LOCAL-POLYNOMIAL ESTIMATION FOR MULTIVARIATE REGRESSION DISCONTINUITY DESIGNS.[1][2]

Preliminary draft. The latest version is available HERE. Version: April 20, 2023.

Takuya Ishihara[a], Daisuke Kurisu[b], Yasumasa Matsuda[a] and Masayuki Sawada[c]

We introduce a multivariate local-linear estimator for regression discontinuity designs. Unlike current local-linear approaches, we handle multivariate designs as multivariate. For that purpose, we develop a novel asymptotic normality for multivariate local-polynomial estimators. Consequently, we overcome the limitations of current local-linear approaches that either contradict the underlying assumptions or have limited interpretation. We demonstrate the effectiveness of our estimator through numerical simulations and an empirical illustration of a Colombian scholarship study by Londoño-Vélez, Rodríguez, and Sánchez (2020). Specifically, our estimates reveal a richer heterogeneity of the treatment effect that is hidden in the original estimates.

Keywords: Regression Discontinuity Designs, Local-Polynomial Estimation, Multiple Running Variables.

1. INTRODUCTION

The regression discontinuity (RD) design takes advantage of a particular treatment assignment mechanism that the eligibility of a program is set by the *running variable*. For example, a scholarship is awarded to applicants whose scores are above a threshold. The eligibility often requires additional requirement, such as the applicants' poverty scores being below a threshold as well. These RD designs are *multivariate* in their running variables. The multivariate RD design is superior to the standard RD design

[1]We thank the participants of the third Tohoku-ISM-UUlm workshop at Tohoku University and the seminar at Hitotsubashi University for their valuable comments.

[2]This work was supported by the JSPS KAKENHI Grant Number 22K13373 (ISHIHARA), the JSPS KAKENHI Grant Number 20K13468 (KURISU), the Grant-in-Aid for Scientific Research (B) 21H03400 (MATSUDA) and the JSPS KAKENHI Grant Number 21K13269 (SAWADA). First Version: Febrary 27, 2023.

[a]Graduate School of Economics and Management, Tohoku University

[b]Graduate School of International Social Sciences, Yokohama National University

[c]Institute of Economic Research, Hitotsubashi University

in its capability to capture heterogeneous treatment effects over the policy boundary. The multivariate design has the policy boundary to explore; however, the scalar RD design has only the single point of the policy cutoff. The frequent practices are dimension-redacted single-variable estimators; however, these practices either contradict the underlying assumptions, which include Assumption 1 (a) of Calonico, Cattaneo, and Titiunik (2014), or have limited interpretation and applicability. As a result, the frequent practices ruin either their flexible interpretation or asymptotic validity.

We achieve flexible interpretation with asymptotic validity by proposing an alternative estimation that takes multivariate RD designs as multivariate. For that purpose, we develop a novel asymptotic theory of the multivariate local-polynomial estimator with dimension-specific bandwidths.

We demonstrate favorable properties of the estimator in simulation and empirical replication studies. In simulation studies, our estimator demonstrates favorable performance against the current practices. We apply our estimates to the data of Londoño-Vélez, Rodríguez, and Sánchez (2020) who study the impact of a Colombian scholarship program on the college attendance rate. In the application, our estimates reveal a new finding on the treatment effects heterogeneity that was hidden in the original estimates. Specifically, the impact of the tuition program is homogeneous across different poverty levels with the same test scores, however, the impact sharply declines among the poor students with particularly high test scores.

Our contribution is in two-folds. First, to propose a multivariate local-linear RD estimation, we complete the asymptotic theory of multivariate local-polynomial estimation. Previously, Ruppert and Wand (1994) show the consistency of the multivariate local-polynomial estimator; Masry (1996) later shows the asymptotic normality of the estimator; however, Masry (1996) imposes that the bandwidths are common across dimensions. In our simulation results of Section 3, allowing for heterogeneous bandwidths is critical for the bias correction procedure in the RD estimation. Hence, our asymptotic result of our first contribution is theoretically important and practically relevant.

Second and more importantly, we fill the missing piece of practices in RD designs, local-linear estimation for the multivariate RD design. For a scalar running variable, the local-linear estimation of Calonico et al. (2014) with its companion package, *rdrobust*, is the first choice because the local-linear estimation is intuitive analogue to the RD identification. Nevertheless, existing local-linear estimators are limited to a uni-variate running variable. Current estimators are either not local-linear, limited in its interpretation, or violating the underlying assumption for the asymptotic normality. Our local-linear estimator is intuitive as much as a scalar-variable RD design, applicable to a variety of designs, and is capable to reveal a rich heterogeneity in treatment effects as demonstrated in our empirical illustration.

Local-linear estimation is the first choice for the RD estimator for a number of reasons. On the one hand, in the identification strategy of RD designs, the treated and control units *around the boundary point* are compared. On the other hand, in the kernel estimation, the kernel-weighted averages of the treated and control units *within a small bandwidth from the boundary point* are compared. Because local-constant estimation has a boundary problem, local-linear estimation is preferred since Fan and Gijbels (1992) and Imbens and Kalyanaraman (2012). Currently, however, multivariate estimations are available only in a non-kernel procedure such as Imbens and Wager (2019) and Kwon and Kwon (2020) with tuning parameters of the worst-case second derivative instead of the bandwidth.

As a result in empirical practices, the applied researchers convert a multivariate problem into a single-dimensional problem by taking either (1) a subsample of all but one requirement being satisfied for treatment or (2) some distance measure from a boundary point. Two strategies are in relation of a trade-off. The former *subsample* strategy has limited applicability by designs and is less capable of capturing heterogeneous effects over the boundary; the latter *distance* strategy can produce different estimates over the boundary; however, we point out its critical modeling issues below.

Matsudaira (2008) is an example of the first *subsample* strategy. Matsudaira (2008) considers the participation of a program based on an either failure of language and math exams. Matsudaira (2008) makes comparisons among two subsamples: first, the language-passing students who are at the boundary of the math exam; second, the math-passing students who are at the boundary of the language exam. These approaches have two issues. First, not all multivariate RD designs can accommodate this *subsample* strategy. Second, these approaches mask the important heterogeneity in treatment effects over the boundary. For example, among students at the language score on the border, the impact of a program may be substantially different by their math scores. Such heterogeneity in treatment effects is academically interesting and policy-relevant. Londoño-Vélez, Rodríguez, and Sánchez (2020) accommodate the same strategy as Matsudaira (2008), and we offer a richer heterogeneity than the original estimates as demonstrated in Section 4 with their data.

In the second *distance* strategy, multivariate running variable is explicitly reduced to a scalar distance measure. For example, Black (1999) computes the closest boundary point for each unit and compares units of the same closest boundary point to achieve the mean effect across the boundary. Furthermore, Keele and Titiunik (2015) propose another approach with the Euclid distance from a particular boundary point. The *distance* approach is capable to estimate heterogeneous effects at each boundary point. A package implementation in stata and R, *rdmulti*, is also offered as a wrapper of *rdrobust* to implement the latter Euclidean distance-based approach (Cattaneo, Titiunik, and Vazquez-Bare, 2020). This second *distance* strategy is straightforward to implement with the standard scalar RD estimator and applicable to a wider range of designs.

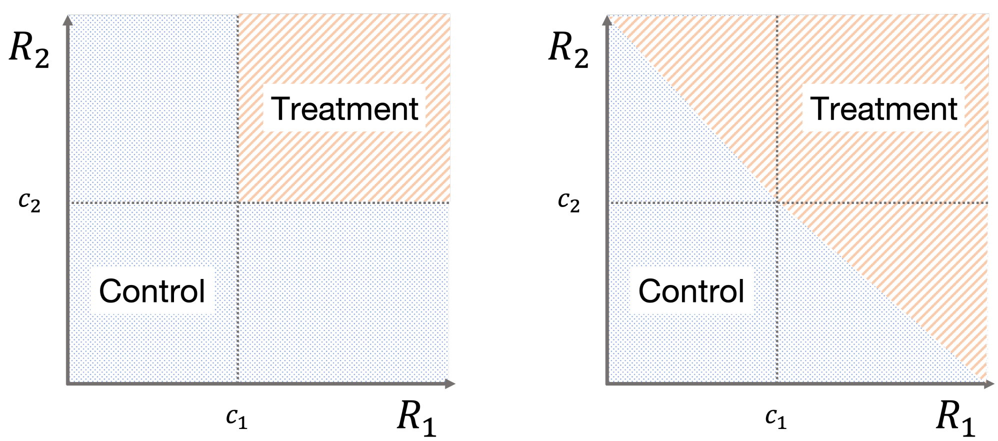
Conversely, there are two critical drawbacks in the distance strategy. First and critically, the value of the density of the Euclid distance converges to 0 as it approaches to the boundary. Consequently, the density, which appears in the denominator of the asymptotic variance, converges to 0 in the limit. As a result, Assumption 1 (a) of Calonico et al. (2014) is violated, and asymptotic normality does not hold. This simple fact is a novel remark in this study. Second, the induced conditional mean function for a point contradicts the other induced mean function from a nearby point on the boundary. We avoid these issues by handling the multivariate design as a multivariate estimation.

In the remainder of the paper, we introduce and motivate our estimator it in Section 2. We evaluate the proposed estimator in a Monte Carlo simulation exercise in Section 3. We demonstrate the added value of our estimator in the empirical study of Londoño-Vélez et al. (2020) in Section 4. Specifically, our estimator reveals a richer heterogeneity that was hidden in the original study, and our estimator is stable in its pattern relative to current practice. We conclude with future challenges in Section 5.

2. METHOD

2.1. *Model and Objective*

Consider a binary treatment $D \in \{0,1\}$ and associated pair of potential outcomes $\{Y(1), Y(0)\}$ such that $Y = DY(1) + (1-D)Y(0)$ for an observed outcome $Y \in \mathbb{R}$. We consider a sharp RD design with a *vector* of running variables $R \in \mathbb{R}^d$ for some integer $d \geq 1$. Specifically, $D = 1\{R \in \mathcal{T}\}$ where $\mathcal{T}$ is the treatment region, which is a subset of the support of R. To fix ideas, consider a pair of scores (R_1, R_2) for a student. For example, a student is eligible for a program when both scores exceed their corresponding thresholds (c_1, c_2). For such a program, the treatment region is $\mathcal{T} = \{(R_1, R_2) \in \mathbb{R}^2 : R_1 \geq c_1, R_2 \geq c_2\}$ (Figure 2.1 (a)). For another example, a student is eligible when the sum of scores exceeds a single threshold $c_1 + c_2$, $\mathcal{T} = \{(R_1, R_2) \in \mathbb{R}^2 : R_1 + R_2 \geq c_1 + c_2\}$ (Figure 2.1 (b)).



Panel (a) Panel (b)

FIGURE 2.1.— Illustration of $\mathcal{T}$. Panel (a) is under $\mathcal{T} = \{(R_1, R_2) \in \mathbb{R}^2 : R_1 \geq c_1, R_2 \geq c_2\}$; Panel (b) is under $\mathcal{T} = \{(R_1, R_2) \in \mathbb{R}^2 : R_1 + R_2 \geq c_1 + c_2\}$.

Let $(Y_i, D_i, R_i)_{i \in \{1,\ldots,n\}}$, $R_i = (R_{i,1}, R_{i,2})$ be the i.i.d. sample of (Y, D, R), $R = (R_1, R_2)$. Let c be a particular point on the boundary of $\mathcal{T}$. Our target parameter is $\theta(c) := \lim_{r \to c, r \in \mathcal{T}} E[Y(1) - Y(0)|R = r] - \lim_{r \to c, r \in \mathcal{T}^C} E[Y(1) - Y(0)|R = r]$. In the following, we focus on the issues in estimating the given identified parameter, $\theta(c)$. Under the following assumption (Hahn, Todd, and der Klaauw, 2001; Keele and Titiunik, 2015), $\theta(c)$ is the average treatment effect (ATE) at each point of the boundary c:

Proposition 2.1 (Keele and Titiunik, 2015, Proposition 1) If $E[Y(1)|R = r]$ and $E[Y(0)|R = r]$ are continuous in r at all points c of the boundary of $\mathcal{T}$; $P(D_i = 1) = 1$ for all i such that $R_i \in \mathcal{T}$; $P(D_i = 1) = 0$ for all i such that $R_i \in \mathcal{T}^C$, then,

$$\theta(c) = E[Y(1) - Y(0)|R = c]$$

for all c in the boundary of $\mathcal{T}$.

2.2. *Issues in the Conventional Estimators*

In Introduction, we describe two major approaches of multivariate RD estimation. The former *subsample* strategy such as Matsudaira (2008) is a single-variate RD design because it restricts its attention to the subsample who satisfy all but one requirement for treatment. Nevertheless, the subsample strategy is limited to a particular assignment mechanism. Furthermore, the subsample strategy dismisses the important merit of the multivariate designs, discovering the treatment effect heterogeneity. We demonstrate this critical merit of our strategy in discovering the heterogeneity in Section 4 with the Londoño-Vélez et al. (2020) data.

In the latter *distance* strategy, multivariate running variable is explicitly reduced to a scalar distance measure. Frequent choice is the Euclidean distance from a point or the closest boundary (Keele and Titiunik, 2015). The *distance* strategy is straightforward to implement in many designs; however, there are two critical drawbacks. First, a pair of points of the same distance from the point on the boundary share the same mean values. Hence, the induced conditional mean functions for different points contradict each other unless the mean function is entirely homogeneous over the boundary. Second and more importantly, the density of the distance running variable shrinks to zero as approaching to the boundary. Consequently, the inference and estimates of the distance strategy is not theoretically guaranteed because Assumption 1 (a) of Calonico et al. (2014) is violated.

To demonstrate the latter claim, consider the treated subsample $R_i \in \mathcal{T}$ and let $Z_i = \|R_i - c\|$ with a boundary point $c = (0,0)$, for simplicity. For $z > 0$, we have

$$\begin{aligned} P(Z_i \leq z) = P(\|R_i\| \leq z) &= \int_{\{r_1^2 + r_2^2 \leq z^2\}} f(r_1, r_2) dr_1 dr_2 \\ &= \int_0^z \int_0^{2\pi} t f(t\cos\theta, t\sin\theta) d\theta dt = \int_0^z \underbrace{t \left(\int_0^{2\pi} f(t\cos\theta, t\sin\theta) d\theta \right)}_{\text{density function of } Z_i} dt \end{aligned}$$

where $f(\cdot,\cdot)$ is the joint density of $R = (R_1, R_2)$. Hence, as long as the density function $f(\cdot,\cdot)$ is bounded, the distance density

$$f_Z(z) \equiv z \cdot \left(\int_0^{2\pi} f(z\cos\theta, z\sin\theta) d\theta \right)$$

shrinks to 0 as R_i approaches to the boundary point $c = (0,0)$. For the valid inference of a scalar RD estimate, Calonico et al. (2014) assumes that the density $f_Z(z)$ is continuous and bounded away from zero (Assumption 1 (a)). Consequently, the asymptotic normality of the local-linear estimation with Z_i is not guaranteed.

In Appendix C, we further show that the kernel density estimation of the distance running variable diminishes to 0 as the bandwidth $h \to 0$. Hence, the issue can be severe with a direct density estimation as in Imbens and Kalyanaraman (2012). The use of *rdrobust* package avoids the direct density estimation, however, the same concern applies in its asymptotic validity.

2.3. *Our Estimator*

We resolve the aforementioned limitations with a new estimator. Our estimator can capture the heterogeneous treatment effect over the boundary unlike the *subsample* strategy; our estimator avoids the issues in its inference unlike the *distance* strategy.

Consider the following local-linear estimator $\hat{\beta}^+(c) = (\hat{\beta}_0^+(c), \hat{\beta}_1^+(c), \hat{\beta}_2^+(c))'$

$$\hat{\beta}^+(c) = \underset{(\beta_0,\beta_1,\beta_2)' \in \mathbb{R}^3}{\arg\min} \sum_{i=1}^n (Y_i - \beta_0 - \beta_1(R_{i,1} - c_1) - \beta_2(R_{i,2} - c_2))^2 K_h(R_i - c) 1\{R_i \in \mathcal{T}\}$$

where $K_h(R_i - c) = K\left(\frac{R_{i,1}-c_1}{h_1}, \frac{R_{i,2}-c_2}{h_2}\right)$ and each h_j is a sequence of positive bandwidths such that $h_j \to 0$ as $n \to \infty$. Unlike Masry (1996), we allow $h_1 \neq h_2$ for the asymptotic normality. Later in Section 3, we demonstrate the importance of allowing heterogeneous bandwidths. Similarly, let $\hat{\beta}^-(c)$ be the estimator using $1\{R_i \in \mathcal{T}^c\}$ subsample. Our multivariate RD estimator at c is $\hat{\beta}_0^+(c) - \hat{\beta}_0^-(c)$.

In the main text below, we demonstrate our theoretical results in a special case of local-linear estimation with two-dimensional running variables. In Appendix A, under the same assumptions shown below, we show the general results for pth order local-polynomial estimation with d-dimensional running variables. These general results are also the basis of the bias correction procedure of our estimator.

Because we consider a random sample, the treated sample is independent of the control sample. Hence, we consider the following nonparametric regression models for each sample:

$$
\begin{aligned}
Y_i =& m_+(R_i) + \varepsilon_{+,i},\ \ E[\varepsilon_{+,i}|R_i] = 0,\ \ i \in \{1, \dots, n : R_i \in \mathcal{T}\} \text{ and} \\
Y_i =& m_-(R_i) + \varepsilon_{-,i},\ \ E[\varepsilon_{-,i}|R_i] = 0,\ \ i \in \{1, \dots, n : R_i \in \mathcal{T}^C\}.
\end{aligned}
$$

For the asymptotic normality, we impose the following regularity conditions that are standard in kernel regression estimations. In Assumption 2.1, we assume the existence of the continuous density function for the running variable R. Assumption 2.2 is the regularity conditions for a kernel function to use. We pick a particular set of kernel functions for our analysis later. Assumption 2.3 imposes a set of smoothness conditions for the mean function m as well as for the moments of the conditional mean residual ε_i. Finally, Assumption 2.4 specifies the rate of convergence of the vector of bandwidths $\{h_1, \dots, h_d\}$ relative to the sample size n.

