Manuscript submitted to *The Econometrics Journal*, pp. 1–22.

Supplement to "Optimal Bandwidth Choice for Robust Bias Corrected Inference in Regression Discontinuity Designs"[1]

Sebastian Calonico†, Matias D. Cattaneo‡ and Max H. Farrell††

†*Department of Health Policy and Management, Columbia University, USA.*
E-mail: sebastian.calonico@columbia.edu
‡*Department of Operations Research and Financial Engineering, Princeton University, USA.*
E-mail: cattaneo@princeton.edu
††*Booth School of Business, University of Chicago, USA.*
E-mail: max.farrell@chicagobooth.edu

Summary
This supplement contains technical details and formulas omitted from the main text, proofs of all theoretical results, further technical and methodological derivations, and details on practical and numerical implementations.

S.1. SETUP, NOTATION, AND ASSUMPTIONS

We assume the researcher observes a random sample $(Y_i, T_i, X_i)'$, $i = 1, 2, \ldots, n$, where Y_i denotes the outcome variable of interest, T_i denotes treatment status, and X_i denotes an observed continuous score or running random variable, which determines treatment assignment for each unit in the sample. In the canonical sharp RD design, all units with X_i not smaller than a known threshold c are assigned to the treatment group and take-up treatment, while all units with X_i smaller than c are assigned to the control group and do not take-up treatment, so that $T_i = \mathbb{1}(X_i \geq c)$. Using the potential outcomes framework, $Y_i = Y_i(0) \cdot (1 - T_i) + Y_i(1) \cdot T_i$, with $Y_i(1)$ and $Y_i(0)$ denoting the potential outcomes with and without treatment, respectively, for each unit.

The parameter of interest in sharp RD designs are either the average treatment effect at the cutoff or its derivatives. Thus, herein we study the generic population parameter, for some integer $\nu \geq 0$:

$$\tau_\nu = \tau_\nu(c) = \frac{\partial^\nu}{\partial x^\nu} \mathbb{E}[Y_i(1) - Y_i(0) | X_i = x] \bigg|_{x=c}, \tag{S.1.1}$$

Here and elsewhere we drop evaluation points of functions when it causes no confusion. With this notation, τ_0 corresponds to the standard sharp RD estimand, while τ_1 denotes the sharp kink RD estimand (up to scale).

[1]A preliminary version of this paper circulated under the title "Coverage Error Optimal Confidence Intervals for Regression Discontinuity Designs" (first draft: June 26, 2016). We thank Josh Angrist, Zhuan Pei, Rocio Titiunik, and Gonzalo Vazquez-Bare for comments. Cattaneo gratefully acknowledges financial support from the National Science Foundation (SES 1357561 and SES 1459931). Farrell gratefully acknowledges financial support from the Richard N. Rosett and John E. Jeuck Fellowships.

S.1.1. Local Polynomial Point Estimation

Will not give a complete treatment of local polynomial estimation here. For background, careful derivations of the results and formulas herein, and further technical details, see the following: Fan and Gijbels (1996) for background, Calonico, Cattaneo, and Titiunik (2014) in the context of RD specifically, and Calonico, Cattaneo, and Farrell (2018, 2019) for further technical details particularly in the context of Edgeworth expansions.

We estimate τ_ν by taking the difference of two local polynomial estimators, from each side of c. Define the coefficients of the (one-sided, weighted) local regressions:

$$\begin{aligned}\hat{\boldsymbol{\beta}}_- = \hat{\boldsymbol{\beta}}_{-,p}(h) &= \underset{\boldsymbol{b}\in\mathbb{R}^{p+1}}{\arg\min} \sum_{i=1}^{n} (Y_i - \boldsymbol{r}_p(X_i - c)'\boldsymbol{b})^2 K_-\left(X_{h,i}\right) = \frac{1}{nh^\nu}\boldsymbol{\Gamma}_{-,p}^{-1}\boldsymbol{\Omega}_{-,p}\boldsymbol{Y}, \\ \hat{\boldsymbol{\beta}}_+ = \hat{\boldsymbol{\beta}}_{-,p}(h) &= \underset{\boldsymbol{b}\in\mathbb{R}^{p+1}}{\arg\min} \sum_{i=1}^{n} (Y_i - \boldsymbol{r}_p(X_i - c)'\boldsymbol{b})^2 K_+\left(X_{h,i}\right) = \frac{1}{nh^\nu}\boldsymbol{\Gamma}_{+,p}^{-1}\boldsymbol{\Omega}_{+,p}\boldsymbol{Y},\end{aligned} \tag{S.1.2}$$

where:

- p is an integer greater than $\min\{1,\nu\}$,
- $\boldsymbol{e}_k$ is a conformable zero vector with a one in the $(k+1)$ position, for example $\boldsymbol{e}_\nu$ is the $(p+1)$-vector with a one in the ν^{th} position and zeros in the rest,
- $\boldsymbol{r}_p(u) = (1, u, u^2, \ldots, u^p)'$,
- h is a positive bandwidth sequence that vanishes as n diverges,
- $X_{h,i} = (X_i - c)/h$, for a bandwidth h and point of interest c,
- $K_-(u) = \mathbb{1}\{u < 0\}K(u)$ and $K_+(u) = \mathbb{1}\{u \geq 0\}K(u)$ for a kernel function $K(u)$, with in particular $K_-(X_{h,i}) = \mathbb{1}(X_i < c)K(X_{h,i})$ and $K_+(X_{h,i}) = \mathbb{1}(c \leq X_i)K(X_{h,i})$,
- to save space, products of functions will often be written together, with only one argument, for example

  $$(K\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i}) := K(X_{h,i})r_p(X_{h,i})r_p(X_{h,i})' = K\left(\frac{X_i - c}{h}\right)\boldsymbol{r}_p\left(\frac{X_i - c}{h}\right)\boldsymbol{r}_p\left(\frac{X_i - c}{h}\right)',$$

- $\boldsymbol{\Gamma}_{-,p} = \frac{1}{nh}\sum_{i=1}^n (K_-\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})$ and $\boldsymbol{\Gamma}_{+,p} = \frac{1}{nh}\sum_{i=1}^n (K_+\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})$,
- $\boldsymbol{\Omega}_{-,p} = h^{-1}[(K_-\boldsymbol{r}_p)(X_{h,1}), \ldots, (K_-\boldsymbol{r}_p)(X_{h,n})]$ and $\boldsymbol{\Omega}_{+,p} = h^{-1}[(K_+\boldsymbol{r}_p)(X_{h,1}), \ldots, (K_+\boldsymbol{r}_p)(X_{h,n})]$, and
- $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$.

We maintain the same bandwidth and kernel function on both sides of the cutoff for notational simplicity. Accommodating different bandwidths, which share a rate of decay, is only a matter of notational burden. At the expense of substantial complication, any aspect of the local polynomial fit on one side can be different from the other, including the bandwidth rate and the order p; all the results will still hold in principle. As this approach is rarely taken in practice, we decide not to introduce the complication.

The standard point estimator of the parameter of interest τ_ν of Equation (S.1.1) is then the difference of the appropriate two entries from the one-sided coefficient vectors:

$$\hat{\tau}_\nu = \hat{\tau}_{\nu,\mathtt{US}} = \hat{\tau}_\nu(h) = \nu!\boldsymbol{e}_\nu'\hat{\boldsymbol{\beta}}_+ - \nu!\boldsymbol{e}_\nu'\hat{\boldsymbol{\beta}}_- = \frac{1}{nh^\nu}\nu!\boldsymbol{e}_\nu'\left(\boldsymbol{\Gamma}_{+,p}^{-1}\boldsymbol{\Omega}_{+,p} - \boldsymbol{\Gamma}_{-,p}^{-1}\boldsymbol{\Omega}_{-,p}\right)\boldsymbol{Y}, \tag{S.1.3}$$

which is also denoted $\hat{\tau}_{\nu,\mathtt{US}}$ to explicitly refer to the fact that undersmoothing is required for valid inference. Compared to the main text, we will often drop the dependence on the bandwidth unless it is required to make a specific point.

S.1.2. Assumptions

Let $g^{(s)}(x) = \partial^\nu g(x)/\partial x^\nu$ for any sufficiently smooth function $g(\cdot)$, with $g(x) = g^{(0)}(x)$ to save notation.

ASSUMPTION S.1.1. (RD) *For some $S > p \geq \min\{1, \nu\}$ and all $x \in [x_l, x_u]$, where $x_l < c < x_u$,*

(a) the Lebesgue density of X_i, denoted by $f(x)$, is positive and continuous,
(b) $\mu_-(x) = \mathbb{E}[Y_i(0)|X_i = x]$ and $\mu_+(x) = \mathbb{E}[Y_i(1)|X_i = x]$ are S times continuously differentiable, with $\mu_-^{(S)}(x)$ and $\mu_+^{(S)}(x)$ Hölder continuous with exponent $a \in (0,1]$, and
(c) $\mathbb{E}[|Y_i(t)|^\delta|X_i = x]$ continuous, for $t \in \{0,1\}$ and $\delta > 8$, with $\sigma_-^2(x) = \mathbb{V}[Y_i(0)|X_i = x]$ and $\sigma_+^2(x) = \mathbb{V}[Y_i(1)|X_i = x]$ positive and continuous, and
(d) the Lebesgue density of $(Y(t), X)$, $f_{y_t x}(\cdot)$, is positive and continuous.

The only difference between this assumption and its counterpart in the main text is that we have defined the function μ_+, μ_-, $\sigma_+^2(x)$, and $\sigma_-^2(x)$, which we will need later. With this notation the parameter of interest is (cf. (S.1.1))

$$\tau_\nu = \tau_\nu(c) = \frac{\partial^\nu}{\partial x^\nu}\mathbb{E}[Y_i(1) - Y_i(0)|X_i = x]\bigg|_{x=c} = \mu_+^{(\nu)}(c) - \mu_-^{(\nu)}(c)$$

and the standard point estimator is (cf. (S.1.3))

$$\hat{\tau}_\nu = \hat{\tau}_{\nu,\mathtt{US}} = \hat{\tau}_\nu(h) = \nu!\boldsymbol{e}_\nu'\hat{\boldsymbol{\beta}}_+ - \nu!\boldsymbol{e}_\nu'\hat{\boldsymbol{\beta}}_- = \hat{\mu}_+^{(\nu)}(c) - \hat{\mu}_-^{(\nu)}(c).$$

The conditions on the kernel function are as follows.

ASSUMPTION S.1.2. (KERNEL) *$K(u) = \mathbb{1}(u < 0)k(-u) + \mathbb{1}(u \geq 0)k(u)$, where $k(\cdot) : [0,1] \mapsto \mathbb{R}$ is bounded and continuous on its support, positive $(0,1)$, zero outside its support, and either is constant or $(1, K(u)\boldsymbol{r}_{3(p+1)}(u)')$ is linearly independent on $(0,1)$.*

S.2. TECHNICAL DETAILS AND FORMULAS OMITTED FROM THE MAIN TEXT

In this section we state formulas and technical details omitted from the main text. These consist of bias and variance terms and their estimators and the terms of the coverage error expansion. Throughout we maintain Assumptions S.1.1 and S.1.2 with $S \geq p + 1$, or, where mentioned, $S \geq p + 2$. Derivations of many of the formulas in the first two subsections can be found in Calonico, Cattaneo, and Titiunik (2014). When sufficient smoothness does not exist, the results of Calonico, Cattaneo, and Farrell (2018, 2019) apply.

Recall from Equation (S.1.3) that the standard point estimator of the parameter of interest τ_ν of Equation (S.1.1) is the difference of the appropriate two entries from the one-sided coefficient vectors,

$$\hat{\tau}_\nu = \hat{\tau}_\nu(h) = \nu!\boldsymbol{e}_\nu'\hat{\boldsymbol{\beta}}_+ - \nu!\boldsymbol{e}_\nu'\hat{\boldsymbol{\beta}}_- = \frac{1}{nh^\nu}\nu!\boldsymbol{e}_\nu'\left(\boldsymbol{\Gamma}_{+,p}^{-1}\boldsymbol{\Omega}_{+,p} - \boldsymbol{\Gamma}_{-,p}^{-1}\boldsymbol{\Omega}_{-,p}\right)\boldsymbol{Y},$$

which will also be denoted $\hat{\tau}_{\nu,\mathtt{US}}$ to explicitly refer to the fact that undersmoothing is required for valid inference.

S.2.1. Bias and Bias Correction

The conditional bias of $\hat{\tau}_\nu$ obeys

$$\begin{aligned} &\mathbb{E}\left[\hat{\tau}_\nu \middle| X_1, \ldots, X_n\right] - \tau_\nu = h^{p+1-\nu}\mathscr{B} + o_P(h^{p+1-\nu}), \\ &\text{where} \qquad \mathscr{B} = \frac{\nu!}{(p+1)!}\boldsymbol{e}_\nu'\left(\boldsymbol{\Gamma}_{+,p}^{-1}\boldsymbol{\Lambda}_{+,p}\mu_+^{(p+1)} - \boldsymbol{\Gamma}_{-,p}^{-1}\boldsymbol{\Lambda}_{-,p}\mu_-^{(p+1)}\right), \end{aligned} \tag{S.2.1}$$

with

- $\boldsymbol{\Lambda}_{+,p} = \boldsymbol{\Omega}_{+,p}\left[X_{h,1}^{p+1}, \ldots, X_{h,n}^{p+1}\right]'/n$ and similarly for $\boldsymbol{\Lambda}_{-,p}$, and
- $\mu_+^{(p+1)} = \frac{\partial^\nu}{\partial x^\nu}\mathbb{E}[Y(1)|X_i = x]\big|_{x=c}$, and similarly for $\mu_-^{(p+1)}$, see Assumption S.1.1.

