Difference Estimation for Regression Discontinuity with Multiple Ordered Thresholds

Vahid Moghani[1] and Saeed Badri[2]

[1]Erasmus School of Economics, Tinbergen Institute, and Erasmus Centre for Health Economics Rotterdam

[2]Vrije Universiteit Amsterdam, Erasmus School of Economics, and Tinbergen Institute

September 12, 2025

Abstract

This paper develops an estimator for regression discontinuity designs (RDD) with multiple thresholds. We construct an integrated–derivative estimator that recovers the outcome difference by numerically integrating nonparametric slope estimates between thresholds. We derive its asymptotic distribution, establishing a central limit theorem with a feasible variance formula. The integrated estimator is asymptotically independent of the standard boundary estimator, which enables an inverse–variance combination that is more efficient. Simulations confirm the theoretical predictions: efficiency gains are modest under uniform designs and more pronounced when data are sparse near the cutoffs.

[1]We thank Chen Zhou, Guido Imbens, Stefan Wager, Peter Hull, and Jonathan Roth for their constructive comments.

1 Introduction

In many empirical settings, researchers face running variables that cross multiple known thresholds, each inducing a discrete change in treatment. Regression Discontinuity Designs (RDD) are often extended to such contexts either by recentering and pooling discontinuities or by estimating the local discontinuity at each threshold separately and then aggregating the results to infer an average effect (Cattaneo et al., 2016; Bertanha, 2020). This approach captures the immediate impact of crossing a threshold but typically ignores how the outcome evolves between thresholds.

This paper proposes a complementary strategy. We study settings with a continuous running variable $z \in \mathbb{R}$ and an outcome $\mathbb{E}[y \mid z] = g(z)$, where $g(\cdot)$ is smooth between a finite number of known discontinuities $\{t_1, t_2, \ldots, t_K\}$. Rather than focusing solely on the jumps at each threshold, we exploit information from the interior of each segment. Specifically, we construct an estimator by numerically integrating a nonparametrically estimated derivative $g'(z)$ between thresholds, thereby capturing the accumulated change in the outcome across the interval. Whereas standard RDD estimators rely on outcome differences near the cutoffs, our approach draws on observations throughout the interior. The two estimators therefore use largely distinct subsets of the data, and we show they are asymptotically independent.

This property enables a natural efficiency gain. Since both estimators target the same parameter but with independent information, we combine them using inverse-variance weights. The resulting estimator is more efficient than either component alone, particularly when $g(z)$ carries information between thresholds or when data are sparse at the boundaries.

Our contributions are threefold. First, we introduce the integrated–derivative estimator, which leverages variation between thresholds in multiple-cutoff RDD designs. Second, we establish its asymptotic properties, including a CLT and variance expression. Third, we show that combining it with the conventional boundary estimator can improve efficiency in both theory and simulations.

The econometric foundations of regression discontinuity (RD) designs were established by Hahn et al. (2001), who showed that treatment effects at a cutoff can be identified as discontinuities in conditional expectations and derived the asymptotic behavior of local polynomial estimators. Subsequent work refined inference by addressing the boundary bias problem: Calonico et al. (2014) proposed robust bias–corrected intervals, and Calonico et al. (2020) developed bandwidth selectors optimized for coverage accuracy. Recent advances further broaden the scope of RD. Imbens and Wager (2019) derive minimax linear estimators that deliver uniformly valid confidence intervals and naturally extend to multivariate settings, while Calonico et al. (2025) provide a framework for analyzing treatment effect heterogeneity. Parallel to these developments, a smaller strand considers designs with multiple thresholds. Cattaneo et al. (2016) interpret pooled multi–cutoff estimators as weighted averages of local effects, and Bertanha (2020) proposes estimators that optimally combine information across cutoffs to target policy–relevant parameters. These approaches, however, remain boundary-based, relying primarily on observations near each cutoff. Our contribution is to complement this literature with an integrated–derivative estimator that, under minimal continuity assumptions and without imposing structural structure, exploits interior variation between thresholds. The estimator delivers a feasible central limit theorem with an explicit variance formula and is asymptotically independent of standard boundary estimators, enabling efficiency gains through inverse–variance combination.

The structure of the paper is as follows. Section 2 introduces the problem setup and reviews the standard multiple-threshold RD framework. Section 3 develops our integrated–derivative estimator in a simple single-segment setting and compares it with boundary-based estimation. Section 4 extends the approach to a general case. Section 5 presents simulation evidence. Section 6 concludes with a discussion of implications and potential extensions.

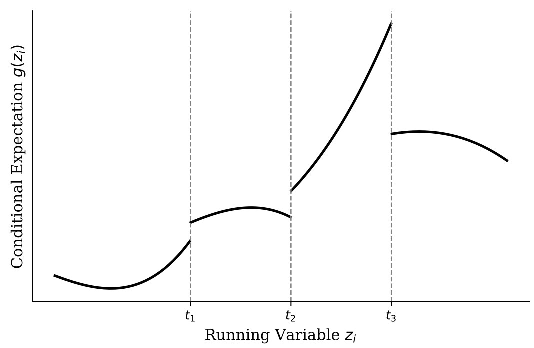

Figure 1: Illustration of Multiple Cutoffs in Regression Discontinuity Design

2 Problem Setup; Regression Discontinuity Design with Multiple Thresholds

Empirical research often employs Regression Discontinuity Designs (RDD) in settings where decisions hinge on multiple cutoffs. For example, children's school-entry dates are determined by birthdate cutoffs: being born just before or after a specific date decides whether a child starts school in one academic year or the next. This cutoff is applied repeatedly across multiple cohorts, creating a series of thresholds. Similarly, neonatal care policies can change depending on whether a baby is born just before or after midnight, influencing the care duration provided at every midnight boundary (see, for example, Almond and Doyle, 2011; Fredriksson and Öckert, 2014; Persson et al., 2021).

These scenarios involve multiple thresholds, $t_1, \ldots, t_K$, along the range of the running variable, z_i, each potentially inducing a discontinuity in the conditional expectation of the outcome, y_i. The conditional expectation is denoted as $g(z_i) = \mathbb{E}[y_i \mid z_i]$, which captures the relationship between the running variable and the expected outcome. For instance, these outcomes might represent students' expected grades affected by school-entry timing or the average health of newborns impacted by changes in neonatal care protocols. Figure 1 shows an illustration of such discontinuities.

To formalize, consider N independent observations (z_i, y_i), where $z_i \in [a, b]$. The function $g(\cdot)$, representing the conditional expectation of y_i given z_i, is smooth (infinitely differentiable or at least C^{p+1} for some p) except at the K known thresholds $t_1, \ldots, t_K$, where:

$$g(t_k^-) \neq g(t_k^+), \quad k = 1, \ldots, K.$$

Each threshold t_k induces a local jump in $g(\cdot)$, defined as:

$$\tau_k = g(t_k^+) - g(t_k^-).$$

Researchers often summarize these local jumps by estimating a weighted average of the discontinuities:

$$\tau = \sum_{k=1}^{K} \omega_k \tau_k, \quad \text{with} \sum_{k=1}^{K} \omega_k = 1.$$

The weights ω_k reflect the relative importance assigned to each threshold. If no specific threshold is prioritized, equal weights $\omega_k = 1/K$ may be used. Alternatively, weights can be proportional to the density of observations at each threshold. This weighting is automatically handled in standard RDD setups by recentering and pooling all thresholds (Cattaneo et al., 2016).

