

Online Supplement for “Treatment Effect Heterogeneity in Regression Discontinuity Designs”

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

July 3, 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	3
SA1.1 Notation	3
SA1.2 Setup	4
SA1.3 Assumptions	7
SA1.4 Mapping between Main Text and Supplement	8
SA1.5 Auxiliary Lemmas and Quantities	8
SA2 Main Results	11
SA2.1 Asymptotic Approximation and Asymptotic Variance	12
SA2.2 Variance Estimation	13
SA2.3 Smoothing Bias and Probability Limit	14
SA2.4 Consistency	16
SA2.5 Causal Interpretation of the Probability Limit	17
SA2.6 MSE Expansions	18
SA3 Empirical Practice Investigating Covariate-Heterogeneity	20
SA4 Proofs	22
SA4.1 Proof of Lemma SA-1	22
SA4.2 Proof of Lemma SA-2	22
SA4.3 Proof of Lemma SA-3	23
SA4.4 Proof of Lemma SA-4	25
SA4.5 Proof of Lemma SA-5	25
SA4.6 Proof of Theorem SA-1	26
SA4.7 Proof of Theorem SA-3	27
SA4.8 Proof of Lemma SA-6	28

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}_ν 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 $\mathbf{1}_k$ 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, $\text{tr}(\cdot)$ is the trace operator, and $\text{diag}(\mathbf{x})$ yields a square diagonal matrix with the elements of \mathbf{x} on its main diagonal. With a slight abuse of notation, we denote with $\mathbf{v}^k = (v_1^k, \dots, v_n^k)'$ 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, $|\mathbf{A}|^2 = \sum_i \sum_j |a_{ij}|^2 = \text{tr}(\mathbf{A}'\mathbf{A})$. 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 $\exists M \in \mathbb{R}_{++} : a_n \leq Mb_n$ for all large n , $a_n = o(b_n)$ if $\lim_{n \rightarrow \infty} a_n b_n^{-1} = 0$, and $a_n \lesssim b_n$ if there exists a constant C such that $a_n \leq Cb_n$ 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 $\forall \epsilon > 0, \lim_{n \rightarrow \infty} \mathbb{P}[|A_n B_n^{-1}| \geq \epsilon] = 0$ and $A_n = O_{\mathbb{P}}(B_n)$ if $\forall \epsilon > 0, \exists M, n_0 \in \mathbb{R}_{++} : \mathbb{P}[|A_n B_n^{-1}| > M] < \epsilon$, 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 $\mathbf{N}(\mathbf{a}, \mathbf{B})$, 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

$$Y_i = T_i \cdot Y_i(1) + (1 - T_i) \cdot Y_i(0),$$

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

In sharp RD designs, $T_i = \mathbb{1}(X_i \geq c)$, where $X_i \in \mathbb{R}$ 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 X_i .

We further define

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

and $\sigma_{\mathbf{W}}^2(x) = \mu_{\mathbf{W}\mathbf{W}'}(x) - \mu_{\mathbf{W}}(x)\mu_{\mathbf{W}}'(x)$.

Let $\mathbf{r}_q(x) = (1, x, \dots, x^q)'$ 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

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

Finally, we also define the following matrices

$$\mathbf{\Lambda}_{-,p,s} := f(c) \int_{-\infty}^0 K(u) \mathbf{r}_p(u) \mathbf{r}_s(u)' du, \quad \mathbf{\Lambda}_{+,p,s} := f(c) \int_0^{\infty} K(u) \mathbf{r}_p(u) \mathbf{r}_s(u)' du,$$

and

$$\mathbf{\Xi}_{-,p,s} = f(c) \int_{-\infty}^0 K^2(u) \mathbf{r}_p(u) \mathbf{r}_s(u)' du, \quad \mathbf{\Xi}_{+,p,s} = f(c) \int_0^{\infty} K^2(u) \mathbf{r}_p(u) \mathbf{r}_s(u)' du.$$

SA1.2 Setup

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

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

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

where

$$\hat{\boldsymbol{\alpha}}_{-,p}(h) = (\hat{\alpha}_{-,0}, \dots, \hat{\alpha}_{-,p})', \quad \hat{\boldsymbol{\alpha}}_{+,p}(h) = (\hat{\alpha}_{+,0}, \dots, \hat{\alpha}_{+,p})',$$

and

$$\begin{aligned} \hat{\boldsymbol{\lambda}}_{-,s}(h) &= (\hat{\lambda}_{-,10}(h), \dots, \hat{\lambda}_{-,1s}(h), \dots, \hat{\lambda}_{-,d0}(h), \dots, \hat{\lambda}_{-,ds}(h))', \\ \hat{\boldsymbol{\lambda}}_{+,s}(h) &= (\hat{\lambda}_{+,10}(h), \dots, \hat{\lambda}_{+,1s}(h), \dots, \hat{\lambda}_{+,d0}(h), \dots, \hat{\lambda}_{+,ds}(h))'. \end{aligned}$$

Note that the vector $\hat{\boldsymbol{\lambda}}_{+,s}(h)$ (similarly, $\hat{\boldsymbol{\lambda}}_{-,s}(h)$) stacks first all the coefficients of the polynomial basis of order s for the first covariate, then for the second, and so on until the d -th covariate.