Assumption 2.1 Let U_r be a neighborhood of $r = (r_1, \dots, r_d)'$.

(a) The random variable R_i has a probability density function f.

(b) The density function f is continuous on U_r and $f(r) > 0$.

Assumption 2.2 Let $K : \mathbb{R}^d \to \mathbb{R}$ be a kernel function such that

(a) $\int K(z)dz = 1$.

(b) The kernel function K is bounded and there exists a constant $C_K > 0$ such that K is supported on $[-C_K, C_K]^d$.

(c) Define $\kappa_0^{(r)} := \int K^r(z)dz$, $\kappa_{j_1,\ldots,j_M}^{(r)} := \int \prod_{\ell=1}^M z_{j_\ell} K^r(z)dz$, and

$$\check{z} := (1, (z)_1', \ldots, (z)_p')', \ (z)_L = \left(\prod_{\ell=1}^L z_{j_\ell}\right)'_{1\le j_1\le\cdots\le j_L\le d}, \ 1 \le L \le p.$$

The matrix $S = \int K(\boldsymbol{z}) \begin{pmatrix} 1 \\ \check{\boldsymbol{z}} \end{pmatrix} (1 \ \check{\boldsymbol{z}}')d\boldsymbol{z}$ is non-singular.

Assumption 2.3 Let U_r be a neighborhood of r.

(a) The mean function m is $(p+1)$-times continuously partial differentiable on U_r and define $\partial_{j_1\ldots j_L} m(r) := \partial m(r)/\partial r_{j_1} \ldots r_{j_L}$, $1 \le j_1, \ldots, j_L \le d$, $0 \le L \le p+1$. When $L = 0$, we set $\partial_{j_1\ldots j_L} m(r) = \partial_{j_0} m(r) = m(r)$.

(b) The variance function $\sigma^2(z) = E[\varepsilon_i^2 | R_i = z]$ is continuous at r.

(c) There exists a constant $\delta > 0$ such that $\sup_{z\in U_r} E[|\varepsilon_1|^{2+\delta}|R_1 = x] \le U(r) < \infty$.

Assumption 2.4 As $n \to \infty$,

(a) $h_j \to 0$ for $1 \le j \le d$,

(b) $nh_1 \cdots h_d \times h_{j_1}^2 \ldots h_{j_p}^2 \to \infty$ for $1 \le j_1 \le \cdots \le j_p \le d$,

(c) $nh_1 \cdots h_d \times h_{j_1}^2 \ldots h_{j_p}^2 h_{j_{p+1}}^2 \to c_{j_1\ldots j_{p+1}} \in [0, \infty)$ for $1 \le j_1 \le \cdots \le j_{p+1} \le d$.

Theorem 2.1 (Asymptotic normality of local-linear estimators) Under Assumptions 2.1, 2.2, 2.3 and 2.4 for $r = c$, the mean function m_+ with $d = 2$ and $p = 1$, the conditional mean residual $\varepsilon_{+,i}$, and the variance function $\sigma_+^2(z) = E[\varepsilon_{+,i}^2 | R_i = z]$, as $n \to \infty$, we have

$$\sqrt{nh_1h_2}\left(H^{ll}\left(\hat{\beta}^+(c) - M_+(c)\right) - S^{-1}B^{(2,1)}M_{+,n}^{(2,1)}(c)\right) \xrightarrow{d} N\left(\mathbf{0}, \frac{\sigma_+^2(c)}{f(c)} S^{-1}\mathcal{K}S^{-1}\right),$$

where

$$H^{ll} = \mathrm{diag}(1, h_1, h_2) \in \mathbb{R}^{3\times 3},$$
$$M_+(c) = (m_+(c), \partial_1 m_+(c), \partial_2 m_+(c))',$$

$$M_{+,n}^{(2,1)}(r) = \left(\frac{\partial_{11}m_+(c)}{2}h_1^2, \partial_{12}m_+(c)h_1h_2, \frac{\partial_{22}m_+(c)}{2}h_2^2\right)', \text{ and}$$

$$B^{(2,1)} = \int \begin{pmatrix} 1 \\ \check{\boldsymbol{z}} \end{pmatrix} (\boldsymbol{z})_2' d\boldsymbol{z},\ \mathcal{K} = \int K^2(\boldsymbol{z}) \begin{pmatrix} 1 \\ \check{\boldsymbol{z}} \end{pmatrix} (1\ \check{\boldsymbol{z}}')d\boldsymbol{z}.$$

The parallel result holds for $\hat{\beta}^-(c)$ under the parallel restrictions.

Consequently from Theorem 2.1, the mean-squared error (MSE) of $\hat{m}_+(c)$ has the following asymptotic expansion, for $e_1 = (1,0,0)'$,

$$\underbrace{\left[e_1 S^{-1} B^{(2,1)} \begin{pmatrix} \partial_{11}m_+(c)\frac{h_1^2}{2} \\ \partial_{12}m_+(c)h_1h_2 \\ \partial_{22}m_+(c)\frac{h_2^2}{2} \end{pmatrix}\right]^2}_{\text{Bias term}} + \underbrace{\frac{\sigma_+^2(c)}{nh_1h_2f(c)} e_1 S^{-1}\mathcal{K}S^{-1}e_1'}_{\text{Variance term}}.$$

Following the standard bandwidth selection procedure in RD designs, we aim to find the pair of (h_1, h_2) that minimizes the above asymptotic MSE.

In general, however, all three coefficients of three partial derivatives $\partial_{11}m_+(c), \partial_{12}m_+(c)$ and $\partial_{22}m_+(c)$ in the bias term are non-zero. This general expression is too complex to have an analytical formula for the optimal bandwidths. Hence, we simplify the above expression by taking particular kernels such that

$$\text{(2.1)} \qquad \kappa_1^{(1,1)} = \kappa_{1,2}^{(1,1,1)} = \kappa_1^{(1,2)} = \kappa_{1,2}^{(1,1,2)} = \kappa_{1,2}^{(1,2,1)} = 0.$$

Among product kernels of the form $K(z_1, z_2) = K_1(z_1)K_2(z_2)$, the above restriction amounts to *rotate* the space so that the boundary becomes either the x or y-axis. For example, the following kernels satisfy the above restrictions:

$$K_1(z) = \begin{cases} (1-|z|)1_{\{|z|\le 1\}} & \text{(two-sided triangular kernel)}, \\ \frac{3}{4}(1-z^2)1_{\{|z|\le 1\}} & \text{(Epanechnikov kernel)}, \end{cases}$$

$$K_2(z) = 2(1-|z|)1_{\{0\le z\le 1\}} \text{ (one-sided triangular kernel)}.$$

or a cone kernel

$$K(z_1, z_2) = \frac{6}{\pi}\left(1 - \sqrt{z_1^2 + z_2^2}\right) 1_{\{z_1^2+z_2^2 \le 1, z_2 \ge 0\}} = \frac{6}{\pi}\left(1 - \|z\|\right) 1_{\{\|z\| \le 1, z_2 \ge 0\}}.$$

where $z = (z_1, z_2)$ and $\|z\| = \sqrt{z_1^2 + z_2^2}$ satisfy (2.1). In the following, we assume that K_1 is the two-sided triangular kernel and K_2 is the one-sided triangular kernel. For example, the design with $\mathcal{T} = \{(R_1, R_2) \in \mathbb{R}^2 : R_1 \ge c_1, R_2 \ge c_2\}$ satisfies the restriction (2.1) as is or with a 90 degrees rotation; the design with $\mathcal{T} = \{(R_1, R_2) \in \mathbb{R}^2 : R_1 + R_2 \ge c_1 + c_2\}$ satisfies the restriction (2.1) with a 45 degrees rotation.

Under (2.1), MSE$(\hat{m}_+(c))$, is simplified to

$$\left\{\frac{h_1^2}{2}\partial_{11} m_+(c)\left(\tilde{s}_1 \kappa_1^{(2,1)} + \tilde{s}_3 \kappa_{1,2}^{(2,1,1)}\right) + \frac{h_2^2}{2}\partial_{22} m_+(c)\left(\tilde{s}_1 \kappa_2^{(2,1)} + \tilde{s}_3 \kappa_2^{(3,1)}\right)\right\}^2$$
$$+ \frac{\sigma_+^2(c)}{f(c) n h_1 h_2} \frac{\kappa_0^{(2)}\left(\kappa_1^{(2,1)}\kappa_2^{(2,1)}\right)^2 - 2\kappa_2^{(1,2)}\left(\kappa_1^{(2,1)}\right)^2 \kappa_2^{(2,1)}\kappa_2^{(1,1)} + \kappa_1^{(2,2)}\left(\kappa_1^{(2,1)}\kappa_2^{(1,1)}\right)^2}{\left(\kappa_0^{(1)}\kappa_1^{(2,1)}\kappa_2^{(2,1)} - \left(\kappa_2^{(1,1)}\right)^2 \kappa_2^{(2,1)}\right)^2}.$$

where

$$\begin{pmatrix} \tilde{s}_1 \\ \tilde{s}_2 \\ \tilde{s}_3 \end{pmatrix} := \frac{1}{\kappa_0^{(1)}\kappa_1^{(2,1)}\kappa_2^{(2,1)} - \left(\kappa_2^{(1,1)}\right)^2 \kappa_2^{(2,1)}} \begin{pmatrix} \kappa_1^{(2,1)}\kappa_2^{(2,1)} \\ 0 \\ -\kappa_1^{(2,1)}\kappa_2^{(1,1)} \end{pmatrix} = S^{-1} e_1.$$

Consequently, the MSE of the estimator $\hat{m}_+(c) - \hat{m}_+(c)$ is

$$\left\{\frac{h_1^2}{2}\left(\partial_{11} m_+(c) - \partial_{11} m_-(c)\right)\left(\tilde{s}_1 \kappa_1^{(2,1)} + \tilde{s}_3 \kappa_{1,2}^{(2,1,1)}\right)\right.$$
$$\left. + \frac{h_2^2}{2}\left(\partial_{22} m_+(c) - \partial_{22} m_-(c)\right)\left(\tilde{s}_1 \kappa_2^{(2,1)} + \tilde{s}_3 \kappa_2^{(3,1)}\right)\right\}^2$$
$$+ \frac{\left(\sigma_+^2(c) + \sigma_-^2(c)\right)}{f(c) n h_1 h_2} e_1 S^{-1} \mathcal{K} S^{-1} e_1'$$

when the same kernels are used for both the treatment and control sides.

We consider the optimal pair of bandwidths (h_1, h_2) that minimizes the above asymptotic MSE. There are two remaining issues to minimize the above asymptotic MSE. The first issue is that two bias terms may vanish when the second derivatives of the treatment and control mean functions are equal. This first issue is an extreme scenario when the second derivatives match exactly.

The optimal bandwidths may remain undetermined yet without the first issue. The second issue is that we can choose a pair (h_1, h_2) such that the bias term equals zero when the sign of the first-dimension $\partial_{11}m_+(c) - \partial_{11}m_-(c)$ differs from that of the second-dimension $\partial_{22}m_+(c) - \partial_{22}m_-(c)$. Unlike the first issue, which requires the exact match of mean function shapes, this second issue is more likely because only the signs of mean functions need to equal.

We attain a simple expression as a starting point under the following restrictions.

$$
\begin{aligned}
&\partial_{11}m_+(c) \neq \partial_{11}m_-(c), \partial_{22}m_+(c) \neq \partial_{22}m_-(c), \text{ and} \\
&sgn\left\{(\partial_{11}m_+(c) - \partial_{11}m_-(c))\left(\tilde{s}_1\kappa_1^{(2,1)} + \tilde{s}_3\kappa_{1,2}^{(2,1,1)}\right)\right\} \\
&\quad = sgn\left\{(\partial_{22}m_+(c) - \partial_{22}m_-(c))\left(\tilde{s}_1\kappa_2^{(2,1)} + \tilde{s}_3\kappa_2^{(3,1)}\right)\right\}.
\end{aligned}
$$

The unique pair of optimal bandwidths is attained by

$$
\frac{h_1}{h_2} = \sqrt{\frac{B_2(c)}{B_1(c)}} \text{ and } h_1^6 = \frac{(\sigma_+^2(c) + \sigma_-^2(c))}{2n} e_1 S^{-1} \mathcal{K} S^{-1} e_1' (B_1^{-5/2}(c) B_2^{1/2}(c))
$$

where

$$
\begin{aligned}
B_1(c) =& (\partial_{11}m_+(c) - \partial_{11}m_-(c))\left(\tilde{s}_1\kappa_1^{(2,1)} + \tilde{s}_3\kappa_{1,2}^{(2,1,1)}\right), \text{ and} \\
B_2(c) =& (\partial_{22}m_+(c) - \partial_{22}m_-(c))\left(\tilde{s}_1\kappa_2^{(2,1)} + \tilde{s}_3\kappa_2^{(3,1)}\right).
\end{aligned}
$$

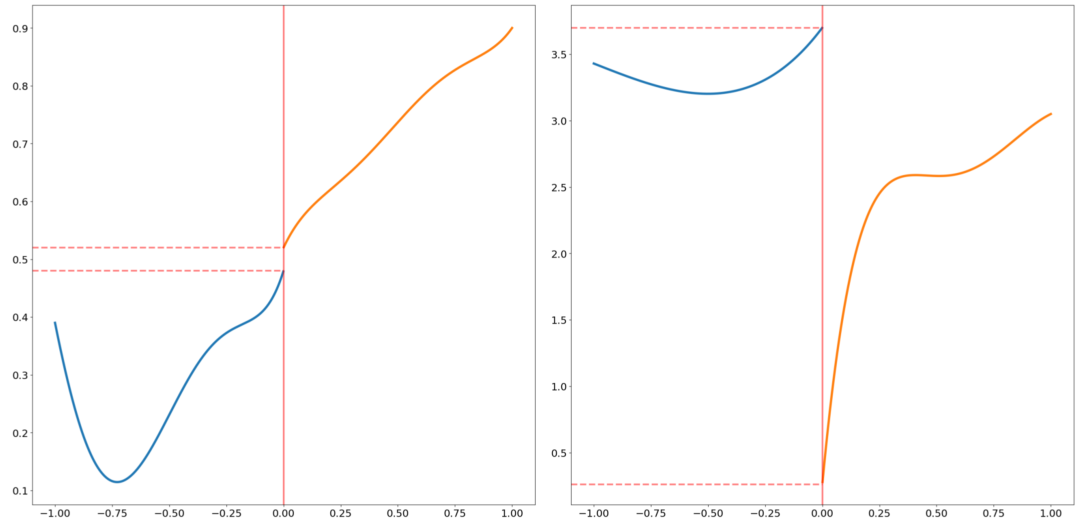
In general, these restrictions can fail. A similar issue arises in the single-variable RD estimation with heterogeneous bandwidths with the treatment and control mean functions (Imbens and Kalyanaraman, 2012). A theoretically possible approach is to follow Arai and Ichimura (2018) who derive the higher-order expansion of the bias terms for the single-variable RD estimation. In Appendix A.2.1, we derive the higher-order expansion of the bias terms. Practically speaking, such a higher-order bias correction is not appropriate for multivariate RD estimations. As shown in Appendix A.2.1, a higher-order bias correction procedure requires a reliable estimation for local estimation of cubic polynomial with 10 coefficients. The higher-order bias correction is theoretically possible; however, such a procedure is practically not reliable. Instead, we follow Imbens and Kalyanaraman (2012) to rely on regularization. In particular, we take the absolute values of the bias terms $B_1(c)$ and $B_2(c)$ as

$$\frac{h_1}{h_2} = \left(\frac{B_2(c)^2}{B_1(c)^2}\right)^{1/4} \text{ and } h_1 = \left[\frac{(\sigma_+^2(c) + \sigma_-^2(c))}{2n} e_1 S^{-1} \mathcal{K} S^{-1} e_1' (|B_1(c)|^{-5/2} |B_2(c)|^{1/2})\right]^{1/6},$$

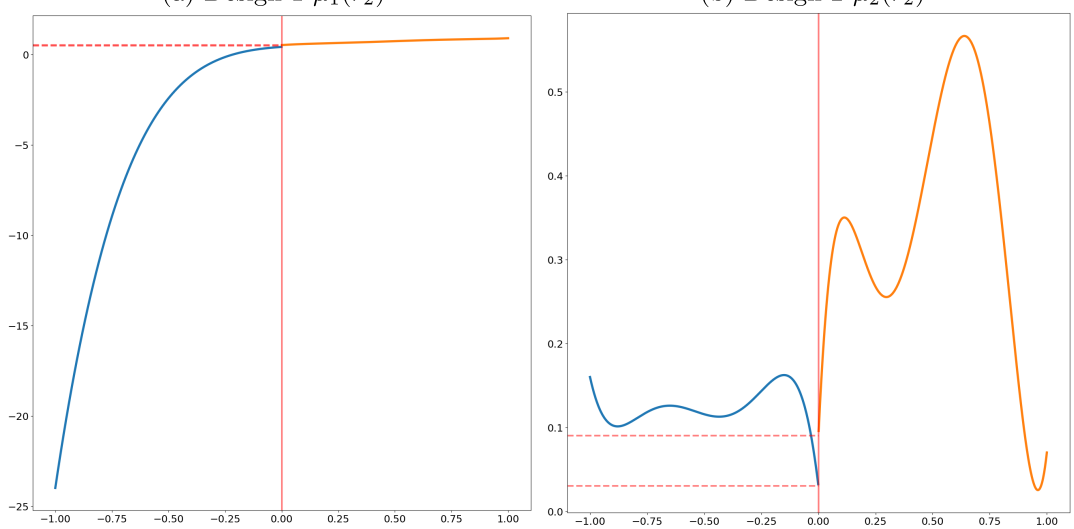
and add regularization terms to $B_1(c)$ and $B_2(c)$ to prevent bandwidths to blow up when the bias terms are zero or close to zero. Note that the optimal bandwidth ratio h_1/h_2 is the same for the optimal inner solution to the minimization as well as for the corner solution of the first-order bias being zero. Given the same bandwidth ratio, we choose h_1 from the above formula when the realized signs of the estimated bias terms are the same. If they are different, then we determine the bandwidths by the regularization, assuming that the bias term disappears. Finally, as is well known for the single-variable RD estimation by Calonico et al. (2014), we need to have a bias correction to have appropriate inference. We propose a plug-in bias correction with the two-dimensional local-quadratic estimation. See Appendix B for these implementation details.