Rewriting τ, we have:

$$\tau = \omega_K g(t_K^+) - \omega_1 g(t_1^-) - \sum_{k=2}^{K} \left[\omega_k g(t_k^-) - \omega_{k-1} g(t_{k-1}^+) \right],$$

or equivalently[2]:

$$\tau = \omega_K g(t_K^+) - \omega_1 g(t_1^-) - \sum_{k=2}^{K} \delta_k, \tag{1}$$

where:

$$\delta_k = \omega_k g(t_k^-) - \omega_{k-1} g(t_{k-1}^+).$$

3 The Estimator for a Single Segment

First, we consider a simplified scenario by estimating the parameter

$$\delta = m(1) - m(0) = \int_0^1 m'(u)du \tag{2}$$

on a single interval normalized to $[0, 1]$. Later, we will extend this discussion to the general setting. We have independent observations $\{(X_i, Y_i)\}_{i=1}^N$ satisfying the model[3]

$$Y_i = m(X_i) + \varepsilon_i, \quad \mathbb{E}[\varepsilon_i \mid X_i] = 0, \quad \mathrm{Var}(\varepsilon_i \mid X_i) = \sigma^2(X_i).$$

Using the integral form of Equation 2, we focus on estimating the derivatives of $m(.)$ nonparametrically by partitioning the interval $[0, 1]$ into B subintervals ("bins") of the same length $2h$ with $h = \frac{1}{2B}$, each centered at points c_b. At each center c_b, we construct a local polynomial estimator of the first derivative $m'(c_b)$ using observations (X_i, Y_i) that lie within the bin $I_b = [c_b - h, c_b + h]$.

[2]As $K \to \infty$, assuming the weights ω_k vanish appropriately (e.g., $\lim_{K\to\infty} \omega_k = 0$), the parameter τ converges to:

$$\lim_{K\to\infty} \tau = -\sum_{k=2}^{\infty} \delta_k.$$

Intuitively, as the number of thresholds increases, τ represents the aggregated sum of changes in the expected outcome between consecutive thresholds. In this way, multiple-threshold RDD scenarios simplify to a series of individual differences across segments, which are weighted and combined to produce the overall parameter of interest.

[3]On the segment $[t_{k-1}, t_k]$, let $\Delta_k := t_k - t_{k-1} > 0$ and normalize $x = (z - t_{k-1})/\Delta_k \in [0, 1]$. We then define

$$m(x) \;=\; g\big(t_{k-1} + x\,\Delta_k\big),$$

so that $m(0) = g(t_{k-1}^+)$, $m(1) = g(t_k^-)$, and $m'(x) = \Delta_k\, g'\big(t_{k-1} + x\,\Delta_k\big)$.

Specifically, we estimate $m'(c_b)$ by solving the weighted least squares problem:

$$\hat{\boldsymbol{\beta}}(c_b) = \arg\min_{\boldsymbol{\beta}} \sum_{X_i \in I_b} \left(Y_i - \sum_{j=0}^{p} \beta_j (X_i - c_b)^j \right)^2 K_h(X_i - c_b),$$

with kernel $K_h(u) = K(u/h)/h$ controlled by bandwidth h. The estimator of interest is the first derivative:

$$\hat{\beta}_1(c_b) = e_2^T \hat{\boldsymbol{\beta}}(c_b) = \sum_{i=1}^{N} w_{i,b} Y_i, \tag{3}$$

where $w_{i,b} = w^T\left(\frac{X_i - c_b}{h}\right)$, and $w^T(u)$ is given by:

$$w^T(u) = e_2^T S_N^{-1} [1, uh, \ldots, (uh)^p]^T \frac{K(u)}{h},$$

with $K(u)$ being a kernel function and S_N a matrix defined as a matrix with its (i,j)th element being $S_{N,i+j}$ with:

$$S_{N,j} = \sum_{i=1}^{N} K_h(X_i - c_b)(X_i - c_b)^j.$$

Integrating these derivative estimates over the interval yields the final estimator:

$$\hat{\beta}_{\text{avg}} = 2h \sum_{b=1}^{B} \hat{\beta}_1(c_b) = \sum_{i=1}^{N} w_{N,i} Y_i,$$

where $w_{N,i}$ is the total weight assigned to observation i.

We impose the following assumptions for our theoretical analysis:

Assumption 1 (Smoothness of m)**.** $m \in C^{\max\{p+1,5\}}([0,1])$ for some integer $p \geq 1$.

Assumption 2 (Design Density)**.** Each X_i takes values in $[0,1]$ with a density f that is continuous, strictly positive on $[0,1]$ (no atoms), and has a bounded second derivative.

Assumption 3 (Bounded Conditional Variance)**.** $\sigma^2(x) = \text{Var}(\varepsilon_i \mid X_i = x)$ is continuous and bounded on $[0,1]$.

Assumption 4 (Kernel Conditions)**.** The kernel K is symmetric, bounded, supported on $[-1,1]$, and satisfies

$$\int_{-1}^{1} K(u)\,du \;=\; 1, \quad \int_{-1}^{1} |u|^{p+1}\,|K(u)|\,du \;<\; \infty.$$

Define the equivalent kernel for the derivative estimator by

$$K^*(u) = e_2^T S^{-1}(1, u, \ldots, u^p)^T K(u),$$

where S is a $(p+1) \times (p+1)$ matrix with entries

$$S_{i,j} = \mu_{i+j-2} \quad \text{and} \quad \mu_j = \int u^j\,K(u)\,du.$$

It follows that K^* is also supported and bounded on $[-1,1]$.

Assumption 5 (Bandwidth)**.** Let $h = h(N) \to 0$ as $N \to \infty$ and assume

$$\frac{Nh}{\log N} \;\to\; \infty, \qquad Nh^2 \;\to\; \infty, \qquad \sqrt{N}\,h^3 \to 0.$$

Assumption 6 (Higher-Order Moment of ε_i)**.** There exists $\delta > 0$ and a constant $C < \infty$ such that

$$\mathbb{E}\big[|\varepsilon_i|^{2+\delta} \;\big|\; X_i = x\big] \;\leq\; C \quad \text{for all } x \in [0,1],$$

uniformly in i.

Theorem 1 (Asymptotic Distribution of the Integrated Derivative Estimator)**.** *Let $\{(X_i, Y_i)\}_{i=1}^N$ be independent observations from the model*

$$Y_i \;=\; m(X_i) + \varepsilon_i, \quad \mathbb{E}\big[\varepsilon_i \mid X_i\big] = 0, \quad \mathrm{Var}\big(\varepsilon_i \mid X_i\big) = \sigma^2\big(X_i\big),$$

where $X_i \in [0,1]$ has density f. Suppose Assumptions 1–6 hold with polynomial degree $p \geq 2$, so that $m(.)$ is $(p+1)$-times continuously differentiable on $[0,1]$, and *we use a degree-p local polynomial estimator to estimate $m'(\cdot)$. Define*

$$\Gamma \;=\; \int_0^1 m'(u)\,\mathrm{d}u \;=\; m(1) - m(0).$$

Partition $[0,1]$ into B subintervals each of length $2h = 1/B$, with midpoints $c_b = (2b-1)\,h$ for $b = 1,\ldots,B$. For each midpoint c_b, let $\hat\beta_1(c_b)$ be the local polynomial estimator of the derivative: $m'(c_b)$. Then define[4]

$$\hat\beta_{\text{avg}} \;=\; 2h \sum_{b=1}^{B} \hat\beta_1(c_b).$$

Then, under Assumptions 1–6, and conditional on the design $X = \{X_1,\ldots,X_N\}$:

$$\sqrt{N\,h^2}\Big(\hat\beta_{\text{avg}} \;-\; \Gamma \;-\; \text{Bias}\Big) \;\xrightarrow{d}\; \mathcal{N}\Big(0,\; 2\int_{-1}^{1}\big[K^*(u)\big]^2\,\mathrm{d}u\;\int_0^1 \frac{\sigma^2(u)}{f(u)}\,\mathrm{d}u\Big),$$

where

- ***If** $p > 2$**,** then the local-polynomial bias is $o(h^2)$, and the dominant bias term comes from the midpoint rule (Lemma 3):*

$$\text{Bias} \;=\; -\frac{h^2}{6}\Big[m''(1) \;-\; m''(0)\Big].$$

- ***If** $p = 2$**,** both the midpoint rule and the local-polynomial estimator contribute to the bias at order h^2. From Lemma 3 and Lemma 4, one obtains*

$$\text{Bias} \;=\; \frac{h^2}{6}\Big(\int_{-1}^{1} u^3\,K^*(u)\,\mathrm{d}u - 1\Big)\,\big[m''(1) - m''(0)\big].$$

[4]The midpoint aggregation induces an $O(h^2)$ quadrature error for $\int_0^1 m'(u)\,du$. If a composite quadrature of global order $q \geq 2$ is used instead, the quadrature error becomes $O(h^q)$. Since the aggregate local–polynomial bias is $O(h^p)$, the deterministic remainder is $O\big(h^{\min\{p,q\}}\big)$. Under the normalization $\sqrt{Nh^2}$, it suffices that

$$\sqrt{N}\,h^{1+\min\{p,q\}} \;\longrightarrow\; 0.$$

Thus, while the theorem is stated under midpoint aggregation ($q = 2$, requiring $\sqrt{N}h^3 \to 0$), the bandwidth restriction is *not* pivotal: with composite Simpson ($q = 4$) the requirement relaxes to $\sqrt{N}h^4 \to 0$ when $p = 3$, and to $\sqrt{N}h^5 \to 0$ when $p \geq 4$. Comparable relaxations obtain with Richardson–extrapolated midpoint (effective $q = 4$) or higher-order Newton–Cotes/Gaussian rules.

