Online Supplement for "Estimation and Inference for Heterogeneous Treatment Effects in Regression Discontinuity Designs"

Sebastian Calonico* Matias D. Cattaneo† Max H. Farrell‡ Filippo Palomba§
Rocío Titiunik¶

March 15, 2025

Abstract

This supplement contains all proofs, additional results, and other technical details about estimation, inference, and identification. Section SA1 describes setup and notation, states the assumptions we rely on, and introduces some auxiliary results. Section SA2 illustrates the main technical results. Section SA3 discusses in detail the common strategies used in empirical analyses. Section SA4 contains all the proofs.

*Graduate School of Management, UC Davis.
†Department of Operations Research and Financial Engineering, Princeton University.
‡Department of Economics, UC Santa Barbara.
§Department of Economics, Princeton University.
¶Department of Politics, Princeton University.

Complete Contents

SA1 Introduction

This section introduces the notation and links it with the one used in the main text, sets up the statistical framework, and enumerates the assumptions we rely on throughout this supplement.

SA1.1 Notation

In this Supplemental Appendix, we use n to denote sample size and $h \equiv h_n$ to denote a bandwidth sequence where we omit the dependence on n to ease notation. Moreover, h denotes a generic bandwidth (e.g. $h = h_-$ or $h = h_+$ depending on the context).

Linear algebra. Throughout the text, $\mathbf{e}_\nu$ denotes a conformable vector of zeros with a 1 in its $(\nu+1)$-th element, which may take different dimensions in different places, $\mathbf{0}_k$ and $\boldsymbol{\iota}_k$ are the k-dimensional zero and one vectors, respectively, $\mathbf{I}_k$ and $\mathbf{0}_{k\times j}$ denote the $k \times k$ identity matrix and a $k \times j$ matrix of zeros, respectively, $\otimes$ indicates the Kronecker product, $\mathrm{tr}(\cdot)$ is the trace operator, and $\mathrm{diag}(\mathbf{x})$ yields a square diagonal matrix with the elements of $\mathbf{x}$ on its main diagonal. With a slight abuse of notation, we denote with $\mathbf{v}^k = (v_1^k, \cdots, v_n^k)'$ the element-wise power for vectors $\mathbf{v} \in \mathbb{R}^n$. The maximum and minimum of two real numbers a and b are denoted by $a \vee b$ and $a \wedge b$, respectively. We let $|\cdot|$ denote the Euclidean norm, $|\mathbf{A}|^2 = \sum_i \sum_j |a_{ij}|^2 = \mathrm{tr}(\mathbf{A}'\mathbf{A})$. Finally, for some $q \in \mathbb{N}$, with $\mathcal{C}^q$ we denote the space of functions that are q-times continuously differentiable.

Asymptotic statements. For two positive sequences $\{a_n\}_n, \{b_n\}_n$, we write $a_n = O(b_n)$ if $\exists\, M \in \mathbb{R}_{++}$: $a_n \leq M b_n$ for all large n, $a_n = o(b_n)$ if $\lim_{n\to\infty} a_n b_n^{-1} = 0$, and $a_n \lesssim b_n$ if there exists a constant C such that $a_n \leq C b_n$ for all large n. For two sequences of random variables $\{A_n\}_n, \{B_n\}_n$, we write $A_n = o_{\mathbb{P}}(B_n)$ if $\forall\, \epsilon > 0, \lim_{n\to\infty} \mathbb{P}[|A_n B_n^{-1}| \geq \varepsilon] = 0$ and $A_n = O_{\mathbb{P}}(B_n)$ if $\forall\, \varepsilon > 0, \exists\, M, n_0 \in \mathbb{R}_{++} : \mathbb{P}[|A_n B_n^{-1}| > M] < \varepsilon$, for $n > n_0$. We denote convergence in probability with $\xrightarrow{\mathbb{P}}$ and convergence in distribution with $\rightsquigarrow$. We denote (possibly multivariate) Gaussian random variable with $\mathsf{N}(\mathbf{a}, \mathbf{B})$, where $\mathbf{a}$ denotes the mean and $\mathbf{B}$ the variance-covariance.

Causal Model. We now describe the population causal model. The outcome variable is

$$Y = T \cdot Y(1) + (1 - T) \cdot Y(0),$$

with $(Y(0), Y(1)) \in \mathbb{R}^2$ denoting the potential outcomes and $T := \mathbb{1}(X \geq c)$ denoting treatment status. We denote the vector of covariates with $\mathbf{W} \in \mathbb{R}^d, d \in \mathbb{N}$. We stress that, as the notation suggests, $\mathbf{W}$ is interpreted as a pretreatment vector of covariates.

In sharp RD designs, $T = \mathbb{1}(X \geq c)$, where $X \in \mathbb{R}$ denotes the running variable and $c \in \mathbb{R}$ is the cutoff. Throughout, $F_R(\cdot)$ denotes the cumulative distribution function (cdf) of a random variable R and $f_R(\cdot)$ is the density of R with respect to the Lebesgue measure. We denote with $f(\cdot)$ the density of the running variable X.

We further define

$$\begin{aligned}
\mathbf{Y} &= [Y_1, \cdots, Y_n]', & \mathbf{W} &= [\mathbf{W}_1, \mathbf{W}_2, \cdots, \mathbf{W}_n]', \\
\boldsymbol{\mu}_W(x) &= \mathbb{E}[\mathbf{W}_i | X_i = x], & \boldsymbol{\mu}_{WW}(x) &= \mathbb{E}[\mathbf{W}_i \mathbf{W}_i' | X_i = x], \\
\mu_{W_\ell}(x) &= \mathbb{E}[W_{i\ell} | X_i = x], & \mu_{W_\ell W_j}(x) &= \mathbb{E}[W_{i\ell} W_{ij} | X_i = x], \quad \ell, j \in \{1, \cdots, d\}, \\
\mu_{Y-}(x, \mathbf{w}) &= \mathbb{E}[Y_i(0) \mid X_i = x, \mathbf{W}_i = \mathbf{w}], & \mu_{Y+}(x, \mathbf{w}) &= \mathbb{E}[Y_i(1) \mid X_i = x, \mathbf{W}_i = \mathbf{w}], \\
\mu_{Y-}^{(\nu)}(x, \mathbf{w}) &= \frac{\partial^\nu}{\partial x^\nu} \mathbb{E}[Y_i(0) \mid X_i = x, \mathbf{W}_i = \mathbf{w}], & \mu_{Y+}(x, \mathbf{w}) &= \frac{\partial^\nu}{\partial x^\nu} \mathbb{E}[Y_i(1) \mid X_i = x, \mathbf{W}_i = \mathbf{w}], \\
\sigma_{Y-}^2(x, \mathbf{w}) &= \mathbb{V}[Y_i(0) \mid X_i = x, \mathbf{W}_i = \mathbf{w}], & \sigma_{Y+}^2(x, \mathbf{w}) &= \mathbb{V}[Y_i(1) \mid X_i = x, \mathbf{W}_i = \mathbf{w}],
\end{aligned}$$

and $\boldsymbol{\sigma}_W^2(x) = \boldsymbol{\mu}_{WW}(x) - \boldsymbol{\mu}_W(x)\boldsymbol{\mu}_W(x)'$.

Let $\mathbf{r}_q(x) = (1, x, \ldots, x^q)'$ be the polynomial basis of order $q \in \mathbb{N}$ and let

$$\mathbf{r}_{p,s}(u, \mathbf{w}) = (\mathbf{r}_p(u)', \mathbf{w}' \otimes \mathbf{r}_s(u)')'.$$

For a generic kernel function $k(\cdot)$ let

$$K(u) = \mathbb{1}(u < 0)k(-u) + \mathbb{1}(u \geq 0)k(u)$$

and

$$K_{\mathbf{h}}(u) = \mathbb{1}(u < 0)k_{h_-}(-u) + \mathbb{1}(u \geq 0)k_{h_+}(u), \quad k_h(u) = k(u/h)/h, \quad \mathbf{h} = (h_-, h_+)'.$$

Finally, we also define the following matrices

$$\boldsymbol{\Lambda}_{-,p,s} := f(c) \int_{-\infty}^{0} K(u)\mathbf{r}_p(u)\mathbf{r}_s(u)' \,\mathrm{d}u, \qquad \boldsymbol{\Lambda}_{+,p,s} := f(c) \int_{0}^{\infty} K(u)\mathbf{r}_p(u)\mathbf{r}_s(u)' \,\mathrm{d}u,$$

and

$$\boldsymbol{\Xi}_{-,p,s} = f(c) \int_{-\infty}^{0} K^2(u)\mathbf{r}_p\left(u\right)\mathbf{r}_s\left(u\right)' \,\mathrm{d}u, \qquad \boldsymbol{\Xi}_{+,p,s} = f(c) \int_{0}^{\infty} K^2(u)\mathbf{r}_p\left(u\right)\mathbf{r}_s\left(u\right)' \,\mathrm{d}u.$$

SA1.2 Setup

Formally, the RD estimators with treatment interactions can be obtained from a "long" regression, i.e.,

$$\widehat{\boldsymbol{\vartheta}}_{p,s}(\mathbf{h}) := \underset{\substack{\mathbf{a}_-,\mathbf{a}_+ \in \mathbb{R}^{1+p},\\ \boldsymbol{\ell}_-,\boldsymbol{\ell}_+ \in \mathbb{R}^{d(1+s)}}}{\arg\min} \sum_{i=1}^{n} \left(Y_i - \begin{bmatrix} \mathbf{r}_p\left(X_i - c\right) \\ \mathbf{r}_p\left(X_i - c\right) \end{bmatrix}' \begin{bmatrix} \mathbf{a}_- \\ \mathbf{a}_+ \end{bmatrix} - \begin{bmatrix} (\mathbf{W}_i \otimes \mathbf{r}_s(X_i - c)) \\ (\mathbf{W}_i \otimes \mathbf{r}_s(X_i - c)) \end{bmatrix}' \begin{bmatrix} \boldsymbol{\ell}_- \\ \boldsymbol{\ell}_+ \end{bmatrix} \right)^2 K_{\mathbf{h}}\left(X_i - c\right)$$

$$\widehat{\boldsymbol{\vartheta}}_{p,s}(\mathbf{h}) = \begin{bmatrix} \widehat{\boldsymbol{\vartheta}}_{-,p,s}(h_-) \\ \widehat{\boldsymbol{\vartheta}}_{+,p,s}(h_+) \end{bmatrix} = \begin{bmatrix} \widehat{\boldsymbol{\alpha}}_{-,p}(h_-) \\ \widehat{\boldsymbol{\lambda}}_{-,s}(h_-) \\ \widehat{\boldsymbol{\alpha}}_{+,p}(h_+) \\ \widehat{\boldsymbol{\lambda}}_{+,s}(h_+) \end{bmatrix}, \qquad p, s \in \mathbb{N}.$$

In what follows, we define $m := d(1 + s)$. The regression above has orthogonal regressors. Indeed, it has a set of regressors that pertains to estimation to the left of the cutoff, i.e., $(\mathbb{1}(X_i < c)\mathbf{r}_p(X_i - c)', \mathbb{1}(X_i \leq c)\mathbf{W}_i' \otimes \mathbf{r}_s(X_i - c)')'$, and a second set of regressors that refers to estimation to the right of the cutoff, i.e., $(\mathbb{1}(X_i \geq c)\mathbf{r}_p(X_i - c)', \mathbb{1}(X_i \geq c)\mathbf{W}_i' \otimes \mathbf{r}_s(X_i - c)')'$. Accordingly, by the partitioned regression theorem, the least squares coefficients in the "long" regression are equivalent to the least squares coefficients in the following "short" regressions:

$$\widehat{\boldsymbol{\vartheta}}_{-,p,s}(h) = \left[\begin{array}{c} \widehat{\boldsymbol{\alpha}}_{-,p}(h) \\ \widehat{\boldsymbol{\lambda}}_{-,s}(h) \end{array} \right] = \underset{\substack{\mathbf{a} \in \mathbb{R}^{1+p},\\ \boldsymbol{\ell} \in \mathbb{R}^m}}{\arg\min} \sum_{i=1}^{n} \mathbb{1}\left(X_i < c\right) \left(Y_i - \mathbf{r}_p\left(X_i - c\right)' \mathbf{a} - \left(\mathbf{W}_i \otimes \mathbf{r}_s(X_i - c)\right)' \boldsymbol{\ell}\right)^2 K_h\left(X_i - c\right),$$

$$\widehat{\boldsymbol{\vartheta}}_{+,p,s}(h) = \left[\begin{array}{c} \widehat{\boldsymbol{\alpha}}_{+,p}(h) \\ \widehat{\boldsymbol{\lambda}}_{+,s}(h) \end{array} \right] = \underset{\substack{\mathbf{a} \in \mathbb{R}^{1+p},\\ \boldsymbol{\ell} \in \mathbb{R}^m}}{\arg\min} \sum_{i=1}^{n} \mathbb{1}\left(X_i \geq c\right) \left(Y_i - \mathbf{r}_p\left(X_i - c\right)' \mathbf{a} - \left(\mathbf{W}_i \otimes \mathbf{r}_s(X_i - c)\right)' \boldsymbol{\ell}\right)^2 K_h\left(X_i - c\right),$$

which can be equivalently written in matrix form as

$$\widehat{\boldsymbol{\vartheta}}_{-,p,s}(h) = \mathbf{H}_{p,s}^{-1}(h)\widehat{\boldsymbol{\Gamma}}_{-,p,s}^{-1}(h)\widehat{\boldsymbol{\Upsilon}}_{-,p,s}(h), \qquad \widehat{\boldsymbol{\vartheta}}_{+,p,s}(h) = \mathbf{H}_{p,s}^{-1}(h)\widehat{\boldsymbol{\Gamma}}_{+,p,s}^{-1}(h)\widehat{\boldsymbol{\Upsilon}}_{+,p,s}(h),$$

where the hessian and score matrices are defined as

$$\widehat{\boldsymbol{\Gamma}}_{-,p,s}(h) = \mathbf{R}_{p,s}(h)'\mathbf{K}_-(h)\mathbf{R}_{p,s}(h)/n, \qquad \widehat{\boldsymbol{\Upsilon}}_{-,p,s}(h) = \mathbf{R}_{p,s}(h)'\mathbf{K}_-(h)\mathbf{Y}/n,$$
$$\widehat{\boldsymbol{\Gamma}}_{+,p,s}(h) = \mathbf{R}_{p,s}(h)'\mathbf{K}_+(h)\mathbf{R}_{p,s}(h)/n, \qquad \widehat{\boldsymbol{\Upsilon}}_{+,p,s}(h) = \mathbf{R}_{p,s}(h)'\mathbf{K}_+(h)\mathbf{Y}/n,$$

where

$$\mathbf{R}_p(h) = \begin{bmatrix} \mathbf{r}_p\left(\frac{X_1-c}{h}\right)' \\ \mathbf{r}_p\left(\frac{X_2-c}{h}\right)' \\ \vdots \\ \mathbf{r}_p\left(\frac{X_n-c}{h}\right)' \end{bmatrix}_{n\times(1+p)}, \quad \mathbf{R}_{p,s}(h) = \begin{bmatrix} \mathbf{r}_p\left(\frac{X_1-c}{h}\right)' & \mathbf{W}_1'\otimes\mathbf{r}_s\left(\frac{X_1-c}{h}\right)' \\ \mathbf{r}_p\left(\frac{X_2-c}{h}\right)' & \mathbf{W}_2'\otimes\mathbf{r}_s\left(\frac{X_2-c}{h}\right)' \\ \vdots & \vdots \\ \mathbf{r}_p\left(\frac{X_n-c}{h}\right)' & \mathbf{W}_n'\otimes\mathbf{r}_s\left(\frac{X_n-c}{h}\right)' \end{bmatrix}_{n\times(1+p+m)},$$

with the scaling and kernel matrices defined as

$$\mathbf{H}_{p,s}(h) = \begin{bmatrix} \mathbf{H}_p(h) & \mathbf{0}_{(1+p)\times m} \\ \mathbf{0}_{m\times(1+p)} & \mathbf{I}_d\otimes\mathbf{H}_s(h) \end{bmatrix}_{(1+p+m)\times(1+p+m)}, \qquad \mathbf{H}_q(h) = \operatorname{diag}(h^\ell : \ell = 0,\dots,q),$$

$$\mathbf{K}_-(h) = \operatorname{diag}\left(\mathbb{1}(X_i<c)k_h(X_i-c) : i=1,\dots,n\right), \quad \mathbf{K}_+(h) = \operatorname{diag}\left(\mathbb{1}(X_i\geq c)k_h(X_i-c) : i=1,\dots,n\right).$$

In what follows, we mostly refer to the RD estimator to the right of the cutoff, $\widehat{\boldsymbol{\vartheta}}_{+,p,s}(h)$, but everything follows symmetrically for the other RD estimator to the left of the cutoff, $\widehat{\boldsymbol{\vartheta}}_{-,p,s}(h)$.

We focus our attention on estimators of the following form:

$$\widehat{\chi}_{\nu,p,s}(\mathbf{h}) := \breve{\mathbf{e}}_\nu'(\widehat{\boldsymbol{\vartheta}}_{+,p,s}(h_+) - \widehat{\boldsymbol{\vartheta}}_{-,p,s}(h_-)), \qquad \nu\in\{0,1,\dots,p\wedge s\},$$

where

$$\breve{\mathbf{e}}_\nu := \begin{bmatrix} \nu\mathbf{I}_{p+1}\mathbf{e}_\nu \\ \nu!\boldsymbol{\iota}_d\otimes\mathbf{I}_{s+1}\mathbf{e}_\nu \end{bmatrix}$$

is the vector which extracts the $(\nu+1)$-th term in the polynomial basis in (X_i-c) and the $(\nu+1)$-th terms in the interaction between such basis and $\mathbf{W}_i$. The identity matrices are there just to stress the different dimensions $\mathbf{e}_\nu$ takes in different places.