3. SIMULATION RESULTS

We demonstrate the numerical properties of our estimator in the following Monte Carlo simulations with four different designs, partially taken from Arai and Ichimura (2018), Calonico et al. (2014), and Imbens and Kalyanaraman (2012). Specifically, we take four designs of Arai and Ichimura (2018) as the base specifications for mean function shapes for one of two dimensions. Figure 3.2 is the shapes of mean functions used in the numerical simulations of Arai and Ichimura (2018). Those specifications are repeatedly used in other RD studies such as Calonico et al. (2014) and Imbens and Kalyanaraman (2012) to evaluate their numerical performances.



(a) Design 1 $\mu_1(r_2)$ (b) Design 2 $\mu_2(r_2)$



(c) Design 3 $\mu_3(r_2)$ (d) Design 4 $\mu_4(r_2)$

FIGURE 3.2.— Basic mean functions taken from Arai and Ichimura (2018). Design 1 is from Lee (2008) Data and Design 3 is a modification of Design 1 by Imbens and Kalyanaraman (2012). Design 2 and 4 are from Ludwig and Miller (2007) Data.

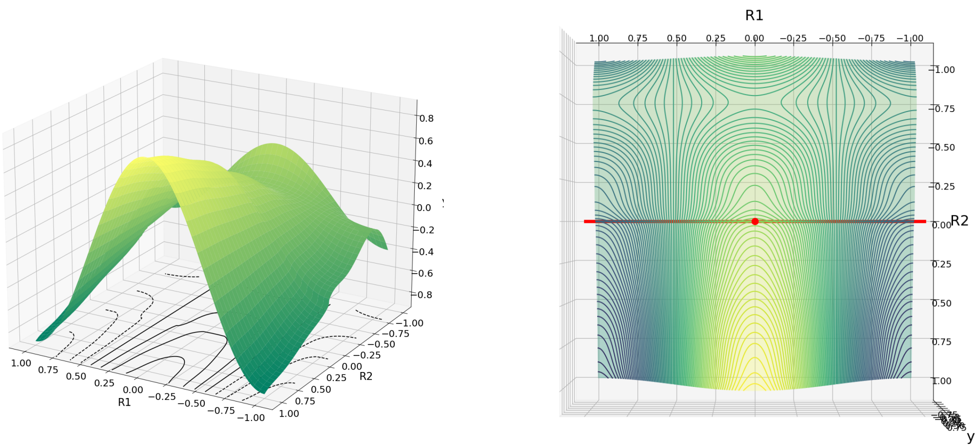


FIGURE 3.3.— Contour plots for Design 1, $m(r_1, r_2) = \mu_1(r_2)\cos(\pi r_1)$ with $\mu_1(r_2)$ as in Figure 3.2 (a). The red line is the boundary; the red circle is the evaluation point.

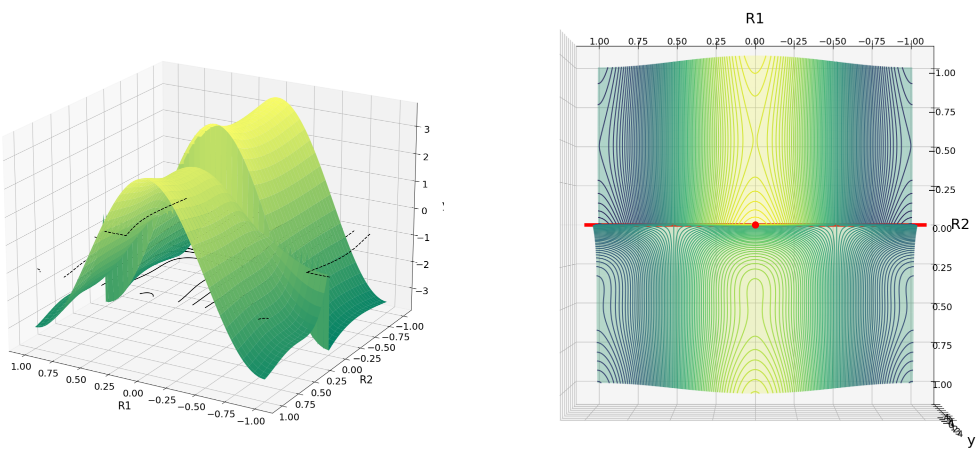


FIGURE 3.4.— Contour plots for Design 2, $m(r_1, r_2) = \mu_2(r_2)\cos(\pi r_1)$ with $\mu_2(r_2)$ as in Figure 3.2 (b). The red line is the boundary; the red circle is the evaluation point.

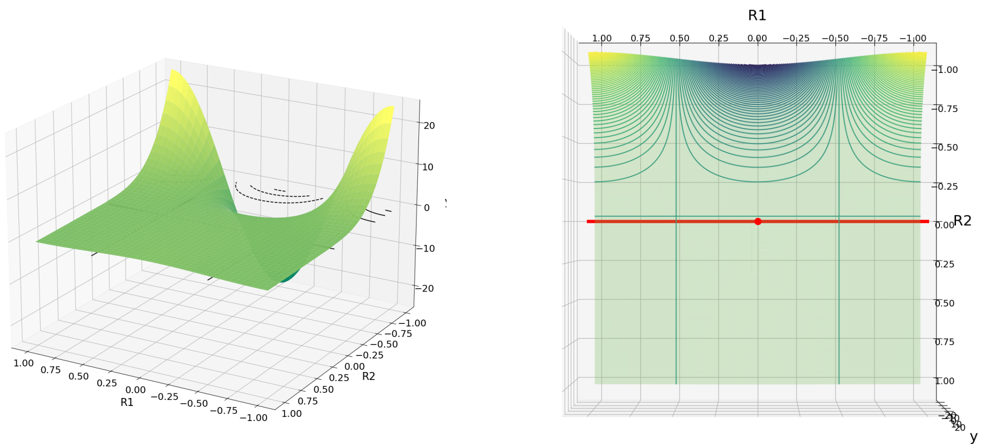


FIGURE 3.5.— Contour plots for Design 3, $m(r_1, r_2) = \mu_3(r_2)\cos(\pi r_1)$ with $\mu_3(r_2)$ as in Figure 3.2 (c). The red line is the boundary; the red circle is the evaluation point.

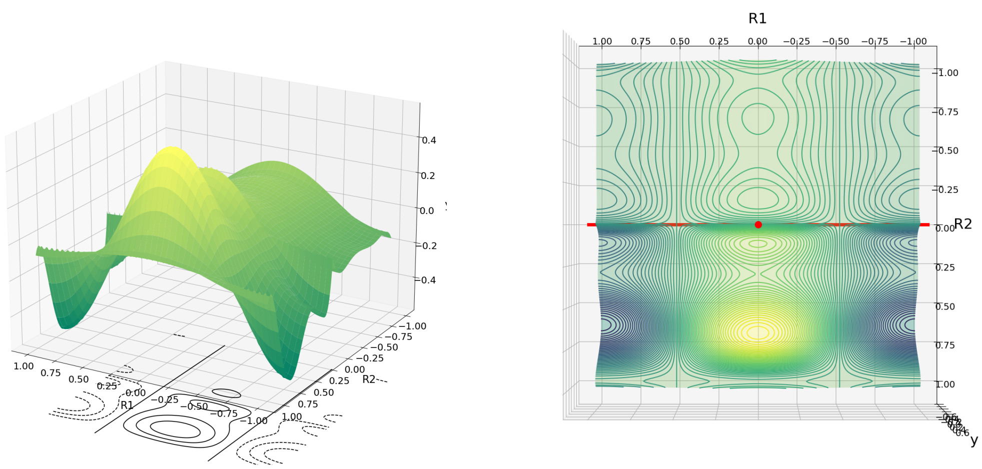
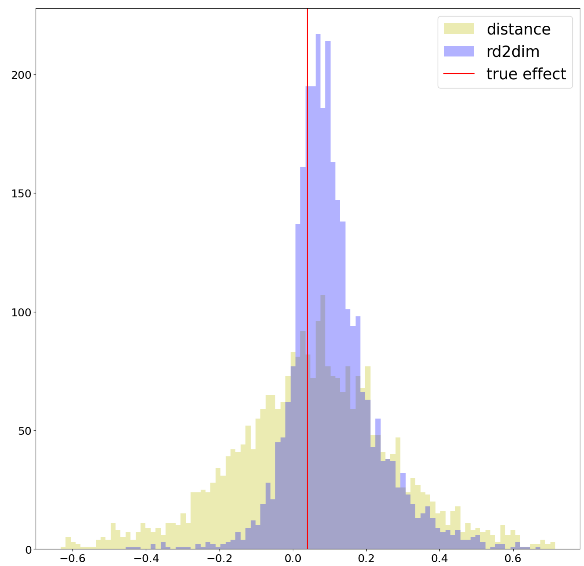
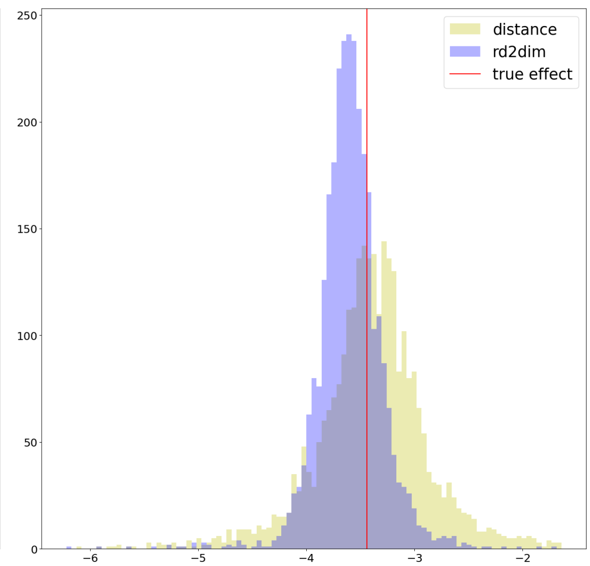


FIGURE 3.6.— Contour plots for Design 4, $m(r_1, r_2) = \mu_4(r_2)\cos(\pi r_1)$ with $\mu_4(r_2)$ as in Figure 3.2 (d). The red line is the boundary; the red circle is the evaluation point.

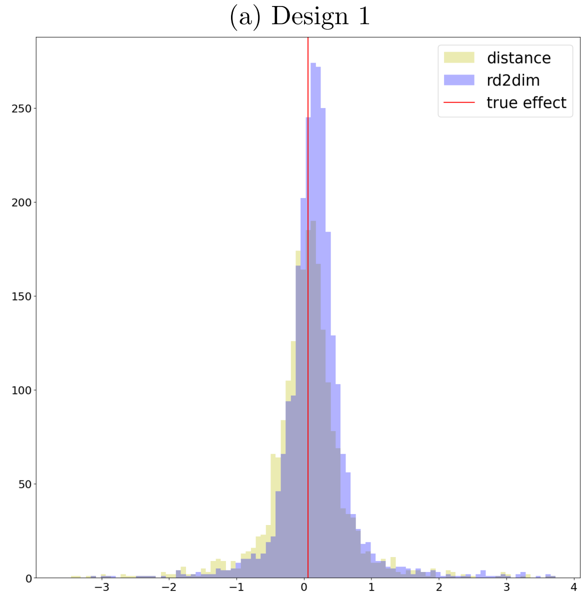
Based on the shapes of the mean function of a single dimension R_1, we multiply a cosine function of the other dimension R_2. Figures 3.3, 3.4, 3.5, and 3.6 are the 3D plots of the mean functions. The cosine function is chosen among trigonometric functions so that their second derivatives in R_1 and R_2 are nonzero. Among the four designs, the shape in Design 1 is relatively moderate compared to other specifications. Design 2 has a massive jump on the boundary with similar shapes on both sides; Design 3 is extremely flat on the control side; Design 4 has a complex shape in the control side.



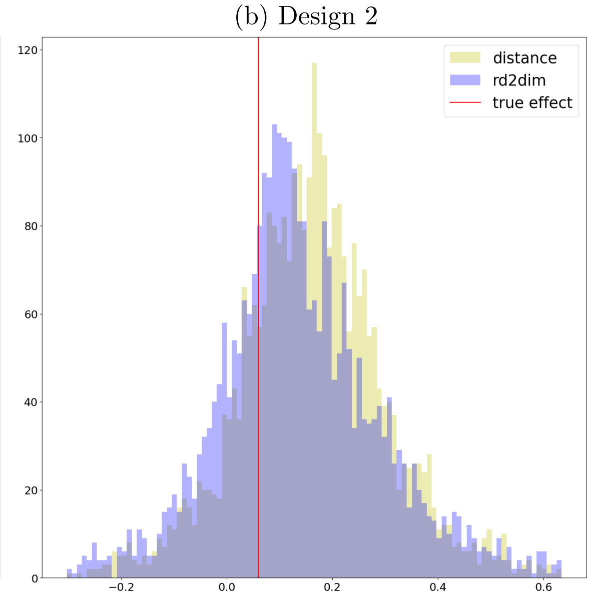
(a) Design 1



(b) Design 2



(c) Design 3



(d) Design 4

FIGURE 3.7.— Histograms of point estimates for three designs with truncation of 1% tail observations. Darker blue distributions are of our preferred estimates; lighter yellow distributions are of *distance* based estimates.

For each draw of a simulation sample, we draw $R_1 \sim 2 \times Beta(2,4) - 1$ and $R_2 \sim U[-1,1]$ independently each other; we generate the outcome variable as $m(R_{i1}, R_{i2}) + \epsilon_i$ where $\epsilon_i \sim N(0, 0.1295^2)$. We compare the quality of our estimator, *rd2dim*, relative to the *distance* estimation using rdrobust in Figures 3.7. Figures 3.7 are histograms of realized estimates of 3000 times replications. The darker blue histograms of *rd2dim* have mainly thinner shapes than the lighter yellow histograms of *distance* estimation using rdrobust. Nevertheless, for some specifications, the distance approach has better bias corrections than ours. The yellow histograms are better centered around the red line of the true effect than the blue histogram.

TABLE 3.1

SIMULATION RESULTS OF THREE ESTIMATORS FOR THREE DESIGNS.

Design	Estimator	Mean length	Median length	Mean bias	Coverage rate	RMSE	Success rate
Design 1	rd2dim	0.47	0.43	0.08	0.92	0.18	1.00
Design 1	common	0.44	0.42	0.11	0.78	0.17	1.00
Design 1	distance	0.88	0.73	0.02	0.93	0.31	1.00
Design 2	rd2dim	1.70	1.47	-0.14	0.96	1.74	0.99
Design 2	common	1.64	1.60	-0.55	0.89	0.59	1.00
Design 2	distance	41123.96	1.99	0.02	0.95	1.60	0.98
Design 3	rd2dim	1.02	0.87	0.24	0.77	5.73	0.88
Design 3	common	0.75	0.69	0.35	0.49	0.69	0.99
Design 3	distance	212563.46	1.34	-0.15	0.96	5.70	0.74
Design 4	rd2dim	0.64	0.64	0.08	0.92	0.24	1.00
Design 4	common	0.52	0.50	0.17	0.74	0.22	1.00
Design 4	distance	0.53	0.48	0.10	0.85	0.19	1.00

Notes: Results are from $3,000$ replication draws of $1,000$ observation samples. *rd2dim* refers to our preferred estimator; *common* is our estimator with imposing the bandwidths being the same for two dimensions; *distance* is the estimator with the Euclidean distance from the boundary point as the running variable. All the implementations are in *Python*. *Mean length* and *Median length* are of generated confidence interval length. *Success* is the rate of successful reporting among replicated $3,000$ samples, counting the failures in positive variance estimation or in singularity of the design matrix.

We have a closer look at the performance comparisons in Table 3.1. Our first observation is that estimations with heterogeneous bandwidths $h_1 \neq h_2$ matter. The *common* estimator is a version of *rd2dim* that imposes $h_1 = h_2$. For all designs, the 95% coverage rate of the true effect size is much worse for *common* compared to *rd2dim*, apparently due to better bias correction with heterogeneous bandwidth selection.

When we compare *rd2dim* against *distance*, RMSE of *rd2dim* is approximately half of *distance* for Design 1; the RSMEs are similar for two estimators for the other designs. We conjecture the reason for the massively superior performance in Design 1 for its sufficient variations in the mean functions over both axes of R_1 and R_2. The other designs are less natural as two-dimensional designs than Design 1 and have extremely flat or extremely dipping shapes. Our *rd2dim* is equally favorable in the 95% coverage rate compared to *distance*. Nevertheless, the lengths of the mean and median confidence intervals are much shorter for *rd2dim* relative to *distance* for most specification. Importantly, our estimator is much more stable than distance based that sometimes fail to report a valid standard error estimate. These tendencies are shown as extreme values in the standard errors and consequently the mean length as well as the successful reporting of the estimates without division by zero error. The instability of the *distance* estimator is natural because the assumption for the valid inference is violated (see Section 2.2 for details).

4. APPLICATION

We illustrate our estimator in an empirical application of a Colombian scholarship, Londoño-Vélez, Rodríguez, and Sánchez (2020). The scholarship of interest is primarily determined by two thresholds: merit-based and need-based. Consequently, there is a policy *boundary* instead of a single cutoff. Our estimator is particularly relevant to their study because of their interest in the heterogeneity over the policy boundary. The outcome of interest is enrollment in any college; hence, the policy impact may be heterogeneous by their poverty level and their level of academic ability.

From 2014 to 2018, the Colombian government operated a large-scale scholarship program called Ser Pilo Paga (SPP). The scholarship loan covers "the full tuition cost of attending *any* four-year or five-year undergraduate program in *any* government-certified "high-quality" university in Colombia." (Londoño-Vélez et al. (2020), pp.194). The scholarship takes the form of a loan, but the loan is forgiven if the recipient graduates the university appropriately. The eligibility of the SPP program is three-fold: first, students must have their scores from a high-school exit exam exceeding a threshold; second, the students must be from a welfare recipient household; third, the students must be admitted by an eligible university. The first merit-based threshold is based on the nationally standardized high school graduation exam, SABER 11. In 2014 of Londoño-Vélez et al. (2020)'s study period, the cutoff was the top 9% of the score distribution. The second need-based threshold is based on the eligibility of a social welfare program, SISBEN. Being eligible for SISBEN means that the family is roughly the poorest 50 percent. Students who exceed two thresholds may still be ineligible for the program due to the third requirement. Hence, the impact of exceeding both thresholds is not the impact of the program itself due to potential noncompliance. The estimand is the impact of the program eligibility, which is the intention-to-treat (ITT) effect.