The bias of (S.2.1) is first-order important without further steps. See the main paper for discussion. Because its asymptotic order is $h^{p+1-\nu}$, undersmoothing relies on a "small" bandwidth choice, i.e. one assumed to vanish rapidly enough to render this bias ignorable. Robust bias correction involves estimating $\mathscr{B}$ and subtracting this estimate from the point estimator $\hat{\tau}_\nu$. The estimator of $\mathscr{B}$ will also be based on one-sided local polynomial regressions, of exactly the same form as (S.1.2) but with the degree of the polynomial one order higher, $q = p + 1$ (see Remark S.2.1), and a bandwidth b defined as $b = \rho^{-1}h$. Specifically,

$$\hat{\tau}_{\nu,\mathtt{BC}} = \hat{\tau}_\nu - h^{p+1-\nu}\hat{\mathscr{B}} = \frac{1}{nh^\nu}\nu!\boldsymbol{e}_\nu'\left(\boldsymbol{\Gamma}_{+,p}^{-1}\boldsymbol{\Omega}_{+,\mathtt{BC}} - \boldsymbol{\Gamma}_{-,p}^{-1}\boldsymbol{\Omega}_{-,\mathtt{BC}}\right)\boldsymbol{Y}, \tag{S.2.2}$$

where

$$\hat{\mathscr{B}} = \frac{\nu!}{(p+1)!}\boldsymbol{e}_\nu'\left(\boldsymbol{\Gamma}_{+,p}^{-1}\boldsymbol{\Lambda}_{+,p}\hat{\mu}_+^{(p+1)} - \boldsymbol{\Gamma}_{-,p}^{-1}\boldsymbol{\Lambda}_{-,p}\hat{\mu}_-^{(p+1)}\right), \tag{S.2.3}$$

and

$$\boldsymbol{\Omega}_{+,\mathtt{BC}} = \boldsymbol{\Omega}_{+,p} - \rho^{p+1}\boldsymbol{\Lambda}_{+,p}\boldsymbol{e}_{p+1}'\boldsymbol{\Gamma}_{+,q}^{-1}\boldsymbol{\Omega}_{+,q} \quad \text{and} \quad \boldsymbol{\Omega}_{-,\mathtt{BC}} = \boldsymbol{\Omega}_{-,p} - \rho^{p+1}\boldsymbol{\Lambda}_{-,p}\boldsymbol{e}_{p+1}'\boldsymbol{\Gamma}_{-,q}^{-1}\boldsymbol{\Omega}_{-,q}$$

stemming from the estimation of the derivatives using local polynomials. That is,

$$\hat{\mu}_+^{(p+1)} = \frac{1}{nb^{p+1}}(p+1)!\boldsymbol{e}_{p+1}'\boldsymbol{\Gamma}_{+,q}^{-1}\boldsymbol{\Omega}_{+,q}\boldsymbol{Y} \quad \text{and} \quad \hat{\mu}_-^{(p+1)} = \frac{1}{nb^{p+1}}(p+1)!\boldsymbol{e}_{p+1}'\boldsymbol{\Gamma}_{-,q}^{-1}\boldsymbol{\Omega}_{-,q}\boldsymbol{Y},$$

with

- an integer $q \geq p$ taken throughout to be $q = p + 1$ (Calonico, Cattaneo, and Farrell (2018) show why $q = p + 1$ is the optimal choice for coverage considerations. See also Remark S.2.1) and
- $b = \rho^{-1}h$ is a positive bandwidth sequence that vanishes as n diverges.

Given these, the rest of the notation is defined analogously to the above, namely:

- $\boldsymbol{r}_q(u) = (1, u, u^2, \ldots, u^q)'$,
- $X_{b,i} = (X_i - c)/b$, for a bandwidth b and point of interest c,
- $\boldsymbol{\Gamma}_{-,q} = \frac{1}{nb}\sum_{i=1}^n(K_-\boldsymbol{r}_q\boldsymbol{r}_q')(X_{b,i})$ and $\boldsymbol{\Omega}_{-,q} = b^{-1}[(K_-\boldsymbol{r}_q)(X_{b,1}), \ldots, (K_-\boldsymbol{r}_q)(X_{b,n})]$ and similarly for $\boldsymbol{\Gamma}_{+,q}$ and $\boldsymbol{\Omega}_{+,q}$.

The bias of $\hat{\tau}_{\nu,\mathtt{BC}}$ itself is an important quantity for the coverage error expansions and feasible inference-optimal bandwidth selectors. This is given by

$$\mathbb{E}\left[\hat{\tau}_{\nu,\mathtt{BC}}\middle|X_1,\dots,X_n\right]-\tau_\nu=\begin{cases}O(h^{S+a-\nu}) & \text{if } S\le p+1\\ h^{p+2-\nu}\mathscr{B}_{\mathtt{BC}}+o_P(h^{p+2-\nu}) & \text{if } S\ge p+2,\end{cases}\tag{S.2.4}$$

where

$$\mathscr{B}_{\mathtt{BC}}=\frac{\mu_+^{(p+2)}}{(p+2)!}\nu!e_\nu'\mathbf{\Gamma}_{+,p}^{-1}\left\{\mathbf{\Lambda}_{+,p,2}-\rho^{-1}\mathbf{\Lambda}_{+,p}e_{p+1}'\mathbf{\Gamma}_{+,q}^{-1}\mathbf{\Lambda}_{+,q}\right\}$$
$$-\frac{\mu_-^{(p+2)}}{(p+2)!}\nu!e_\nu'\mathbf{\Gamma}_{-,p}^{-1}\left\{\mathbf{\Lambda}_{-,p,2}-\rho^{-1}\mathbf{\Lambda}_{-,p}e_{p+1}'\mathbf{\Gamma}_{-,q}^{-1}\mathbf{\Lambda}_{-,q}\right\},$$

using the notation

- $\rho=h/b$,
- $\mathbf{\Lambda}_{+,p,k}=\mathbf{\Omega}_{+,p}\left[X_{h,1}^{p+k},\dots,X_{h,n}^{p+k}\right]'/n$, with $\mathbf{\Lambda}_{+,p,1}=\mathbf{\Lambda}_{+,p}$ in particular, and similarly for $\mathbf{\Lambda}_{-,p,k}$, and
- $\mathbf{\Lambda}_{+,q,k}=\mathbf{\Omega}_{+,q}\left[X_{b,1}^{q+k},\dots,X_{b,n}^{q+k}\right]'/n$, with $\mathbf{\Lambda}_{+,q,1}=\mathbf{\Lambda}_{+,q}$ in particular, and similarly for $\mathbf{\Lambda}_{-,q,k}$.

Remark S.2.1. (Setting $q>p+1$ or $\rho\to\infty$) *It is possible to perform robust bias correction with a polynomial order $q>p+1$ or with $\rho\to\infty$, i.e. a bandwidth b asymptotically smaller than h. However, neither can improve coverage. The former will tend to inflate variance constants and (to be made feasible) requires estimation of higher derivatives, while the latter leads to a slower variance rate. To see why, first, the general form of $\mathscr{B}_{\mathtt{BC}}$, provided all derivatives exist (and if they do not, there is even less point to higher q) will be*

$$\mathscr{B}_{\mathtt{BC}}=\frac{\mu_+^{(p+2)}}{(p+2)!}\nu!e_\nu'\mathbf{\Gamma}_{+,p}^{-1}\mathbf{\Lambda}_{+,p,2}-\rho^{-1}b^{q-p-1}\frac{\mu_+^{(q+1)}}{(q+1)!}\nu!e_\nu'\mathbf{\Gamma}_{+,p}^{-1}\mathbf{\Lambda}_{+,p}e_{p+1}'\mathbf{\Gamma}_{+,q}^{-1}\mathbf{\Lambda}_{+,q}$$
$$-\frac{\mu_-^{(p+2)}}{(p+2)!}\nu!e_\nu'\mathbf{\Gamma}_{-,p}^{-1}\mathbf{\Lambda}_{-,p,2}-\rho^{-1}b^{q-p-1}\frac{\mu_-^{(q+1)}}{(q+1)!}\nu!e_\nu'\mathbf{\Gamma}_{-,p}^{-1}\mathbf{\Lambda}_{-,p}e_{p+1}'\mathbf{\Gamma}_{-,q}^{-1}\mathbf{\Lambda}_{-,q}.$$

The order of the second term of each line decreases for higher q (provided the same h sequence is assumed) because the bias of the bias estimator is decreasing. However, the first term of each line, representing the bias not targeted for bias correction, is unchanged. Thus, in rates, nothing can be gained from $q>p+1$.

Next, suppose that we allow $\rho\to\infty$. Again, the second term in each line is higher order but the first is unchanged, and so the bias rate is not improved (unless $q>p+1$). However, the variance of the estimator will now be determined by $(nb)^{-1}$ instead of $(nh)^{-1}$, that is, the variance of the the derivative estimates $\hat{\mu}_+^{(p+1)}$ and $\hat{\mu}_-^{(p+1)}$ is now the dominant variance portion. Setting a finite, positive ρ balances these two.

See Calonico, Cattaneo, and Farrell (2018) for further discussion and an expansion with general q in the context of local polynomial regression.

S.2.2. Variance and Variance Estimators

To compute the conditional variance define the matrices

- $\Sigma_+ = \text{diag}(\sigma_+^2(X_i) : i = 1, \ldots, n)$, with $\sigma_+^2(x) = \mathbb{V}[Y(1)|X = x]$ defined in Assumption S.1.1, and similarly for Σ_-.

For $\hat{\tau}_\nu$, given in (S.1.3), we find

$$\begin{aligned} \mathbb{V}\left[\hat{\tau}_\nu \middle| X_1, \ldots, X_n\right] &= \frac{1}{nh^{1+2\nu}} \mathscr{V}, \\ \mathscr{V} = \mathsf{v}_{\mathtt{US}}^2 &= \frac{h}{n} \nu!^2 \boldsymbol{e}_\nu' \Big(\boldsymbol{\Gamma}_{+,p}^{-1} \boldsymbol{\Omega}_{+,p} \boldsymbol{\Sigma}_+ \boldsymbol{\Omega}_{+,p}' \boldsymbol{\Gamma}_{+,p}^{-1} + \boldsymbol{\Gamma}_{-,p}^{-1} \boldsymbol{\Omega}_{-,p} \boldsymbol{\Sigma}_- \boldsymbol{\Omega}_{-,p}' \boldsymbol{\Gamma}_{-,p}^{-1} \Big) \boldsymbol{e}_\nu, \end{aligned} \tag{S.2.5}$$

where we simultaneously define $\mathscr{V}$ and $\mathsf{v}_{\mathtt{US}}^2$. These are identical, but it will frequently be convenient to write $\mathsf{v}_{\mathtt{US}}$ rather than $\mathscr{V}^{1/2}$. Compared to the main text, we will often drop the dependence on the bandwidth unless it is required to make a specific point, e.g., we write $\mathscr{V}$ instead of $\mathscr{V}(h)$.

For $\hat{\tau}_{\nu,\mathtt{BC}}$, given in (S.2.2), we find

$$\begin{aligned} \mathbb{V}\left[\hat{\tau}_{\nu,\mathtt{BC}} \middle| X_1, \ldots, X_n\right] &= \frac{1}{nh^{1+2\nu}} \mathscr{V}_{\mathtt{BC}}, \\ \mathscr{V}_{\mathtt{BC}} = \mathsf{v}_{\mathtt{BC}}^2 &= \frac{h}{n} \nu!^2 \boldsymbol{e}_\nu' \Big(\boldsymbol{\Gamma}_{+,p}^{-1} \boldsymbol{\Omega}_{+,\mathtt{BC}} \boldsymbol{\Sigma}_+ \boldsymbol{\Omega}_{+,\mathtt{BC}}' \boldsymbol{\Gamma}_{+,p}^{-1} + \boldsymbol{\Gamma}_{-,p}^{-1} \boldsymbol{\Omega}_{-,\mathtt{BC}} \boldsymbol{\Sigma}_- \boldsymbol{\Omega}_{-,\mathtt{BC}}' \boldsymbol{\Gamma}_{-,p}^{-1} \Big) \boldsymbol{e}_\nu, \end{aligned} \tag{S.2.6}$$

where we simultaneously define $\mathscr{V}_{\mathtt{BC}}$ and $\mathsf{v}_{\mathtt{BC}}^2$. These are identical, but it will frequently be convenient to write $\mathsf{v}_{\mathtt{BC}}$ rather than $\mathscr{V}_{\mathtt{BC}}^{1/2}$. Notice that the only change is replacing $\boldsymbol{\Omega}_{+,\mathtt{BC}}$ and $\boldsymbol{\Omega}_{-,\mathtt{BC}}$ for $\boldsymbol{\Omega}_{+,p}$ and $\boldsymbol{\Omega}_{-,p}$, as expected from comparing (S.2.2) and (S.1.3).

To estimate these variances we need only estimate the diagonal matrices Σ_+ and Σ_-. Define

$$\begin{aligned} \hat{\boldsymbol{\Sigma}}_{+,p} &= \text{diag}\Big((Y_i - \boldsymbol{r}_p(X_i - c)'\hat{\boldsymbol{\beta}}_{+,p})^2 : i = 1, \ldots, n \Big), \\ \hat{\boldsymbol{\Sigma}}_{-,p} &= \text{diag}\Big((Y_i - \boldsymbol{r}_p(X_i - c)'\hat{\boldsymbol{\beta}}_{-,p})^2 : i = 1, \ldots, n \Big), \\ \hat{\boldsymbol{\Sigma}}_{+,\mathtt{BC}} &= \text{diag}\Big((Y_i - \boldsymbol{r}_q(X_i - c)'\hat{\boldsymbol{\beta}}_{+,q})^2 : i = 1, \ldots, n \Big), \end{aligned}$$

and

$$\hat{\boldsymbol{\Sigma}}_{-,\mathtt{BC}} = \text{diag}\Big((Y_i - \boldsymbol{r}_q(X_i - c)'\hat{\boldsymbol{\beta}}_{-,q})^2 : i = 1, \ldots, n \Big),$$

where $\hat{\boldsymbol{\beta}}_{+,p}$ and $\hat{\boldsymbol{\beta}}_{-,p}$ are given in Equation (S.1.2) and $\hat{\boldsymbol{\beta}}_{+,q}$ and $\hat{\boldsymbol{\beta}}_{-,q}$ are the same but with b in place of h and q in place of p.