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:

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

which can be equivalently written in matrix form as

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

where the Hessian and score matrices are defined as

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

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

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

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

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:

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

where for fixed $\mathbf{w} \in \mathbb{R}^d$

$$\check{\mathbf{e}}_\nu \equiv \check{\mathbf{e}}_\nu(\mathbf{w}) := \begin{bmatrix} \nu \mathbf{I}_{p+1} \mathbf{e}_\nu \\ \nu! \mathbf{w} \otimes \mathbf{I}_{s+1} \mathbf{e}_\nu \end{bmatrix}$$

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}_ν takes in different places. Most of the time, we suppress the dependence of $\check{\mathbf{e}}_\nu$ on \mathbf{w} to avoid making the notation even more cumbersome and because in the case of orthogonal covariates, the only sensible evaluation point would be $\mathbf{w} = \boldsymbol{\iota}_d$.¹ The evaluation point \mathbf{w} would only turn out to be useful when

¹This case corresponds to situations where each component is binary, $\mathbf{W}_i \in \{0, 1\}^d$, and at most one component takes the

some elements in \mathbf{W}_i are continuous variables, as the following example clarifies.

Example 1. Suppose $d = 1$, $W_i \in \{0, 1\}$ and $\nu = 0$, then

$$\check{\mathbf{e}}'_0 \widehat{\boldsymbol{\vartheta}}_{+,p,s}(h) = \mathbf{e}'_0 \widehat{\boldsymbol{\alpha}}_{+,p}(h) + \mathbf{e}'_0 \widehat{\boldsymbol{\lambda}}_{+,s}(h) = \underbrace{\widehat{\alpha}_{+,0}(h)}_{\text{coefficient on } \mathbb{1}(X_i \geq c)} + \underbrace{\widehat{\lambda}_{+,10}(h)}_{\text{coefficient on } W_i \cdot \mathbb{1}(X_i \geq c)},$$

and

$$\check{\mathbf{e}}'_0 \widehat{\boldsymbol{\vartheta}}_{-,p,s}(h) = \mathbf{e}'_0 \widehat{\boldsymbol{\alpha}}_{-,p}(h) + \mathbf{e}'_0 \widehat{\boldsymbol{\lambda}}_{-,s}(h) = \underbrace{\widehat{\alpha}_{-,0}(h)}_{\text{coefficient on } \mathbb{1}(X_i < c)} + \underbrace{\widehat{\lambda}_{-,10}(h)}_{\text{coefficient on } W_i \cdot \mathbb{1}(X_i < c)}.$$

If instead $\nu = 1$, then

$$\check{\mathbf{e}}'_1 \widehat{\boldsymbol{\vartheta}}_{+,p,s}(h) = \mathbf{e}'_1 \widehat{\boldsymbol{\alpha}}_{+,p}(h) + \mathbf{e}'_1 \widehat{\boldsymbol{\lambda}}_{+,s}(h) = \underbrace{\widehat{\alpha}_{+,1}(h)}_{\text{coefficient on } \mathbb{1}(X_i \geq c) \cdot (X_i - c)} + \underbrace{\widehat{\lambda}_{+,11}(h)}_{\text{coefficient on } W_i \cdot \mathbb{1}(X_i \geq c) \cdot (X_i - c)},$$

and

$$\check{\mathbf{e}}'_1 \widehat{\boldsymbol{\vartheta}}_{-,p,s}(h) = \mathbf{e}'_1 \widehat{\boldsymbol{\alpha}}_{-,p}(h) + \mathbf{e}'_1 \widehat{\boldsymbol{\lambda}}_{-,s}(h) = \underbrace{\widehat{\alpha}_{-,1}(h)}_{\text{coefficient on } \mathbb{1}(X_i < c) \cdot (X_i - c)} + \underbrace{\widehat{\lambda}_{-,11}(h)}_{\text{coefficient on } W_i \cdot \mathbb{1}(X_i < c) \cdot (X_i - c)}.$$

Now, suppose that W_i is a continuous covariate and $\nu = 0$. In this case, the evaluation point $w = w_0$ does matter, and so

$$\check{\mathbf{e}}'_0 \widehat{\boldsymbol{\vartheta}}_{+,p,s}(h) = \mathbf{e}'_0 \widehat{\boldsymbol{\alpha}}_{+,p}(h) + w_0 \mathbf{e}'_0 \widehat{\boldsymbol{\lambda}}_{+,s}(h) = \underbrace{\widehat{\alpha}_{+,0}(h)}_{\text{coefficient on } \mathbb{1}(X_i \geq c)} + \underbrace{\widehat{\lambda}_{+,10}(h)}_{\text{coefficient on } W_i \cdot \mathbb{1}(X_i \geq c)} \cdot w_0,$$

and

$$\check{\mathbf{e}}'_0 \widehat{\boldsymbol{\vartheta}}_{-,p,s}(h) = \mathbf{e}'_0 \widehat{\boldsymbol{\alpha}}_{-,p}(h) + \mathbf{e}'_0 \widehat{\boldsymbol{\lambda}}_{-,s}(h) = \underbrace{\widehat{\alpha}_{-,0}(h)}_{\text{coefficient on } \mathbb{1}(X_i < c)} + \underbrace{\widehat{\lambda}_{-,10}(h)}_{\text{coefficient on } W_i \cdot \mathbb{1}(X_i < c)} \cdot w_0.$$

◆

Moreover, let

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

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.,

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

value one, implying that $\mathbf{W}_i' \mathbf{W}_i \leq 1$.

and

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

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). $\{(Y_i, X_i, \mathbf{W}_i')\}_{i=1}^n$ is a random sample.