The empirical strategy of Londoño-Vélez et al. (2020) is the *subsample* approach. They run two separate local regressions for the merit-based cutoff among the need-eligible students and for the need-based cutoff among the merit-eligible students. Figure 4.8 is the scatter plot of observations in the space of the need-based criterion (SISBEN) for the x-axis and the merit-based criterion (SAVER11) for the y-axis. Their strategy is to estimate the effect of exceeding the SISBEN threshold for those who are around SABER11 score near 0 and of exceeding the SABER11 threshold among those who are around SISBEN score near 0. For each subsample, they run *rdrobust* package based on Calonico et al. (2014). Londoño-Vélez et al. (2020) prefer this approach because *the discontinuities represent different populations, and the heterogeneity in estimated impacts across these frontiers is informative* (pp.205). Londoño-Vélez et al. (2020) report that the effect of exceeding the merit-based (SABER11) threshold on enrollment in any eligible college is 0.32 with the standard error of 0.012 for the need-based (SISBEN) eligible subsample; the effect of exceeding the need-based (SISBEN) threshold on enrollment in any eligible college is 0.274 with the standard error of 0.027 for the merit-based (SABER11) eligible subsample. Students with the need eligibility in the x-axis boundary of Figure 4.8 have a slightly higher effect than students with the merit eligibility in the y-axis boundary of Figure 4.8. Indeed, their strategy captures certain heterogeneity in the two sub-populations, albeit with richer heterogeneity within.

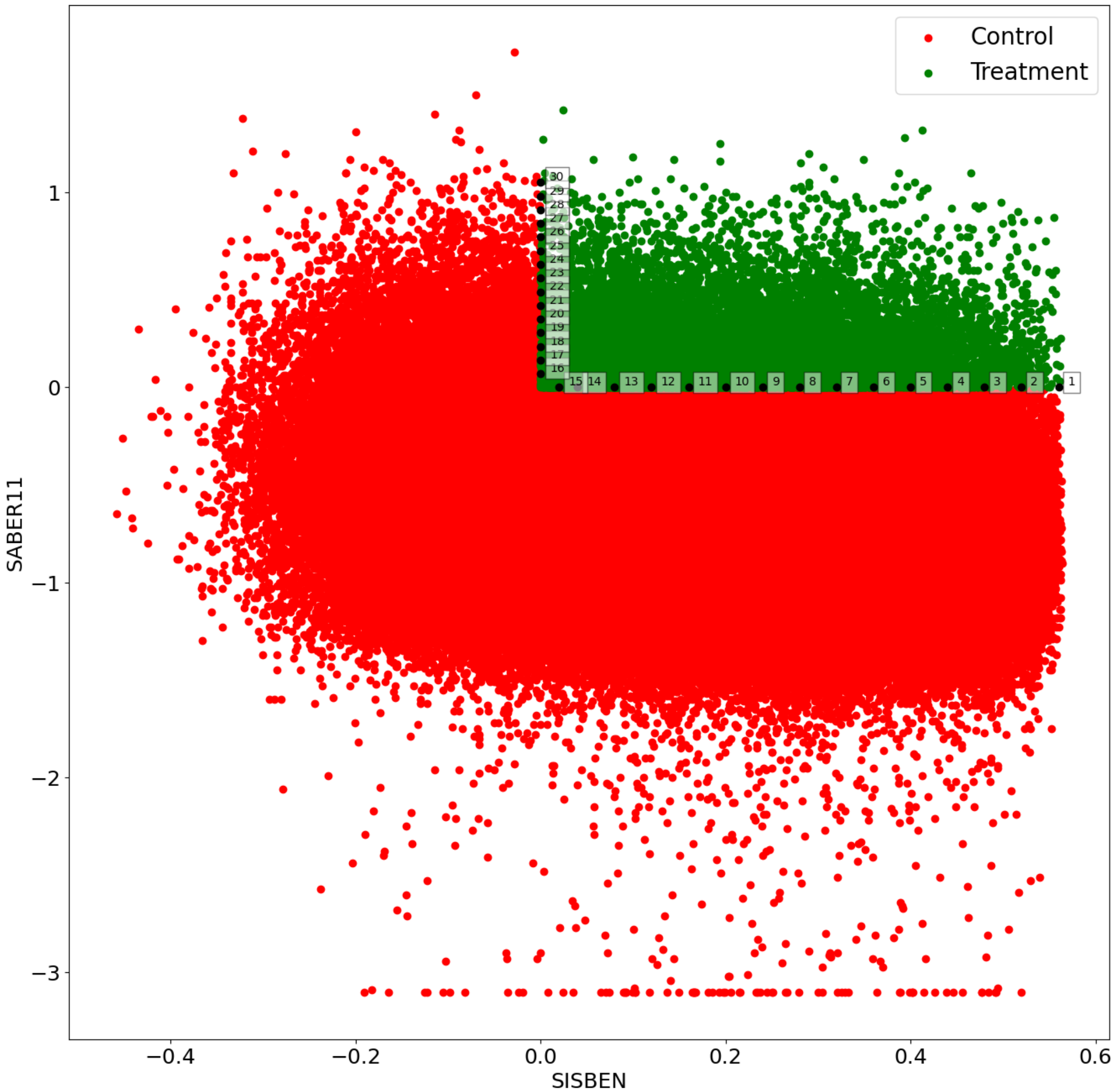


FIGURE 4.8.— Scatter plot of observations. The x-axis represents the distance of SISBEN score from the policy cutoff, divided by 100; the y-axis represents the distance of SABER11 score from the policy cutoff, divided by 100. Positive values in each distance measure imply satisfying one of two policy requirements. The black dots on the boundary are our evaluation points from 1 through 30.

Instead of the subsample approach, we estimate the heterogeneous effects over the whole boundary. We summarize our results in Figure 4.9. The darker blue intervals are the pointwise 95% confidence intervals from our *rd2dim* estimates at each value of the boundary points; the lighter green intervals are the pointwise 95% confidence intervals from the distance-based *rdrobust* estimates at the same values of the boundary points. For the most of points, the pattern of two estimates are similar across the boundary points with a notable difference in the length of the confidence intervals. For the most of the need-based eligible students (point 3 through point 15), our confidence intervals are shorter than the distance-based ones.

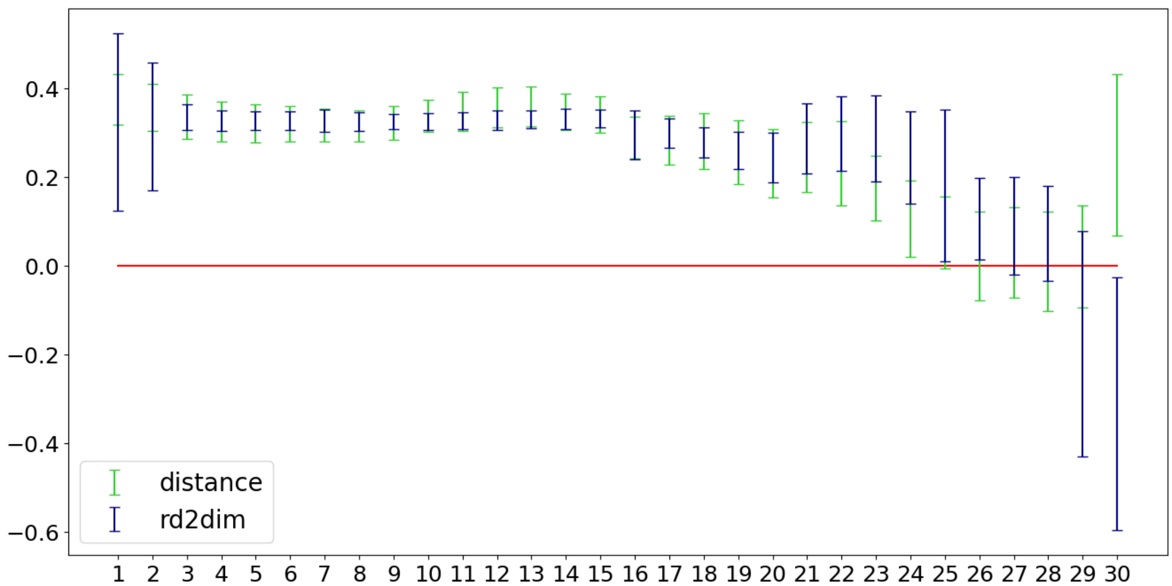


FIGURE 4.9.— Estimation results over the 30 boundary points. Values from 1 through 30 in the x-axis corresponds values in Figure 4.8. Points from 1 through 15 are of exceeding the merit threshold among the need-eligible students; points from 16 through 30 are of exceeding the need threshold among the merit-eligible students.

Both estimates, in particular our boundary-specific estimates, suggest that there are substantial heterogeneity in the effects among the merit-eligible students ($16 \sim 30$) but not among the need-eligible students ($1 \sim 15$). Specifically, the program has similar effects among the majority of students, but the program has no or negative impact for extremely capable students (points 25 through 30).

The zero impact for extremely capable students is reasonable because they would have received other scholarships to attend college without out-of-pocket expenses anyway. The negative impact of the most capable students (points 29 and 30) may be consistent with the definition of the dependent variable. The data is constructed from the administrative SABER11 and SISBEN scores data which is merged with the data from the Ministry of Education of Colombia that tracks students of the postsecondary education system. Hence, the dependent variable of enrollment may not capture the outside options such as enrolling in the selected US schools. The *distance* estimation does not capture this heterogeneity and takes the opposite sign from the other estimates. We conjecture that the *distance* estimation picks the outlier who are away from the boundary because the students of the same distance from the point are compared equally. In fact, this sign-flipping pattern of the *distance* estimation disappears when the relative scale of two axes are adjusted by the absolute maximum values of each axis (Figure 4.10 and 4.11). Finding an appropriate relative scaling of two axes is a difficult task. Our *rd2dim* is free from such a difficult re-scaling task. This is an important merit of our approach that can handle the relative scaling of the two-dimensional data as is.

5. CONCLUSION

We provide an alternative estimator for RD designs with multivariate running variables. Specifically, our estimator does not convert a multivariate RD estimation problem into a scalar RD estimation problem. We estimate the multivariate conditional mean functions as is. For the purpose of RD estimations, we develop a new asymptotic result for the multivariate local-polynomial regression with dimension specific bandwidths. In numerical simulations, we demonstrate favorable performance of our estimator against a frequently used procedure of a distance measure as the scalar running variable. We apply our estimator to the study of Londoño-Vélez et al. (2020) who study the impact of a scholarship program that has two eligibility requirements. In the application, our estimates are consistent with the original estimates and reveal a richer heterogeneity in the program impacts over the policy boundary than the original estimates.

Our contributions are summarized in two ways. First, we demonstrate the issues in the current practices of multivariate RD designs and offer a remedy for the issues. The *distance* approach (Black, 1999, Keele and Titiunik, 2015, for example) of converting a multivariate running variable with the Euclidean distance from a point violates the inference assumption of Calonico et al. (2014) and imposes a nontrivial restriction on the conditional mean function; the *subsample* approach (Matsudaira, 2008, for example) of taking the subsample with eligibility for all but one requirement has limited applicability and capability to capture heterogeneous effects. We provide a strategy that is capable to estimate heterogeneous effects without the dimension reduction.

Second, our asymptotic results complete the theory of multivariate local-polynomial estimates. After Masry (1996) has shown the asymptotic normality of multivariate local-polynomial estimates with common bandwidths between dimensions, no studies have achieved the asymptotic theory with dimension-specific bandwidths. As demonstrated in our simulation results, allowing different bandwidths for each dimension matters substantially for the bias correction procedure, which results in the improved coverage rate of our preferred estimates.

There are a few theoretical and practical issues remaining. First, our consideration is limited to a random sample; hence, spatial RD designs are excluded from our consideration. We defer our focus to spatial designs because of its theoretical and conceptual complexity in addition to the analysis in this study. Nevertheless, we aim to propose a spatial RD estimation based on a newly developed asymptotic results of Kurisu and Matsuda (2022) in a separated study. Second, our theoretical results applies to any finite dimensional RD designs, however, practical performances of such estimators with higher than two dimensions can be limited. Although most RD designs have at most two dimensions, the practical implementation of a higher-dimensional RD estimation is an open question. Similarly, we provide the higher-order bias expressions for our multivariate local-polynomial estimates; however, estimating the derived bias expressions is challenging. A new idea of exploiting these expressions is desirable. Finally, we do not provide any procedure to aggregate heterogeneous estimates over the set of boundary points. We leave these topics for future research questions.

REFERENCES

Arai, Y. and H. Ichimura (2018): "Simultaneous Selection of Optimal Bandwidths for the Sharp Regression Discontinuity Estimator," *Quantitative Economics*, 9, 441–482.

Black, S. E. (1999): "Do Better Schools Matter? Parental Valuation of Elementary Education*," *The Quarterly Journal of Economics*, 114, 577–599.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014): "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 82, 2295–2326.

Cattaneo, M. D., R. Titiunik, and G. Vazquez-Bare (2020): "Analysis of Regression-Discontinuity Designs with Multiple Cutoffs or Multiple Scores," *The Stata Journal: Promoting communications on statistics and Stata*, 20, 866–891.

Fan, J. and I. Gijbels (1992): "Variable Bandwidth and Local Linear Regression Smoothers," *Annals of Statistics*, 20, 2008–2036.

Hahn, J., P. Todd, and W. V. der Klaauw (2001): "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica*, 69, 201–209.

Imbens, G. and K. Kalyanaraman (2012): "Optimal Bandwidth Choice for the Regression Discontinuity Estimator," *The Review of Economic Studies*, 79, 933–959.

Imbens, G. and S. Wager (2019): "Optimized Regression Discontinuity Designs," *The Review of Economics and Statistics*, 101, 264–278.

Keele, L. J. and R. Titiunik (2015): "Geographic Boundaries as Regression Discontinuities," *Political Analysis*, 23, 127–155.

Kurisu, D. and Y. Matsuda (2022): "Local Polynomial Regression for Spatial Data on $\mathbb{R}^d$," *arXiv:2211.13467 [math, stat]*.

Kwon, K. and S. Kwon (2020): "Adaptive Inference in Multivariate Nonparametric Regression Models Under Monotonicity," *arXiv:2011.14219*.

Lee, D. S. (2008): "Randomized Experiments from Non-Random Selection in U.S. House Elections," *Journal of Econometrics*, 142, 675–697.

Londoño-Vélez, J., C. Rodríguez, and F. Sánchez (2020): "Upstream and Downstream Impacts of College Merit-Based Financial Aid for Low-Income Students: Ser Pilo Paga in Colombia," *American Economic Journal: Economic Policy*, 12, 193–227.

Ludwig, J. and D. L. Miller (2007): "Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design*," *The Quarterly Journal of Economics*, 122, 159–208.

Masry, E. (1996): "Multivariate Regression Estimation Local Polynomial Fitting for Time Series," *Stochastic Processes and their Applications*, 65, 81–101.

Matsudaira, J. D. (2008): "Mandatory Summer School and Student Achievement," *Journal of Econometrics*, 142, 829–850.

Ruppert, D. and M. P. Wand (1994): "Multivariate Locally Weighted Least Squares Regression," *The Annals of Statistics*, 22, 1346–1370.

APPENDIX A: ASYMPTOTIC THEORY FOR MULTIVARIATE LOCAL-POLYNOMIAL REGRESSIONS

A.1. *Local-polynomial estimator*

Consider the following nonparametric regression model:

$$Y_i = m(R_i) + \varepsilon_i,\ E[\varepsilon_i|R_i] = 0,\ i = 1, \dots, n,$$

where $\{(Y_i, R_i)\}_{i=1}^n$ is a sequence of i.i.d. random vectors such that $Y_i \in \mathbb{R}$, $R_i = (R_{i,1}, \dots, R_{i,d})' \in \mathbb{R}^d$.