With these in hand, define

$$\begin{aligned} \hat{\mathscr{V}} = \hat{\mathsf{v}}_{\mathtt{US}}^2 &= \frac{h}{n} \nu!^2 \boldsymbol{e}_\nu' \Big(\boldsymbol{\Gamma}_{+,p}^{-1} \boldsymbol{\Omega}_{+,p} \hat{\boldsymbol{\Sigma}}_{+,p} \boldsymbol{\Omega}_{+,p}' \boldsymbol{\Gamma}_{+,p}^{-1} + \boldsymbol{\Gamma}_{-,p}^{-1} \boldsymbol{\Omega}_{-,p} \hat{\boldsymbol{\Sigma}}_{-,p} \boldsymbol{\Omega}_{-,p}' \boldsymbol{\Gamma}_{-,p}^{-1} \Big) \boldsymbol{e}_\nu \\ \hat{\mathscr{V}}_{\mathtt{BC}} = \hat{\mathsf{v}}_{\mathtt{BC}}^2 &= \frac{h}{n} \nu!^2 \boldsymbol{e}_\nu' \Big(\boldsymbol{\Gamma}_{+,p}^{-1} \boldsymbol{\Omega}_{+,\mathtt{BC}} \hat{\boldsymbol{\Sigma}}_{+,\mathtt{BC}} \boldsymbol{\Omega}_{+,\mathtt{BC}}' \boldsymbol{\Gamma}_{+,p}^{-1} + \boldsymbol{\Gamma}_{-,p}^{-1} \boldsymbol{\Omega}_{-,\mathtt{BC}} \hat{\boldsymbol{\Sigma}}_{-,\mathtt{BC}} \boldsymbol{\Omega}_{-,\mathtt{BC}}' \boldsymbol{\Gamma}_{-,p}^{-1} \Big) \boldsymbol{e}_\nu \end{aligned} \tag{S.2.7}$$

Other possibilities for standard errors exist, but retaining the fixed-n form is crucial for good coverage properties. For more discussion, including other options and practical details, see Calonico, Cattaneo, and Farrell (2018, 2019).

S.2.3. Coverage Error Expansion Terms

We now give the precise definition of the terms $\mathscr{Q}_{\mathtt{US},k}$ and $\mathscr{Q}_{\mathtt{RBC},k}$, $k = 1, 2, 3$, appearing the coverage error expansion in the main text. The final formulas appear at the end of this subsection, and require a fair amount of notation to be defined first. See Section S.4.1 for the computation of these terms.

We will maintain, as far as possible, fixed-n calculations. All terms must be nonrandom. First, define the following functions, which depend on n, h, b, ν, p, and K, though this is mostly suppressed notationally. These functions are all calculated in a fixed-n sense and are all bounded and rateless.

$$
\begin{aligned}
\mathcal{L}^0_{\mathtt{US}}(X_i) &= \nu!\boldsymbol{e}'_\nu \left\{ \tilde{\boldsymbol{\Gamma}}^{-1}_{+,p}(K_+\boldsymbol{r}_p)(X_{h,i}) - \tilde{\boldsymbol{\Gamma}}^{-1}_{-,p}(K_-\boldsymbol{r}_p)(X_{h,i}) \right\};\\
\mathcal{L}^0_{\mathtt{RBC}}(X_i) &= \mathcal{L}^0_{\mathtt{US}}(X_i) - \rho^{p+1}\nu!\boldsymbol{e}'_\nu\tilde{\boldsymbol{\Gamma}}^{-1}_{+,p}\tilde{\boldsymbol{\Lambda}}_{+,p}\boldsymbol{e}'_{p+1}\tilde{\boldsymbol{\Gamma}}^{-1}_{+,q}(K_+\boldsymbol{r}_{p+1})(X_{b,i})\\
&\quad + \rho^{p+1}\nu!\boldsymbol{e}'_\nu\tilde{\boldsymbol{\Gamma}}^{-1}_{-,p}\tilde{\boldsymbol{\Lambda}}_{-,p}\boldsymbol{e}'_{p+1}\tilde{\boldsymbol{\Gamma}}^{-1}_{-,q}(K_-\boldsymbol{r}_{p+1})(X_{b,i});\\
\mathcal{L}^1_{\mathtt{US}}(X_i, X_j) &= \nu!\boldsymbol{e}'_\nu\tilde{\boldsymbol{\Gamma}}^{-1}_{+,p}\left(\mathbb{E}[(K_+\boldsymbol{r}_p\boldsymbol{r}'_p)(X_{h,j})] - (K_+\boldsymbol{r}_p\boldsymbol{r}'_p)(X_{h,j})\right)\tilde{\boldsymbol{\Gamma}}^{-1}_{+,p}(K_+\boldsymbol{r}_p)(X_{h,i})\\
&\quad - \nu!\boldsymbol{e}'_\nu\tilde{\boldsymbol{\Gamma}}^{-1}_{-,p}\left(\mathbb{E}[(K_-\boldsymbol{r}_p\boldsymbol{r}'_p)(X_{h,j})] - (K_-\boldsymbol{r}_p\boldsymbol{r}'_p)(X_{h,j})\right)\tilde{\boldsymbol{\Gamma}}^{-1}_{-,p}(K_-\boldsymbol{r}_p)(X_{h,i});\\
\mathcal{L}^1_{\mathtt{RBC}}(X_i, X_j) &= \mathcal{L}^1_{\mathtt{US}}(X_i, X_j)\\
&\quad - \rho^{p+1}\nu!\boldsymbol{e}'_\nu\tilde{\boldsymbol{\Gamma}}^{-1}_{+,p}\Big\{\left(\mathbb{E}[(K_+\boldsymbol{r}_p\boldsymbol{r}'_p)(X_{h,j})] - (K_+\boldsymbol{r}_p\boldsymbol{r}'_p)(X_{h,j})\right)\tilde{\boldsymbol{\Gamma}}^{-1}_{+,p}\tilde{\boldsymbol{\Lambda}}_{+,p}\boldsymbol{e}'_{p+1}\\
&\quad + \left((K_+\boldsymbol{r}_p)(X_{h,j})X^{p+1}_{h,i} - \mathbb{E}[(K_+\boldsymbol{r}_p)(X_{h,j})X^{p+1}_{h,i}]\right)\boldsymbol{e}'_{p+1}\\
&\quad + \tilde{\boldsymbol{\Lambda}}_{+,p}\boldsymbol{e}'_{p+1}\tilde{\boldsymbol{\Gamma}}^{-1}_{+,q}\left(\mathbb{E}[(K_+\boldsymbol{r}_{p+1}\boldsymbol{r}'_{p+1})(X_{b,j})] - (K_+\boldsymbol{r}_{p+1}\boldsymbol{r}'_{p+1})(X_{b,j})\right)\Big\}\tilde{\boldsymbol{\Gamma}}^{-1}_{+,q}(K_+\boldsymbol{r}_{p+1})(X_{b,i})\\
&\quad - \rho^{p+1}\nu!\boldsymbol{e}'_\nu\tilde{\boldsymbol{\Gamma}}^{-1}_{-,p}\Big\{\left(\mathbb{E}[(K_-\boldsymbol{r}_p\boldsymbol{r}'_p)(X_{h,j})] - (K_-\boldsymbol{r}_p\boldsymbol{r}'_p)(X_{h,j})\right)\tilde{\boldsymbol{\Gamma}}^{-1}_{-,p}\tilde{\boldsymbol{\Lambda}}_{-,p}\boldsymbol{e}'_{p+1}\\
&\quad + \left((K_-\boldsymbol{r}_p)(X_{h,j})X^{p+1}_{h,i} - \mathbb{E}[(K_-\boldsymbol{r}_p)(X_{h,j})X^{p+1}_{h,i}]\right)\boldsymbol{e}'_{p+1}\\
&\quad + \tilde{\boldsymbol{\Lambda}}_{-,p}\boldsymbol{e}'_{p+1}\tilde{\boldsymbol{\Gamma}}^{-1}_{-,q}\left(\mathbb{E}[(K_-\boldsymbol{r}_{p+1}\boldsymbol{r}'_{p+1})(X_{b,j})] - (K_-\boldsymbol{r}_{p+1}\boldsymbol{r}'_{p+1})(X_{b,j})\right)\Big\}\tilde{\boldsymbol{\Gamma}}^{-1}_{-,q}(K_-\boldsymbol{r}_{p+1})(X_{b,i}).
\end{aligned}
$$

Further, define

$$
\begin{aligned}
\varepsilon_i &= \mathbb{1}\{X_i < c\}\varepsilon_{-,i} + \mathbb{1}\{X_i \geq c\}\varepsilon_{+,i}\\
v(X_i) &= \mathbb{1}\{X_i < c\}\sigma^2_-(X_i) + \mathbb{1}\{X_i \geq c\}\sigma^2_+(X_i).
\end{aligned}
$$

Let I ("I" for Interval) stand in for either US or RBC.[1] Define $\tilde{\mathsf{v}}^2_I = \mathbb{E}[h^{-1}\mathcal{L}^0_I(X)^2 v(X)]$.

Now we define three functions $Q_{\mathtt{US},k}$ and $Q_{\mathtt{RBC},k}$, $k = 1, 2, 3$ which serve as the main building blocks of the terms of the expansion, capturing in particular all dependence on the data-generating process other than the bias. $Q_{I,1}$ is the most cumbersome notation-

[1]More precisely, with this generic "I" notation, $I = \mathtt{RBC}$ refers to quantities appearing in $\mathscr{Q}_{\mathtt{RBC},k}$, $k = 1, 2, 3$, i.e. those relevant for $I_{\mathtt{RBC}}$, which include notations with a subscript BC, such as $\mathsf{v}_{\mathtt{BC}}$.

ally. Begin with the others. Define

$$Q_{I,2}(z) = -\tilde{\mathsf{v}}_I^{-2}\left\{z/2\right\}$$

and

$$Q_{I,3}(z) = \tilde{\mathsf{v}}_I^{-4}\mathbb{E}[h^{-1}\mathcal{L}_I^0(X_i)^3\varepsilon_i^3]\left\{z^3/3\right\}.$$

For $\mathscr{Q}_{I,1}$, it is not quite as simple to state a generic version. Let $\tilde{\boldsymbol{G}}_+$ stand in for $\tilde{\boldsymbol{\Gamma}}_{+,p}$ or $\tilde{\boldsymbol{\Gamma}}_{+,q}$ and similarly for $\tilde{\boldsymbol{G}}_-$, $\tilde{p}$ stand in for p or $p+1$, and d_n stand in for h or b, all depending on whether $T = T_{\mathtt{US}}$ or $T_{\mathtt{RBC}}$. Note however, that h is still used in many places, in particular for stabilizing fixed-n expectations, for $T_{\mathtt{RBC}}$. Indexes i, j, and k are always distinct (i.e. $X_{h,i} \neq X_{h,j} \neq X_{h,k}$).