Proof. Write

$$\hat{\beta}_{\text{avg}} - \Gamma = \underbrace{2h\sum_{b=1}^{B}\Big(\hat{\beta}_1(c_b) - m'(c_b)\Big)}_{\text{local-polynomial term}} + \underbrace{\Big\{2h\sum_{b=1}^{B} m'(c_b) - \int_0^1 m'(u)\,du\Big\}}_{\text{midpoint (quadrature) term}}.$$

Bias. By Lemma 4,

$$2h\sum_{b=1}^{B}\Big(\hat{\beta}_1(c_b) - m'(c_b)\Big) = \frac{h^p}{(p+1)!}\Big(\int_{-1}^{1} u^{p+1}K^*(u)\,du\Big)\Big[m^{(p)}(1) - m^{(p)}(0)\Big] + o(h^p).$$

By Lemma 3 with $g = m'$,

$$2h\sum_{b=1}^{B} m'(c_b) - \int_0^1 m'(u)\,du = -\frac{h^2}{6}\Big[m''(1) - m''(0)\Big] + O(h^4).$$

Hence:

- If $p > 2$, then $h^p = o(h^2)$ and the leading bias is the midpoint term:

$$\text{Bias} = -\frac{h^2}{6}\big[m''(1) - m''(0)\big], \qquad \big(\hat{\beta}_{\text{avg}} - \Gamma - \text{Bias}\big) = o(h^2).$$

- If $p = 2$, the two h^2 contributions add:

$$\text{Bias} = \frac{h^2}{6}\Big(\int_{-1}^{1} u^3K^*(u)\,du - 1\Big)\big[m''(1) - m''(0)\big], \qquad \big(\hat{\beta}_{\text{avg}} - \Gamma - \text{Bias}\big) = O(h^4).$$

Stochastic term and variance. Let $w_{N,i}$ be the effective weights from Lemma 2, so that

$$\hat{\beta}_{\text{avg}} - \mathbb{E}[\hat{\beta}_{\text{avg}} \mid X] = \sum_{i=1}^{N} w_{N,i}\,\varepsilon_i.$$

By Lemma 5,

$$v_N^2 := \text{Var}\Big(\sum_{i=1}^{N} w_{N,i}\varepsilon_i \;\Big|\; X\Big) = \frac{2}{Nh^2}\Big(\int_{-1}^{1} K^{*2}(u)\,du\Big)\Big(\int_0^1 \frac{\sigma^2(u)}{f(u)}\,du\Big) + o\Big(\tfrac{1}{Nh^2}\Big).$$

Therefore with $\alpha_N = \sqrt{Nh^2}$,

$$\alpha_N^2 v_N^2 \;\to\; \Omega^2 := 2\left(\int_{-1}^{1} K^{*2}(u)\,du\right)\left(\int_0^1 \frac{\sigma^2(u)}{f(u)}\,du\right) \quad \text{in probability.}$$

CLT. The Lyapunov condition for the triangular array $\{w_{N,i}\varepsilon_i\}$ holds by Lemma 2, so

$$\frac{\alpha_N}{\Omega}\left(\hat{\beta}_{\text{avg}} - \mathbb{E}[\hat{\beta}_{\text{avg}} \mid X]\right) \overset{d}{\Rightarrow} \mathcal{N}(0,1) \quad \text{conditionally on } X.$$

Combine this with the bias decomposition and Slutsky's lemma to obtain

$$\sqrt{Nh^2}\big(\hat{\beta}_{\text{avg}} - \Gamma - \text{Bias}\big) \overset{d}{\Rightarrow} \mathcal{N}\Big(0,\ 2\int K^{*2} \int \sigma^2/f\Big).$$

Finally, the residual deterministic term after subtracting Bias satisfies

$$R_N = \begin{cases} O(h^4) + o(h^2), & p = 2, \\ O(h^p) + O(h^4) + o(h^p), & p > 2, \end{cases}$$

so

$$\sqrt{Nh^2}\, R_N = \begin{cases} O(\sqrt{N}h^6) + o(\sqrt{N}h^3), & p = 2, \\ O(\sqrt{N}h^{p+1}) + O(\sqrt{N}h^6) + o(\sqrt{N}h^{p+1}), & p > 2, \end{cases} = o(1)$$

by Assumption 5. This completes the proof. □

The proof mirrors the estimator's decomposition. Lemma 2 yields a CLT for the single linear form in the errors defined by the effective weights (no per–bin CLTs are needed.) Lemma 5 translates the weight geometry into the product–form variance $2\int K^{*2} \cdot \int \sigma^2/f$, which is free of the unknown regression function m. On the deterministic side, Lemma 4 shows that the aggregate local–polynomial error reduces to a boundary term, while Lemma 3 converts the midpoint quadrature error into a boundary expression. Together these yield the leading h^2 bias used for centering in Theorem 1.

3.1 Comparison with Level Estimation at the Boundaries

In this section, we compare the bias and variance of our proposed estimator to an alternative approach that estimates δ in (2) by directly estimating $m(1)$ and $m(0)$ using local polynomial regression at the boundaries and subtracting the results. This boundary-based approach is analogous to the commonly used regression discontinuity design (RDD) estimators, which rely on estimating levels on either side of the threshold and calculating their difference.

From Fan and Yao (2005), the asymptotic behavior of the boundary-based estimator is:

$$\sqrt{Nh}\left[\hat{\delta}_{\text{levels}} - \delta - \frac{h^{p+1}}{(p+1)!}\left(m^{(p+1)}(1^-)\int_{-\infty}^{c} u^{p+1}K^*_{1-c}(u)\,du - m^{(p+1)}(0^+)\int_{-c}^{\infty} u^{p+1}K^*_c(u)\,du\right)\right]$$
$$\xrightarrow{d} \mathcal{N}\left(0, \int_{-\infty}^{c} K^*_{1-c}(u)^2\,du\frac{\sigma^2(1^-)}{f(1^-)} + \int_{-c}^{\infty} K^*_c(u)^2\,du\frac{\sigma^2(0^+)}{f(0^+)}\right), \tag{4}$$

where $\hat{\delta}_{\text{levels}} = \hat{m}(1-ch) - \hat{m}(0+ch)$ for a small positive constant c, and $K^*_c(u) = e_2^T S_c^{-1}(1, u, \ldots, u^p)^T K(u)$. Here, S_c is defined similarly to S, replacing μ_j with $\mu_{j,c} = \int_{-c}^{\infty} u^j K(u)\,du$, while $K^*_{1-c}(u)$ and S_{1-c} are defined analogously, replacing μ_j with $\mu_{j,1-c} = \int_{-\infty}^{c} u^j K(u)\,du$.

From equation (4), compared to our proposed estimator in theorem (1) with the same h and p, the asymptotic bias and variance of the boundary-based estimator decrease at faster rates. Consequently, the boundary estimator $\hat{\delta}_{\text{levels}}$ is asymptotically more efficient, achieving lower bias and variance.

However, when data near the boundaries is sparse or the outcome variance is high in those regions, the integral-based estimator $\hat{\beta}_{\text{avg}}$ may be informative in small samples. By leveraging data from the entire interval, it mitigates the instability typically associated with boundary-based estimates. Additionally, the theoretical variance constants differ between the two estimators, making it nontrivial to determine a priori whether the integral-based estimator will yield a higher or lower variance compared to the level-based estimator in small samples.