Define

$$\begin{aligned} D &= \#\{(j_1, \dots, j_L) : 1 \leq j_1 \leq \cdots \leq j_L \leq d, 0 \leq L \leq p\}, \\ \bar{D} &= \#\{(j_1, \dots, j_{p+1}) : 1 \leq j_1 \leq \cdots \leq j_{p+1} \leq d\}, \end{aligned}$$

and $(s_{j_1 \dots j_L 1}, \dots, s_{j_1 \dots j_L d}) \in \mathbb{Z}^d_{\geq 0}$ such that $s_{j_1 \dots j_L k} = \#\{j_\ell : j_\ell = k, 1 \leq \ell \leq L\}$. Further, define $\boldsymbol{s}_{j_1 \dots j_L}! = s_{j_1 \dots j_L 1}! \dots s_{j_1 \dots j_L d}!$. When $L = 0$, we set $(j_1, \dots, j_L) = j_0 = 0$, $\boldsymbol{s}_{j_1 \dots j_L}! = 1$. Note that $\sum_{j=1}^d s_{j_1 \dots j_L \ell} = L$. The local-polynomial estimator

$$\begin{aligned} \hat{\beta}(r) &= (\hat{\beta}_{j_1, \dots j_L}(r))'_{1 \leq j_1 \leq \cdots \leq j_L \leq d, 0 \leq L \leq p} \\ &:= (\hat{\beta}_0(r), \hat{\beta}_1(r), \dots, \hat{\beta}_d(r), \hat{\beta}_{11}(r), \dots \hat{\beta}_{dd}(r), \dots, \hat{\beta}_{1 \dots 1}(r), \dots, \hat{\beta}_{d \dots d}(r))'. \end{aligned}$$

of

$$\begin{aligned} M(r) &= \left(\frac{1}{\boldsymbol{s}_{j_1 \dots j_L}!} \partial_{j_1, \dots j_L} m(r)\right)'_{1 \leq j_1 \leq \cdots \leq j_L \leq d, 0 \leq L \leq p} \\ &:= \Bigg(m(r), \partial_1 m(r), \dots, \partial_d m(r), \frac{\partial_{11} m(r)}{2!}, \frac{\partial_{12} m(r)}{1!1!}, \dots, \frac{\partial_{dd} m(r)}{2!}, \\ &\qquad \dots, \frac{\partial_{1 \dots 1} m(r)}{p!}, \frac{\partial_{1 \dots 2} m(r)}{(p-1)!1!} \dots, \frac{\partial_{d \dots d} m(r)}{p!}\Bigg)' \end{aligned}$$

is given as a solution of the following problem:

$$\hat{\beta}(r) = \arg\min_{\beta \in \mathbb{R}^D} \sum_{i=1}^{n} \left(Y_i - \sum_{L=0}^{p} \sum_{1 \le j_1 \le \cdots \le j_L \le d} \beta_{j_1 \ldots j_L} \prod_{\ell=1}^{L} (R_{i,j_\ell} - r_{j_\ell}) \right)^2 K_h\left(R_i - r\right) \tag{A.1}$$

where $\beta = (\beta_{j_1 \ldots j_L})'_{1 \le j_1 \le \cdots \le j_L \le d, 0 \le L \le p}$,

$$K_h(R_i - r) = K\left(\frac{R_{i,1} - r_i}{h_1}, \ldots, \frac{R_{i,d} - r_d}{h_d} \right)$$

and each h_j is a sequence of positive constants (bandwidths) such that $h_j \to 0$ as $n \to \infty$. For notational convenience, we interpret $\sum_{1 \le j_1 \le \cdots \le j_L \le d} \beta_{j_1 \ldots j_L} \prod_{\ell=1}^{L} (R_{i,j_\ell} - r_{j_\ell}) = \beta_0$ when $L = 0$. We introduce some notations:

$$\boldsymbol{Y} := \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \ \boldsymbol{W} := \operatorname{diag}\left(K_h\left(R_1 - r\right), \ldots, K_h\left(R_n - r\right) \right),$$

$$\boldsymbol{R} := (\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n) = \begin{pmatrix} 1 & \cdots & 1 \\ (R_1 - r)_1 & \cdots & (R_n - r)_1 \\ \vdots & \ldots & \vdots \\ (R_1 - r)_p & \cdots & (R_n - r)_p \end{pmatrix} = \begin{pmatrix} 1 & \ldots & 1 \\ \check{\boldsymbol{R}}_1 & \ldots & \check{\boldsymbol{R}}_n \end{pmatrix},$$

where

$$(R_i - r)_L = \left(\prod_{\ell=1}^{L} (R_{i,j_\ell} - r_{j_\ell}) \right)'_{1 \le j_1 \le \cdots \le j_L \le d}.$$

The minimization problem (A.1) can be rewritten as

$$\hat{\beta}(r) = \arg\min_{\beta \in \mathbb{R}^D} (\boldsymbol{Y} - \boldsymbol{R}'\beta)' \boldsymbol{W} (\boldsymbol{Y} - \boldsymbol{R}'\beta) = \arg\min_{\beta \in \mathbb{R}^D} Q_n(\beta).$$

Then the first order condition of the problem (A.1) is given by

$$\frac{\partial}{\partial \beta} Q_n(\beta) = -2\boldsymbol{R}\boldsymbol{W}\boldsymbol{Y} + 2\boldsymbol{R}\boldsymbol{W}\boldsymbol{R}'\beta = 0.$$

Hence the solution of the problem (A.1) is given by

$$\begin{aligned}\hat{\beta}(r) &= (\boldsymbol{R}\boldsymbol{W}\boldsymbol{R}')^{-1}\boldsymbol{R}\boldsymbol{W}\boldsymbol{Y} \\ &= \left[\sum_{i=1}^{n} K_h\left(R_i - r\right)\boldsymbol{R}_i\boldsymbol{R}_i'\right]^{-1}\sum_{i=1}^{n} K_h\left(R_i - r\right)\boldsymbol{R}_i Y_i.\end{aligned}$$

Define

$$H := \operatorname{diag}(1, h_1, \ldots, h_d, h_1^2, h_1h_2, \ldots, h_d^2, \ldots, h_1^p, h_1^{p-1}h_2, \ldots, h_d^p) \in \mathbb{R}^{D\times D}.$$

THEOREM A.1 (Asymptotic normality of local-polynomial estimators) Under Assumptions 2.1, 2.2, 2.3 and 2.4, as $n \to \infty$, we have

$$\begin{aligned}&\sqrt{nh_1\cdots h_d}\left(H\left(\hat{\beta}(r) - M(r)\right) - S^{-1}B^{(d,p)}M_n^{(d,p)}(r)\right) \\ &\xrightarrow{d} N\left(\begin{pmatrix}0\\ \vdots \\ 0\end{pmatrix}, \frac{\sigma^2(r)}{f(r)}S^{-1}\mathcal{K}S^{-1}\right),\end{aligned}$$

where

$$\begin{aligned}M_n^{(d,p)}(r) &= \left(\frac{\partial_{j_1\ldots j_{p+1}}m(r)}{\boldsymbol{s}_{j_1\ldots j_{p+1}}!}\prod_{\ell=1}^{p+1}h_{j_\ell}\right)'_{1\le j_1\le\cdots\le j_{p+1}\le d} \\ &= \left(\frac{\partial_{1\ldots 1}m(r)}{(p+1)!}h_1^{p+1}, \frac{\partial_{1\ldots 2}m(r)}{p!}h_1^p h_2, \ldots, \frac{\partial_{d\ldots d}m(r)}{(p+1)!}h_d^{p+1}\right)' \in \mathbb{R}^{\bar{D}},\end{aligned}$$

$$B^{(d,p)} = \int \begin{pmatrix}1\\ \check{\boldsymbol{z}}\end{pmatrix}(\boldsymbol{z})'_{p+1}d\boldsymbol{z} \in \mathbb{R}^{D\times\bar{D}},\ \mathcal{K} = \int K^2(\boldsymbol{z})\begin{pmatrix}1\\ \check{\boldsymbol{z}}\end{pmatrix}(1\ \check{\boldsymbol{z}}')d\boldsymbol{z}.$$

PROOF: Define $h := (h_1, \ldots, h_d)'$ and for $r, y \in \mathbb{R}^d$, let $r \circ y = (r_1y_1, \cdots, r_dy_d)'$ be the Hadamard product. Considering Taylor's expansion of $m(r)$ around $r = (r_1, \ldots, r_d)'$,

$$m(R_i) = (1, \check{\boldsymbol{R}}_i')M(r) + \frac{1}{(p+1)!}\sum_{1\le j_1\le\cdots\le j_{p+1}\le d}\frac{(p+1)!}{\boldsymbol{s}_{j_1\ldots j_{p+1}}!}\partial_{j_1,\ldots,j_{p+1}}m(\tilde{R}_i)$$

$$\times \prod_{\ell=1}^{p+1} (R_{i,j_\ell} - r_{j_\ell}),$$

where $\tilde{R}_i = r + \theta_i(R_i - r)$ for some $\theta_i \in [0,1)$. Then we have

$$\begin{aligned}
&\hat{\beta}(r) - M(r) \\
&= (\boldsymbol{R}\boldsymbol{W}\boldsymbol{R}')^{-1}\boldsymbol{R}\boldsymbol{W}(\boldsymbol{Y} - \boldsymbol{R}'M(r)) \\
&= \left[\sum_{i=1}^{n} K_h\left(R_i - r\right)\begin{pmatrix} 1 \\ \check{\boldsymbol{R}}_i \end{pmatrix}(1\ \check{\boldsymbol{R}}_i')\right]^{-1} \sum_{i=1}^{n} K_h\left(R_i - r\right)\begin{pmatrix} 1 \\ \check{\boldsymbol{R}}_i \end{pmatrix} \\
&\quad \times \left(\varepsilon_i + \sum_{1 \le j_1 \le \cdots \le j_{p+1} \le d} \frac{1}{\boldsymbol{s}_{j_1 \ldots j_{p+1}}!}\partial_{j_1,\ldots,j_{p+1}} m(\tilde{R}_i)\prod_{\ell=1}^{p+1}(R_{i,j_\ell} - r_{j_\ell})\right).
\end{aligned}$$

This yields

$$\sqrt{nh_1 \cdots h_d} H(\hat{\beta}(r) - M(r)) = S_n^{-1}(V_n(r) + B_n(r)),$$

where

$$\begin{aligned}
S_n(r) &= \frac{1}{nh_1 \cdots h_d}\sum_{i=1}^{n} K_h\left(R_i - r\right) H^{-1}\begin{pmatrix} 1 \\ \check{\boldsymbol{R}}_i \end{pmatrix}(1\ \check{\boldsymbol{R}}_i')H^{-1}, \\
V_n(r) &= \frac{1}{\sqrt{nh_1 \cdots h_d}}\sum_{i=1}^{n} K_h\left(R_i - r\right) H^{-1}\begin{pmatrix} 1 \\ \check{\boldsymbol{R}}_i \end{pmatrix}\varepsilon_i \\
&\quad =: (V_{n,j_1 \ldots j_L}(r))'_{1 \le j_1 \le \cdots \le j_L \le d, 0 \le L \le p}, \\
B_n(r) &= \frac{1}{\sqrt{nh_1 \cdots h_d}}\sum_{i=1}^{n} K_h\left(R_i - r\right) H^{-1}\begin{pmatrix} 1 \\ \check{\boldsymbol{R}}_i \end{pmatrix} \\
&\qquad \times \sum_{1 \le j_1 \le \cdots \le j_{p+1} \le d} \frac{1}{\boldsymbol{s}_{j_1 \ldots j_{p+1}}!}\partial_{j_1,\ldots,j_{p+1}} m(\tilde{R}_i)\prod_{\ell=1}^{p+1}(R_{i,j_\ell} - r_{j_\ell}) \\
&\quad =: (B_{n,j_1 \ldots j_L}(\tilde{R}_i))'_{1 \le j_1 \le \cdots \le j_L \le d, 0 \le L \le p}.
\end{aligned}$$

(Step 1) Now we evaluate $S_n(r)$. For $1 \le j_{1,1} \le \cdots \le j_{1,L_1}, j_{2,1}, \ldots, j_{2,L_2} \le d, 0 \le$

$L_1, L_2 \le p$, we define

$$
\begin{aligned}
&I_{n,j_{1,1}\cdots j_{1,L_1},j_{2,1}\cdots j_{2,L_2}} \\
&:= \frac{1}{nh_1\cdots h_d}\sum_{i=1}^{n} K_h\left(R_i - r\right)\prod_{\ell_1=1}^{L_1}\left(\frac{R_{i,j_{\ell_1}} - r_{j_{\ell_1}}}{h_{j_{\ell_1}}}\right)\prod_{\ell_2=1}^{L_2}\left(\frac{R_{i,j_{\ell_2}} - r_{j_{\ell_2}}}{h_{j_{\ell_2}}}\right).
\end{aligned}
$$

Observe that

$$
\begin{aligned}
&E\left[I_{n,j_{1,1}\cdots j_{1,L_1},j_{2,1}\cdots j_{2,L_2}}\right] \\
&= \frac{1}{h_1\cdots h_d}E\left[K_h\left(R_i - r\right)\prod_{\ell_1=1}^{L_1}\left(\frac{R_{i,j_{\ell_1}} - r_{j_{\ell_1}}}{h_{j_{\ell_1}}}\right)\prod_{\ell_2=1}^{L_2}\left(\frac{R_{i,j_{\ell_2}} - r_{j_{\ell_2}}}{h_{j_{\ell_2}}}\right)\right] \\
&= \int\left(\prod_{\ell_1=1}^{L_1} z_{j_{\ell_1}}\right)\left(\prod_{\ell_2=1}^{L_2} z_{j_{\ell_2}}\right)K(z)f(r + h\circ z)dz \\
&= f(r)\kappa^{(1)}_{j_{1,1}\cdots j_{1,L_1}j_{2,1}\cdots j_{2,L_2}} + o(1).
\end{aligned}
$$

For the last equation, we used the dominated convergence theorem.

$$
\begin{aligned}
&\mathrm{Var}(I_{n,j_{1,1}\cdots j_{1,L_1},j_{2,1}\cdots j_{2,L_2}}) \\
&\quad= \frac{1}{n(h_1\cdots h_d)^2}\,\mathrm{Var}\left(K_h\left(R_1 - r\right)\prod_{\ell_1=1}^{L_1}\left(\frac{R_{i,j_{\ell_1}} - r_{j_{\ell_1}}}{h_{j_{\ell_1}}}\right)\prod_{\ell_2=1}^{L_2}\left(\frac{R_{i,j_{\ell_2}} - r_{j_{\ell_2}}}{h_{j_{\ell_2}}}\right)\right) \\
&\quad= \frac{1}{nh_1\cdots h_d}\left\{\int\prod_{\ell_1=1}^{L_1}\left(\frac{R_{i,j_{\ell_1}} - r_{j_{\ell_1}}}{h_{j_{\ell_1}}}\right)^2\prod_{\ell_2=1}^{L_2}\left(\frac{R_{i,j_{\ell_2}} - r_{j_{\ell_2}}}{h_{j_{\ell_2}}}\right)^2 K^2(z)f(r + h\circ z)dz\right. \\
&\qquad\left.-h_1\cdots h_d\left(\int\prod_{\ell_1=1}^{L_1}\left(\frac{R_{i,j_{\ell_1}} - r_{j_{\ell_1}}}{h_{j_{\ell_1}}}\right)\prod_{\ell_2=1}^{L_2}\left(\frac{R_{i,j_{\ell_2}} - r_{j_{\ell_2}}}{h_{j_{\ell_2}}}\right)K(z)f(r + h\circ z)dz\right)^2\right\} \\
&\quad= \frac{1}{nh_1\cdots h_d}\left(f(r)\kappa^{(2)}_{j_{1,1}\cdots j_{1,L_1}j_{2,1}\cdots j_{2,L_2}j_{1,1}\cdots j_{1,L_1}j_{2,1}\cdots j_{2,L_2}} + o(1)\right) \\
&\qquad- \frac{1}{n}(f(r)\kappa^{(1)}_{j_{1,1}\cdots j_{1,L_1}j_{2,1}\cdots j_{2,L_2}} + o(1))^2 \text{ (DCT)} \\
&\quad= \frac{f(r)\kappa^{(2)}_{j_{1,1}\cdots j_{1,L_1}j_{2,1}\cdots j_{2,L_2}j_{1,1}\cdots j_{1,L_1}j_{2,1}\cdots j_{2,L_2}}}{nh_1\cdots h_d} + o\left(\frac{1}{nh_1\cdots h_d}\right).
\end{aligned}
$$

Then for any $\rho > 0$,

$$
\begin{aligned}
&P\left(|I_{n,j_{1,1}\dots j_{1,L_1},j_{2,1}\dots j_{2,L_2}} - f(r)\kappa^{(1)}_{j_{1,1}\dots j_{1,L_1}j_{2,1}\dots j_{2,L_2}}| > \rho\right) \\
&\leq \rho^{-1}\left\{\mathrm{Var}(I_{n,j_{1,1}\dots j_{1,L_1},j_{2,1}\dots j_{2,L_2}}) + \left(E[I_{n,j_{1,1}\dots j_{1,L_1},j_{2,1}\dots j_{2,L_2}}] - f(r)\kappa^{(1)}_{j_{1,1}\dots j_{1,L_1}j_{2,1}\dots j_{2,L_2}}\right)^2\right\} \\
&= O\left(\frac{1}{nh_1\cdots h_d}\right) + o(1) = o(1).
\end{aligned}
$$

This yields $I_{n,j_{1,1}\dots j_{1,L_1},j_{2,1}\dots j_{2,L_2}} \xrightarrow{p} f(r)\kappa^{(1)}_{j_{1,1}\dots j_{1,L_1}j_{2,1}\dots j_{2,L_2}}$. Hence we have $S_n(r) \xrightarrow{p} f(r)S$.

