEN} \geq 0\}\}$, where variable was recentered at its corresponding cutoff. The outcome variable is $Y_i = 1$ if student i attended college or $Y_i = 0$ otherwise. The dataset includes $n = 363,096$ students from the 2014 cohort.

The causal parameter $\tau(\mathbf{b})$ captures the treatment effect of SPP on the probability of college education for students at the margin of program eligibility, at position $\mathbf{b} \in \mathscr{B}$ on the assignment boundary, as determined by their bivariate score $\mathbf{X}_i = (\texttt{SABER11}_i, \texttt{SISBEN}_i)^\top \in \mathscr{B}$. Cattaneo et al. [2025a] already discussed estimation and inference of $\tau(\mathbf{x})$ for this application. Our main contribution in this section is to present estimation and inference for aggregated treatment effects.

Table 1 presents empirical results for different bandwidth choices $h \in \{5, 10, 15, 20\}$. The first part

of the table focuses on the pre-aggregation method in (2). The resulting treatment effect estimators $\widehat{\tau}_{\text{pre}}$ are fairly stable, ranging from 0.38 to 0.43, all highly statistically different from zero. The second part of the table focuses on the post-aggregation method $\widehat{\tau}_w$, which is implemented using 40 evenly-spaced grid points on $\mathscr{B}$ and with equal weighting ($w(\mathbf{b}_j) = 1/J$ for $\mathbf{b}_j \in \mathscr{B}$, $j \in \{1, \ldots, J\}$, and $J = 40$). The estimators are also stable across bandwidth choices, ranging from 0.28 to 0.30, all highly statistically different from zero.

Method	h	Estimate	Z value	p-value	CI
Pre-aggregation	5	0.3813	52.03	0.0000	(0.3669, 0.3957)
	10	0.4062	72.67	0.0000	(0.3952, 0.4171)
	15	0.4191	85.62	0.0000	(0.4095, 0.4287)
	20	0.4327	95.97	0.0000	(0.4239, 0.4416)
Post-aggregation	5	0.2778	6.19	0.0000	(0.1748, 0.3370)
	10	0.2941	13.05	0.0000	(0.2425, 0.3282)
	15	0.2950	17.70	0.0000	(0.2601, 0.3248)
	20	0.3017	20.95	0.0000	(0.2636, 0.3180)

Table 1: Aggregated Treatment Effect Estimates.

6 Conclusion

This paper established formal identification, estimation and inference results for two distinct approaches for estimating aggregated average treatment effects in boundary discontinuity designs. These results elucidate the relative merits of each aggregation approach, providing practical guidance for past and future applications of boundary discontinuity designs. General-purpose companion software is provided to facilitate the used of the methods in empirical work.

References

Sandra E Black. Do better schools matter? parental valuation of elementary education. *Quarterly Journal of Economics*, 114(2):577–599, 1999.

Sebastian Calonico, Matias D. Cattaneo, and Rocio Titiunik. Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326, 2014.

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.

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.

Sebastian Calonico, Matias D. Cattaneo, and Max H. Farrell. Coverage error optimal confidence intervals for local polynomial regression. *Bernoulli*, 28(4):2998–3022, 2022.

David Card and Alan B. Krueger. Minimum wages and employment: A case study of the fast-food industry in new jersey and pennsylvania. *American Economic Review*, 84(4):772–793, 1994.

Matias D. Cattaneo and Rocio Titiunik. Regression discontinuity designs. *Annual Review of Economics*, 14:821–851, 2022.

Matias D Cattaneo, Rocio Titiunik, and Ruiqi Rae Yu. Estimation and inference in boundary discontinuity designs. *arXiv preprint arXiv:2505.05670*, 2025a.

Matias D Cattaneo, Rocio Titiunik, and Ruiqi Rae Yu. `rd2d`: Causal inference in boundary discontinuity designs. *arXiv preprint arXiv:2505.07989*, 2025b.

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.

Melissa Dell. The persistent effects of peru's mining mita. *Econometrica*, 78(6):1863–1903, 2010.

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.

Herbert Federer. *Geometric measure theory*. Springer, 2014.

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.

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.

Ekaterina Jardim, Mark C Long, Robert Plotnick, Jacob Vigdor, and Emma Wiles. Local minimum wage laws, boundary discontinuity methods, and policy spillovers. *Journal of Public Economics*, 234:105131, 2024.

Luke Keele and Rocio Titiunik. Natural experiments based on geography. *Political Science Research and Methods*, 4(1):65–95, 2016.

Luke J. Keele and Rocio Titiunik. Geographic boundaries as regression discontinuities. *Political Analysis*, 23(1):127–155, 2015.

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.

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.

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.

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.

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.