However, the main advantage of our estimator is that it is asymptotically independent from the standard estimator based on the level difference (Lemma 1.) The boundary estimator $\hat{\delta}_{\text{levels}}$ uses data close to the boundaries (within intervals of length h), while the integral-based estimator $\hat{\beta}_{\text{avg}}$ relies on data from the entire unit interval. Consequently, the overlap in data usage between the two estimators diminishes as $h \to 0$. Importantly, this independence allows for combining the two estimators using variance-weighted averages to construct an estimator that is more efficient than either one individually. We demonstrate this approach in Section 5.

Lemma 1 (Asymptotic independence via overlap bound)**.** *Let $\hat{\beta}_{\text{avg}}$ be the integrated-derivative estimator with bandwidth h as in Theorem 1, and write*

$$\hat{\beta}_{\text{avg}} - \mathbb{E}[\hat{\beta}_{\text{avg}} \mid X] = \sum_{i=1}^{N} w_i \, \varepsilon_i, \qquad \max_i |w_i| \le \frac{C}{Nh}, \quad \sum_{i=1}^{N} w_i^2 = \Theta_p\Big(\frac{1}{Nh^2}\Big).$$

Let the boundary levels at 0 and 1 be estimated by one-sided degree-p local polynomials with asymmetric equivalent kernels and bandwidths $h_0 = c_0 h$ (at 0) and $h_1 = c_1 h$ (at 1), where $0 < c_{\min} \le c_0, c_1 \le c_{\max} < \infty$:

$$\hat{\delta}_{\text{levels}} = \hat{m}(1^-) - \hat{m}(0^+), \qquad \hat{\delta}_{\text{levels}} - \mathbb{E}[\hat{\delta}_{\text{levels}} \mid X] = \sum_{i=1}^{N} v_i \, \varepsilon_i,$$

with

$$\max_i |v_i| \le \frac{C}{Nh}, \qquad \sum_{i=1}^{N} v_i^2 = \Theta_p\Big(\frac{1}{Nh}\Big),$$

and $v_i \neq 0$ only if $X_i \in [0, A_0 h] \cup [1 - A_1 h, 1]$, where $A_0, A_1 > 0$ depend only on the one-sided equivalent kernels and on c_0, c_1.

Under Assumptions 1–6 (in particular $Nh \to \infty$ and $h \to 0$ and Assumption 2), conditionally on X,

$$\mathrm{Cov}\big(\hat{\beta}_{\text{avg}}, \hat{\delta}_{\text{levels}} \mid X\big) \;=\; O_p\Big(\frac{A_0 f(0+) + A_1 f(1-)}{Nh}\Big),$$

and

$$\mathrm{Corr}\big(\hat{\beta}_{\mathrm{avg}}, \hat{\delta}_{\mathrm{levels}} \mid X\big) \;=\; O_p\Big(\big(A_0 f(0+) + A_1 f(1-)\big)\, h^{1/2}\Big) \;\longrightarrow\; 0.$$

In particular, together with the marginal CLTs for $\hat{\beta}_{\mathrm{avg}}$ (Theorem 1) and the standard boundary local-polynomial CLT for $\hat{\delta}_{\mathrm{levels}}$, this implies that the jointly scaled vector is asymptotically bivariate normal with a diagonal covariance, i.e. the estimators are asymptotically independent.

Proof. Write the centered forms

$$S_\beta = \hat{\beta}_{\mathrm{avg}} - \mathbb{E}[\hat{\beta}_{\mathrm{avg}} \mid X] = \sum_{i=1}^{N} w_i \varepsilon_i, \qquad S_\delta = \hat{\delta}_{\mathrm{levels}} - \mathbb{E}[\hat{\delta}_{\mathrm{levels}} \mid X] = \sum_{i=1}^{N} v_i \varepsilon_i.$$

Then

$$\mathrm{Cov}(S_\beta, S_\delta \mid X) = \sum_{i=1}^{N} w_i v_i\, \sigma^2(X_i).$$

By construction of the one-sided boundary fits, $v_i \neq 0$ only when $X_i \in \mathcal{O}_h := [0, A_0 h] \cup [1 - A_1 h, 1]$. Since f is bounded with finite one-sided limits and has no atoms at $\{0, 1\}$,

$$\mathbb{P}(X_i \in \mathcal{O}_h) = A_0 h\, f(0+) + A_1 h\, f(1-) + o(h),$$

so

$$\#\{i : X_i \in \mathcal{O}_h\} = N\big(A_0 f(0+) + A_1 f(1-)\big) h + o_p(Nh) = O_p\big(Nh\,(A_0 f(0+) + A_1 f(1-))\big).$$

Using $\max_i |w_i| \le C/(Nh)$, $\max_i |v_i| \le C/(Nh)$, and boundedness of $\sigma^2(\cdot)$,

$$\big|\mathrm{Cov}(S_\beta, S_\delta \mid X)\big| \le \sup_u \sigma^2(u) \sum_{i: X_i \in \mathcal{O}_h} |w_i v_i| \le C \cdot \#\{i : X_i \in \mathcal{O}_h\} \cdot \frac{C}{Nh} \cdot \frac{C}{Nh} = O_p\Big(\frac{A_0 f(0+) + A_1 f(1-)}{Nh}\Big).$$

For the variances, by the standard expansions,

$$\mathrm{Var}(S_\beta \mid X) = \Theta_p\Big(\frac{1}{Nh^2}\Big), \qquad \mathrm{Var}(S_\delta \mid X) = \Theta_p\Big(\frac{1}{Nh}\Big).$$

Hence

$$\mathrm{Corr}(S_\beta, S_\delta \mid X) = \frac{\mathrm{Cov}(S_\beta, S_\delta \mid X)}{\sqrt{\mathrm{Var}(S_\beta \mid X)\,\mathrm{Var}(S_\delta \mid X)}} =$$
$$\frac{O_p\big((A_0 f(0+) + A_1 f(1-))/(Nh)\big)}{\sqrt{(1/(Nh^2))\,(1/(Nh))}} = O_p\Big((A_0 f(0+) + A_1 f(1-))\,h^{1/2}\Big) \to 0.$$

This proves the stated covariance and correlation bounds.

Finally, the marginal CLTs for the linear forms S_β and S_δ (Theorem 1 and the boundary local-polynomial CLT, under the same moment and bandwidth conditions) imply that the jointly normalized vector is asymptotically bivariate normal. Since the off-diagonal term vanishes by the lemma, the limit covariance is diagonal; thus the estimators are asymptotically independent. □

4 The Estimator for General Cases

This section discusses the estimation of the weighted average of expected outcomes at thresholds for a single segment ($z \in [0,1]$) with arbitrary weights:

$$\delta_w = \omega_1 m(1^-) - \omega_0 m(0^+).$$

By the product rule $(\pi m)' = \pi' m + \pi m'$ and integration by parts, we have the representation

$$\delta_w \;=\; \int_0^1 \pi(u)\, m'(u)\, du \;+\; \int_0^1 \pi'(u)\, m(u)\, du, \tag{5}$$

for any function $\pi : [0,1] \to \mathbb{R}$ with $\pi(1^-) = \omega_1$ and $\pi(0^+) = \omega_0$.