$$\begin{aligned}
Q_{I,1}(z) = {} & \tilde{\mathsf{v}}_I^{-6}\mathbb{E}\left[h^{-1}\mathcal{L}_I^0(X_i)^3\varepsilon_i^3\right]^2\left\{z^3/3 + 7z/4 + \tilde{\mathsf{v}}_I^2 z(z^2-3)/4\right\} \\
& + \tilde{\mathsf{v}}_I^{-2}\mathbb{E}\left[h^{-1}\mathcal{L}_I^0(X_i)\mathcal{L}_I^1(X_i,X_i)\varepsilon_i^2\right]\left\{-z(z^2-3)/2\right\} \\
& + \tilde{\mathsf{v}}_I^{-4}\mathbb{E}\left[h^{-1}\mathcal{L}_I^0(X_i)^4(\varepsilon_i^4 - v(X_i)^2)\right]\left\{z(z^2-3)/8\right\} \\
& - \tilde{\mathsf{v}}_I^{-2}\mathbb{E}\left[h^{-1}\mathcal{L}_I^0(X_i)^2\boldsymbol{r}_{\tilde{p}}(X_{d_n,i})'\tilde{\boldsymbol{G}}_+^{-1}(K_+\boldsymbol{r}_{\tilde{p}})(X_{d_n,i})\varepsilon_i^2\right]\left\{z(z^2-1)/2\right\} \\
& - \tilde{\mathsf{v}}_I^{-2}\mathbb{E}\left[h^{-1}\mathcal{L}_I^0(X_i)^2\boldsymbol{r}_{\tilde{p}}(X_{d_n,i})'\tilde{\boldsymbol{G}}_-^{-1}(K_-\boldsymbol{r}_{\tilde{p}})(X_{d_n,i})\varepsilon_i^2\right]\left\{z(z^2-1)/2\right\} \\
& - \tilde{\mathsf{v}}_I^{-4}\mathbb{E}\left[h^{-1}\mathcal{L}_I^0(X_i)^3\boldsymbol{r}_{\tilde{p}}(X_{d_n,i})'\tilde{\boldsymbol{G}}_+^{-1}\varepsilon_i^2\right]\mathbb{E}\left[h^{-1}(K_+\boldsymbol{r}_{\tilde{p}})(X_{d_n,i})\mathcal{L}_I^0(X_i)\varepsilon_i^2\right]\left\{z(z^2-1)\right\} \\
& - \tilde{\mathsf{v}}_I^{-4}\mathbb{E}\left[h^{-1}\mathcal{L}_I^0(X_i)^3\boldsymbol{r}_{\tilde{p}}(X_{d_n,i})'\tilde{\boldsymbol{G}}_-^{-1}\varepsilon_i^2\right]\mathbb{E}\left[h^{-1}(K_-\boldsymbol{r}_{\tilde{p}})(X_{d_n,i})\mathcal{L}_I^0(X_i)\varepsilon_i^2\right]\left\{z(z^2-1)\right\} \\
& + \tilde{\mathsf{v}}_I^{-2}\mathbb{E}\left[h^{-2}\mathcal{L}_I^0(X_i)^2(\boldsymbol{r}_{\tilde{p}}(X_{d_n,i})'\tilde{\boldsymbol{G}}_+^{-1}(K_+\boldsymbol{r}_{\tilde{p}})(X_{d_n,j}))^2\varepsilon_j^2\right]\left\{z(z^2-1)/4\right\} \\
& + \tilde{\mathsf{v}}_I^{-2}\mathbb{E}\left[h^{-2}\mathcal{L}_I^0(X_i)^2(\boldsymbol{r}_{\tilde{p}}(X_{d_n,i})'\tilde{\boldsymbol{G}}_-^{-1}(K_-\boldsymbol{r}_{\tilde{p}})(X_{d_n,j}))^2\varepsilon_j^2\right]\left\{z(z^2-1)/4\right\} \\
& + \tilde{\mathsf{v}}_I^{-4}\mathbb{E}\left[h^{-3}\mathcal{L}_I^0(X_j)^2\boldsymbol{r}_{\tilde{p}}(X_{d_n,j})'\tilde{\boldsymbol{G}}_+^{-1}(K_+\boldsymbol{r}_{\tilde{p}})(X_{d_n,i})\mathcal{L}_I^0(X_i)\boldsymbol{r}_{\tilde{p}}(X_{d_n,j})'\tilde{\boldsymbol{G}}_+^{-1}(K_+\boldsymbol{r}_{\tilde{p}})(X_{d_n,k})\mathcal{L}_I^0(X_k)\varepsilon_i^2\varepsilon_k^2\right] \\
& \qquad\qquad \times \left\{z(z^2-1)/2\right\} \\
& + \tilde{\mathsf{v}}_I^{-4}\mathbb{E}\left[h^{-3}\mathcal{L}_I^0(X_j)^2\boldsymbol{r}_{\tilde{p}}(X_{d_n,j})'\tilde{\boldsymbol{G}}_-^{-1}(K_-\boldsymbol{r}_{\tilde{p}})(X_{d_n,i})\mathcal{L}_I^0(X_i)\boldsymbol{r}_{\tilde{p}}(X_{d_n,j})'\tilde{\boldsymbol{G}}_-^{-1}(K_-\boldsymbol{r}_{\tilde{p}})(X_{d_n,k})\mathcal{L}_I^0(X_k)\varepsilon_i^2\varepsilon_k^2\right] \\
& \qquad\qquad \times \left\{z(z^2-1)/2\right\} \\
& + \tilde{\mathsf{v}}_I^{-4}\mathbb{E}\left[h^{-1}\mathcal{L}_I^0(X_i)^4\varepsilon_i^4\right]\left\{-z(z^2-3)/24\right\} \\
& + \tilde{\mathsf{v}}_I^{-4}\mathbb{E}\left[h^{-1}\left(\mathcal{L}_I^0(X_i)^2 v(X_i) - \mathbb{E}[\mathcal{L}_I^0(X_i)^2 v(X_i)]\right)\mathcal{L}_I^0(X_i)^2\varepsilon_i^2\right]\left\{z(z^2-1)/4\right\} \\
& + \tilde{\mathsf{v}}_I^{-4}\mathbb{E}\left[h^{-2}\mathcal{L}_I^1(X_i,X_j)\mathcal{L}_I^0(X_i)\mathcal{L}_I^0(X_j)^2\varepsilon_j^2 v(X_i)\right]\left\{z(z^2-3)\right\} \\
& + \tilde{\mathsf{v}}_I^{-4}\mathbb{E}\left[h^{-2}\mathcal{L}_I^1(X_i,X_j)\mathcal{L}_I^0(X_i)\left(\mathcal{L}_I^0(X_j)^2 v(X_j) - \mathbb{E}[\mathcal{L}_I^0(X_j)^2 v(X_j)]\right)\varepsilon_i^2\right]\left\{-z\right\} \\
& + \tilde{\mathsf{v}}_I^{-4}\mathbb{E}\left[h^{-1}\left(\mathcal{L}_I^0(X_i)^2 v(X_i) - \mathbb{E}[\mathcal{L}_I^0(X_i)^2 v(X_i)]\right)^2\right]\left\{-z(z^2+1)/8\right\}.
\end{aligned}$$

For computation, note that the tenth and eleventh terms can be rewritten by factoring the expectation, after rearranging the terms using the fact that $\boldsymbol{r}_{\tilde{p}}(X_{d_n,j})'\tilde{\boldsymbol{G}}^{-1}\boldsymbol{r}_{\tilde{p}}(X_{d_n,i})$ is a scalar, as follows:

$$\begin{aligned}
& \mathbb{E}\left[h^{-3}\mathcal{L}_I^0(X_j)^2\boldsymbol{r}_{\tilde{p}}(X_{d_n,j})'\tilde{\boldsymbol{G}}^{-1}(K\boldsymbol{r}_{\tilde{p}})(X_{d_n,i})\mathcal{L}_I^0(X_i)\boldsymbol{r}_{\tilde{p}}(X_{d_n,j})'\tilde{\boldsymbol{G}}^{-1}(K\boldsymbol{r}_{\tilde{p}})(X_{d_n,k})\mathcal{L}_I^0(X_k)\varepsilon_i^2\varepsilon_k^2\right] \\
& \quad = \mathbb{E}\left[h^{-1}\mathcal{L}_I^0(X_i)\varepsilon_i^2(K\boldsymbol{r}_{\tilde{p}}')(X_{d_n,i})\tilde{\boldsymbol{G}}^{-1}\right]\;\mathbb{E}\left[h^{-1}\boldsymbol{r}_{\tilde{p}}(X_{d_n,j})\mathcal{L}_I^0(X_j)^2\boldsymbol{r}_{\tilde{p}}(X_{d_n,j})'\tilde{\boldsymbol{G}}^{-1}\right]
\end{aligned}$$

$$\times\ \mathbb{E}\left[h^{-1}(K\boldsymbol{r}_{\tilde{p}})(X_{d_n,k})\mathcal{L}^0_I(X_k)\varepsilon_k^2\right].$$

This will greatly ease implementation.

The final ingredient required to define the $\mathscr{Q}_{\texttt{US},k}$ and $\mathscr{Q}_{\texttt{RBC},k}$ terms is the bias. The expressions in Equations (S.2.1) and (S.2.4) can not be used as these are random. Instead, their fixed-n analogues will appear. To this end, define

$$\tilde{\mathscr{B}}_{\texttt{US}} = \frac{\nu!}{(p+1)!}\boldsymbol{e}_\nu'\left(\tilde{\boldsymbol{\Gamma}}^{-1}_{+,p}\tilde{\boldsymbol{\Lambda}}_{+,p}\mu_+^{(p+1)} - \tilde{\boldsymbol{\Gamma}}^{-1}_{-,p}\tilde{\boldsymbol{\Lambda}}_{-,p}\mu_-^{(p+1)}\right)$$

and

$$\tilde{\mathscr{B}}_{\texttt{BC}} = \frac{\mu_+^{(p+2)}}{(p+2)!}\nu!e_\nu'\tilde{\boldsymbol{\Gamma}}^{-1}_{+,p}\left\{\tilde{\boldsymbol{\Lambda}}_{+,p,2} - \rho^{-1}\tilde{\boldsymbol{\Lambda}}_{+,p}e_{p+1}'\tilde{\boldsymbol{\Gamma}}^{-1}_{+,q}\tilde{\boldsymbol{\Lambda}}_{+,q}\right\}$$
$$- \frac{\mu_-^{(p+2)}}{(p+2)!}\nu!e_\nu'\tilde{\boldsymbol{\Gamma}}^{-1}_{-,p}\left\{\tilde{\boldsymbol{\Lambda}}_{-,p,2} - \rho^{-1}\tilde{\boldsymbol{\Lambda}}_{-,p}e_{p+1}'\tilde{\boldsymbol{\Gamma}}^{-1}_{-,q}\tilde{\boldsymbol{\Lambda}}_{-,q}\right\},$$

where

- $\tilde{\boldsymbol{\Gamma}}_{+,p} = \mathbb{E}[\boldsymbol{\Gamma}_{+,p}]$, $\tilde{\boldsymbol{\Lambda}}_{+,p} = \mathbb{E}[\boldsymbol{\Lambda}_{+,p}]$, and so forth.

Finally, the $\mathscr{Q}_{\texttt{US},k}$ and $\mathscr{Q}_{\texttt{RBC},k}$ terms are defined as follows, where as usual I stands in for either $I_{\texttt{US}}$ or $I_{\texttt{RBC}}$,

$$\begin{aligned}\mathscr{Q}_{I,1} &= 2\phi(z_{\alpha/2})Q_{I,1}(z_{\alpha/2})\\ \mathscr{Q}_{I,2} &= 2\phi(z_{\alpha/2})Q_{I,2}(z_{\alpha/2})\tilde{\mathscr{B}}_I^2 \\ \mathscr{Q}_{I,3} &= 2\phi(z_{\alpha/2})Q_{I,3}(z_{\alpha/2})\tilde{\mathscr{B}}_I\end{aligned} \tag{S.2.8}$$

S.3. MAIN RESULTS: COVERAGE ERROR AND EDGEWORTH EXPANSIONS

We now state the main theoretical results: Edgeworth expansion for the distributions of the t-statistics

$$T_{\texttt{US}} = \frac{\sqrt{nh^{1+2\nu}}(\hat{\tau}_{\nu,\texttt{US}} - \tau_\nu)}{\hat{\mathtt{v}}_{\texttt{US}}} \qquad \text{and} \qquad T_{\texttt{RBC}} = \frac{\sqrt{nh^{1+2\nu}}(\hat{\tau}_{\nu,\texttt{BC}} - \tau_\nu)}{\hat{\mathtt{v}}_{\texttt{BC}}}. \tag{S.3.1}$$

The point estimators are given in Equations (S.1.3) and (S.2.2) and the standard errors are in (S.2.7).

Before stating the results, more notation is needed. In addition to the terms $Q_{I,k}$, $k = 1, 2, 3$, two other terms appear in the Edgeworth expansion for the t-statistic, which then cancel upon computing coverage error due to symmetry. These are

$$Q_{I,4}(z) = \tilde{\mathtt{v}}_I^{-3}\mathbb{E}\left[h^{-1}\mathcal{L}^0_I(X_i)^3\varepsilon_i^3\right]\left\{(2z^2-1)/6\right\} \qquad \text{and} \qquad Q_{I,5}(z) = -\tilde{\mathtt{v}}_I^{-1}.$$

The coverage error expansions follow immediately from the results below by taking the difference of expansions at $z_{1-\alpha/2}$ and $z_{\alpha/2}$. It is clearest to state separate results for $T_{\texttt{US}}$ and $T_{\texttt{RBC}}$. For the standard, or undersmoothing, approach, we have the following result.

Theorem S.3.1. (Edgeworth Expansion for $T_{\texttt{US}}$) *Suppose Assumptions S.1.1 and S.1.2 hold with $S \geq p+1$. If $nh/\log(nh)^{2+\gamma} \to \infty$ and $\sqrt{nh}h^{p+1}\log(nh)^{1+\gamma} \to 0$, for any $\gamma > 0$, then*

$$\sup_{z\in\mathbb{R}}\left|\mathbb{P}[T_{\texttt{US}} < z] - \Phi(z) - \phi(z)\mathcal{E}_{\texttt{US}}(z)\right| = \epsilon_{\texttt{US}},$$

where $\epsilon_{\texttt{US}} = o((nh)^{-1}) + O(nh^{3+2p+2a} + h^{p+1+a})$ *and*

$$\mathscr{E}_{\texttt{US}}(z) = \frac{1}{nh}Q_{I_{\texttt{US}},1} + nh^{3+2p}Q_{I_{\texttt{US}},2}\tilde{\mathscr{B}}^2_{\texttt{US}} + h^{p+1}Q_{I_{\texttt{US}},3}\tilde{\mathscr{B}}_{\texttt{US}} + \frac{1}{\sqrt{nh}}Q_{I_{\texttt{US}},4}(z) + \sqrt{nh}h^{p+1}\tilde{\mathscr{B}}_{\texttt{US}}Q_{5,I_{\texttt{US}}}(z).$$

This immediately yields the follow result for optimal undersmoothing, analogous to the result for robust bias correction in the paper.

Corollary S.3.1. *Let the conditions of Theorem S.3.1 hold. Then the fastest coverage error decay possible is* $\mathbb{P}[\tau_\nu \in I_{\texttt{US}}(h)] = (1-\alpha) + O\left(n^{-(p+1)/(p+2)}\right)$ *and is attained by choosing* $h \asymp n^{-1/(p+2)}$. *In particular, if* $\mathscr{Q}_{\texttt{US},k} \neq 0, k = 1,2,3$, *the optimal bandwidth is given by*

$$h_{\texttt{US}} = \mathcal{H}_{\texttt{US}} n^{-1/(p+2)}, \qquad \text{with} \qquad \mathcal{H}_{\texttt{US}} = \underset{H>0}{\arg\min}\left|\frac{1}{H}\mathscr{Q}_{I_{\texttt{US}},1} + H^{3+2p}\mathscr{Q}_{I_{\texttt{US}},2} + H^{1+p}\mathscr{Q}_{I_{\texttt{US}},3}\right|.$$

Turning to robust bias correction, we differentiate between the case when $S \geq p+2$, allowing all bias terms to be characterized, and the case when there is not sufficient smoothness to do so.

Theorem S.3.2. (Edgeworth Expansions for $T_{\texttt{RBC}}$) *Suppose Assumptions S.1.1 and S.1.2 hold. Assume* $nh/\log(nh)^{2+\gamma} \to \infty$ *for any* $\gamma > 0$ *and* $\rho = h/b \to \bar{\rho} < \infty$.