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 $k \in \mathbb{R}_{++}$, with $-k < c < k$ and for all $x \in [-k, k]$, $\mathbf{w} \in \mathbb{R}^d$:

- (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) $\mathbb{V}[Y_i(t) \mid X_i = x]$ and $\mathbb{V}[Y_i(t) \mid X_i = x, \mathbf{W}_i = \mathbf{w}]$ are continuous and bounded away from zero.
- (d) $\mathbb{E}[|Y_i(t)|^4 \mid X_i = x, \mathbf{W}_i = \mathbf{w}]$ is continuous in both arguments.
- (e) $\mathbb{E}[|\mathbf{W}_i|^4 \mid X_i = x]$ is continuous.
- (f) $\mathbb{E}[\mathbf{W}_i \mathbb{V}[Y_i(t) \mid X_i, \mathbf{W}_i] \mid X_i = x]$ and $\mathbb{E}[\mathbf{W}_i \mathbf{W}_i' \mathbb{V}[Y_i(t) \mid X_i, \mathbf{W}_i] \mid X_i = x]$ are continuous.
- (g) $\boldsymbol{\Gamma}_{-,p,s}$ and $\boldsymbol{\Gamma}_{+,p,s}$ in Equation (3) are positive definite matrices.

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

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

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

Assumption SA5 (Kernel). The kernel function $k(\cdot) : [0, 1] \rightarrow \mathbb{R}$ 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 $p = s = 1$

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

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, we set up the problem as the long regression

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

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

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

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 results 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 $\check{\mathbf{e}}_\nu = \mathbf{s}$;
- 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

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

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

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

	Main	Supplement
<i>Estimands</i>		
	$\alpha(\cdot)$	$\alpha_-(\cdot)$
	$\theta(\cdot)$	$\alpha_+(\cdot) - \alpha_-(\cdot)$
	$\lambda(\cdot)$	$\lambda_-(\cdot)$
	$\xi(\cdot)$	$\lambda_+(\cdot) - \lambda_-(\cdot)$
	ς	$(\alpha_+(0) - \alpha_-(0), \lambda_+(0)' - \lambda_-(0)')'$
<i>Estimators</i>		
	$\hat{\alpha}$	$\mathbf{e}'_0 \hat{\alpha}_-$
	$\hat{\theta}$	$\mathbf{e}'_0 (\hat{\alpha}_+ - \hat{\alpha}_-)$
	$\hat{\lambda}$	$(\boldsymbol{\nu}_d \otimes \mathbf{I}_{s+1} \mathbf{e}_0)' \hat{\lambda}_-$
	$\hat{\xi}$	$(\boldsymbol{\nu}_d \otimes \mathbf{I}_{s+1} \mathbf{e}_0)' (\hat{\lambda}_+ - \hat{\lambda}_-)$
	$\hat{\varsigma}$	$\check{\mathbf{e}}'_0 (\hat{\boldsymbol{\vartheta}}_{+,1,1} - \hat{\boldsymbol{\vartheta}}_{-,1,1})$
	$\hat{\omega}_1$	$\mathbf{e}'_1 \hat{\alpha}_-$
	$\hat{\omega}_2$	$\mathbf{e}'_1 (\hat{\alpha}_+ - \hat{\alpha}_-)$
	$\hat{\omega}_3$	$(\boldsymbol{\nu}_d \otimes \mathbf{I}_{s+1} \mathbf{e}_1)' \hat{\lambda}_-$
	$\hat{\omega}_4$	$(\boldsymbol{\nu}_d \otimes \mathbf{I}_{s+1} \mathbf{e}_1)' (\hat{\lambda}_+ - \hat{\lambda}_-)$
<i>Constants</i>		
	\mathbf{B}_s	$s' \mathbf{M}(\mathbf{B}_{+,1,1}^{[0]} + \mathbf{B}_{+,1,1}^{[1]} - \mathbf{B}_{-,1,1}^{[0]} + \mathbf{B}_{-,1,1}^{[1]})$
	\mathbf{V}_s	$s' \mathbf{M}(\boldsymbol{\Gamma}_{+,1,1}^{-1} \mathbf{V}_{+,1,1} \boldsymbol{\Gamma}_{+,1,1}^{-1'} + \boldsymbol{\Gamma}_{+,1,1}^{-1} \mathbf{V}_{-,1,1} \boldsymbol{\Gamma}_{-,1,1}^{-1'}) \mathbf{M}' s$