Estimator. Let $\{I_b\}_{b=1}^B$ be the disjoint bins of width $2h$, with midpoints $c_b = (2b-1)h$. For each c_b, let $\hat\beta_1(c_b)$ be the degree-p local-polynomial estimator of $m'(c_b)$ from Theorem 1. Given a (possibly data-chosen) quadratic weight function $\pi(\cdot)$ with $\pi(1^-) = \omega_1$ and $\pi(0^+) = \omega_0$, define

$$\hat{\delta}_w \ := \ 2h \sum_{b=1}^{B} \pi(c_b)\, \hat{\beta}_1(c_b) \ + \ \int_0^1 \pi'(u)\, \hat{m}(u)\, du, \tag{6}$$

where $\hat{m}$ is any estimator satisfying $\int_0^1 q(u)\, \hat{m}(u)\, du = \int_0^1 q(u)\, m(u)\, du + O_p(N^{-1/2})$ for each fixed quadratic q.[5]

Theorem 2 (Weighted integrated derivative)**.** *Suppose the assumptions of Theorem 1 hold (including the bandwidth condition) with local-polynomial degree $p \geq 2$. Let $\pi(\cdot)$ be a quadratic polynomial with $\pi(1^-) = \omega_1$, $\pi(0^+) = \omega_0$ and $\int_0^1 \pi'(u)\, \hat{m}(u)\, du = 0$. Then, conditional on X,*

$$\sqrt{Nh^2}\left(\hat{\delta}_w - \delta_w - \mathrm{Bias}_\pi\right) \ \overset{d}{\Rightarrow} \ \mathcal{N}\Big(0,\ 2\int_{-1}^{1} K^{*2}(u)\, du \ \int_0^1 \frac{\pi^2(u)\, \sigma^2(u)}{f(u)}\, du\Big),$$

where the bias satisfies

$$\mathrm{Bias}_\pi \ = \ \begin{cases} -\dfrac{h^2}{6}\Big[(\pi m')'(1) \ - \ (\pi m')'(0)\Big] \ + \ o(h^2), & \textit{if the local-poly degree } > 2, \\ \dfrac{h^2}{6}\Big(\displaystyle\int_{-1}^{1} u^3 K^*(u)\, du - 1\Big)\Big[(\pi m')'(1) \ - \ (\pi m')'(0)\Big] \ + \ o(h^2), & \textit{if the degree } = 2, \end{cases}$$

with $(\pi m')'(u) = \pi'(u)\, m'(u) + \pi(u)\, m''(u)$.

Proof. Write

$$\hat{\delta}_w - \delta_w = \underbrace{2h \sum_b \pi(c_b)\big(\hat{\beta}_1(c_b) - m'(c_b)\big)}_{A} \ +$$
$$\underbrace{\Big\{2h \sum_b \pi(c_b) m'(c_b) - \int_0^1 \pi(u) m'(u)\, du\Big\}}_{B} \ + \ \underbrace{\int_0^1 \pi'(u)\, \big(\hat{m}(u) - m(u)\big)\, du}_{C}.$$

Term A is handled by Lemma 4 with the factor $\pi(c_b)$ inside the sum: when the

[5] If $\pi(\cdot)$ is selected from the quadratic family by imposing the sample orthogonality $\int_0^1 \pi'(u)\, \hat{m}(u)\, du = 0$ together with the endpoint constraints, then the second term in (6) is zero by construction, and $\hat{\delta}_w = 2h \sum_b \pi(c_b) \hat{\beta}_1(c_b)$ in practice.

local-polynomial degree is > 2, $A = o(h^2)$; when it is 2, A contributes an h^2 term with coefficient $\int u^3 K^*(u)\,du$. Term B is the midpoint quadrature error with $g(u) = \pi(u)m'(u)$; Lemma 3 gives

$$B \;=\; -\frac{h^2}{6}\Big[(\pi m')'(1) - (\pi m')'(0)\Big] \;+\; O(h^4).$$

Term C is a sample moment error of the form $\int q(u)\,\big(\hat{m}(u) - m(u)\big)\,du$ with $q = \pi'$ quadratic; by assumption $C = O_p(N^{-1/2})$, hence

$$\sqrt{Nh^2}\,C \;=\; \sqrt{h^2}\,O_p(1) \;=\; o_p(1) \quad \text{as } h \to 0.$$

Collecting A and B yields the stated Bias_π (the $O(h^4)$ and $o(h^2)$ remainders are negligible under $\sqrt{Nh^2}$), and the variance and CLT follow by repeating the argument of Lemma 5 with the extra factor $\pi(c_b)$ and invoking Lemma 2. This proves the claim. $\square$

Remark 1 (Data-chosen π is first-order negligible). If $\pi(\cdot)$ is selected from the quadratic family by the constraints $\pi(1^-) = \omega_1$, $\pi(0^+) = \omega_0$ and the sample orthogonality $\int_0^1 \pi'(u)\,\hat{m}(u)\,du = 0$, then the second term of (6) is exactly zero in implementation. Moreover, letting π^* denote the population solution with $\int_0^1 \pi^{*\prime}(u)m(u)\,du = 0$, one has $\|\hat{\pi} - \pi^*\|_\infty = O_p(N^{-1/2})$ and

$$\sqrt{Nh^2}\,\Big(\hat{\delta}_w(\hat{\pi}) - \hat{\delta}_w(\pi^*)\Big) = o_p(1).$$

Thus, to first order, the stochastic choice of π does not affect the limit law in Theorem 2.

The estimators for each interval difference in eq. 1 use disjoint observations between thresholds. Under the fixed-design, with independent errors, the segment-level estimators are independent. Each estimator is asymptotically normal, so any finite linear combination is also asymptotically normal with variance being the weighted sum of the segment variances. We therefore focus on a single interior segment in the next section.

5 Simulations and Empirical Demonstration

We compare three estimators of the boundary difference $\delta = m(1) - m(0)$: (i) *Boundary local linear* (LL), $\hat{\delta}_{\text{bnd}} = \hat{m}(1) - \hat{m}(0)$, using one–sided local linear fits at 0 and 1 with triangular kernel $K(u) = \max\{0, 1 - |u|\}$; (ii) an *Integrated–Derivative* estimator that rewrites $\delta = \int_0^1 m'(x)\,dx$ and estimates local slopes $m'(c_b)$ at midpoints c_b via local quadratic regression, aggregated by composite midpoint quadrature; and (iii) a *Combined* estimator, $\hat{\delta}_{\text{comb}} = w_{\text{bnd}}\hat{\delta}_{\text{bnd}} + w_{\text{int}}\hat{\delta}_{\text{int}}$, with inverse–variance weights computed once per setting from asymptotic variance formulas.

We partition $[0, 1]$ into B equal bins ($\Delta = 1/B$) and set $h_0 = \Delta/2$ ($\kappa = 1$). For bin b, we run a local quadratic regression of Y on $(1, X - c_b, (X - c_b)^2)$ with triangular weights on $|X - c_b| \leq h_0$, and take the slope $\hat{\beta}_1(c_b)$ as $\widehat{m'}(c_b)$. The integrated estimator is

$$\hat{\delta}_{\text{int}} = \sum_{b=1}^{B} \Delta\, \hat{\beta}_1(c_b).$$

We use the sample equivalent variance formulas:

$$\Omega^2 \;=\; 2C_K \sum_{b=1}^{B} \Delta\, \frac{\sigma_b^2}{f_b}, \qquad \widehat{\text{SE}}_{\text{int}} = \sqrt{\widehat{\Omega}^2/(nh_0^2)}, \quad \widehat{\Omega}^2 = 2C_K \sum_b \Delta\, \frac{\widehat{\sigma}_b^2}{\widehat{f}_b},$$

where $C_K = \int K^*(u)^2 du$ is the derivative equivalent–kernel constant for local quadratic (computed numerically), $\widehat{f}_b$ is a local density estimate based on counts in $[c_b - h_0, c_b + h_0]$, and $\widehat{\sigma}_b^2$ is the local residual MSE from the derivative fit. For the boundary LL estimator, the plug–in variance uses the one–sided equivalent–kernel constant $C_0 = \int K_c^*(u)^2 du$ with local residual MSEs and boundary densities.

We fix a draw of $X_1, \ldots, X_N$ from either Uniform$[0, 1]$ or a truncated normal $\mathcal{N}(0.5, 0.3^2)$ clipped to $[0, 1]$ and generate $Y_i = m(X_i) + \varepsilon_i$ with $\varepsilon_i \sim \mathcal{N}(0, \sigma^2(X_i))$, where $\sigma(x) = 1$ (homoskedastic) or $\sigma^2(x) = 1 + x$ (heteroskedastic). We repeat $nsim = 500$ times by redrawing the sample, conditional on X. Asymptotic variances are computed from the realized X to form fixed inverse–variance weights for the combined estimator. We report bias, standard deviation, RMSE, and the empirical 95% coverage of the feasible Wald CI for $\hat{\delta}_{\text{int}}$.

We consider $m(x) = 2x + x^2$ (polynomial; $\delta = 3$) and $m(x) = (1 - \cos(\pi x))/2$ (cosine; $\delta = 1$). The former makes composite midpoint quadrature exact; the latter has zero endpoint slopes and substantial curvature, stressing boundary bias and quadrature. In our baseline with $B = 10$ and $h_0 = \Delta/2$, the bias is small, so differences across DGPs mainly reflect design (f, σ^2) rather than $m(\cdot)$. The simulation codes can be accessed Here.