Example 1. Suppose $d=1$ and $\nu=0$, then

$$\breve{\mathbf{e}}_\nu'\widehat{\boldsymbol{\vartheta}}_{+,p,s}(h) = \underbrace{\mathbf{e}_\nu'\widehat{\boldsymbol{\alpha}}_{+,p}(h)}_{\text{coefficient on } (X_i-c)} + \underbrace{\mathbf{e}_\nu'\widehat{\boldsymbol{\lambda}}_{+,s}(h)}_{\text{coefficient on } W_i\cdot\mathbb{1}(X_i\geq c)},$$

and

$$\breve{\mathbf{e}}_\nu'\widehat{\boldsymbol{\vartheta}}_{-,p,s}(h) = \underbrace{\mathbf{e}_\nu'\widehat{\boldsymbol{\alpha}}_{-,p}(h)}_{\text{coefficient on } \mathbb{1}(X_i<c)} + \underbrace{\mathbf{e}_\nu'\widehat{\boldsymbol{\lambda}}_{-,s}(h)}_{\text{coefficient on } W_i\cdot\mathbb{1}(X_i<c)}.$$

If instead $\nu=1$, then

$$\breve{\mathbf{e}}_\nu'\widehat{\boldsymbol{\vartheta}}_{+,p,s}(h) = \underbrace{\mathbf{e}_\nu'\widehat{\boldsymbol{\alpha}}_{+,p}(h)}_{\text{coefficient on } (X_i-c)\cdot\mathbb{1}(X_i\geq c)} + \underbrace{\mathbf{e}_\nu'\widehat{\boldsymbol{\lambda}}_{+,s}(h)}_{\text{coefficient on } W_i\cdot(X_i-c)\cdot\mathbb{1}(X_i\geq c)},$$

and

$$\breve{\mathbf{e}}_\nu' \widehat{\boldsymbol{\vartheta}}_{-,p,s}(h) = \underbrace{\mathbf{e}_\nu' \widehat{\boldsymbol{\alpha}}_{-,p}(h)}_{\text{coefficient on } (X_i - c)\cdot \mathbb{1}(X_i<c)} + \underbrace{\mathbf{e}_\nu' \widehat{\boldsymbol{\lambda}}_{-,s}(h)}_{\text{coefficient on } W_i\cdot(X_i-c)\cdot\mathbb{1}(X_i<c)} .$$

♦

Moreover, let

$$\boldsymbol{\vartheta}^\star_{-,p,s}(h) := \underset{\mathbf{t}\in\mathbb{R}^{1+p+m}}{\arg\min}\ \mathbb{E}\left[\mathbb{1}(X_i<c)K\left(\frac{X_i-c}{h}\right)(Y_i - \mathbf{r}_{p,s}(X_i-c,\mathbf{W}_i)'\mathbf{t})^2\right],$$
$$\boldsymbol{\vartheta}^\star_{+,p,s}(h) := \underset{\mathbf{t}\in\mathbb{R}^{1+p+m}}{\arg\min}\ \mathbb{E}\left[\mathbb{1}(X_i\geq c)K\left(\frac{X_i-c}{h}\right)(Y_i - \mathbf{r}_{p,s}(X_i-c,\mathbf{W}_i)'\mathbf{t})^2\right],$$

be the fixed-h best linear mean square error predictors of $\mathbb{E}[Y_i(d) \mid X_i = c, \mathbf{W}_i], d\in\{0,1\}$, i.e.,

$$\begin{aligned}
\boldsymbol{\vartheta}^\star_{-,p,s}(h) &= \mathbf{H}^{-1}_{p,s}(h)\,\mathbb{E}\left[\mathbb{1}(X_i<c)K\left(\frac{X_i-c}{h}\right)\mathbf{r}_{p,s}\left(\frac{X_i-c}{h},\mathbf{W}_i\right)\mathbf{r}_{p,s}\left(\frac{X_i-c}{h},\mathbf{W}_i\right)'\right]^{-1}\times\\
&\qquad \mathbb{E}\left[\mathbb{1}(X_i<c)K\left(\frac{X_i-c}{h}\right)\mathbf{r}_{p,s}\left(\frac{X_i-c}{h},\mathbf{W}_i\right)\mu_{Y-}(X_i,\mathbf{W}_i)\right]\\
&= \mathbf{H}^{-1}_{p,s}(h)\,\mathbb{E}[\widehat{\boldsymbol{\Gamma}}_{-,p,s}(h)]^{-1}\,\mathbb{E}[\widehat{\boldsymbol{\Upsilon}}_{-,p,s}(h)],
\end{aligned}$$

and

$$\begin{aligned}
\boldsymbol{\vartheta}^\star_{+,p,s}(h) &= \mathbf{H}^{-1}_{p,s}(h)\,\mathbb{E}\left[\mathbb{1}(X_i\geq c)K\left(\frac{X_i-c}{h}\right)\mathbf{r}_{p,s}\left(\frac{X_i-c}{h},\mathbf{W}_i\right)\mathbf{r}_{p,s}\left(\frac{X_i-c}{h},\mathbf{W}_i\right)'\right]^{-1}\times\\
&\qquad \mathbb{E}\left[\mathbb{1}(X_i\geq c)K\left(\frac{X_i-c}{h}\right)\mathbf{r}_{p,s}\left(\frac{X_i-c}{h},\mathbf{W}_i\right)\mu_{Y+}(X_i,\mathbf{W}_i)\right]\\
&= \mathbf{H}^{-1}_{p,s}(h)\,\mathbb{E}[\widehat{\boldsymbol{\Gamma}}_{+,p,s}(h)]^{-1}\,\mathbb{E}[\widehat{\boldsymbol{\Upsilon}}_{+,p,s}(h)].
\end{aligned}$$

SA1.3 Assumptions

In what follows, we state all the assumptions we rely on to prove the results in this supplemental appendix. Assumption 1 in the main text contains Assumptions SA1, SA2, SA3, and SA5, whereas Assumption 2 coincides with SA4.

Assumption SA1 (Sampling)**.** *$\{(Y_i, X_i, \mathbf{W}_i')'\}_{i=1}^n$ are independent draws from $(Y, X, \mathbf{W}')'$.*

Assumption SA2 (Density of Running Variable)**.** *The Lebesgue density of X_i is continuous, bounded, and bounded away from zero.*

Assumption SA3 (DGP)**.** *For $t\in\{0,1\}$ and $\mathsf{k}\in\mathbb{R}_{++}$, with $-\mathsf{k}<c<\mathsf{k}$ and for all $x\in[-\mathsf{k},\mathsf{k}], \mathbf{w}\in\mathbb{R}^d$:*

(a) $\mathbb{E}[\mathbf{W}_i \mid X_i = x]$ is continuous and $\mathbb{E}[\mathbf{W}_i\mathbf{W}_i' \mid X_i = x]$ is continuous and invertible.

(b) $\mathbb{E}[Y_i(t) \mid X_i = x, \mathbf{W}_i = \mathbf{w}]$is q-times continuously differentiable in x and continuous in $\mathbf{w}$ for $q\geq 1$.

(c) $\mathbb{V}[Y_i(t)|X_i = x]$ and $\mathbb{V}[Y_i(t)|X_i = x, \mathbf{W}_i = \mathbf{w}]$ are continuous and bounded away from zero.

(d) $\mathbb{E}[|Y_i(t)|^4|X_i = x, \mathbf{W}_i = \mathbf{w}]$ is continuous in both arguments.

(e) $\mathbb{E}[|\mathbf{W}_i|^4|X_i = x]$ is continuous.

(f) $\mathbb{E}[\mathbf{W}_i\,\mathbb{V}[Y_i(t)|X_i,\mathbf{W}_i] \mid X_i = x]$ and $\mathbb{E}[\mathbf{W}_i\mathbf{W}_i'\,\mathbb{V}[Y_i(t)|X_i,\mathbf{W}_i] \mid X_i = x]$ are continuous.

(g) $\boldsymbol{\Gamma}_{-,p,s}$ and $\boldsymbol{\Gamma}_{+,p,s}$ in Equation (3) are positive definite matrices.

Assumption SA4 (Identification)**.** *For* $\mathsf{k} \in \mathbb{R}$, *with* $-\mathsf{k} < c < \mathsf{k}$ *and for all* $x \in [-\mathsf{k}, \mathsf{k}], \mathbf{w} \in \mathbb{R}^d$ *the expectation of* $Y_i(t), t \in \{0, 1\}$ *conditional on* $(X_i, \mathbf{W}_i)$ *is*

$$\mu_-(x, \mathbf{w}) = \alpha_-(x) + \boldsymbol{\lambda}_-(x)'\mathbf{w}, \qquad \mu_+(x, \mathbf{w}) = \alpha_+(x) + \boldsymbol{\lambda}_+(x)'\mathbf{w},$$

where $\boldsymbol{\lambda}_-(x) := (\lambda_{-,1}(x), \cdots, \lambda_{-,d}(x))', \boldsymbol{\lambda}_+(x) := (\lambda_{+,1}(x), \cdots, \lambda_{+,d}(x))'$, *and* $\alpha_-(x), \alpha_+(x) \in \mathcal{C}^q$, *and* $\lambda_{-,\ell}(x), \lambda_{+,\ell}(x), \in \mathcal{C}^q$ *for all* $\ell \in \{1, \cdots, d\}$ *for some* $q \in \mathbb{N}$.

Assumption SA5 (Kernel)**.** *The kernel function* $k(\cdot) : [0, 1] \to \mathbb{R}$ *is continuous and nonnegative.*

In words, under Assumption SA1 the sample is composed of independent draws from an underlying population of interest. Assumptions SA2, SA3a, SA3c, and SA3f are technical conditions on the data-generating process (DGP) we rely on when we establish convergence rates and characterize the probability limits of interest. Assumption SA3g is a standard assumption in least squares which requires the probability limit of the Gram matrix to be positive definite. Assumption SA3a also establishes that $\mathbf{W}$ can be thought of as a vector of pretreatment covariates. Assumption SA3b is necessary to Taylor expand the conditional expectation of the potential outcomes and characterize the smoothing bias. Assumptions SA3d and SA3e are standard bounded absolute higher-order moment conditions that make the Lyapunov condition satisfied and allow us to invoke a Lindeberg-Feller central limit theorem to show asymptotic normality of the RD estimator. Assumption SA4 imposes a partially linear (in $\mathbf{W}$) structure –locally at the cutoff– on the population conditional expectation of the potential outcomes which we rely on when attaching a causal interpretation to the probability limit of the RD estimator. Assumption SA5 states standard technical conditions on the kernel used in the local polynomial regression.

SA1.4 Mapping between Main Text and Supplement

In the main text, we present all the results using the long regression with $p = s = 1$

$$Y_i \quad \text{onto} \quad (\mathbf{r}_p(X_i - c)', T_i\mathbf{r}_p(X_i - c)', \mathbf{W}_i' \otimes \mathbf{r}_s(X_i - c)', T_i\mathbf{W}_i' \otimes \mathbf{r}_s(X_i - c)')', \tag{1}$$

which aligns with the classical way of thinking about linear regressions. Indeed, this representation allows us to readily interpret the coefficients on the terms interacted with T_i as specific differences between treated and control groups.

In this supplement instead, we set up the problem as the long regression

$$Y_i \quad \text{onto} \quad ((1 - T_i)\mathbf{r}_p(X_i - c)', T_i\mathbf{r}_p(X_i - c)', (1 - T_i)\mathbf{W}_i' \otimes \mathbf{r}_s(X_i - c)', T_i\mathbf{W}_i' \otimes \mathbf{r}_s(X_i - c)')'. \tag{2}$$

This is equivalent to (1) in that the predicted values and the residuals of these two regressions are the same. However, the regression coefficients in (2) have a different interpretation than those in (1). For example, the coefficients on the terms interacted with T_i do not describe differences between treated and control groups but rather capture only moments of the treated population. On the one hand, (1) allows coefficients to be directly interpreted as long as differences between treated and control are of interest. On the other hand, as we already stressed in Section SA1.2, (2) can be written as two separate "short" regressions with orthogonal design matrices, making statements and proofs less cumbersome from a notational standpoint. For this reason, we rely on (1) in the main text and on (2) in this supplement.

Table SA-1 maps the notation used in this supplement with the one used in the main paper. Dependence on h is omitted to simplify notation. For the MSE constants, we will use the relationship

$$\widehat{\varsigma}(\mathbf{h}) = \mathbf{M}\widehat{\boldsymbol{\vartheta}}_{1,1}(\mathbf{h}) = \mathbf{M}\begin{bmatrix}\widehat{\boldsymbol{\vartheta}}_{-,p,s}(h_-) \\ \widehat{\boldsymbol{\vartheta}}_{+,p,s}(h_+)\end{bmatrix}, \qquad \mathbf{M} := \begin{bmatrix}-\mathbf{e}_0'\mathbf{I}_{p+1} & \mathbf{0}_m' \\ \mathbf{0}_{d\times(p+1)} & -(\boldsymbol{\iota}_d \otimes \mathbf{I}_{s+1}\mathbf{e}_0)' \\ \mathbf{e}_0'\mathbf{I}_{p+1} & \mathbf{0}_m' \\ \mathbf{0}_{d\times(p+1)} & (\boldsymbol{\iota}_d \otimes \mathbf{I}_{s+1}\mathbf{e}_0)'\end{bmatrix}.$$

Table SA-1: *Mapping between main text and supplement notation.*

	Main	Supplement
Estimands		
	$\alpha(\cdot)$	$\alpha_-(\cdot)$
	$\theta(\cdot)$	$\alpha_+(\cdot) - \alpha_-(\cdot)$
	$\boldsymbol{\lambda}(\cdot)$	$\boldsymbol{\lambda}_-(\cdot)$
	$\boldsymbol{\xi}(\cdot)$	$\boldsymbol{\lambda}_+(\cdot) - \boldsymbol{\lambda}_-(\cdot)$
	ς	$(\alpha_+(0) - \alpha_-(0), \boldsymbol{\lambda}_+(0)' - \boldsymbol{\lambda}_-(0)')'$
Estimators		
	$\widehat{\alpha}$	$\mathbf{e}_0'\widehat{\boldsymbol{\alpha}}_-$
	$\widehat{\theta}$	$\mathbf{e}_0'(\widehat{\boldsymbol{\alpha}}_+ - \widehat{\boldsymbol{\alpha}}_-)$
	$\widehat{\boldsymbol{\lambda}}$	$(\boldsymbol{\iota}_d \otimes \mathbf{I}_{s+1}\mathbf{e}_0)'\widehat{\boldsymbol{\lambda}}_-$
	$\widehat{\boldsymbol{\xi}}$	$(\boldsymbol{\iota}_d \otimes \mathbf{I}_{s+!}\mathbf{e}_0)'(\widehat{\boldsymbol{\lambda}}_+ - \widehat{\boldsymbol{\lambda}}_-)$
	$\widehat{\varsigma}$	$\breve{\mathbf{e}}_0'(\widehat{\boldsymbol{\vartheta}}_{+,1,1} - \widehat{\boldsymbol{\vartheta}}_{-,1,1})$
	$\widehat{\omega}_1$	$\mathbf{e}_1'\widehat{\boldsymbol{\alpha}}_-$
	$\widehat{\omega}_2$	$\mathbf{e}_1'(\widehat{\boldsymbol{\alpha}}_+ - \widehat{\boldsymbol{\alpha}}_-)$
	$\widehat{\boldsymbol{\omega}}_3$	$(\boldsymbol{\iota}_d \otimes \mathbf{I}_{s+1}\mathbf{e}_1)'\widehat{\boldsymbol{\lambda}}_-$
	$\widehat{\boldsymbol{\omega}}_4$	$(\boldsymbol{\iota}_d \otimes \mathbf{I}_{s+1}\mathbf{e}_1)'(\widehat{\boldsymbol{\lambda}}_+ - \widehat{\boldsymbol{\lambda}}_-)$
Constants		
	B_{s}	$\mathsf{s}'\mathbf{M}(\mathbf{B}^{[0]}_{+,1,1} + \mathbf{B}^{[1]}_{+,1,1} - \mathbf{B}^{[0]}_{-,1,1} + \mathbf{B}^{[1]}_{-,1,1})$
	V_{s}	$\mathsf{s}'\mathbf{M}(\boldsymbol{\Gamma}^{-1}_{+,1,1}\mathbf{V}_{+,1,1}\boldsymbol{\Gamma}^{-1\prime}_{+,1,1} + \boldsymbol{\Gamma}^{-1}_{-,1,1}\mathbf{V}_{-,1,1}\boldsymbol{\Gamma}^{-1\prime}_{-,1,1})\mathbf{M}'\mathsf{s}$

We conclude this section by mapping the results in the main text with the proofs in the supplement:

- Theorem 1 is a particular case of the result proved in Section SA2.4 with $p = s = 1$ and $\nu = 0$;
- Theorem 2 follows from the results in Section SA2.6 with $p = s = 1, \nu = 0$, and $\breve{\mathbf{e}}_\nu = \mathbf{s}$;
- Theorem 3 is a particular case of Corollary SA-5 with $p = s = 1$ and $\nu = 0$.

SA1.5 Auxiliary Lemmas and Quantities

In this subsection, we introduce a series of auxiliary lemmas and quantities we rely upon in the rest of the supplemental appendix. We heavily rely on the next lemma for asymptotic statements.

Lemma SA-1. *Let $\mathbf{A}_n$ be a sequence of random matrices with finite first two moments. Then*

$$\mathbf{A}_n = \mathbb{E}[\mathbf{A}_n] + O_{\mathbb{P}}(|\,\mathbb{V}[\mathbf{A}_n]|^{1/2}).$$

[Proof]

The next lemma handles the Gram matrices

$$\widehat{\boldsymbol{\Lambda}}_{-,p,s}(h) = \mathbf{R}_p(h)'\mathbf{K}_-(h)\mathbf{R}_s(h)/n, \qquad \widehat{\boldsymbol{\Lambda}}_{+,p,s}(h) = \mathbf{R}_p(h)'\mathbf{K}_+(h)\mathbf{R}_s(h)/n,$$

and shows the object concentrates around in probability and the rate at which such concentration occurs.

Lemma SA-2. *Let Assumptions SA1, SA2, and SA5 hold with $\mathsf{k} \geq h$. If $nh \to \infty$ and $h \to 0$, then*

$$\widehat{\boldsymbol{\Lambda}}_{-,p,s}(h) = \widetilde{\boldsymbol{\Lambda}}_{-,p,s}(h) + O_{\mathbb{P}}(1/\sqrt{nh}), \qquad \widehat{\boldsymbol{\Lambda}}_{+,p,s}(h) = \widetilde{\boldsymbol{\Lambda}}_{+,p,s}(h) + O_{\mathbb{P}}(1/\sqrt{nh}),$$

with

$$\widetilde{\boldsymbol{\Lambda}}_{-,p,s}(h) := \mathbb{E}[\widehat{\boldsymbol{\Lambda}}_{-,p,s}(h)] = \int_{-\infty}^{0} K(u)\mathbf{r}_p(u)\mathbf{r}_s(u)' f(uh+c)\,\mathrm{d}u,$$

$$\widetilde{\boldsymbol{\Lambda}}_{+,p,s}(h) := \mathbb{E}[\widehat{\boldsymbol{\Lambda}}_{+,p,s}(h)] = \int_{0}^{\infty} K(u)\mathbf{r}_p(u)\mathbf{r}_s(u)' f(uh+c)\,\mathrm{d}u.$$

[Proof]

In the next lemma, we give conditions for the asymptotic invertibility of $\widehat{\boldsymbol{\Gamma}}_{-,p,s}(h)$ and $\widehat{\boldsymbol{\Gamma}}_{+,p,s}(h)$, thereby making local polynomial estimators well-defined in large samples.