(a) If $S \geq p+2$ *and* $\sqrt{nh}h^{p+2}(1+\rho^{-1})\log(nh)^{1+\gamma} \to 0$ *for any* $\gamma > 0$ *then*

$$\sup_{z\in\mathbb{R}} |\mathbb{P}[T_{\texttt{RBC}} < z] - \Phi(z) - \phi(z)\mathscr{E}_{\texttt{RBC}}(z)| = \epsilon_{\texttt{RBC}},$$

where $\epsilon_{\texttt{RBC}} = o((nh)^{-1}) + O(nh^{5+2p+2a} + h^{p+2+a})$ *and*

$$\mathscr{E}_{\texttt{RBC}}(z) = \frac{1}{nh}\phi(z)Q_{I_{\texttt{RBC}},1} + nh^{5+2p}\phi(z)Q_{I_{\texttt{RBC}},2}\tilde{\mathscr{B}}^2_{\texttt{BC}} + h^{p+2}\phi(z)Q_{I_{\texttt{RBC}},3}\tilde{\mathscr{B}}_{\texttt{BC}} + \frac{1}{\sqrt{nh}}Q_{I_{\texttt{RBC}},4}(z) + \sqrt{nh}h^{p+1}\tilde{\mathscr{B}}_{\texttt{BC}}Q_{5,I_{\texttt{RBC}}}(z).$$

(b) If $S \geq p+1$ *and* $\sqrt{nh}h^{p+1}(1+\rho^{-1})\log(nh)^{1+\gamma} \to 0$ *for any* $\gamma > 0$ *then*

$$\sup_{z\in\mathbb{R}}\left|\mathbb{P}[T_{\texttt{RBC}} < z] - \Phi(z) - \frac{1}{nh}\phi(z)Q_{I_{\texttt{RBC}},1} - \frac{1}{\sqrt{nh}}\phi(z)Q_{I_{\texttt{RBC}},4}(z) - \Psi_{T_{\texttt{RBC}}}Q_{5,I_{\texttt{RBC}}}(z)\right| = \epsilon_{\texttt{RBC}},$$

where $\epsilon_{\texttt{RBC}} = o((nh)^{-1}) + O(nh^{3+2p+2a} + h^{p+1+a})$ *and*

$$\begin{aligned}\Psi_{T_{\texttt{RBC}}} = {} & \sqrt{nh}\ \nu!\boldsymbol{e}_\nu'\tilde{\boldsymbol{\Gamma}}^{-1}_{+,p}\mathbb{E}\Big[\Big\{h^{-1}(K_+\boldsymbol{r}_p)(X_{h,i}) - \rho^{p+1}\tilde{\boldsymbol{\Lambda}}_{+,p}\boldsymbol{e}_{p+1}'\tilde{\boldsymbol{\Gamma}}^{-1}_{+,q}b^{-1}(K_+\boldsymbol{r}_{p+1})(X_{b,i})\Big\} \\ & \times (\mu_+(X_i) - \boldsymbol{r}_{p+1}(X_i - c)'\boldsymbol{\beta}_{+,p+1})\Big] \\ & - \sqrt{nh}\ \nu!\boldsymbol{e}_\nu'\tilde{\boldsymbol{\Gamma}}^{-1}_{-,p}\mathbb{E}\Big[\Big\{h^{-1}(K_-\boldsymbol{r}_p)(X_{h,i}) - \rho^{p+1}\tilde{\boldsymbol{\Lambda}}_{-,p}\boldsymbol{e}_{p+1}'\tilde{\boldsymbol{\Gamma}}^{-1}_{-,q}b^{-1}(K_-\boldsymbol{r}_{p+1})(X_{b,i})\Big\} \\ & \times (\mu_-(X_i) - \boldsymbol{r}_{p+1}(X_i - c)'\boldsymbol{\beta}_{-,p+1})\Big],\end{aligned}$$

with $\boldsymbol{\beta}_{+,k}$ *the* $k+1$ *vector with* $(j+1)$ *element equal to* $\mu_+^{(j)}(c)/j!$ *for* $j = 0,1,\ldots,k$ *as long as* $j \leq S$, *and zero otherwise, and similarly for* $\boldsymbol{\beta}_{-,k}$.

S.4. PROOFS FOR MAIN RESULTS

We now present proofs for the main theoretical results. We present details for Theorem S.3.1, as the proof for Theorem S.3.2 is largely similar; a brief discussion is given. We first prove the validity of the expansion, deferring computation of the terms to a subsection below. We will rely on some technical results from the supplement to Calonico, Cattaneo, and Farrell (2019), which in general contains more detailed proofs, though in the context of nonparametric regression rather than RD.

The first step in the proof is to show that

$$\mathbb{P}\left[T_{\mathtt{US}} < z\right] = \mathbb{P}\left[\check{T} < z\right] + o\left((nh)^{-1} + h^{p+1} + nh^{3+2p}\right), \tag{S.4.1}$$

for a smooth function $\check{T} := \check{T}((nh)^{-1/2}\sum_{i=1}^{n}\boldsymbol{Z}_i)$, where $\boldsymbol{Z}_i$ a random vector consisting of functions of the data, that, among other requirements, obeys Cramér's condition under our assumptions.

Define

- $s_n = \sqrt{nh}$.

The t-statistic at hand is

$$T_{\mathtt{US}} = \frac{\sqrt{nh^{1+2\nu}}(\hat{\tau}_{\nu,\mathtt{US}} - \tau_\nu)}{\hat{\mathtt{v}}_{\mathtt{US}}} = \frac{s_n \nu! \boldsymbol{e}_\nu' \left(\boldsymbol{\Gamma}_{+,p}^{-1}\boldsymbol{\Omega}_{+,p}\left(\boldsymbol{Y} - \boldsymbol{R}\boldsymbol{\beta}_{+,p}\right) - \boldsymbol{\Gamma}_{-,p}^{-1}\boldsymbol{\Omega}_{-,p}\left(\boldsymbol{Y} - \boldsymbol{R}\boldsymbol{\beta}_{-,p}\right)\right)/n}{\hat{\mathtt{v}}_{\mathtt{US}}}.$$

The numerator is already a smooth function of well-behaved random variables (obeying Cramér's condition in particular), therefore the difference between $T_{\mathtt{US}}$ and $\check{T}$ lies in the denominator. Recall from (S.2.7) that

$$\hat{\mathscr{V}} = \hat{\mathtt{v}}_{\mathtt{US}}^2 = \frac{h}{n}\nu!^2 \boldsymbol{e}_\nu'\Big(\boldsymbol{\Gamma}_{+,p}^{-1}\boldsymbol{\Omega}_{+,p}\hat{\boldsymbol{\Sigma}}_{+,p}\boldsymbol{\Omega}_{+,p}'\boldsymbol{\Gamma}_{+,p}^{-1} + \boldsymbol{\Gamma}_{-,p}^{-1}\boldsymbol{\Omega}_{-,p}\hat{\boldsymbol{\Sigma}}_{-,p}\boldsymbol{\Omega}_{-,p}'\boldsymbol{\Gamma}_{-,p}^{-1}\Big)\boldsymbol{e}_\nu.$$

As with the numerator, the $\boldsymbol{\Gamma}_{\bullet,p}$ matrices are already in the appropriate form. We must expand the "meat" portions, $h\boldsymbol{\Omega}_{+,p}\hat{\boldsymbol{\Sigma}}_{+,p}\boldsymbol{\Omega}_{+,p}'/n$ and $h\boldsymbol{\Omega}_{-,p}\hat{\boldsymbol{\Sigma}}_{-,p}\boldsymbol{\Omega}_{-,p}'/n$, and their estimated residuals. The expansions for the two, being additive, can be done separately. We state only the "+" terms. Let $\varepsilon_{+,i} = Y_i(1) - \mu_+(X_i)$. Then expand

$$\begin{aligned}
\frac{h}{n}\boldsymbol{\Omega}_{+,p}\hat{\boldsymbol{\Sigma}}_{+,p}\boldsymbol{\Omega}_{+,p}' &= \frac{1}{nh}\sum_{i=1}^{n}(K_+^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})\left(Y_i - \boldsymbol{r}_p(X_i - c)'\hat{\boldsymbol{\beta}}_+\right)^2 \\
&= \frac{1}{nh}\sum_{i=1}^{n}(K_+^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})\Big(\varepsilon_i + \left[\mu_+(X_i) - \boldsymbol{r}_p(X_i - c)'\boldsymbol{\beta}_{+,p}\right] \\
&\qquad\qquad + \boldsymbol{r}_p(X_i - c)'\left[\boldsymbol{\beta}_{+,p} - \hat{\boldsymbol{\beta}}_+\right]\Big)^2 \\
&= \frac{1}{nh}\sum_{i=1}^{n}(K_+^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})\Big(\varepsilon_i + \left[\mu_+(X_i) - \boldsymbol{r}_p(X_i - c)'\boldsymbol{\beta}_{+,p}\right] \\
&\qquad\qquad - \boldsymbol{r}_p(X_{h,i})'\boldsymbol{\Gamma}_{+,p}^{-1}\boldsymbol{\Omega}_{+,p}\left[\boldsymbol{Y} - \boldsymbol{R}\boldsymbol{\beta}_{+,p}\right]/n\Big)^2.
\end{aligned} \tag{S.4.2}$$

Define

$$V_1^+ = \frac{1}{nh}\sum_{i=1}^{n}(K_+^2 \boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})\varepsilon_i^2,$$

$$V_2^+ = \frac{1}{nh}\sum_{i=1}^{n}(K_+^2 \boldsymbol{r}_p\boldsymbol{r}_p'\boldsymbol{r}_p')(X_{h,i})\varepsilon_i\boldsymbol{\Gamma}_{+,p}^{-1}\boldsymbol{\Omega}_{+,p}\left[\boldsymbol{Y}-\boldsymbol{R}\boldsymbol{\beta}_{+,p}\right]/n,$$

$$V_3^+ = \frac{1}{nh}\sum_{i=1}^{n}(K_+^2 \boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})\left[\mu_+(X_i)-\boldsymbol{r}_p(X_i-c)'\boldsymbol{\beta}_{+,p}\right]^2,$$

$$V_4^+ = \frac{1}{nh}\sum_{i=1}^{n}(K_+^2 \boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})\left\{\varepsilon_i\left[\mu_+(X_i)-\boldsymbol{r}_p(X_i-c)'\boldsymbol{\beta}_{+,p}\right]\right\},$$

$$V_5^+ = \frac{1}{nh}\sum_{i=1}^{n}(K_+^2 \boldsymbol{r}_p\boldsymbol{r}_p'\boldsymbol{r}_p')(X_{h,i})\left[\mu_+(X_i)-\boldsymbol{r}_p(X_i-c)'\boldsymbol{\beta}_{+,p}\right]\boldsymbol{\Gamma}_{+,p}^{-1}\boldsymbol{\Omega}_{+,p}\left[\boldsymbol{Y}-\boldsymbol{R}\boldsymbol{\beta}_{+,p}\right]/n,$$

$$V_6^+ = \frac{1}{nh}\sum_{i=1}^{n}(K_+^2 \boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})\big\{\boldsymbol{r}_p(X_{h,i})'\boldsymbol{\Gamma}_{+,p}^{-1}\boldsymbol{\Omega}_{+,p}\left[\boldsymbol{Y}-\boldsymbol{R}\boldsymbol{\beta}_{+,p}\right]/n\big\}^2,$$

and

$$\breve{V}_5^+ = \sum_{l_i=0}^{p}\sum_{l_j=0}^{p}\left[\boldsymbol{\Gamma}_{+,p}^{-1}\right]_{l_i,l_j}\mathbb{E}\left[(K_+^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})(X_{h,i})^{l_i}\left(\mu_+(X_i)-\boldsymbol{r}_p(X_i-c)'\boldsymbol{\beta}_{+,p}\right)\right]$$
$$\times\frac{1}{nh}\sum_{j=1}^{n}\left\{K_+(X_{h,j})(X_{h,j})^{l_j}\left(Y_j-\boldsymbol{r}_p(X_j-c)'\boldsymbol{\beta}_{+,p}\right)\right\},$$

$$\breve{V}_6^+ = \sum_{l_{i_1}=0}^{p}\sum_{l_{i_2}=0}^{p}\sum_{l_{j_1}=0}^{p}\sum_{l_{j_2}=0}^{p}\left[\boldsymbol{\Gamma}_{+,p}^{-1}\right]_{l_{i_1},l_{j_1}}\left[\boldsymbol{\Gamma}_{+,p}^{-1}\right]_{l_{i_2},l_{j_2}}\mathbb{E}\left[h^{-1}(K_+^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})(X_{h,i})^{l_{i_1}+l_{i_2}}\right]$$
$$\times\frac{1}{(nh)^2}\sum_{j=1}^{n}\sum_{k=1}^{n}K_+(X_{h,j})(X_{h,j})^{l_{j_1}}\left(Y_j-\boldsymbol{r}_p(X_j-c)'\boldsymbol{\beta}_{+,p}\right)K_+(X_{h,k})(X_{h,k})^{l_{j_2}}\left(Y_k-\boldsymbol{r}_p(X_k-c)'\boldsymbol{\beta}_{+,p}\right).$$

where $\left[\boldsymbol{\Gamma}_{+,p}^{-1}\right]_{l_i,l_j}$ is the $\{l_i+1,l_j+1\}$ element of $\boldsymbol{\Gamma}_{+,p}^{-1}$, and similarly define the corresponding "$-$" versions of all these.