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

$$\hat{\Lambda}_{-,p,s}(h) = \tilde{\Lambda}_{-,p,s}(h) + O_{\mathbb{P}}(1/\sqrt{nh}), \quad \hat{\Lambda}_{+,p,s}(h) = \tilde{\Lambda}_{+,p,s}(h) + O_{\mathbb{P}}(1/\sqrt{nh}),$$

with

$$\begin{aligned} \tilde{\Lambda}_{-,p,s}(h) &:= \mathbb{E}[\hat{\Lambda}_{-,p,s}(h)] = \int_{-\infty}^0 K(u) \mathbf{r}_p(u) \mathbf{r}_s(u)' f(uh + c) du, \\ \tilde{\Lambda}_{+,p,s}(h) &:= \mathbb{E}[\hat{\Lambda}_{+,p,s}(h)] = \int_0^{\infty} K(u) \mathbf{r}_p(u) \mathbf{r}_s(u)' f(uh + c) du. \end{aligned}$$

[Proof]

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

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

$$\hat{\Gamma}_{-,p,s}(h) = \tilde{\Gamma}_{-,p,s}(h) + O_{\mathbb{P}}(1/\sqrt{nh}), \quad \hat{\Gamma}_{+,p,s}(h) = \tilde{\Gamma}_{+,p,s}(h) + O_{\mathbb{P}}(1/\sqrt{nh}),$$

where

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

where

$$\begin{aligned}\tilde{\mathbf{G}}_{-,2}(h) &= \int_{-\infty}^0 K(u) [\boldsymbol{\mu}_W(uh+c)' \otimes \mathbf{r}_p(u) \mathbf{r}_s(u)'] f(uh+c) du, \\ \tilde{\mathbf{G}}_{+,2}(h) &= \int_0^{\infty} K(u) [\boldsymbol{\mu}_W(uh+c)' \otimes \mathbf{r}_p(u) \mathbf{r}_s(u)'] f(uh+c) du, \\ \tilde{\mathbf{G}}_{-,3}(h) &= \int_{-\infty}^0 K(u) [\boldsymbol{\mu}_{WW}(uh+c) \otimes \mathbf{r}_s(u) \mathbf{r}_s(u)'] f(uh+c) du, \\ \tilde{\mathbf{G}}_{+,3}(h) &= \int_0^{\infty} K(u) [\boldsymbol{\mu}_{WW}(uh+c) \otimes \mathbf{r}_s(u) \mathbf{r}_s(u)'] f(uh+c) du.\end{aligned}$$

Last,

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

where

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

with

$$\begin{aligned}\boldsymbol{\Lambda}_{-,p,s} &= f(c) \int_{-\infty}^0 K(u) \mathbf{r}_p(u) \mathbf{r}_s(u)' du, & \boldsymbol{\Lambda}_{+,p,s} &= f(c) \int_0^{\infty} K(u) \mathbf{r}_p(u) \mathbf{r}_s(u)' du \\ \mathbf{G}_{-,2} &= \boldsymbol{\mu}'_W \otimes \boldsymbol{\Lambda}_{-,p,s}, & \mathbf{G}_{+,2} &= \boldsymbol{\mu}'_W \otimes \boldsymbol{\Lambda}_{+,p,s}, \\ \mathbf{G}_{-,3} &= \boldsymbol{\mu}_{WW} \otimes \boldsymbol{\Lambda}_{-,s,s}, & \mathbf{G}_{+,3} &= \boldsymbol{\mu}_{WW} \otimes \boldsymbol{\Lambda}_{+,s,s}.\end{aligned}$$

[Proof]

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

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

The following two lemmas characterize the asymptotic properties of $\hat{\boldsymbol{\zeta}}_{-,p,s,a}(h)$, $\hat{\boldsymbol{\zeta}}_{+,p,s,a}(h)$, $\hat{\boldsymbol{\varphi}}_{-,p,s,a}(h)$, and $\hat{\boldsymbol{\varphi}}_{+,p,s,a}(h)$. 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 $k \geq h$. If $nh \rightarrow \infty$ and $h \rightarrow 0$, then*

$$\hat{\boldsymbol{\zeta}}_{-,p,s,a}(h) = \tilde{\boldsymbol{\zeta}}_{-,p,s,a}(h) + O_{\mathbb{P}}(1/\sqrt{nh}), \quad \hat{\boldsymbol{\zeta}}_{+,p,s,a}(h) = \tilde{\boldsymbol{\zeta}}_{+,p,s,a}(h) + O_{\mathbb{P}}(1/\sqrt{nh}),$$

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

with

$$\begin{aligned} \widetilde{\zeta}_{-,p,s,a}(h) &:= \mathbb{E}[\widehat{\zeta}_{-,p,s,a}(h)] = \int_{-\infty}^0 K(u) \mathbf{r}_{p,s}(u, \boldsymbol{\mu}_W(uh+c)) u^{a+1} f(uh+c) du, \\ \widetilde{\zeta}_{+,p,s,a}(h) &:= \mathbb{E}[\widehat{\zeta}_{+,p,s,a}(h)] = \int_0^{\infty} K(u) \mathbf{r}_{p,s}(u, \boldsymbol{\mu}_W(uh+c)) u^{a+1} f(uh+c) du, \\ \widetilde{\varphi}_{-,p,s,a}(h) &:= E[\widehat{\varphi}_{-,p,s,a}(h)] = \int_{-\infty}^0 K(u) \begin{bmatrix} \mathbf{r}_p(u) \boldsymbol{\mu}_W(uh+c)' \\ \boldsymbol{\mu}_{WW}(uh+c) \otimes \mathbf{r}_s(u) \end{bmatrix} u^{a+1} f(uh+c) du, \\ \widetilde{\varphi}_{+,p,s,a}(h) &:= \mathbb{E}[\widehat{\varphi}_{+,p,s,a}(h)] = \int_0^{\infty} K(u) \begin{bmatrix} \mathbf{r}_p(u) \boldsymbol{\mu}_W(uh+c)' \\ \boldsymbol{\mu}_{WW}(uh+c) \otimes \mathbf{r}_s(u) \end{bmatrix} u^{a+1} f(uh+c) du. \end{aligned}$$