Lemma SA-3. *Let Assumptions SA1, SA2, SA3, and SA5 hold with $\mathsf{k} \geq h$. If $nh \to \infty$ and $h \to 0$, then*

$$\widehat{\boldsymbol{\Gamma}}_{-,p,s}(h) = \widetilde{\boldsymbol{\Gamma}}_{-,p,s}(h) + O_{\mathbb{P}}(1/\sqrt{nh}), \qquad \widehat{\boldsymbol{\Gamma}}_{+,p,s}(h) = \widetilde{\boldsymbol{\Gamma}}_{+,p,s}(h) + O_{\mathbb{P}}(1/\sqrt{nh}),$$

where

$$\widetilde{\boldsymbol{\Gamma}}_{-,p,s}(h) := \begin{bmatrix} \widetilde{\boldsymbol{\Lambda}}_{-,p,p}(h) & \widetilde{\mathbf{G}}_{-,2}(h) \\ \widetilde{\mathbf{G}}_{-,2}(h)' & \widetilde{\mathbf{G}}_{-,3}(h) \end{bmatrix}, \qquad \widetilde{\boldsymbol{\Gamma}}_{+,p,s}(h) := \begin{bmatrix} \widetilde{\boldsymbol{\Lambda}}_{+,p,p}(h) & \widetilde{\mathbf{G}}_{+,2}(h) \\ \widetilde{\mathbf{G}}_{+,2}(h)' & \widetilde{\mathbf{G}}_{+,3}(h) \end{bmatrix},$$

where

$$\widetilde{\mathbf{G}}_{-,2}(h) = \int_{-\infty}^{0} K(u)\left[\boldsymbol{\mu}_W(uh+c)' \otimes \mathbf{r}_p(u)\,\mathbf{r}_s(u)'\right] f(uh+c)\,\mathrm{d}u,$$

$$\widetilde{\mathbf{G}}_{+,2}(h) = \int_{0}^{\infty} K(u)\left[\boldsymbol{\mu}_W(uh+c)' \otimes \mathbf{r}_p(u)\,\mathbf{r}_s(u)'\right] f(uh+c)\,\mathrm{d}u,$$

$$\widetilde{\mathbf{G}}_{-,3}(h) = \int_{-\infty}^{0} K(u)\left[\boldsymbol{\mu}_{WW}(uh+c) \otimes \mathbf{r}_s(u)\,\mathbf{r}_s(u)'\right] f(uh+c)\,\mathrm{d}u,$$

$$\widetilde{\mathbf{G}}_{+,3}(h) = \int_{0}^{\infty} K(u)\left[\boldsymbol{\mu}_{WW}(uh+c) \otimes \mathbf{r}_s(u)\,\mathbf{r}_s(u)'\right] f(uh+c)\,\mathrm{d}u.$$

Last,

$$\widetilde{\boldsymbol{\Gamma}}_{-,p,s}(h) = \boldsymbol{\Gamma}_{-,p,s}\{1+o(1)\}, \qquad \widetilde{\boldsymbol{\Gamma}}_{+,p,s}(h) = \boldsymbol{\Gamma}_{+,p,s}\{1+o(1)\},$$

where

$$\boldsymbol{\Gamma}_{-,p,s} = \begin{bmatrix} \boldsymbol{\Lambda}_{-,p,p} & \mathbf{G}_{-,2} \\ \mathbf{G}'_{-,2} & \mathbf{G}_{-,3} \end{bmatrix}, \qquad \boldsymbol{\Gamma}_{+,p,s} = \begin{bmatrix} \boldsymbol{\Lambda}_{+,p,p} & \mathbf{G}_{+,2} \\ \mathbf{G}'_{+,2} & \mathbf{G}_{+,3} \end{bmatrix}, \tag{3}$$

with

$$\boldsymbol{\Lambda}_{-,p,s} = f(c)\int_{-\infty}^{0} K(u)\mathbf{r}_p(u)\mathbf{r}_s(u)'\,\mathrm{d}u, \qquad \boldsymbol{\Lambda}_{+,p,s} = f(c)\int_{0}^{\infty} K(u)\mathbf{r}_p(u)\mathbf{r}_s(u)'\,\mathrm{d}u$$

$$\mathbf{G}_{-,2} = \boldsymbol{\mu}'_W \otimes \boldsymbol{\Lambda}_{-,p,s}, \qquad \mathbf{G}_{+,2} = \boldsymbol{\mu}'_W \otimes \boldsymbol{\Lambda}_{+,p,s},$$

$$\mathbf{G}_{-,3} = \boldsymbol{\mu}_{WW} \otimes \boldsymbol{\Lambda}_{-,s,s}, \qquad \mathbf{G}_{+,3} = \boldsymbol{\mu}_{WW} \otimes \boldsymbol{\Lambda}_{+,s,s}.$$

[Proof]

To have a more compact notation, we introduce the following quantities for integers $p, s, a \in \mathbb{N}_0$:

$$\widehat{\boldsymbol{\zeta}}_{-,p,s,a}(h) = \frac{1}{nh}\sum_{i=1}^{n} \mathbb{1}(X_i < c)K\left(\frac{X_i - c}{h}\right)\mathbf{r}_{p,s}\left(\frac{X_i - c}{h}, \mathbf{W}_i\right)\left(\frac{X_i - c}{h}\right)^{a+1},$$

$$\widehat{\boldsymbol{\zeta}}_{+,p,s,a}(h) = \frac{1}{nh}\sum_{i=1}^{n} \mathbb{1}(X_i \geq c)K\left(\frac{X_i - c}{h}\right)\mathbf{r}_{p,s}\left(\frac{X_i - c}{h}, \mathbf{W}_i\right)\left(\frac{X_i - c}{h}\right)^{a+1},$$

$$\widehat{\boldsymbol{\varphi}}_{-,p,s,a}(h) = \frac{1}{nh}\sum_{i=1}^{n} \mathbb{1}(X_i < c)K\left(\frac{X_i - c}{h}\right)\mathbf{r}_{p,s}\left(\frac{X_i - c}{h}, \mathbf{W}_i\right)\mathbf{W}_i'\left(\frac{X_i - c}{h}\right)^{a+1},$$

$$\widehat{\boldsymbol{\varphi}}_{+,p,s,a}(h) = \frac{1}{nh}\sum_{i=1}^{n} \mathbb{1}(X_i \geq c)K\left(\frac{X_i - c}{h}\right)\mathbf{r}_{p,s}\left(\frac{X_i - c}{h}, \mathbf{W}_i\right)\mathbf{W}_i'\left(\frac{X_i - c}{h}\right)^{a+1}.$$

The following two lemmas characterize the asymptotic properties of $\widehat{\boldsymbol{\zeta}}_{-,p,s,a}(h), \widehat{\boldsymbol{\zeta}}_{+,p,s,a}(h), \widehat{\boldsymbol{\varphi}}_{-,p,s,a}(h)$, and $\widehat{\boldsymbol{\varphi}}_{+,p,s,a}(h)$. These results are used below to guarantee that the "constant terms" in the MSE expansions are asymptotically well-defined.

Lemma SA-4. *Let Assumptions SA1, SA2, SA3, and SA5 hold with $\mathsf{k} \geq h$. If $nh \to \infty$ and $h \to 0$, then*

$$\widehat{\boldsymbol{\zeta}}_{-,p,s,a}(h) = \widetilde{\boldsymbol{\zeta}}_{-,p,s,a}(h) + O_{\mathbb{P}}(1/\sqrt{nh}), \qquad \widehat{\boldsymbol{\zeta}}_{+,p,s,a}(h) = \widetilde{\boldsymbol{\zeta}}_{+,p,s,a}(h) + O_{\mathbb{P}}(1/\sqrt{nh}),$$
$$\widehat{\boldsymbol{\varphi}}_{-,p,s,a}(h) = \widetilde{\boldsymbol{\varphi}}_{-,p,s,a}(h) + O_{\mathbb{P}}(1/\sqrt{nh}), \qquad \widehat{\boldsymbol{\varphi}}_{+,p,s,a}(h) = \widetilde{\boldsymbol{\varphi}}_{+,p,s,a}(h) + O_{\mathbb{P}}(1/\sqrt{nh}),$$

with

$$\widetilde{\boldsymbol{\zeta}}_{-,p,s,a}(h) := \mathbb{E}[\widehat{\boldsymbol{\zeta}}_{-,p,s,a}(h)] = \int_{-\infty}^{0} K(u)\mathbf{r}_{p,s}(u, \boldsymbol{\mu}_W(uh + c))u^{a+1}f(uh + c)\,\mathrm{d}u,$$

$$\widetilde{\boldsymbol{\zeta}}_{+,p,s,a}(h) := \mathbb{E}[\widehat{\boldsymbol{\zeta}}_{+,p,s,a}(h)] = \int_{0}^{\infty} K(u)\mathbf{r}_{p,s}(u, \boldsymbol{\mu}_W(uh + c))u^{a+1}f(uh + c)\,\mathrm{d}u,$$

$$\widetilde{\boldsymbol{\varphi}}_{-,p,s,a}(h) := E[\widehat{\boldsymbol{\varphi}}_{-,p,s,a}(h)] = \int_{-\infty}^{0} K(u)\begin{bmatrix}\mathbf{r}_p(u)\boldsymbol{\mu}_W(uh + c)' \\ \boldsymbol{\mu}_{WW}(uh + c) \otimes \mathbf{r}_s(u)\end{bmatrix}u^{a+1}f(uh + c)\,\mathrm{d}u,$$

$$\widetilde{\boldsymbol{\varphi}}_{+,p,s,a}(h) := \mathbb{E}[\widehat{\boldsymbol{\varphi}}_{+,p,s,a}(h)] = \int_{0}^{\infty} K(u)\begin{bmatrix}\mathbf{r}_p(u)\boldsymbol{\mu}_W(uh + c)' \\ \boldsymbol{\mu}_{WW}(uh + c) \otimes \mathbf{r}_s(u)\end{bmatrix}u^{a+1}f(uh + c)\,\mathrm{d}u.$$

Last

$$\widetilde{\boldsymbol{\zeta}}_{-,p,s,a}(h) = \boldsymbol{\zeta}_{-,p,s,a}\{1 + o(1)\}, \qquad \widetilde{\boldsymbol{\zeta}}_{+,p,s,a}(h) = \boldsymbol{\zeta}_{+,p,s,a}\{1 + o(1)\},$$
$$\widetilde{\boldsymbol{\varphi}}_{-,p,s,a}(h) = \boldsymbol{\varphi}_{-,p,s,a}\{1 + o(1)\}, \qquad \widetilde{\boldsymbol{\varphi}}_{+,p,s,a}(h) = \boldsymbol{\varphi}_{+,p,s,a}\{1 + o(1)\},$$

where

$$\boldsymbol{\zeta}_{-,p,s,a} = f(c)\int_{-\infty}^{0} K(u)\mathbf{r}_{p,s}(u, \boldsymbol{\mu}_W)u^{a+1}\,\mathrm{d}u, \qquad \boldsymbol{\zeta}_{+,p,s,a} = f(c)\int_{0}^{\infty} K(u)\mathbf{r}_{p,s}(u, \boldsymbol{\mu}_W)u^{a+1}\,\mathrm{d}u,$$

$$\boldsymbol{\varphi}_{-,p,s,a} = f(c)\int_{-\infty}^{0} K(u)\begin{bmatrix}\mathbf{r}_p(u)\boldsymbol{\mu}_W' \\ \boldsymbol{\mu}_{WW} \otimes \mathbf{r}_s(u)\end{bmatrix}u^{a+1}\,\mathrm{d}u, \quad \boldsymbol{\varphi}_{+,p,s,a} = f(c)\int_{0}^{\infty} K(u)\begin{bmatrix}\mathbf{r}_p(u)\boldsymbol{\mu}_W' \\ \boldsymbol{\mu}_{WW} \otimes \mathbf{r}_s(u)\end{bmatrix}u^{a+1}\,\mathrm{d}u.$$

[Proof]

SA2 Main Results

In this section, we mostly focus on the RD estimator to the right of the cutoff as everything holds symmetrically for the RD estimator to the left. All the lemmas and main results are reported for both estimators.

The RD estimator with interacted covariates to the right of the cutoff is defined as

$$\widehat{\boldsymbol{\vartheta}}_{+,p,s}(h) := \underset{\mathbf{t}\in\mathbb{R}^{1+p+m}}{\arg\min} \sum_{i=1}^{n} \mathbb{1}(X_i \geq c) K\left(\frac{X_i - c}{h}\right) (Y_i - \mathbf{r}_{p,s}(X_i - c, \mathbf{W}_i)'\mathbf{t})^2.$$

To analyze the statistical properties of the RD estimator with interacted covariates $\widehat{\boldsymbol{\vartheta}}_{+,p,s}(h)$, we take advantage of the following decomposition:

$$\begin{aligned}\widehat{\boldsymbol{\vartheta}}_{+,p,s}(h) &= \mathbf{H}_{p,s}^{-1}(h)\widehat{\boldsymbol{\Gamma}}_{+,p,s}^{-1}(h)\frac{1}{nh}\sum_{i=1}^{n}\mathbb{1}(X_i \geq c)K\left(\frac{X_i - c}{h}\right)\mathbf{r}_{p,s}\left(\frac{X_i - c}{h}, \mathbf{W}_i\right)Y_i,\\ &= \boldsymbol{\vartheta}^{\star}_{+,p,s}(h) + \mathbf{H}_{p,s}^{-1}(h)\widehat{\boldsymbol{\Gamma}}_{+,p,s}^{-1}(h)\frac{1}{nh}\sum_{i=1}^{n}\mathbb{1}(X_i \geq c)K\left(\frac{X_i - c}{h}\right)\mathbf{r}_{p,s}\left(\frac{X_i - c}{h}, \mathbf{W}_i\right)u_{+,i}(h),\end{aligned}$$

where $\boldsymbol{\vartheta}^{\star}_{+,p,s}(h)$ the fixed-h best linear mean square error predictor of $\mathbb{E}[Y_i(1) \mid X_i = c, \mathbf{W}_i]$ and

$$u_{+,i}(h) := Y_i - \mathbf{r}_{p,s}\left(X_i - c, \mathbf{W}_i\right)'\boldsymbol{\vartheta}^{\star}_{+,p,s}(h)$$

is the fixed-h ℓ_2-projection residual. This decomposition gives us

$$\widehat{\boldsymbol{\vartheta}}_{+,p,s}(h) - \boldsymbol{\vartheta}^{\star}_{+,p,s}(h) = \mathbf{H}_{p,s}^{-1}(h)\widehat{\boldsymbol{\Gamma}}_{+,p,s}^{-1}(h)\mathbf{L}_{+,p,s}(h),$$

where

$$\mathbf{L}_{+,p,s}(h) = \frac{1}{nh}\sum_{i=1}^{n}\mathbb{1}(X_i \geq c)K\left(\frac{X_i - c}{h}\right)\mathbf{r}_{p,s}\left(\frac{X_i - c}{h}, \mathbf{W}_i\right)u_{+,i}(h).$$

SA2.1 Asymptotic Approximation and Asymptotic Variance

Lemma SA-3 shows that $\widehat{\boldsymbol{\Gamma}}_{+,p,s}(h)$ is asymptotically invertible. This Lemma takes care of the "denominator" of the RD estimator by showing that

$$\widehat{\boldsymbol{\Gamma}}_{+,p,s}(h) = \boldsymbol{\Gamma}_{+,p,s} + o(1) + O_{\mathbb{P}}(1/\sqrt{nh}).$$

Coming to the $\mathbf{L}_{+,p,s}(h)$ term, first note that $\mathbb{E}[\mathbf{L}_{+,p,s}(h)] = 0$ because of the properties of ℓ_2-projection residuals. Regarding the variance of this term, the next lemma shows that $\mathbb{V}[\mathbf{L}_{+,p,s}(h)]$ is $O(1/\sqrt{nh})$. Then, it follows that $\mathbf{L}_{+,p,s}(h) = O_{\mathbb{P}}(1/\sqrt{nh})$ by Lemma SA-1.

Lemma SA-5. *Let Assumptions SA1, SA2, SA3, and SA5 hold with* $\mathsf{k} \geq h$. *Then, for fixed* $h > 0$

$$\mathbb{V}[\mathbf{L}_{-,p,s}(h)] = \frac{1}{nh}\mathbf{V}_{-,p,s}(h), \qquad \mathbb{V}[\mathbf{L}_{+,p,s}(h)] = \frac{1}{nh}\mathbf{V}_{+,p,s}(h).$$

Furthermore, as $nh \to \infty$ *and* $h \to 0$

$$\mathbf{V}_{-,p,s}(h) = \mathbf{V}_{-,p,s}\{1 + o(1)\}, \quad \mathbf{V}_{+,p,s}(h) = \mathbf{V}_{+,p,s}\{1 + o(1)\},$$

where $|\mathbf{V}_{-,p,s}| < \infty$ *and* $|\mathbf{V}_{+,p,s}| < \infty$.

[Proof]

Typically, we are interested in estimators of the form $\breve{\mathbf{e}}_\nu'\widehat{\boldsymbol{\vartheta}}_{+,p,s}(h_+)$. By Lemma SA-3 and Lemma SA-5 and Slutsky's theorem, we know that

$$\sqrt{nh^{2\nu+1}}\breve{\mathbf{e}}_\nu'(\widehat{\boldsymbol{\vartheta}}_{+,p,s}(h) - \boldsymbol{\vartheta}^\star_{+,p,s}(h)) = \sqrt{nh}\breve{\mathbf{e}}_\nu'\widehat{\boldsymbol{\Gamma}}^{-1}_{+,p,s}(h)\mathbf{L}_{+,p,s}(h) = O_{\mathbb{P}}(1).$$

The next theorem shows that the term above is not only bounded in probability but also converging in distribution to a Normal random variable.

Theorem SA-1. *Let Assumptions SA1, SA2, SA3, and SA5 hold with* $\mathsf{k} \geq h$. *If* $nh \to \infty$ *and* $h \to 0$, *then*

$$\sqrt{nh}\mathbf{H}_{p,s}(h)(\widehat{\boldsymbol{\vartheta}}_{-,p,s}(h) - \boldsymbol{\vartheta}^\star_{-,p,s}(h)) \rightsquigarrow \mathsf{N}(\mathbf{0}_{1+p+d}, \boldsymbol{\Omega}_{-,p,s}),$$
$$\sqrt{nh}\mathbf{H}_{p,s}(h)(\widehat{\boldsymbol{\vartheta}}_{+,p,s}(h) - \boldsymbol{\vartheta}^\star_{+,p,s}(h)) \rightsquigarrow \mathsf{N}(\mathbf{0}_{1+p+d}, \boldsymbol{\Omega}_{+,p,s}),$$

with

$$\boldsymbol{\Omega}_{-,p,s} = \boldsymbol{\Gamma}^{-1}_{-,p,s}\mathbf{V}_{-,p,s}\boldsymbol{\Gamma}^{-1\prime}_{-,p,s}, \qquad \boldsymbol{\Omega}_{+,p,s} = \boldsymbol{\Gamma}^{-1}_{+,p,s}\mathbf{V}_{+,p,s}\boldsymbol{\Gamma}^{-1\prime}_{+,p,s},$$

where $\boldsymbol{\Omega}_{-,p,s}$ *and* $\boldsymbol{\Omega}_{+,p,s}$ *are positive definite matrices.*

[Proof]

Corollary SA-2. *Let the conditions in Theorem SA-1 hold, then for* $\nu \in \{0, \cdots, p \wedge s\}$:

$$\sqrt{nh^{2\nu+1}}\breve{\mathbf{e}}_\nu'(\widehat{\boldsymbol{\vartheta}}_{-,p,s}(h) - \boldsymbol{\vartheta}^\star_{-,p,s}(h)) \rightsquigarrow \mathsf{N}(0, \mathcal{V}_{-,\nu,p,s}),$$
$$\sqrt{nh^{2\nu+1}}\breve{\mathbf{e}}_\nu'(\widehat{\boldsymbol{\vartheta}}_{+,p,s}(h) - \boldsymbol{\vartheta}^\star_{+,p,s}(h)) \rightsquigarrow \mathsf{N}(0, \mathcal{V}_{+,\nu,p,s}),$$

where

$$\mathcal{V}_{-,\nu,p,s} := \breve{\mathbf{e}}_\nu'\boldsymbol{\Omega}_{-,p,s}\breve{\mathbf{e}}_\nu, \qquad \mathcal{V}_{+,\nu,p,s} := \breve{\mathbf{e}}_\nu'\boldsymbol{\Omega}_{+,p,s}\breve{\mathbf{e}}_\nu.$$

SA2.2 Variance Estimation

To estimate the asymptotic variance of the RD estimator, we propose the following *plug-in* estimator

$$\widehat{\mathcal{V}}_{+,\nu,p,s}(h) = \breve{\mathbf{e}}_\nu'\widehat{\boldsymbol{\Gamma}}^{-1}_{+,p,s}(h)\widehat{\mathbf{V}}_{+,p,s}(h)\widehat{\boldsymbol{\Gamma}}^{-1}_{+,p,s}(h)'\breve{\mathbf{e}}_\nu,$$

where

$$\widehat{\mathbf{V}}_{+,p,s}(h) := \frac{1}{nh}\sum_{i=1}^n \sqrt{w_{+,i}(h)} \cdot \mathbb{1}(X_i \geq c)K\left(\frac{X_i - c}{h}\right)\mathbf{r}_{p,s}\left(\frac{X_i - c}{h}, \mathbf{W}_i\right)\mathbf{r}_{p,s}\left(\frac{X_i - c}{h}, \mathbf{W}_i\right)'\widehat{u}^2_{+,i}(h),$$

with $\widehat{u}_{+,i}(h) := Y_i - \mathbf{r}_{p,s}(X_i - c, \mathbf{W}_i)'\widehat{\boldsymbol{\vartheta}}_{+,p,s}(h)$.