With these definitions in hand, rewrite $\hat{\mathscr{V}} = \hat{\mathtt{v}}_{\mathtt{US}}^2$ as

$$\hat{\mathtt{v}}_{\mathtt{US}}^2 = \nu!^2\boldsymbol{e}_\nu'\boldsymbol{\Gamma}_{+,p}^{-1}\Big(V_1^+ + 2V_4^+ - 2V_2^+ + V_3^+ - 2V_5^+ + V_6^+\Big)\boldsymbol{\Gamma}_{+,p}^{-1}\boldsymbol{e}_\nu$$
$$+\nu!^2\boldsymbol{e}_\nu'\boldsymbol{\Gamma}_{-,p}^{-1}\Big(V_1^- + 2V_4^- - 2V_2^- + V_3^- - 2V_5^- + V_6^-\Big)\boldsymbol{\Gamma}_{-,p}^{-1}\boldsymbol{e}_\nu$$

and let

$$\breve{\mathtt{v}}_{\mathtt{US}}^2 = \nu!^2\boldsymbol{e}_\nu'\boldsymbol{\Gamma}_{+,p}^{-1}\Big(V_1^+ - 2V_2^+ + 2V_4^+ - 2\breve{V}_5 + \breve{V}_6\Big)\boldsymbol{\Gamma}_{+,p}^{-1}\boldsymbol{e}_\nu$$
$$+\nu!^2\boldsymbol{e}_\nu'\boldsymbol{\Gamma}_{-,p}^{-1}\Big(V_1^- - 2V_2^- + 2V_4^- - 2\breve{V}_5^- + \breve{V}_6^-\Big)\boldsymbol{\Gamma}_{-,p}^{-1}\boldsymbol{e}_\nu.$$

Then, referring back to Equation (S.4.1), we have

$$\mathbb{P}\left[T_{\mathtt{US}} < z\right] = \mathbb{P}\left[\breve{T} + U_n < z\right],$$

with

$$U_n = \left(\hat{\mathtt{v}}_{\mathtt{US}}^{-1} - \check{\mathtt{v}}_{\mathtt{US}}^{-1}\right) s_n \nu! \boldsymbol{e}_\nu' \left(\boldsymbol{\Gamma}_{+,p}^{-1}\boldsymbol{\Omega}_{+,p}\left(\boldsymbol{Y} - \boldsymbol{R}\boldsymbol{\beta}_{+,p}\right) - \boldsymbol{\Gamma}_{-,p}^{-1}\boldsymbol{\Omega}_{-,p}\left(\boldsymbol{Y} - \boldsymbol{R}\boldsymbol{\beta}_{-,p}\right)\right)/n$$

and

$$\check{T} = \check{\mathtt{v}}_{\mathtt{US}}^{-1} s_n \nu! \boldsymbol{e}_\nu' \left(\boldsymbol{\Gamma}_{+,p}^{-1}\boldsymbol{\Omega}_{+,p}\left(\boldsymbol{Y} - \boldsymbol{R}\boldsymbol{\beta}_{+,p}\right) - \boldsymbol{\Gamma}_{-,p}^{-1}\boldsymbol{\Omega}_{-,p}\left(\boldsymbol{Y} - \boldsymbol{R}\boldsymbol{\beta}_{-,p}\right)\right)/n.$$

As required, $\check{T} := \check{T}(s_n^{-1}\sum_{i=1}^n \boldsymbol{Z}_i)$ is a smooth function of the sample average of $\boldsymbol{Z}_i = (\boldsymbol{Z}_i^{+\prime}, \boldsymbol{Z}_i^{-\prime})'$, where $\boldsymbol{Z}_i^+$ is defined as

$$\begin{aligned}
\boldsymbol{Z}_i^+ = \Bigg(&\Big\{(K_+\boldsymbol{r}_p)(X_{h,i})(Y_i - \boldsymbol{r}_p(X_i - c)'\boldsymbol{\beta}_{+,p})\Big\}', \\
&\operatorname{vech}\Big\{(K_+\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})\Big\}', \\
&\operatorname{vech}\Big\{(K_+^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})\varepsilon_{+,i}^2\Big\}', \\
&\operatorname{vech}\Big\{(K_+^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})(X_{h,i})^0\varepsilon_{+,i}\Big\}', \operatorname{vech}\Big\{(K_+^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})(X_{h,i})^1\varepsilon_{+,i}\Big\}', \\
&\quad \operatorname{vech}\Big\{(K_+^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})(X_{h,i})^2\varepsilon_{+,i}\Big\}', \ldots, \operatorname{vech}\Big\{(K_+^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})(X_{h,i})^p\varepsilon_{+,i}\Big\}', \\
&\operatorname{vech}\Big\{(K_+^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})\big\{\varepsilon_{+,i}\big[\mu(X_i) - \boldsymbol{r}_p(X_i - c)'\boldsymbol{\beta}_{+,p}\big]\big\}\Big\}'\Bigg)',
\end{aligned}$$

and $\boldsymbol{Z}_i^-$ is analogous. In order of their listing above, these pieces come from (i) the "score" portion of the numerator, (ii) the "Gram" matrix $\boldsymbol{\Gamma}_{+,p}$, (iii) V_1^+, (iv) V_2^+, and (v) V_4^+. Notice that $\check{V}_5^+$ and $\check{V}_6^+$ do not add any additional elements to $\boldsymbol{Z}_i$.

Equation (S.4.1) now follows from the Delta method for Edgeworth expansions (see Lemma S.II.1 of Calonico, Cattaneo, and Farrell, 2019 and discussion there) if we can show that

$$r_{I_{\mathtt{US}}}^{-1}\mathbb{P}[|U_n| > r_n] = o(1), \tag{S.4.3}$$

where $r_{I_{\mathtt{US}}} = \max\{s_n^{-2}, nh^{3+2p}, h^{p+1}\}$ and $r_n = o(r_{I_{\mathtt{US}}})$.

For a point $\bar{\mathtt{v}}^2 \in [\check{\mathtt{v}}_{\mathtt{US}}^2, \hat{\mathtt{v}}_{\mathtt{US}}^2]$, a Taylor expansion gives

$$\hat{\mathtt{v}}_{\mathtt{US}}^{-1} - \check{\mathtt{v}}_{\mathtt{US}}^{-1} = -\frac{1}{2}\frac{\hat{\mathtt{v}}_{\mathtt{US}}^2 - \check{\mathtt{v}}_{\mathtt{US}}^2}{\check{\mathtt{v}}_{\mathtt{US}}^3} + \frac{3}{8}\frac{\left(\hat{\mathtt{v}}_{\mathtt{US}}^2 - \check{\mathtt{v}}_{\mathtt{US}}^2\right)^2}{\bar{\mathtt{v}}^5}.$$

Therefore, if $\left|\hat{\mathtt{v}}_{\mathtt{US}}^2 - \check{\mathtt{v}}_{\mathtt{US}}^2\right| = o_p(1)$, the result in (S.4.3) will hold once we have shown that

$$\begin{aligned}
&r_{I_{\mathtt{US}}}^{-1}\mathbb{P}\left[\left|\left(\hat{\mathtt{v}}_{\mathtt{US}}^2 - \check{\mathtt{v}}_{\mathtt{US}}^2\right)\left(s_n\nu!\boldsymbol{e}_\nu'\left(\boldsymbol{\Gamma}_{+,p}^{-1}\boldsymbol{\Omega}_{+,p}\left(\boldsymbol{Y} - \boldsymbol{R}\boldsymbol{\beta}_{+,p}\right) - \boldsymbol{\Gamma}_{-,p}^{-1}\boldsymbol{\Omega}_{-,p}\left(\boldsymbol{Y} - \boldsymbol{R}\boldsymbol{\beta}_{-,p}\right)\right)/n\right)\right| > r_n\right] \\
&\quad = r_{I_{\mathtt{US}}}^{-1}\mathbb{P}\Bigg[\Bigg|\left(\nu!^2\boldsymbol{e}_\nu'\boldsymbol{\Gamma}_{+,p}^{-1}\left(V_3^+ - 2[V_5^+ - \check{V}_5^+] + [V_6^+ - \check{V}_6^+]\right)\boldsymbol{\Gamma}_{+,p}^{-1}\boldsymbol{e}_\nu\right) \\
&\qquad\qquad \times \left(\nu!^2\boldsymbol{e}_\nu'\boldsymbol{\Gamma}_{-,p}^{-1}\left(V_3^- - 2[V_5^- - \check{V}_5^-] + [V_6^- - \check{V}_6^-]\right)\boldsymbol{\Gamma}_{-,p}^{-1}\boldsymbol{e}_\nu\right) \\
&\qquad\qquad \times \left(s_n\nu!\boldsymbol{e}_\nu'\left(\boldsymbol{\Gamma}_{+,p}^{-1}\boldsymbol{\Omega}_{+,p}\left(\boldsymbol{Y} - \boldsymbol{R}\boldsymbol{\beta}_{+,p}\right) - \boldsymbol{\Gamma}_{-,p}^{-1}\boldsymbol{\Omega}_{-,p}\left(\boldsymbol{Y} - \boldsymbol{R}\boldsymbol{\beta}_{-,p}\right)\right)/n\right)\Bigg| > r_n\Bigg] \\
&\quad = o(1).
\end{aligned}$$

Recall that $r_{I_{\mathtt{US}}} = \max\{s_n^{-2}, nh^{3+2p}, h^{p+1}\}$ and $r_n = o(r_{I_{\mathtt{US}}})$. The result then follows by the same argument as Section S.II.5.1 of Calonico, Cattaneo, and Farrell (2019); cf. their Equation (S.II.23) and notice that all products of "+" and "−" are zero because of their respective indicator functions.

Thus we have established Equation (S.4.1). Section S.II.5.2 of Calonico, Cattaneo, and Farrell (2019) shows that $\sum_{i=1}^{n} \mathbb{V}[\boldsymbol{Z}_i]^{-1/2}(\boldsymbol{Z}_i - \mathbb{E}[\boldsymbol{Z}_i])/\sqrt{n}$ obeys an Edgeworth expansion. From this, we deduce that $\breve{T} = \breve{T}\left(\mathbb{V}[\boldsymbol{Z}_i]^{1/2} S_n + n\mathbb{E}[\boldsymbol{Z}_i]/s_n\right)$ has its own expansion by Skovgaard (1986), and the result for $T_{\mathtt{US}}$ holds by combining the expansion for $\breve{T}$ with Equation (S.4.1). This completes the proof of Theorem S.3.1. □

Let us turn to Theorem S.3.2. The starting point of the proof is the same as that of Theorem S.3.1: the t-statistic. Looking at the two t-statistics in (S.3.1), and the definitions of the respective point estimators, (S.1.3) and (S.2.2), and standard errors, (S.2.7), we see that the only substantive differences are the matrices $\boldsymbol{\Omega}_{\pm,\bullet}$. The estimated residuals are of the same form as above, with only the bandwidth and polynomial order changed. These changes are reflected in the expansion already. The key is thus to redo the expansion of (S.4.2) with $\boldsymbol{\Omega}_{\pm,\mathtt{BC}}$ in place of $\boldsymbol{\Omega}_{\pm,p}$. The latter lead to the weights $(K_+^2 \boldsymbol{r}_p \boldsymbol{r}_p')(X_{h,i})$, and these are simply replaced by

$$\Big((K_+\boldsymbol{r}_p)(X_{h,i}) - \rho^{p+1}\boldsymbol{\Lambda}_{+,p}\boldsymbol{e}_{p+1}'\boldsymbol{\Gamma}_{+,q}^{-1}(K_+\boldsymbol{r}_{p+1})(X_{b,i})\Big)\Big((K_+\boldsymbol{r}_p)(X_{h,i}) - \rho^{p+1}\boldsymbol{\Lambda}_{+,p}\boldsymbol{e}_{p+1}'\boldsymbol{\Gamma}_{+,q}^{-1}(K_+\boldsymbol{r}_{p+1})(X_{b,i})\Big)'.$$

The same steps are then repeated and hold exactly as before, with the corresponding changes to the rates and terms of the expansion. These are all built into the notation. For more details, see Section S.II.6 of Calonico, Cattaneo, and Farrell (2019). □

S.4.1. Computing the Terms of the Expansion

Computing the terms of the Edgeworth expansions of Theorems S.3.1 and S.3.2, listed in Section S.2.3, is straightforward but tedious. We give a short summary here, following the essential steps of (Hall, 1992, Chapter 2) and Calonico, Cattaneo, and Farrell (2019). In what follows, will always discard higher order terms and write $A \overset{o}{=} B$ to denote $A = B + o((nh)^{-1} + h^{p+1} + nh^{3+2p})$.

We will need much of the notation defined in Section S.2.3. As there, let $\tilde{\boldsymbol{G}}_+$ stand in for $\tilde{\boldsymbol{\Gamma}}_{+,p}$ or $\tilde{\boldsymbol{\Gamma}}_{+,q}$ and similarly for $\tilde{\boldsymbol{G}}_-$, $\tilde{p}$ stand in for p or $p+1$, and d_n stand in for h or b, all depending on if $T = T_{\mathtt{US}}$ or $T_{\mathtt{RBC}}$. Note however, that h is still used in many places, in particular for stabilizing fixed-n expectations, for $T_{\mathtt{RBC}}$. Indexes i, j, and k are always distinct (i.e. $X_{h,i} \neq X_{h,j} \neq X_{h,k}$).

The steps to compute the expansion are as follows. First, we compute a Taylor expansion of T around nonrandom denominators, including both $\hat{\mathrm{v}}^{-1}$ and $\tilde{\boldsymbol{G}}^{-1}$. The cumulants of this linearized version are the approximate cumulants of T itself, which determine the terms of the expansion, as described by Bhattacharya and Rao (1976) and Hall (1992).

It is important to note that the functions $\mathcal{L}_I^0(X_i)$ and $\mathcal{L}_I^1(X_i, X_j)$ already include terms to the left and right of the cutoff. The same is true of

$$\varepsilon_i = \mathbb{1}\{X_i < c\}\varepsilon_{-,i} + \mathbb{1}\{X_i \geq c\}\varepsilon_{+,i}$$
$$v(X_i) = \mathbb{1}\{X_i < c\}\sigma_-^2(X_i) + \mathbb{1}\{X_i \geq c\}\sigma_+^2(X_i).$$

Notice that, because of the indicator functions for each side, products such as $\mathcal{L}_I^0(X_i)^2$ or $\mathcal{L}_I^0(X_i)\mathcal{L}_I^1(X_i, X_j)$ or $\mathcal{L}_I^0(X_i)\varepsilon_i^2$, etc., are always correct.