The simulation results in Table 1 mirror the large-sample theory with accuracy. The integrated estimator delivers coverage rates tightly clustered around 0.94–0.95, precisely in line with the feasible CLT and the variance expression $Var(\hat{\delta}_{\text{int}}) = \Omega^2/(Nh_0^2)$. The boundary estimator, as expected, follows a faster convergence rate, the familiar $(Nh_0)^{-1}$ rate.

More importantly, the simulations demonstrate the payoff from combining the two approaches. The inverse–variance weighted estimator consistently lowers RMSE relative to the boundary method alone. Under uniform designs the improvements are modest, on the order of two to four percent, but in more irregular designs, such as the truncated normal case, the gains become substantial, reaching seven to nine percent. This is precisely the scenario highlighted by the theory: because the two estimators rely on distinct information sets, boundary levels versus interior slopes, their variances can be efficiently pooled, and with negligible covariance the combined estimator achieves strictly greater precision.

The results also highlight robustness to different data-generating processes. Polynomial and cosine designs yield nearly identical performance at $B = 10$ and $h_0 = \Delta/2$, reflecting that the asymptotic variance depends on the design distribution and noise rather than the functional form of $m(\cdot)$.[6] Together, these patterns underscore that the integrated and combined estimators not only con-

[6] Only under deliberately coarse grids does the cosine's quadrature bias become visible, and such bias can be eliminated entirely by switching to Simpson's rule.

Scenario	Boundary Bias	Boundary SD	Boundary RMSE	Integrated Bias	Integrated SD	Integrated RMSE	Integrated Cov.95	Combined Bias	Combined SD	Combined RMSE
Panel A: Polynomial DGP $m(x) = 2x + x^2$										
Uniform – Homo	−0.001	0.345	0.345	−0.048	1.012	1.012	0.940	−0.005	0.338	0.338
Uniform – Het	0.001	0.404	0.403	−0.084	1.160	1.162	0.954	−0.006	0.388	0.388
Trunc. Normal (sd=0.3) – Homo	−0.013	0.608	0.607	0.010	1.021	1.020	0.946	−0.009	0.550	0.550
Trunc. Normal (sd=0.3) – Het	0.014	0.701	0.701	0.022	1.390	1.389	0.942	0.016	0.651	0.650
Uniform – Homo [B=8]	0.009	0.275	0.275	−0.034	0.787	0.787	0.944	0.004	0.266	0.266
Uniform – Homo [B=12]	−0.009	0.341	0.341	0.116	1.155	1.160	0.942	0.002	0.336	0.336
Uniform – Homo [n=1000]	−0.002	0.475	0.474	−0.023	1.438	1.437	0.948	−0.004	0.467	0.466
Uniform – Homo [n=4000]	0.010	0.221	0.221	−0.044	0.701	0.702	0.952	0.006	0.217	0.217
Panel B: Cosine DGP $m(x) = \frac{1-\cos(\pi x)}{2}$										
Uniform – Homo	0.000	0.345	0.345	−0.045	1.012	1.012	0.940	−0.004	0.338	0.338
Uniform – Het	0.003	0.404	0.403	−0.081	1.160	1.162	0.954	−0.005	0.388	0.388
Trunc. Normal (sd=0.3) – Homo	−0.012	0.608	0.607	0.012	1.021	1.020	0.946	−0.008	0.550	0.550
Trunc. Normal (sd=0.3) – Het	0.016	0.701	0.701	0.025	1.390	1.389	0.942	0.017	0.651	0.650
Uniform – Homo [B=8]	0.011	0.275	0.275	−0.031	0.787	0.787	0.946	0.006	0.266	0.266
Uniform – Homo [B=12]	−0.009	0.341	0.341	0.118	1.155	1.160	0.942	0.003	0.336	0.336
Uniform – Homo [n=1000]	−0.001	0.475	0.474	−0.021	1.438	1.437	0.948	−0.002	0.467	0.466
Uniform – Homo [n=4000]	0.012	0.221	0.221	−0.042	0.701	0.701	0.952	0.007	0.217	0.217

Notes. Each row reports Monte Carlo results over $nsim = 500$ replications with $N = 2000$ as the default. X designs: Uniform on $[0, 1]$ or truncated normal $\mathcal{N}(0.5, 0.3^2)$ clipped to $[0, 1]$ ("Trunc. Normal"). Errors are homoskedastic ($\sigma(x) = 1$) or heteroskedastic ($\sigma^2(x) = 1 + x$). We partition $[0, 1]$ into B equal bins ($\Delta = 1/B$) and set $h_0 = \Delta/2$ ($\kappa = 1$). The *integrated* estimator fits local *quadratic* regressions of m around midpoints using the triangular kernel $K(u) = \max\{0, 1 - |u|\}$ to estimate $m'(x)$ and aggregates via midpoint quadrature; its feasible SE uses

$$\widehat{\Omega}^2 \;=\; 2\,C_K \sum_{b=1}^{B} \Delta\, \frac{\widehat{\sigma}_b^2}{\widehat{f}_b}, \qquad \widehat{\mathrm{SE}}_{\mathrm{Int}} \;=\; \sqrt{\widehat{\Omega}^2/(nh_0^2)}.$$

Here $C_K = \int K^*(u)^2\,du$ is the derivative equivalent–kernel constant for local quadratic (computed numerically); $\widehat{f}_b$ is a local density estimate (bin counts over $2h_0$) and $\widehat{\sigma}_b^2$ a local residual MSE from the derivative fit. The *boundary* estimator is local linear at 0 and 1 with the same base kernel; its plug–in variance uses the one–sided equivalent–kernel constants

$$C_{K,0+} \;=\; \int_{-c}^{\infty} K_c^*(u)^2\,du, \qquad C_{K,1-} \;=\; \int_{-\infty}^{c} K_{1-c}^*(u)^2\,du,$$

together with local residual MSEs and boundary densities at 0^+ and 1^-. (In our implementation we set the offset $c = 1$ and use the same bandwidth multiple at both boundaries.) The *combined* estimator is inverse–variance weighted using design–adaptive asymptotic variances computed from the realized X. "Bias" is empirical mean minus truth; "SD" is the Monte Carlo standard deviation; "Cov.95" is empirical 95% coverage of the integrated estimator's feasible normal CI; "RMSE" is $\sqrt{\mathbb{E}[(\hat{\delta} - \delta)^2]}$.

Table 1: Simulation results: Integrated–Derivative vs Boundary (two DGPs)

form to theory but also deliver practical efficiency gains in finite samples.

6 Discussions and Conclusions

This paper introduced an integrated–derivative estimator for regression discontinuity designs with multiple thresholds. By exploiting interior variation between cutoffs, the estimator complements conventional boundary estimators, is asymptotically independent from them, and can be combined for efficiency gains. Our theoretical results establish a central limit theorem and demonstrate the potential for precision improvements, which are supported by simulation evidence.

The approach also has limitations. A key assumption is that the conditional expectation function is smooth between thresholds, so that derivative estimates can be consistently recovered from the interior. In applications where the running variable may induce unobserved interventions or latent discontinuities between observed cutoffs, this continuity assumption may be restrictive. Moreover, while asymptotic independence ensures efficiency gains in large samples, in finite samples the integrated estimator alone can be noisy, and gains from combination depend on the design and data availability near boundaries. Finally, the method focuses on settings with known and ordered thresholds; extensions to cases with stochastic or endogenous cutoffs remain open for future research. These caveats highlight that our estimator is best viewed as a complement to, rather than a replacement for, conventional boundary-based RD methods.

References

Almond, D. and J. J. Doyle (2011). After midnight: A regression discontinuity design in length of postpartum hospital stays. *3*(3), 1–34.

Bertanha, M. (2020, September). Regression discontinuity design with many thresholds. *Journal of Econometrics 218*(1), 216–241.

Calonico, S., M. D. Cattaneo, and M. H. Farrell (2020). Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. *The Econometrics Journal 23*(2), 192–210.