The weights $w_{+,i}(h)$ allow for different HC-type estimators. Namely,

	HC_0	HC_1	HC_2	HC_3
$w_{-,i}(h)$	1	$\frac{N_-}{N_- - 2\,\mathrm{tr}(\mathbf{Q}_{-,p,s}) + \mathrm{tr}(\mathbf{Q}_{-,p,s}\mathbf{Q}_{-,p,s})}$	$\frac{1}{1-L_{+,i}}$	$\frac{1}{(1-L_{+,i})^2}$
$w_{+,i}(h)$	1	$\frac{N_+}{N_+ - 2\,\mathrm{tr}(\mathbf{Q}_{+,p,s}) + \mathrm{tr}(\mathbf{Q}_{+,p,s}\mathbf{Q}_{+,p,s})}$	$\frac{1}{1-L_{+,i}}$	$\frac{1}{(1-L_{+,i})^2}$

where

$$N_- := \sum_{i=1}^n \mathbb{1}(X_i < c) \quad \text{and} \quad N_+ := \sum_{i=1}^n \mathbb{1}(X_i \geq c),$$

$\mathbf{Q}_{-,p,s}$, and $\mathbf{Q}_{+,p,s}$ are the "projection" matrices used to get the estimated residuals and defined as,

$$\mathbf{Q}_{-,p,s} := \mathbf{R}_{p,s}(h)\widehat{\mathbf{\Gamma}}^{-1}_{-,p,s}\mathbf{R}_{p,s}(h)'\mathbf{K}_{-}(h)/n, \quad \mathbf{Q}_{+,p,s} = \mathbf{R}_{p,s}(h)\widehat{\mathbf{\Gamma}}^{-1}_{+,p,s}\mathbf{R}_{p,s}(h)'\mathbf{K}_{+}(h)/n,$$

and

$$L_{-,i} := \mathbf{e}_i'\mathbf{Q}_{-,p,s}\mathbf{e}_i, \qquad L_{+,i} := \mathbf{e}_i'\mathbf{Q}_{+,p,s}\mathbf{e}_i,$$

which denote the leverage of each observation.

Theorem SA-3. *Let the assumptions of Theorem SA-1 hold then*

$$\widehat{\mathcal{V}}_{-,\nu,p,s}(h) \xrightarrow{\mathbb{P}} \mathcal{V}_{-,\nu,p,s}, \qquad \widehat{\mathcal{V}}_{+,\nu,p,s}(h) \xrightarrow{\mathbb{P}} \mathcal{V}_{+,\nu,p,s}.$$

[Proof]

Define the standard error of the RD estimators as

$$\widehat{\sigma}_{-,\nu,p,s}(h) := \left(\frac{1}{nh^{2\nu+1}}\widehat{\mathcal{V}}_{-,\nu,p,s}(h)\right)^{1/2}, \qquad \widehat{\sigma}_{+,\nu,p,s}(h) := \left(\frac{1}{nh^{2\nu+1}}\widehat{\mathcal{V}}_{+,\nu,p,s}(h)\right)^{1/2}.$$

The lemma above naturally yields the following corollary via Slutsky's theorem.

Corollary SA-4. *Let Assumptions SA1, SA2, SA3, and SA5 hold with $\mathsf{k} \geq h$. If $nh^{2\nu+1} \to \infty$ and $h \to 0$ then for $\nu \in \{0, 1, \ldots, p \wedge s\}$*

$$\widehat{\sigma}^{-1}_{-,\nu,p,s}(h)\breve{\mathbf{e}}_\nu'(\widehat{\boldsymbol{\vartheta}}_{-,p,s}(h) - \boldsymbol{\vartheta}^\star_{-,p,s}(h)) \rightsquigarrow \mathsf{N}(0,1),$$
$$\widehat{\sigma}^{-1}_{+,\nu,p,s}(h)\breve{\mathbf{e}}_\nu'(\widehat{\boldsymbol{\vartheta}}_{+,p,s}(h) - \boldsymbol{\vartheta}^\star_{+,p,s}(h)) \rightsquigarrow \mathsf{N}(0,1).$$

Clustered Data

In the case of clustered data, the extension of the above results is immediate. The only difference would be reflected in the form of the "meat" matrices $\mathbf{V}_{-,\nu,p,s}$, and $\mathbf{V}_{+,\nu,p,s}$, which will ultimately depend on the particular form of clustering being used. For a review on cluster-robust inference see Cameron and Miller (2015) and MacKinnon et al. (2023).

We assume that each unit $i \in \{1, 2, \cdots, n\}$ belongs to a single cluster $s(i) \in \{1, 2, \cdots, G\}$, where $s : \{1, 2, \cdots, n\} \to \{1, 2, \cdots, G\}$. Furthermore, we consider an asymptotic regime where the number of clusters G grows large, $G \to \infty$, and dominates the bandwidth $Gh \to \infty$.

To estimate the asymptotic variance of the RD estimator under clustering, we propose the following *plug-in* estimator

$$\widehat{\mathcal{V}}^{\mathsf{CL}}_{+,\nu,p,s}(h) = \frac{(G-1)n}{(G-1)(n-p-1-d)}\breve{\mathbf{e}}_\nu'\widehat{\mathbf{\Gamma}}^{-1}_{+,p,s}(h)\widehat{\mathbf{V}}^{\mathsf{CL}}_{+,p,s}(h)\widehat{\mathbf{\Gamma}}^{-1}_{+,p,s}(h)'\breve{\mathbf{e}}_\nu,$$

where

$$\widehat{\mathbf{V}}^{\mathsf{CL}}_{+,p,s}(h) := \frac{1}{Gh}\sum_{g=1}^{G}\sum_{i,j=1}^{n}\mathbb{1}(s(i)=s(j))\cdot\mathbb{1}(X_i \geq c, X_j \geq c)\times$$
$$K\left(\frac{X_i-c}{h}\right)K\left(\frac{X_j-c}{h}\right)\mathbf{r}_{p,s}\left(\frac{X_i-c}{h},\mathbf{W}_i\right)\mathbf{r}_{p,s}\left(\frac{X_j-c}{h},\mathbf{W}_j\right)'\widehat{u}_{+,i}(h)\widehat{u}_{+,j}(h).$$

This estimator, as well as many more, is implemented in our `R` and `Stata` software.

SA2.3 Smoothing Bias and Probability Limit

Without further assumptions on the data generating process or on $\mu_{Y-}(X_i, \mathbf{W}_i)$ and $\mu_{Y+}(X_i, \mathbf{W}_i)$, the probability limit of the RD estimator has a generic best linear mean squared error predictor interpretation. We need to leverage extra structure to attach more interpretation to this probability limit. In this spirit, we posit Assumption SA4, which we state again here below for the reader's convenience.

Assumption SA4. *The expectation of $Y_i(t), t \in \{0,1\}$ conditional on $(X_i, \mathbf{W}_i)$ is*

$$\mu_-(X_i, \mathbf{W}_i) = \alpha_-(X_i) + \boldsymbol{\lambda}_-(X_i)'\mathbf{W}_i, \qquad \mu_+(X_i, \mathbf{W}_i) = \alpha_+(X_i) + \boldsymbol{\lambda}_+(X_i)'\mathbf{W}_i,$$

where $\boldsymbol{\lambda}_-(x) := (\lambda_{-,1}(x), \cdots, \lambda_{-,d}(x))', \boldsymbol{\lambda}_+(x) := (\lambda_{+,1}(x), \cdots, \lambda_{+,d}(x))'$, and $\alpha_-(x), \alpha_+(x) \in \mathcal{C}^{q+2}$, and $\lambda_{-,\ell}(x), \lambda_{+,\ell}(x), \in \mathcal{C}^{q+2}$ for all $\ell \in \{1, \cdots, d\}$ for some $q \in \mathbb{N}$.

Let

$$\boldsymbol{\vartheta}_{-,p,s} := \begin{bmatrix} \boldsymbol{\alpha}_{-,p}(c) \\ \boldsymbol{\lambda}_{-,s}(c) \end{bmatrix}, \qquad \boldsymbol{\vartheta}_{+,p,s} := \begin{bmatrix} \boldsymbol{\alpha}_{+,p}(c) \\ \boldsymbol{\lambda}_{+,s}(c) \end{bmatrix}, \tag{4}$$

where

$$\boldsymbol{\alpha}_{-,p}(x) = \begin{bmatrix} \alpha_-(x) \\ \alpha_-^{(1)}(x) \\ \vdots \\ \frac{\alpha_-^{(p)}(x)}{p!} \end{bmatrix}, \quad \boldsymbol{\lambda}_{-,s}(x) = \begin{bmatrix} \lambda_{-,1}(x) \\ \vdots \\ \frac{\lambda_{-,1}^{(s)}(x)}{s!} \\ \vdots \\ \lambda_{-,d}(x) \\ \vdots \\ \frac{\lambda_{-,d}^{(s)}(x)}{s!} \end{bmatrix}, \quad \boldsymbol{\alpha}_{+,p}(x) = \begin{bmatrix} \alpha_+(x) \\ \alpha_+^{(1)}(x) \\ \vdots \\ \frac{\alpha_+^{(p)}(x)}{p!} \end{bmatrix}, \quad \boldsymbol{\lambda}_{+,s}(x) = \begin{bmatrix} \lambda_{+,1}(x) \\ \vdots \\ \frac{\lambda_{+,1}^{(s)}(x)}{s!} \\ \vdots \\ \lambda_{+,d}(x) \\ \vdots \\ \frac{\lambda_{+,d}^{(s)}(x)}{s!} \end{bmatrix},$$

with $\boldsymbol{\alpha}_{-,p} \equiv \boldsymbol{\alpha}_{-,p}(c), \boldsymbol{\lambda}_{-,s} \equiv \boldsymbol{\lambda}_{-,s}(c), \boldsymbol{\alpha}_{+,p} \equiv \boldsymbol{\alpha}_{+,p}(c)$, and $\boldsymbol{\lambda}_{+,s} \equiv \boldsymbol{\lambda}_{+,s}(c)$.

Remark SA-1. Note that Assumption SA4 is without loss of generality whenever $\mathbf{W}_i$ is a vector of binary covariates. However, when $d > 1$ some extra care is needed in constructing $\mathbf{W}_i$. Indeed, suppose that we have two binary covariates V_i and U_i, then $d = 3$ and $\mathbf{W}_i = (V_i, U_i, V_i \cdot U_i)'$. ♣

Consider the following decomposition

$$\mathbf{H}_{p,s}(h)(\widehat{\boldsymbol{\vartheta}}_{+,p,s}(h) - \boldsymbol{\vartheta}_{+,p,s}) = \mathbf{H}_{p,s}(h)(\widehat{\boldsymbol{\vartheta}}_{+,p,s}(h) - \boldsymbol{\vartheta}^\star_{+,p,s}(h)) + \mathbf{H}_{p,s}(h)(\boldsymbol{\vartheta}^\star_{+,p,s}(h) - \boldsymbol{\vartheta}_{+,p,s}).$$

From Theorem SA-1, we know that $\mathbf{H}_{p,s}(h)(\widehat{\boldsymbol{\vartheta}}_{+,p,s}(h) - \boldsymbol{\vartheta}^\star_{+,p,s}(h)) = O_{\mathbb{P}}(1/\sqrt{nh})$. In what follows, we show that the second term is of order $O_{\mathbb{P}}(h^{1+p\wedge s})$. Throughout, we consider an asymptotic regime in which this term is negligible with respect to the first term. It is precisely in this spirit that we will refer to $\boldsymbol{\vartheta}^\star_{+,p,s}(h) - \boldsymbol{\vartheta}_{+,p,s}$ as a "bias" term. The next lemma shows that the leading term of the bias is of order $O(h^{1+p\wedge s})$.

Lemma SA-6 (Smoothing Bias)**.** *Let Assumptions SA1-SA5 hold for some $q \geq 2 + p \vee s$ and with $\mathsf{k} \geq h$. If $nh \to 0$ and $h \to 0$, then*

$$\boldsymbol{\vartheta}^\star_{-,p,s}(h) - \boldsymbol{\vartheta}_{-,p,s} = \mathbf{H}^{-1}_{p,s}(h)\left(h^{1+p}\widetilde{\mathbf{B}}^{[0]}_{-,p}(h) + h^{2+p}\widetilde{\mathbf{B}}^{[0]}_{-,p+1}(h) + h^{1+s}\widetilde{\mathbf{B}}^{[1]}_{-,s}(h) + h^{2+s}\widetilde{\mathbf{B}}^{[1]}_{-,s+1}(h) + o(h^{2+p\wedge s})\right),$$

$$\boldsymbol{\vartheta}^\star_{+,p,s}(h) - \boldsymbol{\vartheta}_{+,p,s} = \mathbf{H}^{-1}_{p,s}(h)\left(h^{1+p}\widetilde{\mathbf{B}}^{[0]}_{+,p}(h) + h^{2+p}\widetilde{\mathbf{B}}^{[0]}_{+,p+1}(h) + h^{1+s}\widetilde{\mathbf{B}}^{[1]}_{+,s}(h) + h^{2+s}\widetilde{\mathbf{B}}^{[1]}_{+,s+1}(h) + o(h^{2+p\wedge s})\right),$$

where

$$\widetilde{\mathbf{B}}^{[0]}_{-,a}(h) = \widetilde{\mathbf{\Gamma}}^{-1}_{-,p,s}(h)\widetilde{\boldsymbol{\zeta}}_{-,p,s,a}(h)\frac{\alpha^{(a+1)}_{-}(c)}{(a+1)!} \to \mathbf{\Gamma}^{-1}_{-,p,s}\boldsymbol{\zeta}_{-,p,s,a}\frac{\alpha^{(a+1)}_{+}(c)}{(a+1)!} =: \mathbf{B}^{[0]}_{-,a},$$
$$\widetilde{\mathbf{B}}^{[0]}_{+,a}(h) = \widetilde{\mathbf{\Gamma}}^{-1}_{+,p,s}(h)\widetilde{\boldsymbol{\zeta}}_{+,p,s,a}(h)\frac{\alpha^{(a+1)}_{+}(c)}{(a+1)!} \to \mathbf{\Gamma}^{-1}_{+,p,s}\boldsymbol{\zeta}_{+,p,s,a}\frac{\alpha^{(a+1)}_{+}(c)}{(a+1)!} =: \mathbf{B}^{[0]}_{+,a},$$
$$\widetilde{\mathbf{B}}^{[1]}_{-,a}(h) = \widetilde{\mathbf{\Gamma}}^{-1}_{-,p,s}(h)\widetilde{\boldsymbol{\varphi}}_{-,p,s,a}(h)\frac{\boldsymbol{\lambda}^{(a+1)}_{-}(c)}{(a+1)!} \to \mathbf{\Gamma}^{-1}_{-,p,s}\boldsymbol{\varphi}_{-,p,s,a}\frac{\boldsymbol{\lambda}^{(a+1)}_{-}(c)}{(a+1)!} =: \mathbf{B}^{[1]}_{-,a},$$
$$\widetilde{\mathbf{B}}^{[1]}_{+,a}(h) = \widetilde{\mathbf{\Gamma}}^{-1}_{+,p,s}(h)\widetilde{\boldsymbol{\varphi}}_{+,p,s,a}(h)\frac{\boldsymbol{\lambda}^{(a+1)}_{+}(c)}{(a+1)!} \to \mathbf{\Gamma}^{-1}_{+,p,s}\boldsymbol{\varphi}_{+,p,s,a}\frac{\boldsymbol{\lambda}^{(a+1)}_{+}(c)}{(a+1)!} =: \mathbf{B}^{[1]}_{+,a}.$$

[Proof]

The following corollary follows naturally from Theorem SA-1 and Lemma SA-6.

Corollary SA-5. *Let the assumptions of Lemma SA-6 hold. If* $nh^{2\nu+1} \to \infty, h \to 0,$ *and* $nh^{2(p\wedge s)+3} \to 0$*, then*

$$\sqrt{nh^{2\nu+1}}\breve{\mathbf{e}}'_{\nu}(\widehat{\boldsymbol{\vartheta}}_{-,p,s}(h) - \boldsymbol{\vartheta}_{-,p,s}) \rightsquigarrow \mathsf{N}(0, \mathcal{V}_{-,\nu,p,s}),$$
$$\sqrt{nh^{2\nu+1}}\breve{\mathbf{e}}'_{\nu}(\widehat{\boldsymbol{\vartheta}}_{+,p,s}(h) - \boldsymbol{\vartheta}_{+,p,s}) \rightsquigarrow \mathsf{N}(0, \mathcal{V}_{+,\nu,p,s}),$$

where

$$\mathcal{V}_{-,\nu,p,s} := \breve{\mathbf{e}}'_{\nu}\mathbf{\Omega}_{-,p,s}\breve{\mathbf{e}}_{\nu}, \qquad \mathcal{V}_{+,\nu,p,s} := \breve{\mathbf{e}}'_{\nu}\mathbf{\Omega}_{+,p,s}\breve{\mathbf{e}}_{\nu},$$

with

$$\mathbf{\Omega}_{-,p,s} = \mathbf{\Gamma}^{-1}_{-,p,s}\mathbf{V}_{-,p,s}\mathbf{\Gamma}^{-1\prime}_{-,p,s}, \qquad \mathbf{\Omega}_{+,p,s} = \mathbf{\Gamma}^{-1}_{+,p,s}\mathbf{V}_{+,p,s}\mathbf{\Gamma}^{-1\prime}_{+,p,s},$$

and where

$$\mathbf{V}_{-,p,s} = f(c)\int_{-\infty}^{0} K^2(u)\begin{bmatrix} \mathbf{r}_p(u)\mathbf{r}_p(u)'\mathbb{E}[\sigma^2_{Y-}(X_i,\mathbf{W}_i) \mid X_i = c] & \mathbb{E}[\mathbf{W}_i'\sigma^2_{Y-}(X_i,\mathbf{W}_i) \mid X_i = c] \otimes \mathbf{r}_p(u)\mathbf{r}_s(u)' \\ \mathbb{E}[\mathbf{W}_i\sigma^2_{Y-}(X_i,\mathbf{W}_i) \mid X_i = c] \otimes \mathbf{r}_s(u)\mathbf{r}_p(u)' & \mathbb{E}[\mathbf{W}_i\mathbf{W}_i'\sigma^2_{Y-}(X_i,\mathbf{W}_i) \mid X_i = c] \otimes \mathbf{r}_s(u)\mathbf{r}_s(u)' \end{bmatrix} \mathrm{d}u,$$
$$\mathbf{V}_{+,p,s} = f(c)\int_{0}^{\infty} K^2(u)\begin{bmatrix} \mathbf{r}_p(u)\mathbf{r}_p(u)'\mathbb{E}[\sigma^2_{Y+}(X_i,\mathbf{W}_i) \mid X_i = c] & \mathbb{E}[\mathbf{W}_i'\sigma^2_{Y+}(X_i,\mathbf{W}_i) \mid X_i = c] \otimes \mathbf{r}_p(u)\mathbf{r}_s(u)' \\ \mathbb{E}[\mathbf{W}_i\sigma^2_{Y+}(X_i,\mathbf{W}_i) \mid X_i = c] \otimes \mathbf{r}_s(u)\mathbf{r}_p(u)' & \mathbb{E}[\mathbf{W}_i\mathbf{W}_i'\sigma^2_{Y+}(X_i,\mathbf{W}_i) \mid X_i = c] \otimes \mathbf{r}_s(u)\mathbf{r}_s(u)' \end{bmatrix} \mathrm{d}u$$

are positive-definite matrices.