The Taylor expansion is

$$T \overset{o}{=} \left\{1 - \frac{1}{2\tilde{\mathsf{v}}_I^2}\left(W_{I,1} + W_{I,2} + W_{I,3}\right) + \frac{3}{8\tilde{\mathsf{v}}_I^4}\left(W_{I,1} + W_{I,2} + W_{I,3}\right)^2\right\} \\ \times\ \tilde{\mathsf{v}}_I^{-1}\left\{E_{I,1} + E_{I,2} + E_{I,3} + B_{I,1}\right\},$$

where

$$\begin{aligned}
W_{I,1} &= \frac{1}{nh}\sum_{i=1}^{n}\left\{\mathcal{L}_I^0(X_i)^2\left(\varepsilon_i^2 - v(X_i)\right)\right\} \\
&\quad - 2\frac{1}{n^2h^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\{\mathcal{L}_I^0(X_i)^2\boldsymbol{r}_{\tilde{p}}(X_{d_n,i})'\left(\tilde{\boldsymbol{G}}_+^{-1} + \tilde{\boldsymbol{G}}_-^{-1}\right)((K_+ + K_-)\boldsymbol{r}_{\tilde{p}})(X_{d_n,i})\varepsilon_i\varepsilon_j\right\} \\
&\quad + \frac{1}{n^3h^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\left\{\mathcal{L}_I^0(X_i)^2\boldsymbol{r}_{\tilde{p}}(X_{d_n,i})'\left(\tilde{\boldsymbol{G}}_+^{-1} + \tilde{\boldsymbol{G}}_-^{-1}\right)((K_+ + K_-)\boldsymbol{r}_{\tilde{p}})(X_{d_n,i})\varepsilon_j\varepsilon_k\right\}, \\
W_{I,2} &= \frac{1}{nh}\sum_{i=1}^{n}\left\{\mathcal{L}_I^0(X_i)^2v(X_i)^2 - \mathbb{E}[\mathcal{L}_I^0(X_i)^2v(X_i)^2]\right\} + 2\frac{1}{n^2h^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathcal{L}_I^2(X_i, X_j)\mathcal{L}_I^0(X_i)v(X_i), \\
W_{I,3} &= \frac{1}{n^3h^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\mathcal{L}_I^1(X_i, X_j)\mathcal{L}_I^1(X_i, X_k)v(X_i) + 2\frac{1}{n^3h^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\mathcal{L}_I^2(X_i, X_j, X_k)\mathcal{L}_I^0(X_i)v(X_i), \\
B_{I,1} &= s_n\frac{1}{nh}\sum_{i=1}^{n}\mathcal{L}_I^0(X_i)\left([\mu_+(X_i) - \boldsymbol{r}_{\tilde{p}}(X_i - x)'\boldsymbol{\beta}_{+,\tilde{p}}] - [\mu_-(X_i) - \boldsymbol{r}_{\tilde{p}}(X_i - x)'\boldsymbol{\beta}_{-,\tilde{p}}]\right), \\
E_{I,1} &= s_n\frac{1}{nh}\sum_{i=1}^{n}\mathcal{L}_I^0(X_i)\varepsilon_i, \\
E_{I,2} &= s_n\frac{1}{(nh)^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathcal{L}_I^1(X_i, X_j)\varepsilon_i, \\
E_{I,3} &= s_n\frac{1}{(nh)^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\mathcal{L}_I^2(X_i, X_j, X_k)\varepsilon_i,
\end{aligned}$$

with the final line defining $\mathcal{L}_I^2(X_i, X_j, X_k)$ in the obvious way following $\mathcal{L}_I^1$, i.e. taking account of the next set of remainders. Terms involving $\mathcal{L}_I^2(X_i, X_j, X_k)$ are higher-order, which is why $\mathcal{L}_I^2$ is not needed in Section S.2.3.

Straightforward moment calculations yield, where "$\mathbb{E}[T] \overset{o}{=}$" denotes moments of the Taylor expansion above,

$$\mathbb{E}[T] \overset{o}{=} \tilde{\mathsf{v}}_I^{-1}\mathbb{E}\left[B_{I,1}\right] - \frac{1}{2\tilde{\mathsf{v}}_I^2}\mathbb{E}\left[W_{I,1}E_{I,1}\right],$$

$$\begin{aligned}
\mathbb{E}[T^2] &\overset{o}{=} \frac{1}{\tilde{\mathsf{v}}_I^2}\mathbb{E}\left[E_{I,1}^2 + E_{I,2}^2 + 2E_{I,1}E_{I,2} + 2E_{I,1}E_{I,3}\right] \\
&\quad - \frac{1}{\tilde{\mathsf{v}}_I^4}\mathbb{E}\left[W_{I,1}E_{I,1}^2 + W_{I,2}E_{I,1}^2 + W_{I,3}E_{I,1}^2 + 2W_{I,2}E_{I,1}E_{I,2}\right] \\
&\quad + \frac{1}{\tilde{\mathsf{v}}_I^6}\mathbb{E}\left[W_{I,1}^2E_{I,1}^2 + W_{I,2}^2E_{I,1}^2\right] + \frac{1}{\tilde{\mathsf{v}}_I^2}\mathbb{E}\left[B_{I,1}^2\right] - \frac{1}{\tilde{\mathsf{v}}_I^4}\mathbb{E}\left[W_{I,1}E_{I,1}B_{I,1}\right],
\end{aligned}$$

$$\mathbb{E}[T^3] \overset{o}{=} \frac{1}{\tilde{\mathsf{v}}_I^3}\mathbb{E}\left[E_{I,1}^3\right] - \frac{3}{2\tilde{\mathsf{v}}_I^5}\mathbb{E}\left[W_{I,1}E_{I,1}^3\right] + \frac{3}{\tilde{\mathsf{v}}_I^3}\mathbb{E}\left[E_{I,1}^2 B_{I,1}\right],$$

and

$$\begin{aligned}\mathbb{E}[T^4] \overset{o}{=} & \frac{1}{\tilde{\mathsf{v}}_I^4}\mathbb{E}\left[E_{I,1}^4 + 4E_{I,1}^3E_{I,2} + 4E_{I,1}^3E_{I,3} + 6E_{I,1}^2E_{I,3}^2\right] \\ & - \frac{2}{\tilde{\mathsf{v}}_I^6}\mathbb{E}\left[W_{I,1}E_{I,1}^4 + W_{I,2}E_{I,1}^4 + 4W_{I,2}E_{I,1}^3E_{I,2} + W_{I,3}E_{I,1}\right] \\ & + \frac{3}{\tilde{\mathsf{v}}_I^8}\mathbb{E}\left[W_{I,1}^2E_{I,1}^4 + W_{I,2}^2E_{I,1}^4\right] \\ & + \frac{4}{\tilde{\mathsf{v}}_I^4}\mathbb{E}\left[E_{I,1}^3B_{I,1}\right] - \frac{8}{\tilde{\mathsf{v}}_I^6}\mathbb{E}\left[W_{I,1}E_{I,1}^3B_{I,1}\right] + \frac{6}{\tilde{\mathsf{v}}_I^4}\mathbb{E}\left[E_{I,1}^2B_{I,1}^2\right].\end{aligned}$$

Computing each factor, we get the following results. For these terms below, indexes i, j, and k are always distinct (i.e. $X_{h,i} \neq X_{h,j} \neq X_{h,k}$). First, $\mathbb{E}\left[B_{I,1}\right]$ is simply the fixed-n version of the bias terms.

$$\begin{aligned}\mathbb{E}\left[W_{I,1}E_{I,1}\right] &\overset{o}{=} s_n^{-1}\mathbb{E}\left[h^{-1}\mathcal{L}_I^0(X_i)^3\varepsilon_i^3\right], \\ \mathbb{E}\left[E_{I,1}^2\right] &\overset{o}{=} \tilde{\mathsf{v}}_I^2, \\ \mathbb{E}\left[E_{I,1}E_{I,2}\right] &\overset{o}{=} s_n^{-2}\mathbb{E}\left[h^{-1}\mathcal{L}_I^1(X_i,X_i)\mathcal{L}_I^0(X_i)\varepsilon_i^2\right], \\ \mathbb{E}\left[E_{I,2}^2\right] &\overset{o}{=} s_n^{-1}\mathbb{E}\left[h^{-2}\mathcal{L}_I^1(X_i,X_j)^2\varepsilon_i^2\right], \\ \mathbb{E}\left[E_{I,2}E_{I,3}\right] &\overset{o}{=} s_n^{-2}\mathbb{E}\left[h^{-2}\mathcal{L}_v^2(X_i,X_j,X_j)\mathcal{L}_I^0(X_i)\varepsilon_i^2\right], \\ \mathbb{E}\left[W_{I,1}E_{I,1}^2\right] &\overset{o}{=} s_n^{-2}\Bigg\{\mathbb{E}\left[h^{-1}\mathcal{L}_I^0(X_i)^4\left(\varepsilon_i^4 - v(X_i)^2\right)\right] \\ & - 2\tilde{\mathsf{v}}_I^2\mathbb{E}\left[h^{-1}\mathcal{L}_I^0(X_i)^2\boldsymbol{r}_{\tilde{p}}(X_{d_n,i})'\left(\tilde{\boldsymbol{G}}_+^{-1}+\tilde{\boldsymbol{G}}_-^{-1}\right)\tilde{\boldsymbol{G}}^{-1}((K_+ + K_-)\boldsymbol{r}_{\tilde{p}})(X_{d_n,i})\varepsilon_i^2\right] \\ & - 4\mathbb{E}\left[h^{-1}\mathcal{L}_I^0(X_i)^4\boldsymbol{r}_{\tilde{p}}(X_{d_n,i})'\left(\tilde{\boldsymbol{G}}_+^{-1}+\tilde{\boldsymbol{G}}_-^{-1}\right)\varepsilon_i^2\right]\mathbb{E}\left[h^{-1}((K_+ + K_-)\boldsymbol{r}_{\tilde{p}})(X_{d_n,i})\mathcal{L}_I^0(X_i)\varepsilon_i^2\right] \\ & + \tilde{\mathsf{v}}_I^2\mathbb{E}\left[h^{-2}\mathcal{L}_I^0(X_i)^2\left(\boldsymbol{r}_{\tilde{p}}(X_{d_n,i})'\left(\tilde{\boldsymbol{G}}_+^{-1}+\tilde{\boldsymbol{G}}_-^{-1}\right)((K_+ + K_-)\boldsymbol{r}_{\tilde{p}})(X_{d_n,j})\right)^2\varepsilon_j^2\right] \\ & + 2\mathbb{E}\Bigg[h^{-1}\mathcal{L}_I^0(X_j)^2\Bigg(\mathbb{E}\Big[h^{-1}\boldsymbol{r}_{\tilde{p}}(X_{d_n,j})'\left(\tilde{\boldsymbol{G}}_+^{-1}+\tilde{\boldsymbol{G}}_-^{-1}\right) \\ & \qquad \times ((K_+ + K_-)\boldsymbol{r}_{\tilde{p}})(X_{d_n,i})\mathcal{L}_I^0(X_i)\varepsilon_i^2|X_j\Big]\Bigg)^2\Bigg]\Bigg\}, \\ \mathbb{E}\left[W_{I,2}E_{I,1}^2\right] &\overset{o}{=} s_n^{-2}\Big\{\mathbb{E}\left[h^{-1}\left(\mathcal{L}_I^0(X_i)^2v(X_i) - \mathbb{E}[\mathcal{L}_I^0(X_i)^2v(X_i)]\right)\mathcal{L}_I^0(X_i)^2\varepsilon_i^2\right] \\ & + 2\tilde{\mathsf{v}}_I^2\mathbb{E}\left[h^{-1}\mathcal{L}_I^1(X_i,X_i)\mathcal{L}_I^0(X_i)v(X_i)\right]\Big\}, \\ \mathbb{E}\left[W_{I,2}E_{I,1}E_{I,2}\right] &\overset{o}{=} s_n^{-2}\Big\{\mathbb{E}\left[h^{-2}\left(\mathcal{L}_I^0(X_j)^2v(X_j) - \mathbb{E}[\mathcal{L}_I^0(X_j)^2v(X_j)]\right)\mathcal{L}_I^1(X_i,X_j)\mathcal{L}_I^0(X_i)\varepsilon_i^2\right] \\ & + 2\mathbb{E}\left[h^{-3}\mathcal{L}_I^1(X_i,X_j)\mathcal{L}_I^1(X_k,X_j)\mathcal{L}_I^0(X_i)\mathcal{L}_I^0(X_k)v(X_i)\varepsilon_k^2\right]\Big\}, \\ \mathbb{E}\left[W_{I,3}E_{I,1}^2\right] &\overset{o}{=} s_n^{-2}\Big\{\tilde{\mathsf{v}}_I^2\mathbb{E}\left[h^{-2}\left(\mathcal{L}_I^1(X_i,X_j)^2 + 2\mathcal{L}_I^2(X_i,X_j,X_j)\right)v(X_i)\right]\Big\}, \\ \mathbb{E}\left[W_{I,1}^2E_{I,1}^2\right] &\overset{o}{=} s_n^{-2}\Big\{\tilde{\mathsf{v}}_I^2\mathbb{E}\left[h^{-1}\mathcal{L}_I^0(X_i)^4\left(\varepsilon_i^4 - v(X_i)^2\right)\right] + 2\mathbb{E}\left[h^{-1}\mathcal{L}_I^0(X_i)^3\varepsilon_i^3\right]^2\Big\},\end{aligned}$$