Calonico, S., M. D. Cattaneo, M. H. Farrell, F. Palomba, and R. Titiunik (2025). Treatment effect heterogeneity in regression discontinuity designs. *arXiv preprint arXiv:2503.13696*.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *82*(6), 2295–2326. Publisher: Wiley Online Library.

Cattaneo, M. D., L. Keele, R. Titiunik, and G. Vazquez-Bare (2016). Interpreting regression discontinuity designs with multiple cutoffs. *78*(4), 1229–1248. Publisher: The University of Chicago Press.

Fan, J. and Q. Yao (2005). *Nonlinear Time Series: Nonparametric and Parametric Methods* (1 ed.). Springer Series in Statistics. New York: Springer New York, NY. Springer Book Archive.

Fredriksson, P. and B. Öckert (2014). Life-cycle effects of age at school start. *124*(579), 977–1004.

Hahn, J., P. Todd, and W. Van der Klaauw (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica 69*(1), 201–209.

Imbens, G. and S. Wager (2019). Optimized regression discontinuity designs. *Review of Economics and Statistics 101*(2), 264–278.

Persson, P., X. Qiu, and M. Rossin-Slater (2021). Family spillover effects of marginal diagnoses: The case of ADHD.

Appendix

A Proofs

Lemma 2 (Convergence in Distribution)**.** *Under Assumptions 1–6, for each sample size N, consider the triangular array*

$$\{w_{N,i}\,\varepsilon_i : i = 1, \ldots, N\},$$

where $w_{N,i}$ is the effective weight for the outcome of the ith observation Y_i when using local-polynomial first-derivative estimates, with bin width $2h = 2h(N)$. Define

$$T_N = \sum_{i=1}^{N} w_{N,i}\,\varepsilon_i, \qquad v_N^2 = \operatorname{Var}\bigl(T_N \mid (X_1, \ldots, X_N)\bigr) = \sum_{i=1}^{N} w_{N,i}^2\,\sigma^2(X_i).$$

Suppose there exists a deterministic sequence $\alpha_N > 0$ such that

$$\alpha_N^2\, v_N^2 \xrightarrow[N\to\infty]{p} \Omega^2 > 0.$$

Then the Lyapunov condition holds conditionally on $\{X_i\}$, i.e.,

$$\lim_{N\to\infty} \frac{1}{(v_N^2)^{1+\delta/2}} \sum_{i=1}^{N} \mathbb{E}\Bigl[|w_{N,i}\varepsilon_i|^{2+\delta} \mid (X_1, \ldots, X_N)\Bigr] = 0,$$

and thus, by the triangular-array Lyapunov Central Limit Theorem,

$$\frac{\alpha_N T_N}{\sqrt{\Omega^2}} \xrightarrow{d} \mathcal{N}(0, 1).$$

Proof. For each bin center c_b, the local-polynomial fit for $m'(c_b)$ yields weights for observation i in bin I_b:

$$w_{i,b} = \frac{1}{Nh^2 f(c_b)}\, K^*\left(\frac{X_i - c_b}{h}\right) \left\{1 + O_p\Bigl(h + \sqrt{\tfrac{\log N}{Nh}}\Bigr)\right\}.$$

Since each X_i belongs to exactly one bin $I_{b(i)}$, the integrated-estimator weight is

$$w_{N,i} = 2h\, w_{i,b(i)} = \frac{2}{Nh\, f(c_{b(i)})}\, K^*\left(\frac{X_i - c_{b(i)}}{h}\right) \left\{1 + O_p\Big(h + \sqrt{\tfrac{\log N}{Nh}}\Big)\right\}.$$

Hence, using the boundedness of K^* and positivity of $f(\cdot)$,

$$|w_{N,i}| \leq \frac{C_1}{Nh}\left(1 + O_p\left(h + \sqrt{\frac{\log N}{Nh}}\right)\right), \quad \text{and} \quad \sum_{i=1}^{N} w_{N,i}^2 = O_p\left(\frac{1}{Nh^2}\right).$$

Conditional on X, the weights $\{w_{N,i}\}$ are deterministic and the summands $\{w_{N,i}\varepsilon_i\}$ are independent with mean zero.

Verification of the Lyapunov condition. By Assumption 6, there exist $\delta > 0$ and $C < \infty$ such that uniformly in i,

$$\mathbb{E}[|\varepsilon_i|^{2+\delta} \mid X_i] \leq C.$$

Thus,

$$\sum_{i=1}^{N} \mathbb{E}[|w_{N,i}\varepsilon_i|^{2+\delta} \mid X_1, \ldots, X_N] \leq C \sum_{i=1}^{N} |w_{N,i}|^{2+\delta}.$$

Using the bounds above,

$$\sum_{i=1}^{N} |w_{N,i}|^{2+\delta} = O_p\left(\frac{1}{N^{1+\delta}h^{2+\delta}}\right).$$

Since $\sigma^2(\cdot)$ is bounded,

$$v_N^2 = \sum_{i=1}^{N} w_{N,i}^2\, \sigma^2(X_i) = O_p\left(\frac{1}{Nh^2}\right),$$

and hence

$$(v_N^2)^{1+\delta/2} = O_p\left(\frac{1}{N^{1+\delta/2}h^{2+\delta}}\right).$$

Therefore,
$$\frac{\sum_{i=1}^{N}|w_{N,i}|^{2+\delta}}{(v_N^2)^{1+\delta/2}} = O_p\left(\frac{1}{N^{\delta/2}}\right) \xrightarrow[N\to\infty]{p} 0.$$

This verifies the Lyapunov condition. By the triangular-array Lyapunov theorem, conditionally on $(X_1, \ldots, X_N)$,

$$\frac{T_N}{\sqrt{v_N^2}} = \frac{\sum_{i=1}^{N} w_{N,i}\varepsilon_i}{\sqrt{\sum_{i=1}^{N} w_{N,i}^2 \sigma^2(X_i)}} \xrightarrow{d} \mathcal{N}(0,1).$$

Given $\alpha_N^2 v_N^2 \xrightarrow{p} \Omega^2$, Slutsky's lemma yields

$$\frac{\alpha_N T_N}{\sqrt{\Omega^2}} \xrightarrow{d} \mathcal{N}(0,1).$$

This completes the proof. □

Lemma 3 (Midpoint Rule Approximation Error)**.** *Let $g \in C^4([0,1])$. Partition $[0,1]$ into B subintervals of length $2h = 1/B$, with midpoints $c_b = (2b-1)h$ for $b = 1, \ldots, B$. Define*

$$E \;:=\; \int_0^1 g(u)\,du \;-\; 2h\sum_{b=1}^{B} g(c_b).$$

Then
$$E \;=\; \frac{1}{3}\,h^3 \sum_{b=1}^{B} g''(c_b) \;+\; O(h^4).$$

Moreover, as $h \to 0$,

$$E \;=\; \frac{h^2}{6}\left[g'(1) - g'(0)\right] \;+\; O(h^4),$$

so in particular $E \to 0$.

Proof. Partition $[0,1]$ into B subintervals of width $2h$ with $2h\,B = 1$. Let $a_b = (b-1)\cdot 2h$ and $c_b = a_b + h$. On $[a_b, a_b + 2h]$,

$$\int_{a_b}^{a_b+2h} g(u)\,du = 2h\,g(c_b) + \frac{1}{3}\,g''(c_b)\,h^3 + \frac{1}{60}\,g^{(4)}(\eta_b)\,h^5,$$

for some $\eta_b \in [a_b, a_b + 2h]$. Summing gives

$$E \;=\; \frac{1}{3}\, h^3 \sum_{b=1}^{B} g''(c_b) \;+\; \frac{1}{60}\, h^5 \sum_{b=1}^{B} g^{(4)}(\eta_b).$$

Since $g^{(4)}$ is continuous (hence bounded) and $B = 1/(2h)$, $\frac{1}{60} h^5 \sum_{b=1}^{B} g^{(4)}(\eta_b) = O(h^4)$. Thus

$$E \;=\; \frac{1}{3}\, h^3 \sum_{b=1}^{B} g''(c_b) \;+\; O(h^4).$$

By the midpoint Riemann-sum error for $g'' \in C^2$,

$$2h \sum_{b=1}^{B} g''(c_b) \;=\; \int_0^1 g''(u)\, du + O(h^2) \;=\; g'(1) - g'(0) + O(h^2),$$

so $\sum_{b=1}^{B} g''(c_b) = \frac{1}{2h}[g'(1) - g'(0)] + O(h)$. Substituting,

$$E \;=\; \frac{1}{3} h^3 \Big(\frac{1}{2h} [g'(1) - g'(0)] + O(h) \Big) + O(h^4) = \frac{h^2}{6} [g'(1) - g'(0)] + O(h^4).$$