SA2.4 Consistency

Using Lemma SA-5 and Lemma SA-6, we have

$$\mathbf{H}_{p,s}(h)\left(\widehat{\boldsymbol{\vartheta}}_{-,p,s}(h) - \boldsymbol{\vartheta}_{-,p,s}\right) = O_{\mathbb{P}}\left(\frac{1}{\sqrt{nh}} + h^{1+p\wedge s}\right),$$
$$\mathbf{H}_{p,s}(h)\left(\widehat{\boldsymbol{\vartheta}}_{+,p,s}(h) - \boldsymbol{\vartheta}_{+,p,s}\right) = O_{\mathbb{P}}\left(\frac{1}{\sqrt{nh}} + h^{1+p\wedge s}\right),$$

thus if $nh \to \infty$ and $h \to 0$, then $\widehat{\boldsymbol{\vartheta}}_{-,p,s}(h) \xrightarrow{\mathbb{P}} \boldsymbol{\vartheta}_{-,p,s}$ and $\widehat{\boldsymbol{\vartheta}}_{+,p,s}(h) \xrightarrow{\mathbb{P}} \boldsymbol{\vartheta}_{+,p,s}$.

Typically, we are interested in estimators of the form $\breve{\mathbf{e}}_\nu'\widehat{\boldsymbol{\vartheta}}_{-,p,s}(h)$ or $\breve{\mathbf{e}}_\nu'\widehat{\boldsymbol{\vartheta}}_{+,p,s}(h)$. In this case, we get

$$\breve{\mathbf{e}}_\nu'\left(\widehat{\boldsymbol{\vartheta}}_{-,p,s}(h)-\boldsymbol{\vartheta}_{-,p,s}\right)=O_{\mathbb{P}}\left(\frac{1}{\sqrt{nh^{2\nu+1}}}+h^{1+p\wedge s-\nu}\right),$$
$$\breve{\mathbf{e}}_\nu'\left(\widehat{\boldsymbol{\vartheta}}_{+,p,s}(h)-\boldsymbol{\vartheta}_{+,p,s}\right)=O_{\mathbb{P}}\left(\frac{1}{\sqrt{nh^{2\nu+1}}}+h^{1+p\wedge s-\nu}\right).$$

In an asymptotic regime where also $n\min\{h_-,h_+\}^{2\nu+1}\to\infty$ and $h\to 0$, then $\breve{\mathbf{e}}_\nu'\widehat{\boldsymbol{\vartheta}}_{-,p,s}(h)\xrightarrow{\mathbb{P}}\breve{\mathbf{e}}_\nu'\boldsymbol{\vartheta}_{-,p,s}$ and $\breve{\mathbf{e}}_\nu'\widehat{\boldsymbol{\vartheta}}_{+,p,s}(h)\xrightarrow{\mathbb{P}}\breve{\mathbf{e}}_\nu'\boldsymbol{\vartheta}_{+,p,s}$.

Define

$$\widehat{\chi}_{\nu,p,s}(\mathbf{h}):=\breve{\mathbf{e}}_\nu'\left(\widehat{\boldsymbol{\vartheta}}_{+,p,s}(h_+)-\widehat{\boldsymbol{\vartheta}}_{-,p,s}(h_-)\right),\quad \chi_{\nu,p,s}:=\breve{\mathbf{e}}_\nu'\left(\boldsymbol{\vartheta}_{+,p,s}-\boldsymbol{\vartheta}_{-,p,s}\right).$$

Then

$$\widehat{\chi}_{\nu,p,s}(\mathbf{h})-\chi_{\nu,p,s}=O_{\mathbb{P}}\left(\frac{1}{\sqrt{n\min\{h_-,h_+\}^{2\nu+1}}}+\max\{h_-,h_+\}^{1+p\wedge s-\nu}\right),$$

and $\widehat{\chi}_{\nu,p,s}(\mathbf{h})\xrightarrow{\mathbb{P}}\chi_{\nu,p,s}$.

SA2.5 Causal Interpretation of the Probability Limit

Before delving into this section, we redefine the "extractor" vector as

$$\breve{\mathbf{e}}_{\nu_x,\nu_w}=\begin{bmatrix}\nu_x\mathbf{I}_{p+1}\mathbf{e}_{\nu_x}\\ \nu_w!\boldsymbol{\iota}_d\otimes\mathbf{I}_{s+1}\mathbf{e}_{\nu_w}\end{bmatrix}$$

to be the vector extracting the (ν_x+1)-th term in the polynomial basis in (X_i-c) and the (ν_w+1)-th terms in the interaction between such basis and $\mathbf{W}_i$. Furthermore, we define $\breve{\mathbf{e}}_{\nu_x,\cdot}\equiv\mathbf{I}_{1+p+d}\mathbf{e}_{\nu_x}$, to be the vector that extracts only the (ν_x+1)-th term in the polynomial basis in (X_i-c). We also define accordingly

$$\chi_{\nu_x,\nu_w,p,s}:=\breve{\mathbf{e}}_{\nu_x,\nu_w}'\left(\boldsymbol{\vartheta}_{+,p,s}-\boldsymbol{\vartheta}_{-,p,s}\right).$$

The quantity $\chi_{\nu_x,\nu_w,p,s}$ has a natural causal interpretation in many cases which depends on the elements extracted:

1. When $d=1$ and W_i is binary, we have that

   $$\chi_{\nu_x,\cdot,p,s}=\frac{\partial^{\nu_x}}{\partial x^{\nu_x}}\mathbb{E}[Y_i(1)-Y_i(0)\mid X_i=x,W_i=0]\Big|_{x=c},$$

   and

   $$\chi_{\nu,\nu,p,s}=\frac{\partial^{\nu}}{\partial x^{\nu}}\left(\mathbb{E}[Y_i(1)-Y_i(0)\mid X_i=x,W_i=1]-\mathbb{E}[Y_i(1)-Y_i(0)\mid X_i=x,W_i=0]\right)\Big|_{x=c}.$$

   In other words, if a researcher is interested in the conditional average treatment effect (CATE) for the two sub-populations defined by $W_i\in\{0,1\}$, then the first element of $\widehat{\boldsymbol{\vartheta}}_{+,p,s}(h_+)-\widehat{\boldsymbol{\vartheta}}_{-,p,s}(h_-)$ identifies the CATE for the baseline $(W_i=0)$ group, whereas the $(2+p)$-th element identifies the difference in CATEs between in the sub-population with $W_i=1$ and the baseline one.

2. When there is more than one binary covariate, say U_i and V_i, then $\mathbf{W}_i=(U_i,V_i,U_i\cdot V_i)'$ to correctly identify the CATEs in all the sub-populations.

3. If $d=1$ and the covariate of interest W_i is categorical (e.g., race) or multi-valued discrete (e.g., age) –i.e., takes on J distinct values $(1,\dots,J)$– then the same interpretation offered above holds as long as

the covariate is dummied out. In other words, it means that

$$\mathbf{W}_i = (I_2, I_3, \dots, I_J)', \qquad I_j := \mathbb{1}(W_i = j), j = 2, \dots, J.$$

4. If $d = 1$ and W_i is a continuous covariate (e.g., parental income), then we still have

$$\chi_{\nu_x,\cdot,p,s} = \frac{\partial^{\nu_x}}{\partial x^{\nu_x}} \mathbb{E}[Y_i(1) - Y_i(0) \mid X_i = x, W_i = 0]\Big|_{x=c},$$

but the interpretation of the coefficients on the interaction terms between $(X_i - c)$ and W_i changes as follows

$$\chi_{\nu,p,s} = \frac{\partial^{\nu_x}}{\partial x^{\nu_x}} \frac{\partial^{\nu_w}}{\partial w^{\nu_w}} \mathbb{E}[Y_i(1) - Y_i(0) \mid X_i = x, W_i = w]\Big|_{x=c},$$

where the evaluation point of the derivative in the W-dimension needs not to be specified as this derivative is constant in the W-dimension by Assumption SA4. The coefficient on the interaction terms simply captures the change in the CATEs due to a marginal change in W_i. This change is assumed to be linear in virtue of Assumption SA4.

SA2.6 MSE Expansions

We now provide first-order expansions for the MSE of the RD estimator. This is crucial to then obtain formulas for MSE-optimal bandwidths. We first start by providing a Nagar expansion of the unconditional MSE. Then we provide an approximation for the bias and the variance of the RD estimator, where the approximation comes from disregarding higher-order terms.

First of all, recall that the RD estimator and the best linear mean square error predictor are

$$\widehat{\boldsymbol{\vartheta}}_{+,p,s}(h) = \mathbf{H}_{p,s}^{-1}(h)\widehat{\boldsymbol{\Gamma}}_{+,p,s}^{-1}(h)\widehat{\boldsymbol{\Upsilon}}_{+,p,s}(h), \qquad \boldsymbol{\vartheta}^\star_{+,p,s}(h) = \mathbf{H}_{p,s}^{-1}(h)\widetilde{\boldsymbol{\Gamma}}_{+,p,s}^{-1}(h)\widetilde{\boldsymbol{\Upsilon}}_{+,p,s}(h).$$

Consider the following decomposition

$$\begin{aligned}
&\breve{\mathbf{e}}_\nu'\widehat{\boldsymbol{\vartheta}}_{+,p,s}(h) - \mu_+^{(\nu)}(c, \mathbf{W}_i) \\
&\quad = h^{-\nu}\breve{\mathbf{e}}_\nu'\widehat{\boldsymbol{\Gamma}}_{+,p,s}^{-1}(h)\left(\widehat{\boldsymbol{\Upsilon}}_{+,p,s}(h) - \widehat{\boldsymbol{\Gamma}}_{+,p,s}(h)\mathbf{H}_{p,s}(h)\boldsymbol{\vartheta}^\star_{+,p,s}(h)\right) + \breve{\mathbf{e}}_\nu'\boldsymbol{\vartheta}^\star_{+,p,s}(h) - \mu_+^{(\nu)}(c, \mathbf{W}_i) \\
&\quad = \underbrace{h^{-\nu}\breve{\mathbf{e}}_\nu'\widetilde{\boldsymbol{\Gamma}}_{+,p,s}^{-1}(h)\mathbf{L}_{+,p,s}(h)}_{\mathsf{L}_{+,\nu,p,s}(h)} + \underbrace{h^{-\nu}\breve{\mathbf{e}}_\nu'\left(\widehat{\boldsymbol{\Gamma}}_{+,p,s}^{-1}(h) - \widetilde{\boldsymbol{\Gamma}}_{+,p,s}^{-1}(h)\right)\mathbf{L}_{+,p,s}(h)}_{\mathsf{Q}_{+,\nu,p,s}(h)} + \underbrace{\breve{\mathbf{e}}_\nu'\boldsymbol{\vartheta}^\star_{+,p,s}(h) - \mu_+^{(\nu)}(c, \mathbf{W}_i)}_{\mathsf{B}_{+,\nu,p,s}(h)},
\end{aligned}$$

where we refer to $\mathsf{L}_{+,\nu,p,s}(h)$ as the "linear" term, to $\mathsf{Q}_{+,\nu,p,s}(h)$ as the "quadratic" term, and to $\mathsf{B}_{+,\nu,p,s}(h)$ as the "bias" term. By Theorem SA-1 and Lemma SA-3, we know that $\mathsf{L}_{+,\nu,p,s}(h) = O_\mathbb{P}(1/\sqrt{nh^{2\nu+1}})$ and $\mathsf{Q}_{+,\nu,p,s}(h) = O_\mathbb{P}(1/\sqrt{n^2h^{2\nu+1}}) = o_\mathbb{P}(\mathsf{L}_{+,\nu,p,s}(h))$. Moreover, under Assumption SA4, we have that $\mu_+^{(\nu)}(c, \mathbf{W}_i) = \breve{\mathbf{e}}'\boldsymbol{\vartheta}_{+,p,s}(h)$, thus by Lemma SA-6 we have that $\mathsf{B}_{+,\nu,p,s}(h) = O(h^{1+p\wedge s-\nu})$. In an asymptotic regime in which $nh^{2+p\wedge s} \to \infty$, we have that $\mathsf{Q}_{+,\nu,p,s}(h) = o_\mathbb{P}(\mathsf{B}_{+,\nu,p,s}(h))$. Therefore, under Assumptions SA1-SA4, as $n \to \infty$ and $nh^{2+p} \to \infty$ we have

$$\breve{\mathbf{e}}_\nu'(\widehat{\boldsymbol{\vartheta}}_{+,p,s}(h) - \boldsymbol{\vartheta}_{+,p,s}(h)) = \mathsf{L}_{+,\nu,p,s}(h) + \mathsf{B}_{+,\nu,p,s}(h) + o_\mathbb{P}(\min\{\mathsf{L}_{+,\nu,p,s}(h), \mathsf{B}_{+,\nu,p,s}(h)\}).$$

With this result at hand, we define the first-order approximation of the mean squared error as

$$\mathsf{MSE}[\breve{\mathbf{e}}_\nu'(\widehat{\boldsymbol{\vartheta}}_{+,p,s}(h) - \boldsymbol{\vartheta}_{+,p,s})] = \mathbb{E}\left[(\mathsf{L}_{+,\nu,p,s}(h) + \mathsf{B}_{+,\nu,p,s}(h))^2\right] = \mathbb{V}[\mathsf{L}_{+,\nu,p,s}(h)] + \mathsf{B}_{+,\nu,p,s}(h)^2,$$

where the second equality follows from the fact that $\mathsf{B}_{+,\nu,p,s}(h)$ is non-random and $\mathsf{L}_{+,\nu,p,s}(h)$ is unconditionally mean-zero. Similarly, we define

$$\mathsf{MSE}[\breve{\mathbf{e}}_\nu'(\widehat{\boldsymbol{\vartheta}}_{-,p,s}(h) - \boldsymbol{\vartheta}_{-,p,s})] = \mathbb{V}[\mathsf{L}_{-,\nu,p,s}(h)] + \mathsf{B}_{-,\nu,p,s}(h)^2$$

for the left side of the cutoff, whereas for the difference we define

$$\begin{aligned}&\mathsf{MSE}[\breve{\mathbf{e}}_\nu'(\widehat{\boldsymbol{\vartheta}}_{+,p,s}(h)-\widehat{\boldsymbol{\vartheta}}_{-,p,s}(h)-(\boldsymbol{\vartheta}_{+,p,s}(h)-\boldsymbol{\vartheta}_{-,p,s}))]\\&\qquad\qquad=\mathbb{V}[\mathsf{L}_{+,\nu,p,s}(h)]+\mathbb{V}[\mathsf{L}_{-,\nu,p,s}(h)]+(\mathsf{B}_{+,\nu,p,s}(h)-\mathsf{B}_{-,\nu,p,s}(h))^2.\end{aligned}$$

Under Assumptions SA1-SA5 and relying on Lemma SA-6, the bias approximations for the RD estimators are given by

$$\begin{aligned}\mathsf{B}_{-,\nu,p,s}(h)&=h^{1+p-\nu}\mathcal{B}^{[0]}_{-,\nu,p,s}(h)+h^{1+s-\nu}\mathcal{B}^{[1]}_{-,\nu,p,s}(h)+o(h^{1+(p\wedge s)-\nu}),\\ \mathsf{B}_{+,\nu,p,s}(h)&=h^{1+p-\nu}\mathcal{B}^{[0]}_{+,\nu,p,s}(h)+h^{1+s-\nu}\mathcal{B}^{[1]}_{+,\nu,p,s}(h)+o(h^{1+(p\wedge s)-\nu}),\end{aligned}$$

where

$$\begin{aligned}\mathcal{B}^{[0]}_{-,\nu,p,s}(h)&=\breve{\mathbf{e}}_\nu'\mathbf{B}^{[0]}_{-,p,s}(h)\to\breve{\mathbf{e}}_\nu'\mathbf{B}^{[0]}_{-,p,s}=:\mathcal{B}^{[0]}_{-,\nu,p,s},\quad \mathcal{B}^{[1]}_{-,\nu,p,s}(h)=\breve{\mathbf{e}}_\nu'\mathbf{B}^{[1]}_{-,p,s}(h)\to\breve{\mathbf{e}}_\nu'\mathbf{B}^{[1]}_{-,p,s}(h)=:\mathcal{B}^{[1]}_{-,\nu,p,s},\\ \mathcal{B}^{[0]}_{+,\nu,p,s}(h)&=\breve{\mathbf{e}}_\nu'\mathbf{B}^{[0]}_{+,p,s}(h)\to\breve{\mathbf{e}}_\nu'\mathbf{B}^{[0]}_{-,p,s}=:\mathcal{B}^{[0]}_{+,\nu,p,s},\quad \mathcal{B}^{[1]}_{+,\nu,p,s}(h)=\breve{\mathbf{e}}_\nu'\mathbf{B}^{[1]}_{+,p,s}(h)\to\breve{\mathbf{e}}_\nu'\mathbf{B}^{[1]}_{+,p,s}(h)=:\mathcal{B}^{[1]}_{+,\nu,p,s}.\end{aligned}$$

Furthermore, we define

$$\begin{aligned}\mathcal{B}_{-,\nu,p,s}(h)&:=\mathbb{1}(p\leq s)\mathcal{B}^{[0]}_{-,\nu,p,s}(h)+\mathbb{1}(p\geq s)\mathcal{B}^{[1]}_{-,\nu,p,s}(h),\qquad \mathcal{B}_{-,\nu,p,s}:=\mathbb{1}(p\leq s)\mathcal{B}^{[0]}_{-,\nu,p,s}+\mathbb{1}(p\geq s)\mathcal{B}^{[1]}_{-,\nu,p,s},\\ \mathcal{B}_{+,\nu,p,s}(h)&:=\mathbb{1}(p\leq s)\mathcal{B}^{[0]}_{+,\nu,p,s}(h)+\mathbb{1}(p\geq s)\mathcal{B}^{[1]}_{+,\nu,p,s}(h),\qquad \mathcal{B}_{+,\nu,p,s}:=\mathbb{1}(p\leq s)\mathcal{B}^{[0]}_{+,\nu,p,s}+\mathbb{1}(p\geq s)\mathcal{B}^{[1]}_{+,\nu,p,s}.\end{aligned}$$

Under Assumptions SA1-SA5 and relying on Corollary SA-5

$$\begin{aligned}\mathbb{V}[\mathsf{L}_{-,\nu,p,s}(h)]&=\frac{1}{nh^{2\nu+1}}\mathcal{V}_{-,\nu,p,s}(h),\qquad \mathcal{V}_{-,\nu,p,s}(h):=\breve{\mathbf{e}}_\nu'\widetilde{\boldsymbol{\Gamma}}^{-1}_{-,p,s}(h)\mathbf{V}_{-,p,s}(h)\widetilde{\boldsymbol{\Gamma}}^{-1\prime}_{-,p,s}(h)\breve{\mathbf{e}}_\nu,\\ \mathbb{V}[\mathsf{L}_{+,\nu,p,s}(h)]&=\frac{1}{nh^{2\nu+1}}\mathcal{V}_{+,\nu,p,s}(h),\qquad \mathcal{V}_{+,\nu,p,s}(h):=\breve{\mathbf{e}}_\nu'\widetilde{\boldsymbol{\Gamma}}^{-1}_{+,p,s}(h)\mathbf{V}_{+,p,s}(h)\widetilde{\boldsymbol{\Gamma}}^{-1\prime}_{+,p,s}(h)\breve{\mathbf{e}}_\nu.\end{aligned}$$

The variance approximations for the RD estimators are given by

$$\mathcal{V}_{-,\nu,p,s}(h)\to\mathcal{V}_{-,\nu,p,s}:=\breve{\mathbf{e}}_\nu'\boldsymbol{\Gamma}^{-1}_{-,p,s}\mathbf{V}_{-,p,s}\boldsymbol{\Gamma}^{-1\prime}_{-,p,s}\breve{\mathbf{e}}_\nu,\qquad \mathcal{V}_{+,\nu,p,s}(h)\to\mathcal{V}_{+,\nu,p,s}:=\breve{\mathbf{e}}_\nu'\boldsymbol{\Gamma}^{-1}_{+,p,s}\mathbf{V}_{+,p,s}\boldsymbol{\Gamma}^{-1\prime}_{+,p,s}\breve{\mathbf{e}}_\nu.$$

With these approximations, the MSE expansions of the RD estimators are given by

$$\begin{aligned}\mathsf{MSE}[\breve{\mathbf{e}}_\nu'(\widehat{\boldsymbol{\vartheta}}_{-,p,s}(h)-\boldsymbol{\vartheta}_{-,p,s})]&=\frac{1}{nh^{2\nu+1}}\mathcal{V}_{-,\nu,p,s}+h^{2(1+p\wedge s-\nu)}\mathcal{B}^2_{-,\nu,p,s},\\ \mathsf{MSE}[\breve{\mathbf{e}}_\nu'(\widehat{\boldsymbol{\vartheta}}_{+,p,s}(h)-\boldsymbol{\vartheta}_{+,p,s})]&=\frac{1}{nh^{2\nu+1}}\mathcal{V}_{+,\nu,p,s}+h^{2(1+p\wedge s-\nu)}\mathcal{B}^2_{+,\nu,p,s}.\end{aligned}$$

Accordingly, we define an MSE-optimal bandwidth as the minimizer of the Nagar expansion of the unconditional MSE.