(Step 2) Now we evaluate $V_n(r)$. For any $t = (t_0, t_1, \dots, t_d, t_{11}, \dots, t_{dd}, \dots, t_{1\dots 1}, \dots, t_{d\dots d})' \in \mathbb{R}^D$, we define

$$
\begin{aligned}
R_{n,i,j_1\dots j_L} &:= \frac{1}{\sqrt{nh_1\cdots h_d}} K_h\left(R_i - r\right)\prod_{\ell=1}^{L}\left(\frac{R_{i,j_\ell} - r_{j_\ell}}{h_{j_\ell}}\right)\varepsilon_i,\ 1 \leq j_1, \dots, j_L \leq d, \\
Z_{n,i} &:= \sum_{L=0}^{p}\sum_{1\leq j_1\leq\cdots\leq j_L\leq d} t_{j_1\dots j_L} R_{n,i,j_1\dots j_L}.
\end{aligned}
$$

Observe that

$$
\begin{aligned}
\sigma^2_{n,j_1\dots j_L} &:= \mathrm{Var}\left(\sum_{i=1}^{n} R_{n,i,j_1\dots j_L}\right) = \frac{1}{h_1\cdots h_d}E\left[\varepsilon_i^2 K_h^2\left(R_1 - r\right)\prod_{\ell=1}^{L}\left(\frac{R_{1,j_\ell} - r_{j_\ell}}{h_{j_\ell}}\right)^2\right] \\
&= \frac{1}{h_1\cdots h_d}E\left[\sigma^2(R_i)K_h^2\left(R_1 - r\right)\prod_{\ell=1}^{L}\left(\frac{R_{1,j_\ell} - r_{j_\ell}}{h_{j_\ell}}\right)^2\right] \\
&= \int \sigma^2(r + h\circ z)\left(\prod_{\ell=1}^{L} z_{j_\ell}^2\right)K^2(z)f(r + h\circ z)dz \\
&= \sigma^2(r)f(r)\kappa^{(2)}_{j_1\dots j_L j_1\dots j_L} + o(1).
\end{aligned}
$$

For the last equation, we used the dominated convergence theorem. Moreover, for $1 \leq j_{1,1} \leq \cdots \leq j_{1,L_1} \leq d$ and $1 \leq j_{2,1} \leq \cdots \leq j_{2,L_2} \leq d$, we have

$$
\begin{aligned}
&\mathrm{Cov}(V_{n,j_{1,1}\dots j_{1,L_1}}(r), V_{n,j_{2,1}\dots j_{2,L_2}}(r)) \\
&= \frac{1}{h_1\cdots h_d}E\left[\sigma^2(R_i)K_h^2\left(R_i - r\right)\prod_{\ell_1=1}^{L_1}\left(\frac{R_{i,j_{1,\ell_1}} - r_{j_{1,\ell_1}}}{h_{j_{1,\ell_1}}}\right)\prod_{\ell_2=1}^{L_2}\left(\frac{R_{i,j_{2,\ell_2}} - r_{j_{2,\ell_2}}}{h_{j_{2,\ell_2}}}\right)\right]
\end{aligned}
$$

$$= \int \sigma^2(r + h \circ z) \left(\prod_{\ell_1=1}^{L_1} z_{j_{1,\ell_1}} \right) \left(\prod_{\ell_2=1}^{L_2} z_{j_{2,\ell_2}} \right) K^2(z) f(r + h \circ z) dz$$
$$= \sigma^2(r) f(r) \kappa^{(2)}_{j_{1,1} \cdots j_{1,L_1} j_{2,1} \cdots j_{2,L_2}} + o(1).$$

For the last equation, we used the dominated convergence theorem. For sufficiently large n, we have

$$\sum_{i=1}^{n} E[|Z_{n,i}|^{2+\delta}]$$
$$= \frac{1}{n^{\delta/2}(h_1 \cdots h_d)^{1+\delta/2}} E\left[|\varepsilon_i|^{2+\delta} \left| K_h\left(R_i - r\right)\right|^{2+\delta} \right.$$
$$\left. \times \left| \sum_{L=0}^{p} \sum_{1 \le j_1 \le \cdots \le j_L \le d} t_{j_1 \ldots j_L} \prod_{\ell=1}^{L} \left(\frac{R_{i,j_\ell} - r_{j_\ell}}{h_{j_\ell}} \right) \right|^{2+\delta} \right]$$
$$\le \frac{U(r)}{(nh_1 \cdots h_d)^{\delta/2}} \int \left| \sum_{L=0}^{p} \sum_{1 \le j_1 \le \cdots \le j_L \le d} t_{j_1 \ldots j_L} \prod_{\ell=1}^{L} z_{j_\ell} \right|^{2+\delta} |K(z)|^{2+\delta} f(r + h \circ z) dz$$
$$= \frac{U(r) f(r)}{(nh_1 \cdots h_d)^{\delta/2}} \int \left| \sum_{L=0}^{p} \sum_{1 \le j_1 \le \cdots \le j_L \le d} t_{j_1 \ldots j_L} \prod_{\ell=1}^{L} z_{j_\ell} \right|^{2+\delta} |K(z)|^{2+\delta} dz + o(1)$$
$$= o(1).$$

For the second equation, we used the dominated convergence theorem. Thus, Lyapounov's condition is satisfied for $\sum_{i=1}^{n} Z_{n,i}$. Therefore, by Cramér-Wold device, we have

$$V_n(r) \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \sigma^2(r) f(r) \mathcal{K} \right).$$

(Step 3) Now we evaluate $B_n(r)$. Decompose

$$B_{n,j_1 \ldots j_L}(\tilde{R}_i) = \left\{ B_{n,j_1 \ldots j_L}(\tilde{R}_i) - B_{n,j_1 \ldots j_L}(r) - E\left[B_{n,j_1 \ldots j_L}(\tilde{R}_i) - B_{n,j_1 \ldots j_L}(r) \right] \right\}$$
$$+ E\left[B_{n,j_1 \ldots j_L}(\tilde{R}_i) - B_{n,j_1 \ldots j_L}(r) \right]$$

$$
\begin{aligned}
&+\{B_{n,j_1\dots j_L}(r)-E\left[B_{n,j_1\dots j_L}(r)\right]\}\\
&+E\left[B_{n,j_1\dots j_L}(r)\right]\\
=:&\sum_{\ell=1}^{4}B_{n,j_1\dots j_L\ell}.
\end{aligned}
$$

Define $N_r(h) := \prod_{j=1}^{d}[r_j - C_K h_j, r_j + C_K h_j]$. For $B_{n,j_1\dots j_L 1}$,

$$
\begin{aligned}
&\mathrm{Var}(B_{n,j_1\dots j_L 1})\\
&\leq \frac{1}{\{(p+1)!\}^2 h_1\cdots h_d}E\left[K_h^2\left(R_i-r\right)\prod_{\ell=1}^{L}\left(\frac{R_{i,j_\ell}-r_{j_\ell}}{h_{j_\ell}}\right)^2\right.\\
&\quad\times\sum_{1\leq j_{1,1}\leq\cdots\leq j_{1,p+1}\leq d,1\leq j_{2,1}\leq\cdots\leq j_{2,p+1}\leq d}\frac{1}{\boldsymbol{s}_{j_{1,1}\dots j_{1,p+1}}!}\frac{1}{\boldsymbol{s}_{j_{2,1}\dots j_{2,p+1}}!}\\
&\quad\times(\partial_{j_{1,1}\dots j_{1,p+1}}m(\tilde{R}_i)-\partial_{j_{1,1}\dots j_{1,p+1}}m(r))(\partial_{j_{2,1}\dots j_{2,p+1}}m(\tilde{R}_i)-\partial_{j_{2,1}\dots j_{2,p+1}}m(r))\\
&\quad\left.\times\prod_{\ell_1=1}^{p+1}(R_{i,j_{1,\ell_1}}-r_{j_{1,\ell_1}})\prod_{\ell_2=1}^{p+1}(R_{i,j_{2,\ell_2}}-r_{j_{2,\ell_2}})\right]\\
&\leq\frac{1}{\{(p+1)!\}^2}\max_{1\leq j_1\leq\cdots\leq j_{p+1}\leq d}\sup_{y\in N_r(h)}|\partial_{j_1\dots j_{p+1}}m(y)-\partial_{j_1\dots j_{p+1}}m(r)|^2\\
&\qquad\times\sum_{1\leq j_{1,1}\leq\cdots\leq j_{1,p+1}\leq d,1\leq j_{2,1}\leq\cdots\leq j_{2,p+1}\leq d}\prod_{\ell_1=1}^{p+1}h_{j_{1,\ell_1}}\prod_{\ell_2=1}^{p+1}h_{j_{2,\ell_2}}\\
&\qquad\times\int\left(\prod_{\ell=1}^{L}|z_{j_\ell}|\prod_{\ell_1=1}^{p+1}|z_{j_{1,\ell_1}}|\prod_{\ell_2=1}^{p+1}|z_{j_{2,\ell_2}}|\right)K^2(z)f(r+h\circ z)dz\\
&=o\left(\sum_{1\leq j_{1,1}\leq\cdots\leq j_{1,p+1}\leq d,1\leq j_{2,1}\leq\cdots\leq j_{2,p+1}\leq d}\prod_{\ell_1=1}^{p+1}h_{j_{1,\ell_1}}\prod_{\ell_2=1}^{p+1}h_{j_{2,\ell_2}}\right).
\end{aligned}
\tag{A.2}
$$

Then we have $B_{n,j_1\dots j_L 1} = o_p(1)$.

For $B_{n,j_1\dots j_L 2}$,

$$
\begin{aligned}
&|B_{n,j_1\dots j_L 2}|\\
&\leq\frac{1}{(p+1)!}\max_{1\leq j_1,\dots,j_{p+1}\leq d}\sup_{y\in N_r(h)}|\partial_{j_1\dots j_{p+1}}m(y)-\partial_{j_1\dots j_{p+1}}m(r)|
\end{aligned}
$$

$$\times \sqrt{nh_1 \cdots h_d} \sum_{1 \le j_{1,1} \le \cdots \le j_{1,p+1} \le d} \prod_{\ell_1=1}^{p+1} h_{j_{1,\ell_1}} \int \left(\prod_{\ell=1}^{L} |z_{j_\ell}| \prod_{\ell_1=1}^{p+1} |z_{j_{1,\ell_1}}| \right) |K(z)| f(r + h \circ z) dz$$

$$= o(1). \tag{A.3}$$

For $B_{n,j_1 \dots j_L 3}$,

$$\begin{aligned} &\mathrm{Var}(B_{n,j_1 \dots j_L 3}) \\ &\le \frac{1}{\{(p+1)!\}^2} \sum_{1 \le j_{1,1} \le \cdots \le j_{1,p+1} \le d, 1 \le j_{2,1} \le \cdots \le j_{2,p+1} \le d} \partial_{j_{1,1} \dots j_{1,p+1}} m(r) \partial_{j_{2,1} \dots j_{2,p+1}} m(r) \\ &\quad \times \prod_{\ell_1=1}^{p+1} h_{j_{1,\ell_1}} \prod_{\ell_2=1}^{p+1} h_{j_{2,\ell_2}} \int \left(\prod_{\ell=1}^{L} z_{j_\ell}^2 \prod_{\ell_1}^{p+1} |z_{j_{1,\ell_1}}| \prod_{\ell_2=1}^{p+1} |z_{j_{2,\ell_2}}| \right) K^2(z) f(r + h \circ z) dz \\ &= o(1). \end{aligned} \tag{A.4}$$

Then we have $B_{n,j_1 \dots j_L 3} = o_p(1)$.

For $B_{n,j_1 \dots j_L 4}$,

$$\begin{aligned} &B_{n,j_1 \dots j_L 4} \\ &= \sqrt{nh_1 \cdots h_d} \sum_{1 \le j_{1,1} \le \cdots \le j_{1,p+1} \le d} \frac{\partial_{j_{1,1} \dots j_{1,p+1}} m(r)}{\boldsymbol{s}_{j_{1,1} \dots j_{1,p+1}}!} \\ &\quad \times \prod_{\ell_1=1}^{p+1} h_{j_{1,\ell_1}} \int \left(\prod_{\ell=1}^{L} z_{j_\ell} \prod_{\ell_1=1}^{p+1} z_{j_{1,\ell_1}} \right) K(z) f(r + h \circ z) dz \\ &= f(r) \sqrt{nh_1 \cdots h_d} \sum_{1 \le j_{1,1} \le \cdots \le j_{1,p+1} \le d} \frac{\partial_{j_{1,1} \dots j_{1,p+1}} m(r)}{\boldsymbol{s}_{j_{1,1} \dots j_{1,p+1}}!} \prod_{\ell_1=1}^{p+1} h_{j_{1,\ell_1}} \kappa^{(1)}_{j_1 \dots j_L j_{1,1} \dots j_{1,p+1}} + o(1). \end{aligned} \tag{A.5}$$

Combining (A.2)-(A.5),

$$\begin{aligned} B_{n,j_1 \dots j_L}(\tilde{R}_i) = f(r) \sqrt{nh_1 \cdots h_d} &\sum_{1 \le j_{1,1} \le \cdots \le j_{1,p+1} \le d} \frac{\partial_{j_{1,1} \dots j_{1,p+1}} m(r)}{\boldsymbol{s}_{j_{1,1} \dots j_{1,p+1}}!} \\ &\times \prod_{\ell_1=1}^{p+1} h_{j_{1,\ell_1}} \kappa^{(1)}_{j_1 \dots j_L j_{1,1} \dots j_{1,p+1}} + o_p(1). \end{aligned}$$

(Step 4) Combining the results in Steps1-3, we have

$$
\begin{aligned}
A_n(r) &:= V_n(r) + \Big(B_n(r) - f(r)\sqrt{nh_1 \cdots h_d}\, (b_{n,j_1 \dots j_L}(r))'_{1 \le j_1 \le \cdots \le j_L \le d, 0 \le L \le p} \Big) \\
&\xrightarrow{d} N\left(\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \sigma^2(r) f(r) \mathcal{K} \right).
\end{aligned}
$$

This yields the desired result. *Q.E.D.*

Remark A.1 (General form of the MSE of $\widehat{\partial_{j_1 \dots j_L} m}(r)$) Define

$$
\begin{aligned}
\boldsymbol{b}_n^{(d,p)}(r) &:= B^{(d,p)} M_n^{(d,p)}(r) \\
&= (b_{n,0}(r), b_{n,1}(r), \dots, b_{n,d}(r), \\
&\qquad b_{n,11}(r), b_{n,12}(r), \dots, b_{n,dd}(r), \dots, b_{n,1\dots,1}(r), b_{n,1\dots 2}(r), \dots, b_{n,d\dots d}(r))'
\end{aligned}
$$

and let $e_{j_1 \dots j_L} = (0, \dots, 0, 1, 0, \dots, 0)'$ be a D-dimensional vector such that $e'_{j_1 \dots j_L} \boldsymbol{b}_n^{(d,p)}(r) = b_{j_1 \dots j_L}(r)$. Theorem A.1 yields that

$$
b_{n,j_1,\dots,j_L}(r) := \sum_{1 \le j_{1,1} \le \cdots \le j_{1,p+1} \le d} \frac{\partial_{j_{1,1} \dots j_{1,p+1}} m(r)}{\boldsymbol{s}_{j_{1,1} \dots j_{1,p+1}}!} \prod_{\ell_1 = 1}^{p+1} h_{j_{1,\ell_1}} \kappa^{(1)}_{j_1 \dots j_L j_{1,1} \dots j_{1,p+1}},
$$

for $1 \le j_1 \le \cdots \le j_L \le d$, $0 \le L \le p$ and

$$
\begin{aligned}
&\mathrm{MSE}(\widehat{\partial_{j_1 \dots j_L} m}(r)) \\
&= \left\{ \boldsymbol{s}_{j_1 \dots j_L}! \frac{(S^{-1} e_{j_1 \dots j_L})' B^{(d,p)} M_n^{(d,p)}(r)}{\prod_{\ell=1}^{L} h_{j_\ell}} \right\}^2 \\
&\quad + (\boldsymbol{s}_{j_1 \dots j_L}!)^2 \frac{\sigma^2(r)}{nh_1 \cdots h_d \times \left(\prod_{\ell=1}^{L} h_{j_\ell} \right)^2 f(r)} e'_{j_1 \dots j_L} S^{-1} \mathcal{K} S^{-1} e_{j_1 \dots j_L}.
\end{aligned}
$$

A.2. *Higher-order bias*

In this section, we derive higher-order biases of local-polynomial estimators. Suppose that Assumptions 2.1, 2.2, 2.3 and 2.4 hold. Further, we assume that

- the density function f is continuously differentiable on U_r.
- the mean function m is $(p+2)$-times continuously differentiable on U_r.

Recall that

$$\sqrt{nh_1\cdots h_d}H(\hat{\beta}(r)-M(r)) = S_n^{-1}(V_n(r)+B_n(r)),$$

where

$$\begin{aligned}
S_n(r) &= \frac{1}{nh_1\cdots h_d}\sum_{i=1}^n K_h\left(R_i - r\right)H^{-1}\begin{pmatrix} 1 \\ \check{\boldsymbol{R}}_i \end{pmatrix}(1\ \check{\boldsymbol{R}}_i')H^{-1},\\
V_n(r) &= \frac{1}{\sqrt{nh_1\cdots h_d}}\sum_{i=1}^n K_h\left(R_i - r\right)H^{-1}\begin{pmatrix} 1 \\ \check{\boldsymbol{R}}_i \end{pmatrix}\varepsilon_i =: (V_{n,j_1\dots j_L}(r))'_{1\le j_1\le\cdots\le j_L\le d, 0\le L\le p},\\
B_n(r) &= \frac{1}{\sqrt{nh_1\cdots h_d}}\sum_{i=1}^n K_h\left(R_i - r\right)H^{-1}\begin{pmatrix} 1 \\ \check{\boldsymbol{R}}_i \end{pmatrix}\\
&\quad\times\left\{\sum_{1\le j_1\le\cdots\le j_{p+1}\le d}\frac{1}{\boldsymbol{s}_{j_1\dots j_{p+1}}!}\partial_{j_1,\dots,j_{p+1}}m(r)\prod_{\ell=1}^{p+1}(R_{i,j_\ell}-r_{j_\ell})\right.\\
&\qquad\left.+\sum_{1\le j_1\le\cdots\le j_{p+2}\le d}\frac{1}{\boldsymbol{s}_{j_1\dots j_{p+2}}!}\partial_{j_1,\dots,j_{p+2}}m(\tilde{R}_i)\prod_{\ell=1}^{p+2}(R_{i,j_\ell}-r_{j_\ell})\right\}\\
&=: (B_{n,j_1\dots j_L}(\tilde{R}))'_{1\le j_1\le\cdots\le j_L\le d, 0\le L\le p}.
\end{aligned}$$

Now we focus on $B_{n,j_1\dots j_L}(\tilde{R})$.

$$\begin{aligned}
&B_{n,j_1\dots j_L}(\tilde{R})\\
&= \frac{1}{\sqrt{nh_1\cdots h_d}}\sum_{i=1}^n K_h\left(R_i - r\right)\left(\prod_{\ell=1}^{L}\frac{R_{i,j_\ell}-r_{j_\ell}}{h_{j_\ell}}\right)\\
&\quad\times\left\{\sum_{1\le j_{1,1}\le\cdots\le j_{1,p+1}\le d}\frac{1}{\boldsymbol{s}_{j_{1,1}\dots j_{1,p+1}}!}\partial_{j_{1,1},\dots,j_{1,p+1}}m(r)\prod_{\ell_1=1}^{p+1}(R_{i,j_{1,\ell_1}}-r_{j_{1,\ell_1}})\right.
\end{aligned}$$

$$
\begin{aligned}
&+ \sum_{1 \le j_{1,1} \le \cdots \le j_{1,p+2} \le d} \frac{1}{\boldsymbol{s}_{j_{1,1} \ldots j_{1,p+2}}!} \partial_{j_{1,1},\ldots,j_{1,p+2}} m(\tilde{R}_i) \prod_{\ell_1=1}^{p+2} (R_{i,j_{1,\ell_1}} - r_{j_{1,\ell_1}}) \Bigg\} \\
&=: \mathbb{B}_{n,1}(r) + \mathbb{B}_{n,2}(\tilde{R}).
\end{aligned}
$$