Last

$$\begin{aligned} \widetilde{\zeta}_{-,p,s,a}(h) &= \zeta_{-,p,s,a} \{1 + o(1)\}, & \widetilde{\zeta}_{+,p,s,a}(h) &= \zeta_{+,p,s,a} \{1 + o(1)\}, \\ \widetilde{\varphi}_{-,p,s,a}(h) &= \varphi_{-,p,s,a} \{1 + o(1)\}, & \widetilde{\varphi}_{+,p,s,a}(h) &= \varphi_{+,p,s,a} \{1 + o(1)\}, \end{aligned}$$

where

$$\begin{aligned} \zeta_{-,p,s,a} &= f(c) \int_{-\infty}^0 K(u) \mathbf{r}_{p,s}(u, \boldsymbol{\mu}_W) u^{a+1} du, & \zeta_{+,p,s,a} &= f(c) \int_0^{\infty} K(u) \mathbf{r}_{p,s}(u, \boldsymbol{\mu}_W) u^{a+1} du, \\ \varphi_{-,p,s,a} &= f(c) \int_{-\infty}^0 K(u) \begin{bmatrix} \mathbf{r}_p(u) \boldsymbol{\mu}_W' \\ \boldsymbol{\mu}_{WW} \otimes \mathbf{r}_s(u) \end{bmatrix} u^{a+1} du, & \varphi_{+,p,s,a} &= f(c) \int_0^{\infty} K(u) \begin{bmatrix} \mathbf{r}_p(u) \boldsymbol{\mu}_W' \\ \boldsymbol{\mu}_{WW} \otimes \mathbf{r}_s(u) \end{bmatrix} u^{a+1} du. \end{aligned}$$

[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

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

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:

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

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

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

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

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

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 $k \geq h$. Then, for fixed $h > 0$*

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

Furthermore, as $nh \rightarrow \infty$ and $h \rightarrow 0$

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

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

[Proof]

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

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

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 $k \geq h$. If $nh \rightarrow \infty$ and $h \rightarrow 0$, then*

$$\begin{aligned} \sqrt{nh} \mathbf{H}_{p,s}(h) (\widehat{\boldsymbol{\vartheta}}_{-,p,s}(h) - \boldsymbol{\vartheta}_{-,p,s}^*(h)) &\rightsquigarrow \mathbf{N}(\mathbf{0}_{1+p+d}, \boldsymbol{\Omega}_{-,p,s}), \\ \sqrt{nh} \mathbf{H}_{p,s}(h) (\widehat{\boldsymbol{\vartheta}}_{+,p,s}(h) - \boldsymbol{\vartheta}_{+,p,s}^*(h)) &\rightsquigarrow \mathbf{N}(\mathbf{0}_{1+p+d}, \boldsymbol{\Omega}_{+,p,s}), \end{aligned}$$

with

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

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 $\nu \in \{0, \dots, p \wedge s\}$:*

$$\begin{aligned}\sqrt{nh^{2\nu+1}}\check{\mathbf{e}}'_\nu(\widehat{\boldsymbol{\vartheta}}_{-,p,s}(h) - \boldsymbol{\vartheta}_{-,p,s}^*(h)) &\rightsquigarrow \mathbf{N}(0, \mathcal{V}_{-, \nu, p, s}), \\ \sqrt{nh^{2\nu+1}}\check{\mathbf{e}}'_\nu(\widehat{\boldsymbol{\vartheta}}_{+,p,s}(h) - \boldsymbol{\vartheta}_{+,p,s}^*(h)) &\rightsquigarrow \mathbf{N}(0, \mathcal{V}_{+, \nu, p, s}),\end{aligned}$$

where

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

SA2.2 Variance Estimation

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

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

where

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

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

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 are defined as,

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

and

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

which denote the leverage of each observation.

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

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

[Proof]

Define the standard error of the RD estimators as

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

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

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

$$\begin{aligned}\hat{\sigma}_{-, \nu, p, s}^{-1}(h) \check{\mathbf{e}}_{\nu}'(\hat{\boldsymbol{\vartheta}}_{-, p, s}(h) - \boldsymbol{\vartheta}_{-, p, s}^*(h)) &\rightsquigarrow \mathbf{N}(0, 1), \\ \hat{\sigma}_{+, \nu, p, s}^{-1}(h) \check{\mathbf{e}}_{\nu}'(\hat{\boldsymbol{\vartheta}}_{+, p, s}(h) - \boldsymbol{\vartheta}_{+, p, s}^*(h)) &\rightsquigarrow \mathbf{N}(0, 1).\end{aligned}$$