One-sided Optimal Bandwidths. The MSE-optimal bandwidths are defined as

$$\begin{aligned}h^\star_{-,\nu,p,s}&:=\underset{h>0}{\arg\min}\left[\frac{1}{nh^{2\nu+1}}\mathcal{V}_{-,\nu,p,s}+h^{2((p\wedge s)+1-\nu)}\mathcal{B}^2_{-,\nu,p,s}\right],\\ h^\star_{+,\nu,p,s}&:=\underset{h>0}{\arg\min}\left[\frac{1}{nh^{2\nu+1}}\mathcal{V}_{+,\nu,p,s}+h^{2((p\wedge s)+1-\nu)}\mathcal{B}^2_{+,\nu,p,s}\right],\end{aligned}$$

so, under the additional assumption that $\mathcal{B}_{-,\nu,p,s}\neq 0\neq\mathcal{B}_{+,\nu,p,s}$, we get

$$h^\star_{-,\nu,p,s}=\left[\frac{1+2\nu}{2(1+(p\wedge s)-\nu)n}\frac{\mathcal{V}_{-,\nu,p,s}}{\mathcal{B}^2_{-,\nu,p,s}}\right]^{\frac{1}{3+2(p\wedge s)}},\qquad h^\star_{+,\nu,p,s}=\left[\frac{1+2\nu}{2(1+(p\wedge s)-\nu)n}\frac{\mathcal{V}_{+,\nu,p,s}}{\mathcal{B}^2_{+,\nu,p,s}}\right]^{\frac{1}{3+2(p\wedge s)}}.$$

Two-sided Optimal Bandwidth. In this case a single bandwidth is chosen, so $h = h_+ = h_-$ and the optimal bandwidth is defined as

$$h^\star_{\nu,p,s} := \arg\min_{h>0} \left[\frac{1}{nh^{2\nu+1}} (\mathcal{V}_{+,\nu,p,s} + \mathcal{V}_{-,\nu,p,s}) + h^{2((p\wedge s)+1-\nu)} (\mathcal{B}_{+,\nu,p,s} - \mathcal{B}_{-,\nu,p,s})^2 \right],$$

so, under the additional assumption that $\mathcal{B}_{+,\nu,p,s} - \mathcal{B}_{-,\nu,p,s} \neq 0$, we get

$$h^\star_{\nu,p,s} = \left[\frac{1+2\nu}{2(1+(p\wedge s)-\nu)n} \frac{\mathcal{V}_{+,\nu,p,s} + \mathcal{V}_{-,\nu,p,s}}{(\mathcal{B}_{+,\nu,p,s} - \mathcal{B}_{-,\nu,p,s})^2} \right]^{\frac{1}{3+2(p\wedge s)}}.$$

SA2.7 On the Inconsistency of the RD Estimator with Interacted Covariates

Let $c = 0$ and suppose the true model is

$$Y_i = \mathbf{r}_1(X_i)'\boldsymbol{\alpha} + W_i \mathbf{r}_1(X_i)'\boldsymbol{\lambda} + W_i^2 \psi + \epsilon_i, \qquad \mathbb{E}[\epsilon_i] = \mathbb{E}[W_i\epsilon_i] = \mathbb{E}[X_i\epsilon_i] = 0,$$

or, more explicitly,

$$Y_i = \beta_0 + X_i\beta_1 + W_i\gamma_0 + W_iX_i\gamma_1 + W_i^2\psi + \epsilon_i.$$

We run the regression of Y_i onto $(\mathbf{r}_1(X_i)', W_i\mathbf{r}_1(X_i)') = (1, X_i, W_i, X_iW_i)$ with weights $K(X_i/h)$. We are interested in estimating $\boldsymbol{\vartheta} := (\boldsymbol{\alpha}', \boldsymbol{\lambda}')'$. Define $\mathbf{D}_i := (1, X_i/h, W_i, W_iX_i/h)'$. The local polynomial estimator (around 0) gives

$$\widehat{\boldsymbol{\vartheta}} = \mathbf{H}^{-1}(h)\widehat{\boldsymbol{\Gamma}}^{-1}\boldsymbol{\Upsilon}(h),$$

where $\mathbf{H}(h) = \text{diag}(1, h, 1, h)$ and

$$\widehat{\boldsymbol{\Gamma}}(h) = \frac{1}{nh}\sum_{i=1}^n K(X_i/h)\mathbf{D}_i\mathbf{D}_i', \qquad \widehat{\boldsymbol{\Upsilon}}(h) = \frac{1}{nh}\sum_{i=1}^n K(X_i/h)\mathbf{D}_iY_i.$$

Define $\widetilde{\mathbf{D}}_i := (\mathbf{D}_i', W_i^2)'$ and $\widetilde{\boldsymbol{\vartheta}} := (\boldsymbol{\vartheta}', \psi)'$. Note that $Y_i = \widetilde{\mathbf{D}}_i'\mathbf{H}(h)\widetilde{\boldsymbol{\vartheta}}$. We get that

$$\begin{aligned}
\mathbf{H}(h)(\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}) &= \widehat{\boldsymbol{\Gamma}}^{-1}(h)\left(\boldsymbol{\Upsilon}(h) - \widehat{\boldsymbol{\Gamma}}(h)\mathbf{H}(h)\boldsymbol{\vartheta}\right) \\
&= \widehat{\boldsymbol{\Gamma}}^{-1}(h)\left(\frac{1}{nh}\sum_{i=1}^n K(X_i/h)\mathbf{D}_iY_i - \frac{1}{nh}\sum_{i=1}^n K(X_i/h)\mathbf{D}_i\mathbf{D}_i'\mathbf{H}(h)\boldsymbol{\vartheta}\right).
\end{aligned}$$

As $\widehat{\boldsymbol{\Gamma}}(h)$ is asymptotically invertible, $\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}$ converges in probability to 0 if and only if the term in brackets converges to 0 in probability. The expectation of such term is equal to

$$\begin{aligned}
&\frac{1}{h}\mathbb{E}\left[K(X_i/h)\mathbf{D}_iY_i - K(X_i/h)\mathbf{D}_i\mathbf{D}_i'\mathbf{H}(h)\boldsymbol{\vartheta}\right] \\
&\quad = \frac{1}{h}\mathbb{E}\left[K(X_i/h)\mathbf{D}_i\widetilde{\mathbf{D}}_i'\widetilde{\mathbf{H}}(h)\widetilde{\boldsymbol{\vartheta}} - K(X_i/h)\mathbf{D}_i\mathbf{D}_i'\mathbf{H}(h)\boldsymbol{\vartheta}\right], \qquad (5)
\end{aligned}$$

where $\widetilde{\mathbf{H}}(h) = \text{diag}(1, h, 1, h, 1)$. Note that if $\psi = 0$, then $\widetilde{\mathbf{D}}_i'\widetilde{\mathbf{H}}(h)\widetilde{\boldsymbol{\vartheta}} = \mathbf{D}_i'\mathbf{H}(h)\boldsymbol{\vartheta}$ and so the expectation would be 0 implying $\mathbf{H}(h)(\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}) = o_{\mathbb{P}}(1)$. Let's further simplify the problem by letting $\mu_W(x) \equiv \mu_W, \mu_{W^2}(x) \equiv \mu_{W^2}$, and $\mu_{W^3}(x) \equiv \mu_{W^3}$ for all $x \in \mathbb{R}$. In general, we have that (5), after the change of variables $u = x/h$, is equal to

$$\int_0^\infty K(u)f(uh)\begin{bmatrix} 1 & u & \mu_W & u\mu_W & \mu_{W^2} \\ u & u^2 & u\mu_W & u^2\mu_W & u\mu_{W^2} \\ \mu_W & u\mu_W & \mu_{W^2} & u\mu_{W^2} & \mu_{W^3} \\ u\mu_W & u^2\mu_W & u\mu_{W^2} & u^2\mu_{W^2} & u\mu_{W^3} \end{bmatrix} \mathrm{d}u\widetilde{\mathbf{H}}\widetilde{\boldsymbol{\vartheta}}$$
$$- \int_0^\infty K(u)f(uh)\begin{bmatrix} 1 & u & \mu_W & u\mu_W \\ u & u^2 & u\mu_W & u^2\mu_W \\ \mu_W & u\mu_W & \mu_{W^2} & u\mu_{W^2} \\ u\mu_W & u^2\mu_W & u\mu_{W^2} & u^2\mu_{W^2} \end{bmatrix} \mathrm{d}u\mathbf{H}\boldsymbol{\vartheta}. \tag{6}$$

Without loss of generality, suppose that we rely on a uniform kernel, i.e., $K(u) = \mathbb{1}(0 \le u \le 1)$. Then, by solving the integrals in (6), we obtain that their difference is equal to

$$\begin{bmatrix} \mu_{W^2}\psi & \frac{1}{2}\mu_{W^2}\psi & \mu_{W^3}\psi & \frac{1}{2}\mu_{W^3}\psi \end{bmatrix}',$$

which shows that $\widehat{\boldsymbol{\vartheta}}$ is in general inconsistent. Again, if $\psi = 0$, then consistency would be restored. Notably, in a regression discontinuity, one would have a separate local polynomial regression on each side of the cutoff and would be interested in the difference in the estimated coefficients. More formally, let the models for the potential outcomes be

$$\begin{aligned} Y_i(0) &= \beta_{-,0} + X_i\beta_{-,1} + W_i\gamma_{-,0} + W_iX_i\gamma_{-,1} + W_i^2\psi_- + \epsilon_{-,i}, \qquad \mathbb{E}[\epsilon_{-,i}] = \mathbb{E}[W_i\epsilon_{-,i}] = \mathbb{E}[X_i\epsilon_{-,i}] = 0, \\ Y_i(1) &= \beta_{+,0} + X_i\beta_{+,1} + W_i\gamma_{+,0} + W_iX_i\gamma_{+,1} + W_i^2\psi_+ + \epsilon_{+,i}, \qquad \mathbb{E}[\epsilon_{+,i}] = \mathbb{E}[W_i\epsilon_{+,i}] = \mathbb{E}[X_i\epsilon_{+,i}] = 0, \end{aligned}$$

with $\mathbb{E}[\epsilon_{-,i}\epsilon_{+,i}] = 0$ and let the observed outcome be $Y_i = T_iY_i(1) + (1-T_i)Y_i(0)$ with $T_i = \mathbb{1}(X_i \ge 0)$. Then, suppose we run the regression of Y_i onto $(\mathbf{r}_1(X_i)', W_i\mathbf{r}_1(X_i)') = (1, X_i, W_i, X_iW_i)$ with weights $K(X_i/h)$ separately to the left and to the right of the cutoff. Denote the regression coefficients with $\widehat{\boldsymbol{\vartheta}}_-$ and $\widehat{\boldsymbol{\vartheta}}_+$, respectively, and define $\boldsymbol{\vartheta}_- := (\beta_{-,0}, \beta_{-,1}, \gamma_{-,0}, \gamma_{-,1})'$ and $\boldsymbol{\vartheta}_+ := (\beta_{+,0}, \beta_{+,1}, \gamma_{+,0}, \gamma_{+,1})'$. Then, using an argument similar to the one above, we get

$$\mathbf{H}(h)\left(\widehat{\boldsymbol{\vartheta}}_+ - \widehat{\boldsymbol{\vartheta}}_- - (\boldsymbol{\vartheta}_+ - \boldsymbol{\vartheta}_-)\right) = \begin{bmatrix} \mu_{W^2}\psi_+ - \mu_{W^2}\psi_- \\ \frac{1}{2}(\mu_{W^2}\psi_+ - \mu_{W^2}\psi_-) \\ \mu_{W^3}\psi_+ - \mu_{W^3}\psi_- \\ \frac{1}{2}(\mu_{W^3}\psi_+ - \mu_{W^3}\psi_-) \end{bmatrix}.$$

To restore the consistency of the RD estimator, it is necessary that $\psi_+ = \psi_-$. Therefore, taking the difference of the local polynomial regression coefficients provides a slightly larger set of conditions for restoring consistency compared to the case where the two regressions are considered separately (where we would need $\psi_+ = 0 = \psi_-$). Additionally, we emphasize that another crucial simplifying assumption is that the first three conditional moments of W are constant functions of the running variable. This condition can be justified by assuming $W \perp\!\!\!\perp X$, which is equivalent to assuming that treatment assignment is independent of W.

We conclude this section with an important remark on local polynomial regressions. Since we localize solely in the X dimension and not in the W dimension, omitting higher-order terms in W has severe consequences for the consistency of the RD estimator. This issue does not occur when higher-order terms in X are omitted, due to the localization induced by the kernel in the X dimension. More specifically, if Y depends on powers of X up to the power of $p+1$ and we include only $\mathbf{r}_p(X)$ in our local polynomial regression, then we incur a bias of order $O_{\mathbb{P}}(h^{1+p})$, but the estimator remains consistent. However, if Y depends on powers of W up to the power of $s+1$, then including only $(W, \cdots, W^s)$ in the regression makes the estimator inconsistent, as follows from standard omitted variable theory. Moreover, if Y depends on powers of WX up to order $q+1$, including only $\mathbf{r}_q(X)W$ introduces a bias of order $O_{\mathbb{P}}(h^{1+q})$, yet consistency is preserved due to localization in X.

Table SA-1: *Consequence of the omission of different regressors.*

True DGP depends on	Included regressors	Consequence
$(X, \cdots, X^{p+1})$	$(X, \cdots, X^{p})$	$O_{\mathbb{P}}(h^{1+p})$ bias
$(X, \cdots, X^{s+1})W$	$(X, \cdots, X^{s})W$	$O_{\mathbb{P}}(h^{1+s})$ bias
$(W, \cdots, W^{q+1})$	$(W, \cdots, W^{q})$	Inconsistency

SA3 Empirical Practice Investigating Covariate-Heterogeneity

Table SA-2: *Empirical Practice for RD-HTE*

Paper	*Outcome*	*Treatment*	*Running Variable*	*Heterogeneity*			*Estimation*					*Inference*	
				Disc.	*Time*	*Cont.*	*BW*	*HTE BW*	*Local*	*Joint*	*Covs/FE*	*Robust*	*Cluster SE*
Adams et al. (2022)	Healthcare utilization	Financial Assistance	Poverty level		✓		None						
Akhtari et al. (2022)	Municipal Bureaucracy	Political turnover	Close Elections	✓			MSE	✓	✓		✓		✓
Asher and Novosad (2020)	Economic development	Road construction	Population size	✓			MSE		✓		✓		✓
Brollo et al. (2013)	Political corruption, quality	Government revenues	Population size	✓		✓	None			✓	✓		✓
Dell (2015)	Drug-related violence	Drug enforcement	Close Elections	✓	✓		Manual		✓	✓	✓		✓
García-Miralles and Leganza (2024)	Savings, Retirement behavior	Public pension benefits	Age	✓	✓		MSE		✓		✓	✓	
Han et al. (2020)	Healthcare Utilization	Patient Cost-Sharing	Age	✓			Manual		✓				✓
Huh and Reif (2021)	Mortality, risky behaviors	Teenage Driving	Age	✓			MSE		✓			✓	
Jones et al. (2022)	Irrigation Adoption, Profits	Access to water	Spatial Disc.	✓			Manual		✓		✓		✓
Lindo et al. (2010)	Academic Performance	Academic probation	Test Scores	✓			Manual		✓		✓		✓
McEwan et al. (2021)	Economics Major Choice	Higher letter grade	Test Scores	✓			MSE		✓	✓	✓	✓	
Miglino et al. (2023)	Health outcomes	Financial Assistance	Age	✓	✓		Manual		✓		✓		✓
Pop-Eleches and Urquiola (2013)	Academic Performance	Access to better schools	Test Scores	✓	✓		Manual		✓		✓		✓
Shigeoka (2014)	Utilization, health	Patient Cost Sharing	Age	✓	✓		Manual		✓		✓		✓
Silliman and Virtanen (2022)	Labor market returns	Vocational Education	Test Scores	✓	✓		Manual		✓		✓		✓
Zimmerman (2019)	Better jobs, Income	Access to better schools	Test Scores	✓			Manual		✓	✓	✓		✓

SA4 Proofs

SA4.1 Proof of Lemma SA-1

Proof. Here, for simplicity, we just prove the lemma for the scalar case. By Markov inequality we have

$$\forall\, M > 0, \quad \mathbb{P}(|A_n - \mathbb{E}[A_n]| \geq M) \leq \frac{\mathbb{E}[|A_n - \mathbb{E}[A_n]|]}{M} \leq \frac{\mathbb{V}[A_n]^{1/2}}{M},$$

which implies that $\mathbb{V}[A_n]^{-1/2}(A_n - \mathbb{E}[A_n]) = O_{\mathbb{P}}(1)$ which was to be shown. The matrix case follows using $\mathbb{P}(\mathbf{A}_n \preceq \mathbf{I}) \leq \operatorname{tr}(\mathbb{E}[\mathbf{A}_n])$, where $\mathbf{A}_n \preceq \mathbf{I}$ means that $\mathbf{A}_n - \mathbf{I}$ is negative semi-definite. ■

SA4.2 Proof of Lemma SA-2

Proof. The proof covers the case to the right of the cutoff. Everything follows symmetrically for the other case. First, for any $p, s \in \mathbb{N}$

$$\begin{aligned}
\mathbb{E}[\widehat{\boldsymbol{\Lambda}}_{+,p,s}(h)] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(X_i \geq c)\mathbf{r}_p\left(\frac{X_i - c}{h}\right)\mathbf{r}_s\left(\frac{X_i - c}{h}\right)' K\left(\frac{X_i - c}{h}\right)\frac{1}{h}\right] && \text{(definition)}\\
&= \frac{1}{h}\,\mathbb{E}\left[\mathbb{1}(X_i \geq c)\mathbf{r}_p\left(\frac{X_i - c}{h}\right)\mathbf{r}_s\left(\frac{X_i - c}{h}\right)' K\left(\frac{X_i - c}{h}\right)\right] && \text{(Assumption SA1)}\\
&= \frac{1}{h}\int_c^{\infty} \mathbf{r}_p\left(\frac{x - c}{h}\right)\mathbf{r}_s\left(\frac{x - c}{h}\right)' K\left(\frac{x - c}{h}\right) f(x)\,\mathrm{d}x\\
&= \int_0^{\infty} \mathbf{r}_p(u)\mathbf{r}_s(u)' K(u) f(hu + c)\,\mathrm{d}u =: \widetilde{\boldsymbol{\Gamma}}_{+,p,s}(h). && (u = (x - c)/h)
\end{aligned}$$