For $\mathbb{B}_{n,1}(r)$,

$$
\begin{aligned}
E[\mathbb{B}_{n,1}(r)] &= \sqrt{\frac{n}{h_1 \cdots h_d}} E\left[K_h(R_1 - r) \left(\prod_{\ell=1}^{L} \frac{R_{1,j_\ell} - r_{j_\ell}}{h_{j_\ell}} \right) \right. \\
&\qquad \left. \times \sum_{1 \le j_{1,1} \le \cdots \le j_{1,p+1} \le d} \frac{1}{\boldsymbol{s}_{j_{1,1} \ldots j_{1,p+1}}!} \partial_{j_{1,1},\ldots,j_{1,p+1}} m(r) \prod_{\ell_1=1}^{p+1} (R_{1,j_{1,\ell_1}} - r_{j_{1,\ell_1}}) \right] \\
&= \sqrt{n h_1 \cdots h_d} \sum_{1 \le j_{1,1} \le \cdots \le j_{1,p+1} \le d} \frac{1}{\boldsymbol{s}_{j_{1,1} \ldots j_{1,p+1}}!} \partial_{j_{1,1},\ldots,j_{1,p+1}} m(r) \prod_{\ell_1=1}^{p+1} h_{j_{1,\ell_1}} \\
&\qquad \times \int \prod_{\ell=1}^{L} z_{j_\ell} \prod_{\ell_1=1}^{p+1} z_{j_{1,\ell_1}} K(z) f(r + h \circ z) dz \\
&= \sqrt{n h_1 \cdots h_d} \sum_{1 \le j_{1,1} \le \cdots \le j_{1,p+1} \le d} \frac{1}{\boldsymbol{s}_{j_{1,1} \ldots j_{1,p+1}}!} \partial_{j_{1,1},\ldots,j_{1,p+1}} m(r) \prod_{\ell_1=1}^{p+1} h_{j_{1,\ell_1}} \\
&\qquad \times \left(f(r) \int \prod_{\ell=1}^{L} z_{j_\ell} \prod_{\ell_1=1}^{p+1} z_{j_{1,\ell_1}} K(z) dz \right. \\
&\qquad\quad \left. + \sum_{k=1}^{d} \partial_k f(r) h_k \int z_k \prod_{\ell=1}^{L} z_{j_\ell} \prod_{\ell_1=1}^{p+1} z_{j_{1,\ell_1}} K(z) dz \right) (1 + o(1)).
\end{aligned} \tag{A.6}
$$

$$
\begin{aligned}
&\mathrm{Var}(\mathbb{B}_{n,1}(r)) \\
&\le \sum_{1 \le j_{1,1} \le \cdots \le j_{1,p+1} \le d, 1 \le j_{2,1} \le \cdots \le j_{2,p+1} \le d} \partial_{j_{1,1} \ldots j_{1,p+1}} m(r) \partial_{j_{2,1} \ldots j_{2,p+1}} m(r) \\
&\quad \times \prod_{\ell_1=1}^{p+1} h_{j_{1,\ell_1}} \prod_{\ell_2=1}^{p+1} h_{j_{2,\ell_2}} \int \left(\prod_{\ell=1}^{L} z_{j_\ell}^2 \prod_{\ell_1=1}^{p+1} |z_{j_{1,\ell_1}}| \prod_{\ell_2=1}^{p+1} |z_{j_{2,\ell_2}}| \right) K^2(z) f(r + h \circ z) dz \\
&= O\left(\left(\sum_{1 \le j_1 \le \cdots \le j_{p+1} \le d} \prod_{\ell=1}^{p+1} h_{j_\ell} \right)^2 \right).
\end{aligned} \tag{A.7}
$$

For $\mathbb{B}_{n,2}(\tilde{R})$,

$$\begin{aligned}\mathbb{B}_{n,2}(\tilde{R}) &= \left\{\mathbb{B}_{n,2}(\tilde{R}) - \mathbb{B}_{n,2}(r) - E[\mathbb{B}_{n,2}(\tilde{R}) - \mathbb{B}_{n,2}(r)]\right\} \\ &\quad + E[\mathbb{B}_{n,2}(\tilde{R}) - \mathbb{B}_{n,2}(r)] \\ &\quad + \mathbb{B}_{n,2}(r) - E[\mathbb{B}_{n,2}(r)] \\ &\quad + E[\mathbb{B}_{n,2}(r)] \\ &=: \sum_{\ell=1}^{4} \mathbb{B}_{n,2\ell}.\end{aligned}$$

Define $N_r(h) := \prod_{j=1}^{d}[r_j - C_K h_j, r_j + C_K h_j]$. For $\mathbb{B}_{n,21}$,

$$\begin{aligned}&\mathrm{Var}(\mathbb{B}_{n,21}) \\ &\leq \frac{1}{h_1 \cdots h_d} E\left[K_h^2(R_i - r) \prod_{\ell=1}^{L} \left(\frac{R_{i,j_\ell} - r_{j_\ell}}{h_{j_\ell}}\right)^2\right. \\ &\quad \times \sum_{1 \leq j_{1,1} \leq \cdots \leq j_{1,p+2} \leq d, 1 \leq j_{2,1} \leq \cdots \leq j_{2,p+2} \leq d} \frac{1}{\boldsymbol{s}_{j_{1,1} \ldots j_{1,p+2}}!} \frac{1}{\boldsymbol{s}_{j_{2,1} \ldots j_{2,p+2}}!} \\ &\quad \times (\partial_{j_{1,1} \ldots j_{1,p+2}} m(\tilde{R}_i) - \partial_{j_{1,1} \ldots j_{1,p+2}} m(r))(\partial_{j_{2,1} \ldots j_{2,p+2}} m(\tilde{R}_i) - \partial_{j_{2,1} \ldots j_{2,p+2}} m(r)) \\ &\quad \left. \times \prod_{\ell_1=1}^{p+2} (R_{i,j_{1,\ell_1}} - r_{j_{1\ell_1}}) \prod_{\ell_2=1}^{p+2} (R_{i,j_{2,\ell_2}} - r_{j_{2\ell_2}})\right] \\ &\leq \max_{1 \leq j_1 \leq \cdots \leq j_{p+2} \leq d} \sup_{y \in N_r(h)} |\partial_{j_1 \ldots j_{p+2}} m(y) - \partial_{j_1 \ldots j_{p+2}} m(r)|^2 \\ &\qquad \times \sum_{1 \leq j_{1,1} \leq \cdots \leq j_{1,p+2} \leq d, 1 \leq j_{2,1} \leq \cdots \leq j_{2,p+2} \leq d} \prod_{\ell_1=1}^{p+2} h_{j_{1,\ell_1}} \prod_{\ell_2=1}^{p+2} h_{j_{2,\ell_2}} \\ &\qquad \times \int \left(\prod_{\ell=1}^{L} |z_{j_\ell}| \prod_{\ell_1=1}^{p+2} |z_{j_{1,\ell_1}}| \prod_{\ell_2=1}^{p+2} |z_{j_{2,\ell_2}}|\right) K^2(z) f(r + h \circ z) dz \\ &= o\left(\left(\sum_{1 \leq j_1 \leq \cdots \leq j_{p+2} \leq d} \prod_{\ell=1}^{p+2} h_{j_\ell}\right)^2\right). \qquad \text{(A.8)}\end{aligned}$$

For $\mathbb{B}_{n,22}$,

$$|\mathbb{B}_{n,22}|$$

$$\le \max_{1\le j_1,\dots,j_{p+2}\le d} \sup_{y\in N_r(h)} |\partial_{j_1\dots j_{p+2}}m(y) - \partial_{j_1\dots j_{p+2}}m(r)|$$

$$\times \sqrt{nh_1\cdots h_d} \sum_{1\le j_{1,1}\le\cdots\le j_{1,p+2}\le d} \prod_{\ell_1=1}^{p+2} h_{j_{1,\ell_1}} \int \left(\prod_{\ell=1}^{L} |z_{j_\ell}| \prod_{\ell_1=1}^{p+2} |z_{j_{1,\ell_1}}|\right) |K(z)| f(r+h\circ z)dz$$

$$= o\left(\sqrt{nh_1\cdots h_d} \sum_{1\le j_{1,1}\le\cdots\le j_{1,p+2}\le d} \prod_{\ell_1=1}^{p+2} h_{j_{1,\ell_1}}\right). \tag{A.9}$$

For $\mathbb{B}_{n,23}$,

$$\mathrm{Var}(\mathbb{B}_{n,23})$$

$$\le \sum_{1\le j_{1,1}\le\cdots\le j_{1,p+2}\le d, 1\le j_{2,1}\le\cdots\le j_{2,p+2}\le d} \partial_{j_{1,1}\dots j_{1,p+2}}m(r)\partial_{j_{2,1}\dots j_{2,p+2}}m(r)$$

$$\times \prod_{\ell_1=1}^{p+2} h_{j_{1,\ell_1}} \prod_{\ell_2=1}^{p+2} h_{j_{2,\ell_2}} \int \left(\prod_{\ell=1}^{L} z_{j_\ell}^2 \prod_{\ell_1}^{p+2} |z_{j_{1,\ell_1}}| \prod_{\ell_2=1}^{p+2} |z_{j_{2,\ell_2}}|\right) K^2(z) f(r+h\circ z)dz$$

$$= O\left(\left(\sum_{1\le j_1\le\cdots\le j_{p+2}\le d} \prod_{\ell=1}^{p+2} h_{j_\ell}\right)^2\right). \tag{A.10}$$

For $\mathbb{B}_{n,24}$,

$$\mathbb{B}_{n,24} = \sqrt{nh_1\cdots h_d} \sum_{1\le j_{1,1}\le\cdots\le j_{1,p+2}\le d} \frac{\partial_{j_{1,1}\dots j_{1,p+2}}m(r)}{\boldsymbol{s}_{j_{1,1}\dots j_{1,p+2}}!}$$

$$\times \prod_{\ell_1=1}^{p+2} h_{j_{1,\ell_1}} \int \left(\prod_{\ell=1}^{L} z_{j_\ell} \prod_{\ell_1=1}^{p+2} z_{j_{1,\ell_1}}\right) K(z) f(r+h\circ z)dz$$

$$= f(r)\sqrt{nh_1\cdots h_d}$$

$$\times \left(\sum_{1\le j_{1,1}\le\cdots\le j_{1,p+2}\le d} \frac{\partial_{j_{1,1}\dots j_{1,p+2}}m(r)}{\boldsymbol{s}_{j_{1,1}\dots j_{1,p+2}}!} \prod_{\ell_1=1}^{p+2} h_{j_{1,\ell_1}} \int \left(\prod_{\ell=1}^{L} z_{j_\ell} \prod_{\ell_1=1}^{p+2} z_{j_{1,\ell_1}}\right) K(z)dz\right)(1+o(1)). \tag{A.11}$$

Combining (A.6)-(A.11),

$$B_{n,j_1\dots j_L}(\tilde{R})$$

$$= \sqrt{nh_1\cdots h_d} \sum_{1\le j_{1,1}\le\cdots\le j_{1,p+1}\le d} \frac{1}{\boldsymbol{s}_{j_{1,1}\dots j_{1,p+1}}!} \partial_{j_{1,1},\dots,j_{1,p+1}}m(r) \prod_{\ell_1=1}^{p+1} h_{j_{1,\ell_1}}$$

$$\times \left(f(r) \int \prod_{\ell=1}^{L} z_{j_\ell} \prod_{\ell_1=1}^{p+1} z_{j_{1,\ell_1}} K(z)dz + \sum_{k=1}^{d} \partial_k f(r) h_k \int \left(z_k \prod_{\ell=1}^{L} z_{j_\ell} \prod_{\ell_1=1}^{p+1} z_{j_{1,\ell_1}} \right) K(z)dz \right) (1+o(1)).$$

$$+ \sqrt{nh_1 \cdots h_d}$$

$$\times \left(f(r) \sum_{1 \le j_{1,1} \le \cdots \le j_{1,p+2} \le d} \frac{\partial_{j_{1,1} \ldots j_{1,p+2}} m(r)}{\boldsymbol{s}_{j_{1,1} \ldots j_{1,p+2}}!} \prod_{\ell_1=1}^{p+2} h_{j_{1,\ell_1}} \int \left(\prod_{\ell=1}^{L} z_{j_\ell} \prod_{\ell_1=1}^{p+2} z_{j_{1,\ell_1}} \right) K(z)dz \right) (1+o(1)).$$

A.2.1. *Higher-order bias of the local-linear estimator*

For local-linear estimators (i.e., $d = 2, p = 1$), we have

$$\begin{aligned} b_{n,0} = & \frac{f(r)}{2} \sum_{j,k=1}^{2} \partial_{jk} m(r) h_j h_k \int z_k z_j K(z)dz \\ & + \sum_{\ell=1}^{2} \frac{\partial_\ell f(r)}{2} \sum_{j,k=1}^{2} \partial_{jk} m(r) h_j h_k h_\ell \int z_j z_k z_\ell K(z)dz \\ & + \frac{f(r)}{6} \sum_{j,k,\ell=1}^{2} \partial_{jk\ell} m(r) h_j h_k h_\ell \int z_j z_k z_\ell K(z)dz, \end{aligned}$$

$$\begin{aligned} b_{n,1} = & \frac{f(r)}{2} \sum_{j,k=1}^{2} \partial_{jk} m(r) h_j h_k \int z_1 z_k z_j K(z)dz \\ & + \sum_{\ell=1}^{2} \frac{\partial_\ell f(r)}{2} \sum_{j,k=1}^{2} \partial_{jk} m(r) h_j h_k h_\ell \int z_1 z_j z_k z_\ell K(z)dz \\ & + \frac{f(r)}{6} \sum_{j,k,\ell=1}^{2} \partial_{jk\ell} m(r) h_j h_k h_\ell \int z_1 z_j z_k z_\ell K(z)dz, \end{aligned}$$

$$\begin{aligned} b_{n,2} = & \frac{f(r)}{2} \sum_{j,k=1}^{2} \partial_{jk} m(r) h_j h_k \int z_2 z_k z_j K(z)dz \\ & + \sum_{\ell=1}^{2} \frac{\partial_\ell f(r)}{2} \sum_{j,k=1}^{2} \partial_{jk} m(r) h_j h_k h_\ell \int z_2 z_j z_k z_\ell K(z)dz \\ & + \frac{f(r)}{6} \sum_{j,k,\ell=1}^{2} \partial_{jk\ell} m(r) h_j h_k h_\ell \int z_2 z_j z_k z_\ell K(z)dz. \end{aligned}$$

When $K(z) = K_1(z_1)K_2(z_2)$ where $K_1(z_1) = (1-|z_1|)1_{\{|z_1|\leq 1\}}$ and $K_2(z_2) = 2(1-z_2)1_{\{0\leq z_2\leq 1\}}$, we have

$$
\begin{aligned}
b_{n,0} = &\frac{f(r)}{2}\left\{h_1^2\partial_{11}m(r)\kappa_1^{(2,1)} + h_2^2\partial_{22}m(r)\kappa_2^{(2,1)}\right\} \\
&+ \frac{\partial_1 f(r)}{2}\left(2h_1^2h_2\partial_{12}m(r)\kappa_{1,2}^{(2,1,1)}\right) \\
&+ \frac{\partial_2 f(r)}{2}\left(h_1^2h_2\partial_{11}m(r)\kappa_{1,2}^{(2,1,1)} + h_2^3\partial_{22}m(r)\kappa_2^{(3,1)}\right) \\
&+ \frac{f(r)}{6}\left(3h_1^2h_2\partial_{112}m(r)\kappa_{1,2}^{(2,1,1)} + h_2^3\partial_{222}m(r)\kappa_2^{(3,1)}\right),
\end{aligned}
$$

$$
\begin{aligned}
b_{n,1} = &\frac{f(r)}{2}\left(2h_1h_2\partial_{12}m(r)\kappa_{1,2}^{(2,1,1)}\right) \\
&+ \frac{\partial_1 f(r)}{2}\left(h_2^3\partial_{11}m(r)\kappa_1^{(4,1)} + h_1^2h_2\partial_{22}m(r)\kappa_{1,2}^{(2,2,1)}\right) \\
&+ \frac{\partial_2 f(r)}{2}\left(2h_1h_2^2\partial_{12}m(r)\kappa_{1,2}^{(2,2,1)}\right) \\
&+ \frac{f(r)}{6}\left(h_1^3\partial_{111}m(r)\kappa_1^{(4,1)} + 3h_1h_2^2\partial_{122}m(r)\kappa_{1,2}^{(2,2,1)}\right),
\end{aligned}
$$

$$
\begin{aligned}
b_{n,2} = &\frac{f(r)}{2}\left(h_1^2\partial_{11}m(r)\kappa_{1,2}^{(2,1,1)} + h_2^2\partial_{22}m(r)\kappa_2^{(3,1)}\right) \\
&+ \frac{\partial_1 f(r)}{2}\left(2h_1^2h_2\partial_{12}m(r)\kappa_{1,2}^{(2,2,1)}\right) \\
&+ \frac{\partial_2 f(r)}{2}\left(h_1^2h_2\partial_{11}m(r)\kappa_{1,2}^{(2,2,1)} + h_2^3\partial_{22}m(r)\kappa_2^{(4,1)}\right) \\
&+ \frac{f(r)}{6}\left(3h_1^2h_2\partial_{112}m(r)\kappa_{1,2}^{(2,2,1)} + h_2^3\partial_{222}m(r)\kappa_2^{(4,1)}\right).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\mathrm{Bias}(\hat{m}(r)) \\
&= \tilde{s}_1 b_{n,0} + \tilde{s}_3 b_{n,2} \\
&= \left\{\frac{h_1^2}{2}\partial_{11}m(r)(\tilde{s}_1\kappa_1^{(2,1)} + \tilde{s}_3\kappa_{1,2}^{(2,1,1)}) + \frac{h_2^2}{2}\partial_{22}m(r)(\tilde{s}_1\kappa_2^{(2,1)} + \tilde{s}_3\kappa_2^{(3,1)})\right\} \\
&\quad + h_1^2h_2\left(\frac{\partial_{11}m(r)}{2}\frac{\partial_2 f(r)}{f(r)} + \partial_{12}m(r)\frac{\partial_1 f(r)}{f(r)} + \frac{\partial_{112}m(r)}{2}\right)(\tilde{s}_1\kappa_{1,2}^{(2,1,1)} + \tilde{s}_3\kappa_{1,2}^{(2,2,1)})
\end{aligned}
$$

$$+ h_2^3 \left(\frac{1}{2} \partial_{22} m(r) \frac{\partial_2 f(r)}{f(r)} + \frac{1}{6} \partial_{222} m(r) \right) (\tilde{s}_1 \kappa_2^{(3,1)} + \tilde{s}_3 \kappa_2^{(4,1)}).$$

APPENDIX B: IMPLEMENTATION DETAILS

In section 2.3, we propose our optimal bandwidth selection from the following formula:

$$\frac{h_1}{h_2} = \left(\frac{B_2(c)^2}{B_1(c)^2} \right)^{1/4}$$

and

$$h_1 = \left[\frac{(\sigma_+^2(c) + \sigma_-^2(c))}{2n} e_1 S^{-1} \mathcal{K} S^{-1} e_1' |B_1(c)|^{-5/2} |B_2(c)|^{-1/2}) \right]^{1/6}$$

and our RD estimate prior to the bias correction is $\hat{\beta}_0^+(c) - \hat{\beta}_0^-(c)$ where these intercept terms of the local-polynomial estimates $\{\hat{\beta}_0^+(c), \hat{\beta}_0^-(c)\}$ are computed with the bandwidths specified above. Nevertheless, to compute the optimal bandwidth, we need to estimate the bias terms $B_1(c)$ and $B_2(c)$ as well as the residual variances $\{\sigma_+^2(c), \sigma_-^2(c)\}$. We follow Calonico et al., 2014, Section 5) in estimation of the residual variances at the boundary point c. For the bias terms, as in Calonico et al. (2014), we set a pair of pilot bandwidths with the local-quadratic regression. The key complication of our study is that the local-quadratic regression is also multivariate.