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

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

$$\hat{\mathbf{V}}_{+, \nu, p, s}^{\text{CL}}(h) = \frac{(G-1)n}{(G-1)(n-p-1-d)} \check{\mathbf{e}}_{\nu}' \hat{\boldsymbol{\Gamma}}_{+, p, s}^{-1}(h) \hat{\mathbf{V}}_{+, p, s}^{\text{CL}}(h) \hat{\boldsymbol{\Gamma}}_{+, p, s}^{-1}(h) \check{\mathbf{e}}_{\nu},$$

where

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

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*

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

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

Let

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

where

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

with $\alpha_{-,p} \equiv \alpha_{-,p}(c)$, $\lambda_{-,s} \equiv \lambda_{-,s}(c)$, $\alpha_{+,p} \equiv \alpha_{+,p}(c)$, and $\lambda_{+,s} \equiv \lambda_{+,s}(c)$.

Consider the following decomposition

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

From Theorem SA-1, we know that $\mathbf{H}_{p,s}(h)(\hat{\boldsymbol{\vartheta}}_{+,p,s}(h) - \boldsymbol{\vartheta}_{+,p,s}^*(h)) = O_{\mathbb{P}}(1/\sqrt{nh})$. 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 $\boldsymbol{\vartheta}_{+,p,s}^*(h) - \boldsymbol{\vartheta}_{+,p,s}$ 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 $k \geq h$. If $nh \rightarrow 0$ and $h \rightarrow 0$, then*

$$\begin{aligned} \boldsymbol{\vartheta}_{-,p,s}^*(h) - \boldsymbol{\vartheta}_{-,p,s} &= \\ \mathbf{H}_{p,s}^{-1}(h) &\left(h^{1+p} \tilde{\mathbf{B}}_{-,p}^{[0]}(h) + h^{2+p} \tilde{\mathbf{B}}_{-,p+1}^{[0]}(h) + h^{1+s} \tilde{\mathbf{B}}_{-,s}^{[1]}(h) + h^{2+s} \tilde{\mathbf{B}}_{-,s+1}^{[1]}(h) + o(h^{2+p \wedge s}) \right), \\ \boldsymbol{\vartheta}_{+,p,s}^*(h) - \boldsymbol{\vartheta}_{+,p,s} &= \\ \mathbf{H}_{p,s}^{-1}(h) &\left(h^{1+p} \tilde{\mathbf{B}}_{+,p}^{[0]}(h) + h^{2+p} \tilde{\mathbf{B}}_{+,p+1}^{[0]}(h) + h^{1+s} \tilde{\mathbf{B}}_{+,s}^{[1]}(h) + h^{2+s} \tilde{\mathbf{B}}_{+,s+1}^{[1]}(h) + o(h^{2+p \wedge s}) \right), \end{aligned}$$

where

$$\begin{aligned} \tilde{\mathbf{B}}_{-,a}^{[0]}(h) &= \tilde{\mathbf{\Gamma}}_{-,p,s}^{-1}(h) \tilde{\boldsymbol{\zeta}}_{-,p,s,a}(h) \frac{\alpha_{-}^{(a+1)}(c)}{(a+1)!} \rightarrow \mathbf{\Gamma}_{-,p,s}^{-1} \boldsymbol{\zeta}_{-,p,s,a} \frac{\alpha_{-}^{(a+1)}(c)}{(a+1)!} =: \mathbf{B}_{-,a}^{[0]}, \\ \tilde{\mathbf{B}}_{+,a}^{[0]}(h) &= \tilde{\mathbf{\Gamma}}_{+,p,s}^{-1}(h) \tilde{\boldsymbol{\zeta}}_{+,p,s,a}(h) \frac{\alpha_{+}^{(a+1)}(c)}{(a+1)!} \rightarrow \mathbf{\Gamma}_{+,p,s}^{-1} \boldsymbol{\zeta}_{+,p,s,a} \frac{\alpha_{+}^{(a+1)}(c)}{(a+1)!} =: \mathbf{B}_{+,a}^{[0]}, \\ \tilde{\mathbf{B}}_{-,a}^{[1]}(h) &= \tilde{\mathbf{\Gamma}}_{-,p,s}^{-1}(h) \tilde{\boldsymbol{\varphi}}_{-,p,s,a}(h) \frac{\lambda_{-}^{(a+1)}(c)}{(a+1)!} \rightarrow \mathbf{\Gamma}_{-,p,s}^{-1} \boldsymbol{\varphi}_{-,p,s,a} \frac{\lambda_{-}^{(a+1)}(c)}{(a+1)!} =: \mathbf{B}_{-,a}^{[1]}, \\ \tilde{\mathbf{B}}_{+,a}^{[1]}(h) &= \tilde{\mathbf{\Gamma}}_{+,p,s}^{-1}(h) \tilde{\boldsymbol{\varphi}}_{+,p,s,a}(h) \frac{\lambda_{+}^{(a+1)}(c)}{(a+1)!} \rightarrow \mathbf{\Gamma}_{+,p,s}^{-1} \boldsymbol{\varphi}_{+,p,s,a} \frac{\lambda_{+}^{(a+1)}(c)}{(a+1)!} =: \mathbf{B}_{+,a}^{[1]}. \end{aligned}$$