By Lemma SA-1

$$\widehat{\boldsymbol{\Lambda}}_{+,p,s}(h) = \mathbb{E}[\widehat{\boldsymbol{\Lambda}}_{+,p,s}(h)] + O_{\mathbb{P}}(|\,\mathbb{V}[\widehat{\boldsymbol{\Lambda}}_{+,p,s}(h)]|).$$

The generic element of $\widehat{\boldsymbol{\Lambda}}_{+,p,s}(h)$ is

$$\frac{1}{nh}\sum_{i=1}^{n} \mathbb{1}(X_i \geq c)\left(\frac{X_i - c}{h}\right)^j K\left(\frac{X_i - c}{h}\right), \quad j \in \{0, \cdots, p + s\}.$$

Fix $j \in \{0, \cdots, p + s\}$, then

$$\begin{aligned}
&\mathbb{V}\left[\frac{1}{nh}\sum_{i=1}^{n} \mathbb{1}(X_i \geq c)\left(\frac{X_i - c}{h}\right)^j K\left(\frac{X_i - c}{h}\right)\right]\\
&\qquad = \frac{1}{nh^2}\,\mathbb{V}\left[\mathbb{1}(X_i \geq c)\left(\frac{X_i - c}{h}\right)^j K\left(\frac{X_i - c}{h}\right)\right] && \text{(Assumption SA1)}\\
&\qquad \leq \frac{1}{nh^2}\,\mathbb{E}\left[\mathbb{1}(X_i \geq c)\left(\frac{X_i - c}{h}\right)^{2j} K\left(\frac{X_i - c}{h}\right)^2\right]\\
&\qquad = \frac{1}{nh}\int_0^{\infty} u^{2j} K(u)^2 f(uh + c)\,\mathrm{d}u = O(n^{-1}h^{-1}), && (u = (x - c)/h)
\end{aligned}$$

where the last equality follows because $u^{2j}K(u)^2 f(uh + c)$ is integrable due to Assumptions SA2, SA5, and

$\mathsf{k} \geq h$. Hence $|\mathbb{V}[\widehat{\mathbf{\Lambda}}_{+,p,s}(h)]| = O(n^{-1}h^{-1})$. Finally, we have $\widehat{\mathbf{\Lambda}}_{+,p,s}(h) = \mathbb{E}[\widehat{\mathbf{\Lambda}}_{+,p,s}(h)] + O_{\mathbb{P}}(1/\sqrt{nh})$, which was to be shown. ■

SA4.3 Proof of Lemma SA-3

Proof. The proof covers the case to the right of the cutoff. Throughout the proof, we maintain that n is large enough so that $\mathsf{k} \leq c + h$ in Assumption SA3. Everything follows symmetrically for the other case. First, note that

$$\widehat{\mathbf{\Gamma}}_{+,p,s}(h) = \begin{bmatrix} \widehat{\mathbf{\Lambda}}_{+,p,p}(h) & \widehat{\mathbf{G}}_{+,2}(h) \\ \widehat{\mathbf{G}}_{+,2}(h)' & \widehat{\mathbf{G}}_{+,3}(h) \end{bmatrix},$$

where

$$\widehat{\mathbf{\Lambda}}_{+,p,p}(h) = \frac{1}{nh}\sum_{i=1}^{n} \mathbb{1}(X_i \geq c)\mathbf{r}_p\left(\frac{X_i - c}{h}\right)\mathbf{r}_p\left(\frac{X_i - c}{h}\right)' K\left(\frac{X_i - c}{h}\right),$$

$$\widehat{\mathbf{G}}_{+,2}(h) = \frac{1}{nh}\sum_{i=1}^{n} \mathbb{1}(X_i \geq c)\left[\mathbf{W}_i' \otimes K\left(\frac{X_i - c}{h}\right)\mathbf{r}_p\left(\frac{X_i - c}{h}\right)\mathbf{r}_s\left(\frac{X_i - c}{h}\right)'\right],$$

$$\widehat{\mathbf{G}}_{+,3}(h) = \frac{1}{nh}\sum_{i=1}^{n} \mathbb{1}(X_i \geq c)\left[\mathbf{W}_i\mathbf{W}_i' \otimes K\left(\frac{X_i - c}{h}\right)\mathbf{r}_s\left(\frac{X_i - c}{h}\right)\mathbf{r}_s\left(\frac{X_i - c}{h}\right)'\right].$$

Then, from Lemma SA-2 we immediately have

$$\widehat{\mathbf{\Lambda}}_{+,p,p}(h) = \widetilde{\mathbf{\Lambda}}_{+,p,p}(h) + O_{\mathbb{P}}(1/\sqrt{nh}).$$

Consider the expectation of $\widehat{\mathbf{G}}_{+,2}(h)$. Using Assumption SA1 and the change of variable $u = (x - c)/h$:

$$\begin{aligned}\mathbb{E}[\widehat{\mathbf{G}}_{+,2}(h)] &= \mathbb{E}\left[\frac{1}{h}\mathbb{1}(X_i \geq c)\left\{\boldsymbol{\mu}_W(X_i)' \otimes K\left(\frac{X_i - c}{h}\right)\mathbf{r}_p\left(\frac{X_i - c}{h}\right)\mathbf{r}_s\left(\frac{X_i - c}{h}\right)'\right\}\right] \\ &= \int_0^\infty \boldsymbol{\mu}_W(uh + c)' \otimes \mathbf{r}_p(u)\,\mathbf{r}_s(u)'\,K(u) f(uh + c)\,\mathrm{d}u =: \widetilde{\mathbf{G}}_{+,2}(h),\end{aligned}$$

The generic element of $\widehat{\mathbf{G}}_{+,2}(h)$ is

$$\frac{1}{nh}\sum_{i=1}^{n} \mathbb{1}(X_i \geq c)W_{i\ell} \cdot \left(\frac{X_i - c}{h}\right)^j K\left(\frac{X_i - c}{h}\right), \qquad \ell \in \{1, \cdots, d\},\ j \in \{0, \cdots, p + s\}.$$

Fix $\ell \in \{1, \cdots, d\}$ and $j \in \{0, \cdots, p + s\}$. Then

$$\begin{aligned}&\mathbb{V}\left[\frac{1}{nh}\sum_{i=1}^{n} \mathbb{1}(X_i \geq c)W_{i\ell} \cdot \left(\frac{X_i - c}{h}\right)^j K\left(\frac{X_i - c}{h}\right)\right] \\ &\quad = \frac{1}{nh^2}\,\mathbb{V}\left[\mathbb{1}(X_i \geq c)W_{i\ell} \cdot \left(\frac{X_i - c}{h}\right)^j K\left(\frac{X_i - c}{h}\right)\right] && \text{(Assumption SA1)} \\ &\quad \leq \frac{1}{nh^2}\,\mathbb{E}\left[\mathbb{1}(X_i \geq c)W_{i\ell}^2 \cdot \left(\frac{X_i - c}{h}\right)^{2j} K\left(\frac{X_i - c}{h}\right)^2\right] \\ &\quad = \frac{1}{nh}\int_0^\infty \mu_{W_\ell^2}(uh + c)u^{2j}K(u)^2 f(uh + c)\,\mathrm{d}u = O(n^{-1}h^{-1}), && (u = (x - c)/h)\end{aligned}$$

where the last equality follows because of Assumptions SA2, SA5, and the fact that $\mu_{W_\ell^2}(\cdot)$ is bounded in a neighborhood of c, which is granted by Assumption SA3a and $\mathsf{k} \geq h$. Hence, by Lemma SA-1

$$\widehat{\mathbf{G}}_{+,2}(h) = \widetilde{\mathbf{G}}_{+,2}(h) + O_{\mathbb{P}}(1/\sqrt{nh}).$$

Consider $\widehat{\mathbf{G}}_{+,3}(h)$. Using a similar logic as above

$$\begin{aligned}
\mathbb{E}[\widehat{\mathbf{G}}_{+,3}(h)] &= \frac{1}{h}\,\mathbb{E}\left[\boldsymbol{\mu}_{WW}(X_i) \otimes \mathbb{1}(X_i \geq c) K\left(\frac{X_i - c}{h}\right) \mathbf{r}_s\left(\frac{X_i - c}{h}\right) \mathbf{r}_s\left(\frac{X_i - c}{h}\right)'\right] \\
&= \int_0^\infty \boldsymbol{\mu}_{WW}(uh + c) \otimes K\left(u\right) \mathbf{r}_s\left(u\right) \mathbf{r}_s\left(u\right)' f(uh + c)\,\mathrm{d}u =: \widetilde{\mathbf{G}}_{+,3}(h)
\end{aligned}$$

and for $\ell, \ell' \in \{1, \cdots, d\}$ and $j \in \{0, \cdots, 2s\}$. Then

$$\begin{aligned}
&\mathbb{V}\left[\frac{1}{nh}\sum_{i=1}^n \mathbb{1}(X_i \geq c) W_{i\ell} W_{i\ell'} \cdot \left(\frac{X_i - c}{h}\right)^j K\left(\frac{X_i - c}{h}\right)\right] \\
&\qquad = \frac{1}{nh^2}\,\mathbb{V}\left[\mathbb{1}(X_i \geq c) W_{i\ell} W_{i\ell'} \cdot \left(\frac{X_i - c}{h}\right)^j K\left(\frac{X_i - c}{h}\right)\right] && \text{(Assumption SA1)} \\
&\qquad \leq \frac{1}{nh^2}\,\mathbb{E}\left[\mathbb{1}(X_i \geq c) W_{i\ell}^2 W_{i\ell'}^2 \cdot \left(\frac{X_i - c}{h}\right)^{2j} K\left(\frac{X_i - c}{h}\right)^2\right] \\
&\qquad = \frac{1}{nh}\int_0^\infty \mathbb{E}[W_{i\ell}^2 W_{i\ell'}^2 \mid X_i = uh + c] u^{2j} K(u)^2 f(uh + c)\,\mathrm{d}u = O(n^{-1}h^{-1}), && (u = (x - c)/h)
\end{aligned}$$

where the last equality follows because of Assumptions SA2, SA5, and the fact that $\mathbb{E}[W_{i\ell}^2 W_{i\ell'}^2 \mid X_i = uh + c]$ is bounded in $[0, c + h]$, which is granted by Assumption SA3e and $\mathsf{k} \geq h$.

Last, we prove that $\widehat{\boldsymbol{\Gamma}}_{+,p,s}(h)$ is asymptotically invertible. To do so, note that by taking the limit as $h \to 0$, the continuity of $f(\cdot)$ (Assumption SA2), $\boldsymbol{\mu}_W(\cdot)$, and $\boldsymbol{\mu}_{WW}(\cdot)$ (Assumption SA3a) give us that

$$\begin{aligned}
\widetilde{\boldsymbol{\Lambda}}_{+,p,p}(h) &= f(c)\int_0^\infty K(u)\mathbf{r}_p(u)\mathbf{r}_p(u)'\,\mathrm{d}u + o(1) = \boldsymbol{\Lambda}_{+,p,p} + o(1), \\
\widetilde{\mathbf{G}}_{+,2}(h) &= \boldsymbol{\mu}_W' \otimes f(c)\int_0^\infty K(u)\mathbf{r}_p(u)\mathbf{r}_s(u)'\,\mathrm{d}u + o(1) = \boldsymbol{\mu}_W' \otimes \boldsymbol{\Lambda}_{+,p,s} + o(1) = \mathbf{G}_{+,3} + o(1), \\
\widetilde{\mathbf{G}}_{+,3}(h) &= \boldsymbol{\mu}_{WW} \otimes f(c)\int_0^\infty K(u)\mathbf{r}_s(u)\mathbf{r}_s(u)'\,\mathrm{d}u + o(1) = \boldsymbol{\mu}_{WW} \otimes \boldsymbol{\Lambda}_{+,s,s} + o(1) = \mathbf{G}_{+,3} + o(1).
\end{aligned}$$

Therefore,

$$\widehat{\boldsymbol{\Gamma}}_{+,p,s}(h) = \boldsymbol{\Gamma}_{+,p,s} + o(1) + O_{\mathbb{P}}(1/\sqrt{nh}), \quad \text{where} \quad \boldsymbol{\Gamma}_{+,p,s} = \begin{bmatrix} \boldsymbol{\Lambda}_{+,p,p} & \mathbf{G}_{+,2} \\ \mathbf{G}_{+,2}' & \mathbf{G}_{+,3} \end{bmatrix},$$

which is non-singular by Assumption SA3g. ■

SA4.4 Proof of Lemma SA-4

Proof. First,

$$\begin{aligned}
\mathbb{E}[\widehat{\boldsymbol{\zeta}}_{+,p,s,a}(h)] &= \mathbb{E}\left[\frac{1}{nh}\sum_{i=1}^{n}\mathbf{r}_{p,s}\left(\frac{X_i-c}{h},\mathbf{W}_i\right)K\left(\frac{X_i-c}{h}\right)\left(\frac{X_i-c}{h}\right)^{a+1}\right] \\
&= \int_0^\infty K(u)\mathbf{r}_{p,s}(u,\boldsymbol{\mu}_W(uh+c))u^{a+1}f(c+hu)\,\mathrm{d}u \\
&= f(c)\int_0^\infty K(u)\mathbf{r}_{p,s}(u,\boldsymbol{\mu}_W)u^{a+1}\,\mathrm{d}u + O(h).
\end{aligned}$$

Moreover, the generic element of $\widehat{\boldsymbol{\zeta}}_{+,p,s,a}(h)$ is either

$$\frac{1}{nh}\sum_{i=1}^{n}\mathbb{1}(X_i\geq c)K\left(\frac{X_i-c}{h}\right)\left(\frac{X_i-c}{h}\right)^{j+a+1}, \qquad j\in\{0,1,\cdots,p\},$$

or

$$\frac{1}{nh}\sum_{i=1}^{n}\mathbb{1}(X_i\geq c)K\left(\frac{X_i-c}{h}\right)W_{i\ell}\cdot\left(\frac{X_i-c}{h}\right)^{j+a+1}, \qquad j\in\{0,1,\cdots,s\},\ \ \ell\in\{1,\cdots,d\}.$$

The variances of these elements are of order $O(n^{-1}h^{-1})$. Indeed, the variance of the latter is

$$\begin{aligned}
&\mathbb{V}\left[\frac{1}{nh}\sum_{i=1}^{n}\mathbb{1}(X_i\geq c)K\left(\frac{X_i-c}{h}\right)W_{i\ell}\cdot\left(\frac{X_i-c}{h}\right)^{j+a+1}\right] \\
&\quad\leq \frac{1}{nh^2}\,\mathbb{E}\left[\mathbb{1}(X_i\geq c)K^2\left(\frac{X_i-c}{h}\right)W_{i\ell}^2\cdot\left(\frac{X_i-c}{h}\right)^{2(j+a+1)}\right] \\
&\quad= \frac{1}{nh}\int_0^\infty \mu_{W_\ell}(uh+c)u^{2(j+a+1)}K^2(u)f(uh+c) = O\left(\frac{1}{nh}\right),
\end{aligned}$$

where for the last equality we used Assumptions SA1 and SA5 and $\mathsf{k}\geq h$. Then, $\widehat{\boldsymbol{\zeta}}_{+,p,s,a}(h)=\mathbb{E}[\widehat{\boldsymbol{\zeta}}_{+,p,s,a}(h)]+O_{\mathbb{P}}(1/\sqrt{nh})$ by Lemma SA-1. A similar argument goes through for $\widehat{\boldsymbol{\varphi}}_{+,p,s,a}(h)$.

The last part of the lemma follows by taking the limit as $h\to 0$ and using the continuity of $f(\cdot)$ (Assumption SA2), $\boldsymbol{\mu}_W(\cdot)$, and $\boldsymbol{\mu}_{WW}(\cdot)$ (Assumption SA3a). ■

SA4.5 Proof of Lemma SA-5

Proof. Fix $h>0$ and define

$$\mathbf{w}_{+,i}(h) := \mathbb{1}(X_i\geq c)K\left(\frac{X_i-c}{h}\right)\mathbf{r}_{p,s}\left(\frac{X_i-c}{h},\mathbf{W}_i\right),$$

so that we can write

$$\mathbf{L}_{+,p,s}(h) = \frac{1}{nh}\sum_{i=1}^{n}\mathbf{w}_{+,i}(h)u_{+,i}(h).$$

By the properties of ℓ_2-residuals, we get that $\mathbb{E}[\mathbf{w}_{+,i}(h)u_{+,i}(h)] = 0$ and so $\mathbb{E}[\mathbf{L}_{+,p,s}(h)] = 0$. The variance of $\mathbf{L}_{+,p,s}(h)$ is

$$\begin{aligned}
\mathbb{V}\left[\frac{1}{nh}\sum_{i=1}^{n}\mathbf{w}_{+,i}(h)u_{+,i}(h)\right] &= \frac{1}{n^2h^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{C}\mathrm{ov}(\mathbf{w}_{+,i}(h)u_{+,i}(h), \mathbf{w}_{+,j}(h)u_{+,j}(h)) \\
&= \frac{1}{nh^2}\,\mathbb{V}[\mathbf{w}_{+,i}(h)u_{+,i}(h)] && \text{(Assumption SA1)} \\
&= \frac{1}{nh}\underbrace{\frac{1}{h}\,\mathbb{E}[\mathbf{w}_{+,i}(h)\mathbf{w}_{+,i}(h)'u_{+,i}^2(h)]}_{:=\mathbf{V}_{+,p,s}(h)}. && (\mathbb{E}[\mathbf{w}_{+,i}(h)u_{+,i}(h)] = 0)
\end{aligned}$$

By the law of iterated expectations and changing variables, we get

$$\mathbf{V}_{+,p,s}(h) =$$
$$\int_0^\infty f(uh+c)K^2(u)\begin{bmatrix} \mathbf{r}_p(u)\,\mathbf{r}_p(u)'\,\mathbb{E}[u_{+,i}^2(h)\mid X_i = uh+c] & \mathbb{E}[\mathbf{W}_i'u_{+,i}^2(h)\mid X_i = uh+c]\otimes\mathbf{r}_p(u)\,\mathbf{r}_s(u)' \\ \mathbb{E}[\mathbf{W}_iu_{+,i}^2(h)\mid X_i = uh+c]\otimes\mathbf{r}_s(u)\,\mathbf{r}_p(u)' & \mathbb{E}[\mathbf{W}_i\mathbf{W}_i'u_{+,i}^2(h)\mid X_i = uh+c]\otimes\mathbf{r}_s(u)\,\mathbf{r}_s(u)' \end{bmatrix}\mathrm{d}u,$$

Finally, the last part of the lemma follows by taking the limit as $h \to 0$, using the properties of compact kernels (Assumption SA5), the continuity of $f(\cdot)$ (Assumption SA2), and the bounded conditional fourth moments of $(Y_i, \mathbf{W}_i')'$ (Assumptions SA3d-SA3e), together with the Cauchy-Schwarz inequality and $\mathsf{k} \geq h$. ■

SA4.6 Proof of Theorem SA-1

Proof. Fix an arbitrary $\mathbf{c} \in \mathbb{R}^{1+p+d}\setminus\{\mathbf{0}_{1+p+d}\}$. Consider

$$\phi_{+,p,s}(h) := \sqrt{nh}\mathbf{c}'\mathbf{H}_{p,s}(h)(\widehat{\boldsymbol{\vartheta}}_{+,p,s}(h) - \boldsymbol{\vartheta}^\star_{+,p,s}(h)) = \sqrt{nh}\mathbf{c}'\widehat{\boldsymbol{\Gamma}}^{-1}_{+,p,s}(h)\mathbf{L}_{+,p,s}(h).$$

By Lemma SA-3 and Lemma SA-5, we get

$$\phi_{+,p,s}(h) = \dot{\phi}_{+,p,s}(h) + o_{\mathbb{P}}(1), \qquad \dot{\phi}_{+,p,s}(h) = \sqrt{nh}\mathbf{c}'\widetilde{\boldsymbol{\Gamma}}^{-1}_{+,p,s}(h)\mathbf{L}_{+,p,s}(h).$$