The expression of the bias terms involve a pair of partial derivatives $(\partial_{11} m_+(c), \partial_{22} m_+(c))$ for the treated and $(\partial_{11} m_-(c), \partial_{22} m_-(c))$ for the control. Given a pair of pilot bandwidths b_+ and b_- for the treated and the control, we run the local-quadratic estimation

$$\begin{aligned}
\hat{\gamma}^+(c) = \underset{(\gamma_0,\ldots,\gamma_5)' \in \mathbb{R}^6}{\arg\min} \sum_{i=1}^n & (Y_i - \gamma_0 - \gamma_1 (R_{i,1} - c_1) \\
& - \gamma_2 (R_{i,2} - c_2) - \gamma_3 (R_{i,1} - c_2)^2 \\
& - \gamma_4 (R_{i,1} - c_1)(R_{i,2} - c_2) \\
& - \gamma_5 (R_{i,2} - c_2)^2)^2 K_b (R_i - c) 1\{R_i \in \mathcal{T}\}
\end{aligned}$$

and

$$\hat{\gamma}^{-}(c) = \underset{(\gamma_0,\ldots,\gamma_5)'\in\mathbb{R}^6}{\arg\min} \sum_{i=1}^{n} (Y_i - \gamma_0 - \gamma_1(R_{i,1} - c_1)$$

$$- \gamma_2(R_{i,2} - c_2) - \gamma_3(R_{i,1} - c_2)^2$$

$$- \gamma_4(R_{i,1} - c_1)(R_{i,2} - c_2)$$

$$- \gamma_5(R_{i,2} - c_2)^2)^2 K_b(R_i - c)\, 1\{R_i \in \mathcal{T}^C\}$$

where $K_b(R_i - c) = K\left(\frac{R_{i,1}-c_1}{b}, \frac{R_{i,2}-c_2}{b}\right)$ to obtain these partial derivatives. These pilot bandwidths (b_+, b_-) are chosen from minimizing the mean squared error of estimating the bias term, which involves the local cubic regression. [1]

Given the pilot bandwidths, we estimate the bias terms $B_1(c)$ and $B_2(c)$. Let $\hat{B}_1(c)$ and $\hat{B}_2(c)$ be their estimates. In the optimal bandwidth selection, we follow Imbens and Kalyanaraman (2012) to regularize the bias term which appears in the denominator. Specifically, we employ their result that the inverse of bias term estimation error is approximated by 3 times their variance. If the estimated signs of the bias terms are the same, $sgn(\hat{B}_1(c)\hat{B}_2(c)) \geq 0$, then the optimal bandwidths should be chosen from the first-order condition: we set

$$h_1 = \left[\frac{(\hat{\sigma}_+^2(c) + \hat{\sigma}_-^2(c))}{2n} e_1 S^{-1}\mathcal{K}S^{-1}e_1'(\hat{B}_1(c)^2 + 3\hat{\boldsymbol{V}}(\hat{B}_1(c))^{-1}\left(\frac{\hat{B}_2(c)^2}{\hat{B}_1(c)^2 + 3\hat{\boldsymbol{V}}(\hat{B}_1(c))}\right)^{1/4}\right]^{1/6}$$

and

$$h_2 = \left[\frac{(\hat{\sigma}_+^2(c) + \hat{\sigma}_-^2(c))}{2n} e_1 S^{-1}\mathcal{K}S^{-1}e_1'(\hat{B}_2(c)^2 + 3\hat{\boldsymbol{V}}(\hat{B}_2(c))^{-1}\left(\frac{\hat{B}_1(c)^2}{\hat{B}_2(c)^2 + 3\hat{\boldsymbol{V}}(\hat{B}_2(c))}\right)^{1/4}\right]^{1/6}$$

separately for each subsample of the treated and control, where $\hat{\boldsymbol{V}}(\hat{B}_1(c))$ and $\hat{\boldsymbol{V}}(\hat{B}_2(c))$ are variance estimates from the bias estimation with the pilot bandwidths. If the estimated signs of the bias terms are different, $sgn(\hat{B}_1(c)\hat{B}_2(c)) < 0$, then we use the same

[1] Furthermore, we choose the preliminary bandwidth for the local cubic regression from minimizing the mean squared error of estimating the bias term for the pilot bandwidth. This preliminary bandwidth selection involves the global 4th order polynomial regressions.

bandwidth ratio h_1/h_2, but the first-order bias can be eliminated. Hence, we set

$$h_1 = \left[\frac{(\hat{\sigma}_+^2(c) + \hat{\sigma}_-^2(c))}{2n} e_1 S^{-1} \mathcal{K} S^{-1} e_1' (3\hat{\boldsymbol{V}}(\hat{B}_1(c)))/2)^{-1} \left(\frac{\hat{B}_2(c)^2}{\hat{B}_1(c)^2 + 3\hat{\boldsymbol{V}}(\hat{B}_1(c))} \right)^{1/4} \right]^{1/6}$$

and

$$h_2 = \left[\frac{(\hat{\sigma}_+^2(c) + \hat{\sigma}_-^2(c))}{2n} e_1 S^{-1} \mathcal{K} S^{-1} e_1' (3\hat{\boldsymbol{V}}(\hat{B}_2(c))/2)^{-1} \left(\frac{\hat{B}_1(c)^2}{\hat{B}_2(c)^2 + 3\hat{\boldsymbol{V}}(\hat{B}_2(c)))} \right)^{1/4} \right]^{1/6}$$

where the bias terms are replaced with the regularization terms.

APPENDIX C: CONSEQUENCE OF CONVERTING TWO-DIMENSIONAL DATA TO ONE DIMENSION.

Let $Z_i = \|R_i\|$ and $K_1(r) = 2(1-r)1_{\{0 \le r \le 1\}}$. Define

$$\check{f}(\mathbf{0}) = \frac{1}{\check{n}h} \sum_{i=1}^n K_1(Z_i/h) 1_{\{R_{i,2} \ge 0\}}, \ \check{n} = \sum_{i=1}^n 1_{\{R_{i,2} \ge 0\}}.$$

Note that $\frac{\check{n}}{n} = P(R_{1,2} \ge 0) + O_p(n^{-1/2})$ and

$$\begin{aligned}
\check{f}(\mathbf{0}) &= \left(\frac{1}{(\check{n}/n)} - \frac{1}{P(R_{1,2} \ge 0)} + \frac{1}{P(R_{1,2} \ge 0)} \right) \frac{1}{nh} \sum_{i=1}^n K_1(Z_i/h) 1_{\{R_{i,2} \ge 0\}} \\
&= \frac{1}{P(R_{1,2} \ge 0)} \frac{1}{nh} \sum_{i=1}^n K_1(Z_i/h) 1_{\{R_{i,2} \ge 0\}} + O_p(n^{-1/2}) \\
&=: \frac{1}{P(R_{1,2} \ge 0)} \tilde{f}(\mathbf{0}) + O_p(n^{-1/2}).
\end{aligned}$$

Further,

$$\begin{aligned}
E[\tilde{f}(\mathbf{0})] &= \frac{2}{h} E[K_1(Z_1/h) 1_{\{R_{1,2} \ge 0\}}] \\
&= \frac{2}{h} \int (1 - \|(r_1/h, r_2/h)\|) 1\{\|(r_1/h, r_2/h)\| \le 1\} 1_{\{r_2 \ge 0\}} f(r) dr \\
&= \frac{2}{h} \int (1 - \|(r_1/h, r_2/h)\|) 1\{\|(r_1/h, r_2/h)\| \le 1\} 1_{\{r_2/h \ge 0\}} f(r) dr
\end{aligned}$$

$$
\begin{aligned}
&= 2h \int (1 - \|z\|) 1_{\{\|z\| \leq 1, z_2 \geq 0\}} f(hz_1, hz_2) dz \\
&= 2h \left(f(\mathbf{0}) \int (1 - \|z\|) 1_{\{\|z\| \leq 1, z_2 \geq 0\}} dz + o(1) \right) \\
&= 2h \left(f(\mathbf{0}) \int_0^1 (1 - r) r dr \int_0^\pi d\theta + o(1) \right) \\
&= 2h \left(\frac{\pi}{6} f(\mathbf{0}) + o(1) \right)
\end{aligned}
$$

where we used the dominated convergence theorem for the fifth equation, and

$$
\begin{aligned}
\mathrm{Var}(\tilde{f}(\mathbf{0})) &\leq \frac{1}{nh^2} E\left[K_1^2(Z_1/h) 1_{\{R_{1,2} \geq 0\}} \right] \\
&= \frac{4}{n} \int (1 - \|z\|)^2 1_{\{\|z\| \leq 0, z_2 \geq 0\}} f(hz_1, hz_2) dz \\
&= \frac{4}{n} \left(f(\mathbf{0}) \int (1 - \|z\|)^2 1_{\{\|z\| \leq 1, z_2 \geq 0\}} dz + o(1) \right) \\
&= \frac{4}{n} \left(f(\mathbf{0}) \int_0^1 (1 - r)^2 r dr \int_0^\pi d\theta + o(1) \right) \\
&= \frac{4}{n} \left(\frac{\pi}{12} f(\mathbf{0}) + o(1) \right)
\end{aligned}
$$

where we used the dominated convergence theorem for the second equation. Then we have

$$
\check{f}(\mathbf{0}) = \frac{\pi h}{3P(R_{1,2} \geq 0)} f(\mathbf{0}) + o(h) + O_p(n^{-1/2}).
$$

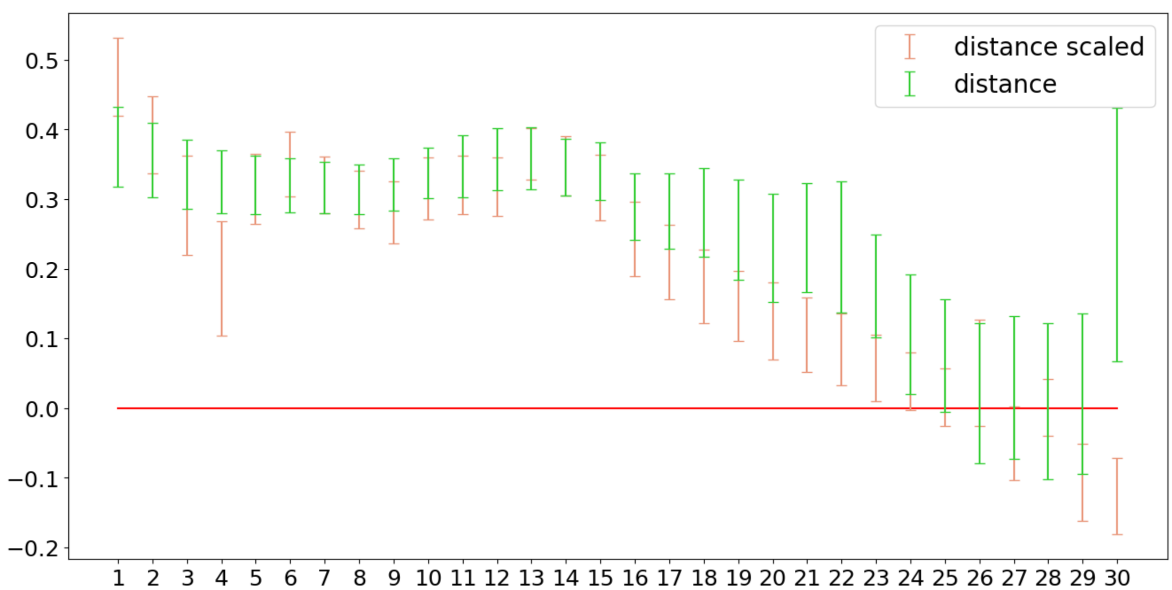


FIGURE 4.10.— Estimation results over the 30 boundary points comparing two *distance* estimates with and without modifying the relative scale of two axes. Values from 1 through 30 in the x-axis corresponds values in Figure 4.8. Points from 1 through 15 are of exceeding the merit threshold among the need-eligible students; points from 16 through 30 are of exceeding the need threshold among the merit-eligible students.

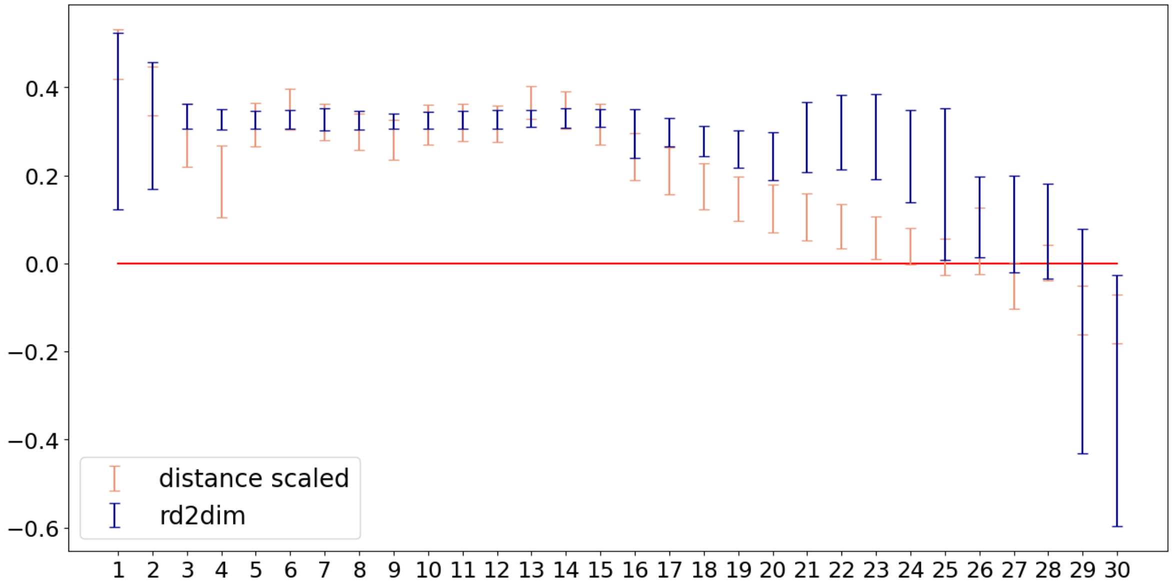


FIGURE 4.11.— The same estimates as Figure 4.10, comparing the scaled *distance* estimates against the non-scaled *rd2dim* estimates. Values from 1 through 30 in the x-axis corresponds values in Figure 4.8. Points from 1 through 15 are of exceeding the merit threshold among the need-eligible students; points from 16 through 30 are of exceeding the need threshold among the merit-eligible students.

APPENDIX E: ADDITIONAL TABLES

spec	mean bias band	mean opt band	mean opt band2	mean prelim band	var bias band	var opt band	var opt band2	var prelim band
common	0.794076	0.296791	NA	1.073086	0.030233	0.009093	NA	0.079046
rd2dim	0.794076	0.487928	0.451023	1.073086	0.030233	0.180021	0.120483	0.079046
distance	0.365445	0.175759	NA	NA	0.001036	0.000615	NA	NA

TABLE 5.2

Bandwidths values in simulation studies: Design 1, Lee shape.

spec	mean bias band	mean opt band	mean opt band2	mean prelim band	var bias band	var opt band	var opt band2	var prelim band
common	0.764508	0.172198	NA	1.043899	0.006529	0.000332	NA	0.018939
rd2dim	0.764849	0.166120	0.082935	1.044551	0.006539	0.000088	0.000153	0.019033
distance	0.311965	0.144708	NA	NA	0.000486	0.000298	NA	NA

TABLE 5.3

Bandwidths values in simulation studies: Design 2, LM shape.

spec	mean bias band	mean opt band	mean opt band2	mean prelim band	var bias band	var opt band	var opt band2	var prelim band
common	0.452217	0.117801	NA	0.649366	0.001988	0.000145	NA	0.007845
rd2dim	0.451275	0.161877	0.050484	0.647143	0.001857	0.000623	0.000060	0.006993
distance	0.217524	0.101804	NA	NA	0.000181	0.000218	NA	NA

TABLE 5.4

Bandwidths values in simulation studies: Design 3, additive shape.

spec	mean bias band	mean opt band	mean opt band2	mean prelim band	var bias band	var opt band	var opt band2	var prelim band
common	0.664811	0.221003	NA	1.073762	0.013142	0.005277	NA	0.079971
rd2dim	0.665097	0.265174	0.169662	1.073617	0.013127	0.051824	0.031091	0.080807
distance	0.554709	0.292187	NA	NA	0.011424	0.005015	NA	NA

TABLE 5.5

Bandwidths values in simulation studies: Design 4, LM2 shape.