□

Lemma 4 (Aggregate Bias of Local Polynomial Estimator)**.** *Under Assumptions 1–6, let $\hat{\beta}_1(c_b)$ denote the degree-p local-polynomial estimator of $m'(c_b)$. Partition $[0, 1]$ into B subintervals of width $2h = 1/B$ with midpoints $c_b = (2b - 1)h$, and define*

$$R \;:=\; 2h \sum_{b=1}^{B} \hat{\beta}_1(c_b) \;-\; 2h \sum_{b=1}^{B} m'(c_b).$$

Let the aggregate bias *be $\mathcal{B}_N(h) := \mathbb{E}[R \mid X]$. Then*

$$\mathcal{B}_N(h) \;=\; \frac{2\, h^{p+1}}{(p+1)!} \left(\int_{-1}^{1} u^{p+1}\, K^*(u)\, du \right) \sum_{b=1}^{B} m^{(p+1)}(c_b) \;+\; r_N(h), \qquad (7)$$

where, with $\omega_{m^{(p+1)}}(t) := \sup_{|x-y| \le t} |m^{(p+1)}(x) - m^{(p+1)}(y)|$,

$$|r_N(h)| \;\le\; C\, h^p \Big\{ \omega_{m^{(p+1)}}(h) + h \Big\} \;=\; o(h^p) \qquad (h \to 0).$$

Moreover, by continuity of $m^{(p+1)}$,

$$2h \sum_{b=1}^{B} m^{(p+1)}(c_b) \;\to\; \int_0^1 m^{(p+1)}(u)\,du \;=\; m^{(p)}(1) - m^{(p)}(0),$$

hence

$$\mathcal{B}_N(h) \;=\; \frac{h^p}{(p+1)!} \left(\int_{-1}^{1} u^{p+1} K^*(u)\,du \right) \left[m^{(p)}(1) - m^{(p)}(0) \right] \;+\; o(h^p). \tag{8}$$

Here $K^*(\cdot)$ is the equivalent kernel defined in Assumption 4.

Proof. Fix an interior midpoint $c_b \in [h, 1-h]$. The standard equivalent-kernel expansion for local-polynomial derivative estimators (e.g. Fan and Yao, 2005) yields, conditionally on X,

$$\mathbb{E}[\hat{\beta}_1(c_b) \mid X] - m'(c_b) \;=\; \frac{h^p}{(p+1)!}\, m^{(p+1)}(c_b) \int_{-1}^{1} u^{p+1} K^*(u)\,du \;+\; h^p\,\rho_b(h),$$

where the remainder satisfies the bound

$$|\rho_b(h)| \;\le\; C\Big\{ \omega_{m^{(p+1)}}(h) + h \Big\}.$$

(The term $\omega_{m^{(p+1)}}(h)$ controls the within-window variation of $m^{(p+1)}$; the $+h$ term captures smooth design effects via f and is standard under Assumption 2.)

Summing over $b = 1, \ldots, B$ and multiplying by $2h$,

$$\mathcal{B}_N(h) = 2h \sum_{b=1}^{B} \Big(\mathbb{E}[\hat{\beta}_1(c_b) \mid X] - m'(c_b) \Big) = \frac{2\,h^{p+1}}{(p+1)!} \Big(\int u^{p+1} K^*(u)\,du \Big) \sum_{b=1}^{B} m^{(p+1)}(c_b) + 2h \sum_{b=1}^{B} h^p \rho_b(h).$$

Because $2h \sum_{b=1}^{B} 1 = 1$, the aggregate remainder satisfies

$$\Big| 2h \sum_{b=1}^{B} h^p \rho_b(h) \Big| \;\le\; h^p \sup_b |\rho_b(h)| \;\le\; C\,h^p \{ \omega_{m^{(p+1)}}(h) + h \} \;=\; o(h^p),$$

which proves (7). Finally, since $m^{(p+1)}$ is continuous on $[0,1]$, the midpoint

Riemann sums converge to the integral:

$$2h \sum_{b=1}^{B} m^{(p+1)}(c_b) \;\to\; \int_0^1 m^{(p+1)}(u)\, du = m^{(p)}(1) - m^{(p)}(0),$$

and substituting into (7) yields (8). ☐

Lemma 5 (Variance of the Estimator)**.** *Under Assumptions 1–6, let $\hat{\beta}_1(c_b)$ be the degree-p local-polynomial estimator of $m'(c_b)$. Partition $[0,1]$ into B subintervals of width $2h = 1/B$, with midpoints $c_b = (2b-1)\,h$ for $b = 1, \dots, B$, and define*

$$\hat{\beta}_{\text{avg}} \;:=\; 2h \sum_{b=1}^{B} \hat{\beta}_1(c_b).$$

Then, conditional on the design X,

$$\operatorname{Var}\big[\hat{\beta}_{\text{avg}}\big] \;=\; \frac{2}{N\,h^2} \int_{-1}^{1} K^{*2}(u)\, du \;\int_0^1 \frac{\sigma^2(u)}{f(u)}\, du \;+\; o\Big(\tfrac{1}{N\,h^2}\Big).$$

Proof. Because the kernel has support $[-1,1]$ and each local fit at c_b uses only observations with $|X_i - c_b| \le h$, while bins have width $2h$ and are disjoint, the windows $[c_b - h, c_b + h]$ coincide with bins. Thus, given X, the sets of observations used to form $\{\hat{\beta}_1(c_b)\}_{b=1}^{B}$ are disjoint; with independent errors, the estimators are independent conditionally on X. Hence

$$\operatorname{Var}\big[\hat{\beta}_{\text{avg}}\big] = \operatorname{Var}\Big(2h \sum_{b=1}^{B} \hat{\beta}_1(c_b)\Big) = 4h^2 \sum_{b=1}^{B} \operatorname{Var}\big[\hat{\beta}_1(c_b) \mid X\big].$$

For interior points $c_b \in [h, 1-h]$, a standard local-polynomial variance expansion (e.g. Fan and Yao, 2005) gives, *uniformly in b*,

$$\operatorname{Var}\big[\hat{\beta}_1(c_b) \mid X\big] = \frac{\sigma^2(c_b)}{N\,h^3\, f(c_b)} \int_{-1}^{1} K^{*2}(u)\, du \;+\; o\Big(\tfrac{1}{N\,h^3}\Big).$$

Summing over b and extracting $4h^2$,

$$\operatorname{Var}\big[\hat{\beta}_{\text{avg}}\big] = \frac{4}{N\,h}\Big(\int_{-1}^{1} K^{*2}(u)\,du\Big)\sum_{b=1}^{B}\frac{\sigma^2(c_b)}{f(c_b)} \;+\; o\Big(\tfrac{1}{N\,h^2}\Big),$$

where the $o((Nh^2)^{-1})$ remainder uses the uniformity and $4h^2 \cdot B = 2h$.

Let $q(u) := \sigma^2(u)/f(u)$. By Assumptions 2 and 3, q is continuous on $[0,1]$. Therefore its midpoint Riemann sums converge:

$$2h\sum_{b=1}^{B} q(c_b) \;=\; \int_0^1 q(u)\,du \;+\; o(1),$$

so

$$\sum_{b=1}^{B}\frac{\sigma^2(c_b)}{f(c_b)} \;=\; \frac{1}{2h}\int_0^1 \frac{\sigma^2(u)}{f(u)}\,du \;+\; o\Big(\tfrac{1}{h}\Big).$$

Substituting back yields

$$\operatorname{Var}\big[\hat{\beta}_{\text{avg}}\big] = \frac{2}{N\,h^2}\int_{-1}^{1} K^{*2}(u)\,du\int_0^1 \frac{\sigma^2(u)}{f(u)}\,du \;+\; o\Big(\tfrac{1}{N\,h^2}\Big),$$

as claimed. □