[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 $nh^{2\nu+1} \rightarrow \infty$, $h \rightarrow 0$, and $nh^{2(p \wedge s)+3} \rightarrow 0$,*

then

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

where

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

with

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

and where

$$\begin{aligned}\mathbf{V}_{-,p,s} &= f(c) \int_{-\infty}^0 K^2(u) \begin{bmatrix} \mathbf{r}_p(u) \mathbf{r}_p(u)' \mathbb{E}[\sigma_{Y-}^2(X_i, \mathbf{W}_i) | X_i = c] & \mathbb{E}[\mathbf{W}_i' \sigma_{Y-}^2(X_i, \mathbf{W}_i) | X_i = c] \otimes \mathbf{r}_p(u) \mathbf{r}_s(u)' \\ \mathbb{E}[\mathbf{W}_i \sigma_{Y-}^2(X_i, \mathbf{W}_i) | X_i = c] \otimes \mathbf{r}_s(u) \mathbf{r}_p(u)' & \mathbb{E}[\mathbf{W}_i \mathbf{W}_i' \sigma_{Y-}^2(X_i, \mathbf{W}_i) | X_i = c] \otimes \mathbf{r}_s(u) \mathbf{r}_s(u)' \end{bmatrix} du, \\ \mathbf{V}_{+,p,s} &= f(c) \int_0^\infty K^2(u) \begin{bmatrix} \mathbf{r}_p(u) \mathbf{r}_p(u)' \mathbb{E}[\sigma_{Y+}^2(X_i, \mathbf{W}_i) | X_i = c] & \mathbb{E}[\mathbf{W}_i' \sigma_{Y+}^2(X_i, \mathbf{W}_i) | X_i = c] \otimes \mathbf{r}_p(u) \mathbf{r}_s(u)' \\ \mathbb{E}[\mathbf{W}_i \sigma_{Y+}^2(X_i, \mathbf{W}_i) | X_i = c] \otimes \mathbf{r}_s(u) \mathbf{r}_p(u)' & \mathbb{E}[\mathbf{W}_i \mathbf{W}_i' \sigma_{Y+}^2(X_i, \mathbf{W}_i) | X_i = c] \otimes \mathbf{r}_s(u) \mathbf{r}_s(u)' \end{bmatrix} du\end{aligned}$$

are positive-definite matrices.

SA2.4 Consistency

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

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

thus if $nh \rightarrow \infty$ and $h \rightarrow 0$, 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 $\check{\mathbf{e}}'_\nu \widehat{\boldsymbol{\vartheta}}_{-,p,s}(h)$ or $\check{\mathbf{e}}'_\nu \widehat{\boldsymbol{\vartheta}}_{+,p,s}(h)$. In this case, we get

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

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

Define

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

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

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

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

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

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

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

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 $\hat{\boldsymbol{\vartheta}}_{+, p, s}(h_+) - \hat{\boldsymbol{\vartheta}}_{-, p, s}(h_-)$ 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

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

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 obtain formulas for MSE-optimal bandwidths. We first start by giving 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

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

Consider the following decomposition for fixed $\mathbf{w} \in \mathbb{R}^d$

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

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

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

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

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

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

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

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

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

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

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

where

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

Furthermore, we define

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

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

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

The variance approximations for the RD estimators are given by

$$\mathcal{V}_{-, \nu, p, s}(h) \rightarrow \mathcal{V}_{-, \nu, p, s} := \check{\mathbf{e}}_{\nu}' \mathbf{\Gamma}_{-, p, s}^{-1} \mathbf{V}_{-, p, s} \mathbf{\Gamma}_{-, p, s}^{-1'} \check{\mathbf{e}}_{\nu}, \quad \mathcal{V}_{+, \nu, p, s}(h) \rightarrow \mathcal{V}_{+, \nu, p, s} := \check{\mathbf{e}}_{\nu}' \mathbf{\Gamma}_{+, p, s}^{-1} \mathbf{V}_{+, p, s} \mathbf{\Gamma}_{+, p, s}^{-1'} \check{\mathbf{e}}_{\nu}.$$

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

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

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

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

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

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

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

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

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

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

SA3 Empirical Practice Investigating Covariate-Heterogeneity

To examine common empirical practices regarding heterogeneous effects, we sampled papers from AEA journals over the past decade. The papers are listed in Table SA-2 below, along with a summarization of how treatment effect heterogeneity was studied in each, including the empirical setting, the type of heterogeneity, and details on how estimation and inference were conducted. For full details of each study, please refer to the original work.