Then, by Lemma SA-5

$$\mathbb{V}[\sqrt{nh}\mathbf{c}'\widetilde{\boldsymbol{\Gamma}}^{-1}_{+,p,s}(h)\mathbf{L}_{+,p,s}(h)] = \mathbf{c}'\boldsymbol{\Omega}_{+,p,s}\mathbf{c} + o(1), \qquad \boldsymbol{\Omega}_{+,p,s} := \boldsymbol{\Gamma}^{-1}_{+,p,s}\mathbf{V}_{+,p,s}\boldsymbol{\Gamma}^{-1\prime}_{+,p,s}.$$

Positive definiteness of $\boldsymbol{\Omega}_{+,p,s}$ follows from the fact that it is a product of symmetric positive definite matrices (Lemma SA-3 and Lemma SA-5). Finally, by appropriately scaling $\dot{\phi}_{+,p,s}(h)$ by its variance, we can rewrite it as

$$\sum_{i=1}^{n} w_{+,i}(h)u_{+,i}(h),$$

with

$$w_{+,i}(h) := (\mathbf{c}'\boldsymbol{\Omega}_{+,p,s}\mathbf{c})^{-1/2}\,\mathbf{c}'\widetilde{\boldsymbol{\Gamma}}^{-1}_{+,p,s}(h)\mathbb{1}(X_i \geq c)K\left(\frac{X_i - c}{h}\right)\mathbf{r}_{p,s}\left(\frac{X_i - c}{h}, \mathbf{W}_i\right)/\sqrt{nh}.$$

Therefore, $\{w_{+,i}(h)u_{+,i}(h)\}_{i=1}^n$ is a triangular array of mean-zero row-wise independent summands with variance converging to 1 to which we can apply a Lindeberg-Feller CLT. To do so, we verify the Lindeberg condition by showing that the stronger Lyapunov condition holds for the fourth moment. To see this, by Assumption SA3

$$\sum_{i=1}^{n}\mathbb{E}[|w_{+,i}(h)u_i(h)|^4] \lesssim \frac{1}{nh^2}\cdot\int_c^\infty\left|\breve{\mathbf{e}}_\nu'\widetilde{\boldsymbol{\Gamma}}^{-1}_{+,p,s}(h)K\left(\frac{x-c}{h}\right)\mathbf{r}_{p,s}\left(\frac{x-c}{h}, \boldsymbol{\mu}_W(x)\right)\right|^4 f(x)\,\mathrm{d}x = O\left(\frac{1}{nh}\right),$$

where the final order comes from the change of variables $u = \frac{x-c}{h}$ and the fact that the resulting integral is bounded by Assumptions SA2, SA3, and SA5. Therefore, we can conclude that

$$\sqrt{nh}\mathbf{c}'\mathbf{H}_{p,s}(h)\widehat{\boldsymbol{\Gamma}}_{+,p,s}^{-1}(h)\mathbf{L}_{+,p,s}(h) \rightsquigarrow \mathsf{N}(0, \mathbf{c}'\boldsymbol{\Omega}_{+,p,s}\mathbf{c}).$$

As $\mathbf{c}$ was chosen arbitrarily, by the Cramér-Wold device we get

$$\sqrt{nh}\mathbf{H}_{p,s}(h)\widehat{\boldsymbol{\Gamma}}_{+,p,s}^{-1}(h)\mathbf{L}_{+,p,s}(h) \rightsquigarrow \mathsf{N}(\mathbf{0}_{1+p+d}, \boldsymbol{\Omega}_{+,p,s}),$$

which was to be shown. ■

SA4.7 Proof of Theorem SA-3

Proof. First, note that under the assumptions of Theorem SA-1, we get that

$$\max_{i\in[n]} L_{-,i} = o_{\mathbb{P}}(1) \qquad \text{and} \qquad \max_{i\in[n]} L_{+,i} = o_{\mathbb{P}}(1). \tag{7}$$

The result above follows from Assumptions SA1, SA3, SA5, and Lemma SA-3.

Define $\omega_{+,i}(h) := \mathbb{1}(X_i \geq 0)K\left(\frac{X_i-c}{h}\right)\sqrt{w_{+,i}(h)}$. First, note that

$$\widehat{u}_{+,i}(h) = u_{+,i}(h) + \widehat{v}_{+,i}(h),$$

where $u_{+,i}(h) := Y_i - \mathbf{r}_{p,s}(X_i - c, \mathbf{W}_i)'\boldsymbol{\vartheta}^\star_{+,p,s}(h)$ and $\widehat{v}_{+,i}(h) := \mathbf{r}_{p,s}(X_i - c, \mathbf{W}_i)'(\boldsymbol{\vartheta}^\star_{+,p,s}(h) - \widehat{\boldsymbol{\vartheta}}_{+,p,s}(h))$. Then, focusing on the "meat" part of the estimator, we get

$$\begin{aligned}
\widehat{\mathbf{V}}_{+,p,s}(h) &= \frac{1}{nh}\sum_{i=1}^n \omega_{+,i}(h)\mathbf{r}_{p,s}\left(\frac{X_i-c}{h}, \mathbf{W}_i\right)\mathbf{r}_{p,s}\left(\frac{X_i-c}{h}, \mathbf{W}_i\right)'\widehat{u}^2_{+,i}(h) \\
&= \frac{1}{nh}\sum_{i=1}^n \omega_{+,i}(h)\mathbf{r}_{p,s}\left(\frac{X_i-c}{h}, \mathbf{W}_i\right)\mathbf{r}_{p,s}\left(\frac{X_i-c}{h}, \mathbf{W}_i\right)'u^2_{+,i}(h) && (:= \mathbf{V}_{+,1}(h)) \\
&\quad + \frac{1}{nh}\sum_{i=1}^n \omega_{+,i}(h)\mathbf{r}_{p,s}\left(\frac{X_i-c}{h}, \mathbf{W}_i\right)\mathbf{r}_{p,s}\left(\frac{X_i-c}{h}, \mathbf{W}_i\right)'\widehat{v}^2_{+,i}(h), && (:= \mathbf{V}_{+,2}(h)) \\
&\quad + \frac{2}{nh}\sum_{i=1}^n \omega_{+,i}(h)\mathbf{r}_{p,s}\left(\frac{X_i-c}{h}, \mathbf{W}_i\right)\mathbf{r}_{p,s}\left(\frac{X_i-c}{h}, \mathbf{W}_i\right)'\widehat{v}_{+,i}(h)u_{+,i}(h). && (:= \mathbf{V}_{+,3}(h))
\end{aligned}$$

For convenience, let $\mathbf{V}^{\texttt{HC0}}_{+,1}(h)$ denote $\mathbf{V}_{+,1}(h)$ when $w_{+,i}(h) = 1$. Note that $\mathbf{V}^{\texttt{HC0}}_{+,1}(h) \xrightarrow{\mathbb{P}} \mathbf{V}_{+,p,s}$ by the weak law of large numbers. Then, it is immediate to see that $\mathbf{V}^{\texttt{HC1}}_{+,1}(h) = \frac{n-k}{n}\mathbf{V}^{\texttt{HC0}}_{+,1}(h)$ and so $\mathbf{V}^{\texttt{HC1}}_{+,1}(h) \xrightarrow{\mathbb{P}} \mathbf{V}_{+,p,s}$. Regarding the HC2-type estimator, it suffices to note that

$$|\mathbf{V}^{\texttt{HC2}}_{+,1}(h) - \mathbf{V}^{\texttt{HC0}}_{+,1}(h)| \leq \left(\frac{1}{nh}\sum_{i=1}^n \left|\mathbf{r}_{p,s}\left(\frac{X_i-c}{h}, \mathbf{W}_i\right)\right|^2 u^2_{+,i}(h)\right) \cdot \left|\frac{1}{1-\max_{i\in[n]} L_i} - 1\right|,$$

where the first term is $O_{\mathbb{P}}(1)$ by Assumptions SA1-SA3 and the second term is $o_{\mathbb{P}}(1)$ due to (7). A similar argument works for the HC3-type estimator.

At this point, if we show that $\mathbf{V}_{+,j}(h) \xrightarrow{\mathbb{P}} 0, j = 2, 3$, then we have established consistency of our estimator for the variance of the RD estimator. Let's start with $\mathbf{V}_{+,2}(h)$:

$$|\mathbf{V}_{+,2}(h)| \leq \frac{1}{nh}\sum_{i=1}^n |\omega_{+,i}(h)| \left|\mathbf{r}_{p,s}\left(\frac{X_i-c}{h}, \mathbf{W}_i\right)\right|^2 \left|\mathbf{r}_{p,s}\left(\frac{X_i-c}{h}, \mathbf{W}_i\right)\right|^2 \left|\mathbf{H}_{p,s}(h)\left(\widehat{\boldsymbol{\vartheta}}_{+,p,s}(h) - \boldsymbol{\vartheta}^\star_{+,p,s}(h)\right)\right|^2$$

$$\leq \left|\mathbf{H}_{p,s}(h)\left(\widehat{\boldsymbol{\vartheta}}_{+,p,s}(h)-\boldsymbol{\vartheta}^{\star}_{+,p,s}(h)\right)\right|^2 \cdot \max_{i\in[n]}|\omega_{+,i}(h)| \cdot \frac{1}{nh}\sum_{i=1}^{n}\left|\mathbf{r}_{p,s}\left(\frac{X_i-c}{h},\mathbf{W}_i\right)\right|^4.$$

The first term on the right-hand side is $O_{\mathbb{P}}(n^{-1}h^{-1})$ by Theorem SA-1, the second term is $O_{\mathbb{P}}(1)$ by (7), whereas the third term is $O_{\mathbb{P}}(1)$ as it converges in probability to its expectation which is

$$\int_0^{\infty} f(uh+c)K(u)\left|\mathbf{r}_{p,s}(u,\boldsymbol{\mu}_W(uh+c))\right|^4 \,\mathrm{d}u,$$

which is finite because of Assumptions SA2, SA3, and SA5. Therefore, $\mathbf{V}_{+,2}(h)=o_{\mathbb{P}}(1)$.

Then

$$|\mathbf{V}_{+,3}(h)| \leq \left|\mathbf{H}_{p,s}(h)\left(\widehat{\boldsymbol{\vartheta}}_{+,p,s}(h)-\boldsymbol{\vartheta}^{\star}_{+,p,s}(h)\right)\right| \cdot \max_{i\in[n]}|\omega_{+,i}(h)| \cdot \frac{1}{nh}\sum_{i=1}^{n}\left|\mathbf{r}_{p,s}\left(\frac{X_i-c}{h},\mathbf{W}_i\right)\right|^3 |u_{+,i}(h)|.$$

The first term on the right-hand side is $O_{\mathbb{P}}(1/\sqrt{nh})$ by Theorem SA-1. The second term is $O_{\mathbb{P}}(1)$ by(7). To see that the third term is $O_{\mathbb{P}}(1)$, note that by the Hölder's inequality

$$\mathbb{E}\left[\left|\mathbf{r}_{p,s}\left(\frac{X_i-c}{h},\mathbf{W}_i\right)\right|^3 |u_{+,i}(h)|\right] \leq \left(\int_0^{\infty} f(uh+c)K(u)\left|\mathbf{r}_{p,s}\left(u,\boldsymbol{\mu}_W(uh+c)\right)\right|^3 \,\mathrm{d}u\right)^{3/4} \mathbb{E}[u^4_{+,i}(h)]^{1/4},$$

which is finite because of Assumptions SA2, SA3, and SA5. Therefore, $\mathbf{V}_{+,3}(h)=O_{\mathbb{P}}(1/\sqrt{nh})=o_{\mathbb{P}}(1)$.

In conclusion, we showed that

$$\widehat{\mathbf{V}}_{+,p,s}(h)=\sum_{\ell=1}^{3}\mathbf{V}_{+,\ell}(h),$$

where $\mathbf{V}_{+,1}(h)$ converges in probability to $\mathbf{V}_{+,p,s}(h)$ and the other terms are all $o_{\mathbb{P}}(1)$. Moreover, Lemma SA-3 gives us $\widehat{\boldsymbol{\Gamma}}^{-1}_{+,p,s}(h)\xrightarrow{\mathbb{P}}\boldsymbol{\Gamma}^{-1}_{+,p,s}$. Hence, by Slutsky's theorem

$$\widehat{\mathcal{V}}_{+,\nu,p,s}(h)\xrightarrow{\mathbb{P}}\mathcal{V}_{+,\nu,p,s}. \qquad \blacksquare$$

SA4.8 Proof of Lemma SA-6

Proof. Let $q:=p\vee s$. First of all, if we take a $(q+2)$-th order Taylor expansion of $\mu_+(\cdot,\mathbf{w})$ around c, we get

$$\begin{aligned}
\boldsymbol{\vartheta}^{\star}_{+,p,s}(h) &= \mathbf{H}^{-1}_{p,s}(h)\widetilde{\boldsymbol{\Gamma}}^{-1}_{+,p,s}(h)\,\mathbb{E}\left[\frac{1}{h}\mathbb{1}(X_i\geq c)K\left(\frac{X_i-c}{h}\right)\mathbf{r}_{p,s}\left(\frac{X_i-c}{h},\mathbf{W}_i\right)\mu(X_i,\mathbf{W}_i)\right] \\
&= \boldsymbol{\vartheta}_{+,p,s}+\mathbf{H}^{-1}_{p,s}(h)\widetilde{\boldsymbol{\Gamma}}_{+,p,s}(h)^{-1}\left(h^{p+1}\widetilde{\boldsymbol{\zeta}}_{+,p,s,p}(h)\frac{\alpha^{(p+1)}_+(c)}{(p+1)!}+h^{p+2}\widetilde{\boldsymbol{\zeta}}_{+,p,s,p+1}(h)\frac{\alpha^{(p+2)}_+(c)}{(p+2)!}\right) \\
&\quad + \mathbf{H}^{-1}_{p,s}(h)\widetilde{\boldsymbol{\Gamma}}_{+,p,s}(h)^{-1}\left(h^{s+1}\widetilde{\boldsymbol{\varphi}}_{+,p,s,s}(h)\frac{\boldsymbol{\lambda}^{(s+1)}_+(c)}{(s+1)!}+h^{s+2}\widetilde{\boldsymbol{\varphi}}_{+,p,s,s+1}(h)\frac{\boldsymbol{\lambda}^{(s+2)}_+(c)}{(s+2)!}\right) \\
&\quad + o(h^{2+p\wedge s}),
\end{aligned}$$

where

$$\widetilde{\boldsymbol{\zeta}}_{+,p,s,a}(h)=\int_0^{\infty}K(u)\mathbf{r}_{p,s}(u,\boldsymbol{\mu}_W(uh+c))u^{a+1}f(uh+c)\,\mathrm{d}u,$$

$$\widetilde{\boldsymbol{\varphi}}_{+,p,s,a}(h) = \int_0^\infty K(u) \begin{bmatrix} \mathbf{r}_p(u)\boldsymbol{\mu}_W(uh+c)' \\ \boldsymbol{\mu}_{WW}(uh+c) \otimes \mathbf{r}_s(u) \end{bmatrix} u^{a+1} f(uh+c)\,\mathrm{d}u.$$

The $o(h^{2+p\wedge s})$ term comes from the integral form of the remainder of the Taylor expansion. Suppose $p = s$. The remainder is of the form

$$\mathbf{v}_+(h) = \int_0^\infty \int_{\mathbb{R}^d} K(u)\,\mathbf{r}_{p,s}(u,\mathbf{w}) \int_c^{uh+c} \left(\frac{\mu_+^{(p+3)}(t,\mathbf{w})}{(p+2)!}(uh+c-t)^{p+2} \right) \mathrm{d}t f_{X,W}(uh+c,\mathbf{w})\,\mathrm{d}\mathbf{w}\,\mathrm{d}u.$$

Being $\mu_+^{(p+3)}(x,\mathbf{w})$ continuous by Assumption SA3b, $\mathsf{k} \geq h$, and using Assumptions SA2 and SA5, we get $\mathbf{v}_+(h) = o(h^{2+p})$.

Whenever $s \neq p$, the remainder $\mathbf{v}_+(h)$ includes additional terms we ignored in the linearization above. These terms will always include either $\alpha^{(\ell)}(c), \ell \geq p+3$ or $\lambda^{(\ell)}(c), \ell \geq s+3$ and be of order $o(h^{2+p})$ and $o(h^{2+s})$, respectively.

The last part of the lemma follows by taking the limit as $nh \to \infty$ and $h \to 0$ and relying on Lemma SA-3 and SA-4. ■

Supplement References

Adams, A., Kluender, R., Mahoney, N., Wang, J., Wong, F., and Yin, W. (2022), "The Impact of Financial Assistance Programs on Health Care Utilization: Evidence from Kaiser Permanente," American Economic Review: Insights, 4, 389–407.

Akhtari, M., Moreira, D., and Trucco, L. (2022), "Political Turnover, Bureaucratic Turnover, and the Quality of Public Services," American Economic Review, 112, 442–93.

Asher, S., and Novosad, P. (2020), "Rural Roads and Local Economic Development," American Economic Review, 110, 797–823.

Brollo, F., Nannicini, T., Perotti, R., and Tabellini, G. (2013), "The Political Resource Curse," American Economic Review, 103, 1759–96.

Cameron, A. C., and Miller, D. L. (2015), "A practitioner's guide to cluster-robust inference," Journal of human resources, 50, 317–372.

Dell, M. (2015), "Trafficking Networks and the Mexican Drug War," American Economic Review, 105, 1738–79.

García-Miralles, E., and Leganza, J. M. (2024), "Public Pensions and Private Savings," American Economic Journal: Economic Policy, 16, 366–405.

Han, H.-W., Lien, H.-M., and Yang, T.-T. (2020), "Patient Cost-Sharing and Healthcare Utilization in Early Childhood: Evidence from a Regression Discontinuity Design," American Economic Journal: Economic Policy, 12, 238–78.

Huh, J., and Reif, J. (2021), "Teenage Driving, Mortality, and Risky Behaviors," American Economic Review: Insights, 3, 523–39.

Jones, M., Kondylis, F., Loeser, J., and Magruder, J. (2022), "Factor Market Failures and the Adoption of Irrigation in Rwanda," American Economic Review, 112, 2316–52.

Lindo, J. M., Sanders, N. J., and Oreopoulos, P. (2010), "Ability, Gender, and Performance Standards: Evidence from Academic Probation," American Economic Journal: Applied Economics, 2, 95–117.

MacKinnon, J. G., Nielsen, M. Ø., and Webb, M. D. (2023), "Cluster-Robust Inference: A Guide to Empirical Practice," Journal of Econometrics, 232, 272–299.

McEwan, P. J., Rogers, S., and Weerapana, A. (2021), "Grade Sensitivity and the Economics Major at a Women's College," AEA Papers and Proceedings, 111, 102–06.

Miglino, E., Navarrete H., N., Navarrete H., G., and Navarrete H., P. (2023), "Health Effects of Increasing Income for the Elderly: Evidence from a Chilean Pension Program," American Economic Journal: Economic Policy, 15, 370–93.

Pop-Eleches, C., and Urquiola, M. (2013), "Going to a Better School: Effects and Behavioral Responses," American Economic Review, 103, 1289–1324.

Shigeoka, H. (2014), "The Effect of Patient Cost Sharing on Utilization, Health, and Risk Protection," American Economic Review, 104, 2152–84.

Silliman, M., and Virtanen, H. (2022), "Labor Market Returns to Vocational Secondary Education," American Economic Journal: Applied Economics, 14, 197–224.

Zimmerman, S. D. (2019), "Elite Colleges and Upward Mobility to Top Jobs and Top Incomes," American Economic Review, 109, 1–47.