Table SA-2: Empirical Practice for RD-HTE

<i>Paper</i>	<i>Outcome</i>	<i>Treatment</i>	<i>Running Var.</i>	<i>Heterogeneity</i>			<i>Estimation</i>			<i>Std. Errs.</i>	
				<i>Disc.</i>	<i>Time</i>	<i>Cont.</i>	<i>Bandwidth</i>	<i>Local</i>	<i>Joint</i>	<i>Covs/FE</i>	<i>Robust Cluster</i>
Adams et al. (2022)	Healthcare utilization	Financial Assistance	Poverty level	✓	✓		None	✓		✓	✓
Akhtari et al. (2022)	Municipal Bureaucracy	Political turnover	Vote Margin	✓			MSE-HTE	✓			
Asher and Novosad (2020)	Economic development	Road construction	Population	✓			MSE	✓		✓	✓
Brollo et al. (2013)	Political corruption, quality	Government revenues	Population	✓		✓	None	✓	✓		✓
Dell (2015)	Drug-related violence	Drug enforcement	Vote Margin	✓	✓		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	✓			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	✓	✓	✓	✓

Notes:

- (i) The left four columns give the citation and the empirical setting in terms of the outcome variable, treatment studied, and running variable.
- (ii) The type of heterogeneity is studied is listed in columns 5–7, showing, respectively, if the covariates used for heterogeneity are discrete (“Disc.”), time-related (“Time”), or continuous (“Cont.”).
- (iii) For estimation, the “Bandwidth” column reports the method of bandwidth selection used, where “Manual” indicates a hand-selected bandwidth, “MSE” for a mean squared error optimal bandwidth, and “MSE-HTE” if the bandwidth was chosen particularly for the heterogeneity analysis. The “Local” indicates that the analysis is conducted local to the cutoff; “Joint” if the heterogeneity was conducted jointly across levels of the variable, and “Covs/FE” if additional covariates or fixed effects are included for efficiency.
- (iv) Standard errors are described in the final two columns, indicating if robust standard errors are used or if the standard errors are clustered.

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 $\mathbb{P}(\mathbf{A}_n \preceq \mathbf{I}) \leq \text{tr}(\mathbb{E}[\mathbf{A}_n])$, 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 $p, s \in \mathbb{N}$

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

By Lemma SA-1

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

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

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

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

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

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

$k \geq h$. 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. \blacksquare

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 $k \leq c + h$ 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

$$\begin{aligned} \widehat{\mathbf{\Lambda}}_{+,p,p}(h) &= \frac{1}{nh} \sum_{i=1}^n \mathbb{1}(X_i \geq c) \mathbf{r}_p \left(\frac{X_i - c}{h} \right) \mathbf{r}_p \left(\frac{X_i - c}{h} \right)' K \left(\frac{X_i - c}{h} \right), \\ \widehat{\mathbf{G}}_{+,2}(h) &= \frac{1}{nh} \sum_{i=1}^n \mathbb{1}(X_i \geq c) \left[\mathbf{W}_i' \otimes K \left(\frac{X_i - c}{h} \right) \mathbf{r}_p \left(\frac{X_i - c}{h} \right) \mathbf{r}_s \left(\frac{X_i - c}{h} \right)' \right], \\ \widehat{\mathbf{G}}_{+,3}(h) &= \frac{1}{nh} \sum_{i=1}^n \mathbb{1}(X_i \geq c) \left[\mathbf{W}_i \mathbf{W}_i' \otimes K \left(\frac{X_i - c}{h} \right) \mathbf{r}_s \left(\frac{X_i - c}{h} \right) \mathbf{r}_s \left(\frac{X_i - c}{h} \right)' \right]. \end{aligned}$$

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 $u = (x - c)/h$:

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

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

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

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

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

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 $k \geq h$. 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

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

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

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

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 $k \geq h$.

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

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

Therefore,

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

which is non-singular by Assumption [SA3g](#). ■

SA4.4 Proof of Lemma SA-4

Proof. First,

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

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

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

or

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

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

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

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

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

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

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

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

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

Finally, the last part of the lemma follows by taking the limit as $h \rightarrow 0$, 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 $k \geq h$. \blacksquare

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

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

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

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

Then, by Lemma SA-5

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

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

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

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

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

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

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

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

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

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

$$\max_{i \in [n]} L_{-,i} = o_{\mathbb{P}}(1) \quad \text{and} \quad \max_{i \in [n]} L_{+,i} = o_{\mathbb{P}}(1). \quad (5)$$

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

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

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

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

where the first term is $O_{\mathbb{P}}(1)$ by Assumptions SA1-SA3 and the second term is $o_{\mathbb{P}}(1)$ due to (5). 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 the consistency of our estimator for the variance of the RD estimator. Let's start with $\mathbf{V}_{+,2}(h)$:

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

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

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 (5), whereas the third term is $O_{\mathbb{P}}(1)$ as it converges in probability to its expectation which is

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

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

Then

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

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 (5). To see that the third term is $O_{\mathbb{P}}(1)$, note that by the Hölder's inequality

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

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 $\widehat{\boldsymbol{\Gamma}}_{+,p,s}^{-1}(h) \xrightarrow{\mathbb{P}} \boldsymbol{\Gamma}_{+,p,s}^{-1}$. Hence, by Slutsky's theorem

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

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

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

where

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

$$\tilde{\varphi}_{+,p,s,a}(h) = \int_0^\infty K(u) \left[\begin{array}{c} \mathbf{r}_p(u) \boldsymbol{\mu}_W(uh+c)' \\ \boldsymbol{\mu}_{WW}(uh+c) \otimes \mathbf{r}_s(u) \end{array} \right] u^{a+1} f(uh+c) du.$$

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

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

Being $\mu_+^{(p+3)}(x, \mathbf{w})$ continuous by Assumption SA3b, $k \geq h$, 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 $nh \rightarrow \infty$ and $h \rightarrow 0$ 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.