$$
\begin{aligned}
\mathbb{E}\left[W_{I,2}^2 E_{I,1}^2\right] &\overset{o}{=} s_n^{-2}\tilde{\mathrm{v}}_I^2\Big\{\mathbb{E}\left[h^{-1}\left(\mathcal{L}_I^0(X_i)^2 v(X_i) - \mathbb{E}[\mathcal{L}_I^0(X_i)^2 v(X_i)]\right)^2\right] \\
&\quad + 4\mathbb{E}\left[h^{-2}\left(\mathcal{L}_I^0(X_i)^2 v(X_i) - \mathbb{E}[\mathcal{L}_I^0(X_i)^2 v(X_i)]\right)\mathcal{L}_I^1(X_j, X_i)\mathcal{L}_I^0(X_j)v(X_j)\right] \\
&\quad + 4\mathbb{E}\left[h^{-3}\mathcal{L}_I^1(X_i, X_j)\mathcal{L}_I^0(X_i)v(X_i)\mathcal{L}_I^1(X_k, X_j)\mathcal{L}_I^0(X_k)v(X_k)\right]\Big\}, \\
\mathbb{E}\left[W_{I,1}E_{I,1}B_{I,1}\right] &\overset{o}{=} \mathbb{E}\left[W_{I,1}E_{I,1}\right]\mathbb{E}\left[B_{I,1}\right], \\
\mathbb{E}\left[E_{I,1}^3\right] &\overset{o}{=} s_n^{-1}\mathbb{E}\left[h^{-1}\mathcal{L}_I^0(X_i)^3\varepsilon_i^3\right], \\
\mathbb{E}\left[W_{I,1}E_{I,1}^3\right] &\overset{o}{=} \mathbb{E}\left[E_{I,1}^2\right]\mathbb{E}\left[W_{I,1}E_{I,1}\right], \\
\mathbb{E}\left[E_{I,1}^4\right] &\overset{o}{=} 3\tilde{\mathrm{v}}_I^4 + s_n^{-2}\mathbb{E}\left[h^{-1}\mathcal{L}_I^0(X_i)^4\varepsilon_i^3\right], \\
\mathbb{E}\left[E_{I,1}^3 E_{I,2}\right] &\overset{o}{=} s_n^{-2}6\tilde{\mathrm{v}}_I^2\mathbb{E}\left[h^{-1}\mathcal{L}_I^1(X_i, X_i)\mathcal{L}_I^0(X_i)\varepsilon_i^2\right], \\
\mathbb{E}\left[E_{I,1}^3 E_{I,3}\right] &\overset{o}{=} s_n^{-2}3\tilde{\mathrm{v}}_I^2\mathbb{E}\left[h^{-2}\mathcal{L}_I^2(X_i, X_j, X_j)\mathcal{L}_I^0(X_i)\varepsilon_i^2\right], \\
\mathbb{E}\left[E_{I,1}^2 E_{I,2}^2\right] &\overset{o}{=} s_n^{-2}\Big\{\tilde{\mathrm{v}}_I^2\mathbb{E}\left[h^{-2}\mathcal{L}_I^1(X_i, X_j)^2\varepsilon_i^2\right] + 2\mathbb{E}\left[h^{-3}\mathcal{L}_I^1(X_i, X_j)\mathcal{L}_I^1(X_k, X_j)\mathcal{L}_I^0(X_i)\mathcal{L}_I^0(X_k)\varepsilon_i^2\varepsilon_k^2\right]\Big\}, \\
\mathbb{E}\left[W_{I,1}E_{I,1}^4\right] &\overset{o}{=} s_n^{-2}\Big\{\mathbb{E}\left[h^{-1}\mathcal{L}_I^0(X_i)^3\varepsilon_i^3\right]\mathbb{E}\left[h^{-1}\mathcal{L}_I^0(X_i)^3\varepsilon_i^3\right] + 6\mathbb{E}\left[E_{I,1}^2\right]\mathbb{E}\left[W_{I,1}E_{I,1}^2\right]\Big\}, \\
\mathbb{E}\left[W_{I,2}E_{I,1}^4\right] &\overset{o}{=} s_n^{-2}\tilde{\mathrm{v}}_I^2 6\Big\{\mathbb{E}\left[h^{-1}\left(\mathcal{L}_I^0(X_i)^2 v(X_i) - \mathbb{E}[\mathcal{L}_I^0(X_i)^2 v(X_i)]\right)\mathcal{L}_I^0(X_i)^2\varepsilon_i^2\right] \\
&\quad + 2\mathbb{E}\left[h^{-2}\mathcal{L}_I^1(X_i, X_j)\mathcal{L}_I^0(X_i)\mathcal{L}_I^0(X_j)^2\varepsilon_j^2 v(X_i)\right] + \mathbb{E}\left[h^{-1}\mathcal{L}_I^1(X_i, X_i)\mathcal{L}_I^0(X_i)v(X_i)\right]\Big\}, \\
\mathbb{E}\left[W_{I,2}E_{I,1}^3 E_{I,2}\right] &\overset{o}{=} 3\mathbb{E}\left[E_{I,1}^2\right]\mathbb{E}\left[W_{I,2}E_{I,1}E_{I,2}\right], \\
\mathbb{E}\left[W_{I,3}E_{I,1}^4\right] &\overset{o}{=} 3\mathbb{E}\left[E_{I,1}^2\right]\mathbb{E}\left[W_{I,3}E_{I,1}^2\right], \\
\mathbb{E}\left[W_{I,1}^2 E_{I,1}^4\right] &\overset{o}{=} 3\mathbb{E}\left[E_{I,1}^2\right]\mathbb{E}\left[W_{I,1}^2 E_{I,1}^2\right], \\
\mathbb{E}\left[W_{I,2}^2 E_{I,1}^4\right] &\overset{o}{=} 3\mathbb{E}\left[E_{I,1}^2\right]\mathbb{E}\left[W_{I,2}^2 E_{I,1}^2\right].
\end{aligned}
$$

The so-called approximate cumulants of T, denoted here by $\kappa_{I,k}$ for the k^{th} cumulant, can now be directly calculated from these approximate moments using standard formulas, such as Equation (2.6) of Hall (1992) which then become the terms of the expansion. See Hall (1992) for the general case and Calonico, Cattaneo, and Farrell (2018, 2019) in the context of nonparametric regression.

S.5. DETAILS OF PRACTICAL IMPLEMENTATION

We now give details on practical issues that are discussed in the main text. These include the direct plug-in (DPI) rule to implement the coverage-error optimal bandwidth, variance estimation (bias estimation is discussed in Section S.2.1), and the optimal choices ρ^*. These methods are implemented in `R` and `STATA` via the `rdrobust` package, available from `http://sites.google.com/site/rdpackages/rdrobust`.

S.5.1. Bandwidth Choice: Direct Plug-In (DPI)

In order to implement the plug-in bandwidth $\hat{h}_{\texttt{RBC}}$, we always set $K = L$ and $q = p + 1$. The main steps are:

(1) As a pilot bandwidth, use $\hat{h}_{\texttt{MSE}}$: any data-driven version of $h_{\texttt{MSE}}$.

(2) Using this bandwidth, estimate $\hat{\boldsymbol{\beta}}_{+,q}$ and $\hat{\boldsymbol{\beta}}_{-,q}$ on each side of the threshold. Then, form $\hat{\varepsilon}_{+,i} = Y_i - \boldsymbol{r}_q(X_i - c)'\hat{\boldsymbol{\beta}}_{+,q}$ and $\hat{\varepsilon}_{-,i} = Y_i - \boldsymbol{r}_q(X_i - c)'\hat{\boldsymbol{\beta}}_{-,q}$.

(3) Using the pilot bandwidth and a choice of ρ, estimate the terms $\mathscr{Q}_{\texttt{RBC},k}$, $k = 1, 2, 3$. As discussed more just below, from the formulas in Section S.2.3, the estimates are defined by replacing:

   (i) h with $\hat{h}_{\texttt{MSE}}$,
   (ii) population expectations with sample averages,
   (iii) residuals ε_i with $\hat{\varepsilon}_i$, and
   (iv) limiting matrices with the corresponding sample versions using the pilot bandwidth.

(4) To estimate the bias constants $\tilde{\mathscr{B}}_{\texttt{BC}}$, we follow Fan and Gijbels (1996, Section 4.2) and estimate derivatives $\mu^{(p+2)}$ using a global least squares polynomial fit of order $p + 4$ on each side of the threshold.

(5) Finally we obtain:

$$\hat{h}_{\texttt{RBC}} = \hat{\mathscr{H}}\, n^{-1/(3+p)}, \qquad \hat{\mathscr{H}} = \underset{H>0}{\arg\min} \left| \frac{1}{H}\hat{\mathscr{Q}}_{\texttt{RBC},1} + H^{5+2p}\hat{\mathscr{Q}}_{\texttt{RBC},2} + H^{2+p}\hat{\mathscr{Q}}_{\texttt{RBC},3} \right|,$$

Consistency of this bandwidth, meaning $\hat{h}_{\texttt{RBC}}/h_{\texttt{RBC}} \to_{\mathbb{P}} 1$, will follow under natural conditions. In particular, all that is required is consistent estimates for the constants appearing in $\mathscr{Q}_{\texttt{RBC},k}$, $k = 1, 2, 3$, as listed in Section S.2.3. The constants involved are fixed-n computations, and so by "consistent" we mean $\hat{\mathscr{Q}}_{\texttt{RBC},1}/\mathscr{Q}_{\texttt{RBC},k} \to_{\mathbb{P}} 1$. All of the constants involved are kernel-weighted population averages, which may or may not involve $\mu_+(x)$ and $\mu_-(x)$ or their derivatives. Using pilot bandwidths these can be consistently estimated by sample analogues.

For example, the obvious estimator of $\tilde{\boldsymbol{\Gamma}}_{-,p}(h) = \mathbb{E}[h^{-1}(K_-\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})]$ is, for some pilot bandwidth $\bar{h}$, $\boldsymbol{\Gamma}_{-,p}(\bar{h}) = \sum_{i=1}^n \big(K_-\boldsymbol{r}_p\boldsymbol{r}_p'\big)\big((X_i - c)/\bar{h}\big)/n\bar{h}$. If $n\bar{h} \to \infty$, a law of large numbers yields that $\boldsymbol{\Gamma}_{-,p}(\bar{h})$ is consistent for its fixed-n expectation, as in $\boldsymbol{\Gamma}_{-,p}(\bar{h})/\mathbb{E}[\boldsymbol{\Gamma}_{-,p}(\bar{h})] \to_{\mathbb{P}} 1$. If $h \vee \bar{h} \to 0$ then the limits of both fixed-n expectations agree, $\mathbb{E}[\boldsymbol{\Gamma}_{-,p}(\bar{h})]/\tilde{\boldsymbol{\Gamma}}_{-,p}(h) \to 1$. This yields the desired result.

The logic for all the remaining terms is similar, with the possible addition of a consistent estimator for μ_+ or μ_-, and the associated estimated residuals, variances, and biases. These are also easily formed based on pilot bandwidths, for example using rule-of-thumb implementations of the respective MSE-optimal choice for the specific problem. As an example, consider estimating $\mathscr{Q}_{\texttt{RBC},3} = 2\phi(z_{\alpha/2})Q_{\texttt{RBC},3}(z_{\alpha/2})\tilde{\mathscr{B}}_{\texttt{BC}}$. This requires estimates of $Q_{\texttt{RBC},3}(z_{\alpha/2})$ and $\tilde{\mathscr{B}}_{\texttt{BC}}$. The former term is $Q_{\texttt{RBC},3}(z) = \tilde{\mathsf{v}}_I^{-4}\mathbb{E}[h^{-1}\mathcal{L}_I^0(X_i)^3\varepsilon_i^3]\left\{z^3/3\right\}$. First, $\tilde{\mathsf{v}}_I^{-4}$ can be estimated by employing $\hat{\mathsf{v}}_T^2$ following Section S.2.2: all that is required is a pilot bandwidth that delivers consistent estimates of μ_+ and μ_-, for which any ROT MSE choice will do, and estimates of other sample averages, which follow as above and can use the same pilot bandwidth. Notice that $\tilde{\mathsf{v}}_I^2 = \mathbb{E}[h^{-1}\mathcal{L}_I^0(X)^2 v(X)]$, and so if we can estimate this quantity it is obvious that replacing the squaring with cubing estimates the factor $\mathbb{E}[h^{-1}\mathcal{L}_I^0(X_i)^3\varepsilon_i^3]$, and altogether we find that $\hat{Q}_{\texttt{RBC},3}(z_{\alpha/2})(\bar{h})/Q_{\texttt{RBC},3}(z_{\alpha/2})(h) \to_{\mathbb{P}} 1$. Estimation of the bias term follows the same way, and we follow Fan and Gijbels (1996, Section 4.2).

S.5.2. Alternative Standard Errors

We consider two alternative estimates of $\boldsymbol{\Sigma}_+$ and $\boldsymbol{\Sigma}_-$ than those presented in Section S.2.2. First, motivated by the fact that the least-squares residuals are on average too small, we propose HCk heteroskedasticity-consistent estimators; see MacKinnon (2013) for details and a recent review. Calonico, Cattaneo, Farrell, and Titiunik (2019) discuss how they can be applied in the context of local polynomial estimation to construct $\hat{\mathsf{v}}_I^2$-HCk, $k = 0, 1, 2, 3$, where $\hat{\mathsf{v}}_I^2$-HC0 is the original estimator presented above and the others use different weights based on projection matrices.

A second option is to use a nearest-neighbor-based variance estimators with a fixed number of neighbors, following the ideas of Muller and Stadtmuller (1987) and Abadie and Imbens (2008). To define these, let J be a fixed number and $j(i)$ be the j-th closest observation to X_i, $j = 1, \ldots, J$, and set $\hat{\varepsilon}_{+,i} = \mathbb{1}(X_i \geq c)\sqrt{\frac{J}{J+1}}(Y_i - \sum_{j=1}^{J} Y_{j(i)}/J)$, $\hat{\varepsilon}_{-,i} = \mathbb{1}(X_i < c)\sqrt{\frac{J}{J+1}}(Y_i - \sum_{j=1}^{J} Y_{j(i)}/J)$.

As discussed in Calonico, Cattaneo, and Farrell (2018), both types of residual estimators could be handled in our results under natural modifications.

S.5.3. Equivalent Kernels

We discuss how to optimize the asymptotic variance constant featuring the length of the RBC confidence interval estimator using the *equivalent kernel representation* of local polynomials; see Section 3.2.2 of Fan and Gijbels (1996). Detailed derivations are found there.

For simplicity, consider the one-sided bias-corrected estimate of μ_+, i.e., half of $\hat{\tau}_{0,\mathtt{BC}} = \hat{\tau}_0 - h^{p+1}\hat{\mathscr{B}}$. The same of course holds for the "$-$" half of $\hat{\tau}_{0,\mathtt{BC}}$. Recall the definitions in and around (S.2.2) and that $q = p + 1$. Then we consider

$$\hat{\mu}_{+,\mathtt{BC}}^{(0)}(c) = \hat{\mu}_{+,\mathtt{BC}} = \frac{1}{n}\boldsymbol{e}_0'\boldsymbol{\Gamma}_{+,p}^{-1}\boldsymbol{\Omega}_{+,\mathtt{BC}}\boldsymbol{Y} = \frac{1}{n}\boldsymbol{e}_0'\boldsymbol{\Gamma}_{+,p}^{-1}\left(\boldsymbol{\Omega}_{+,p} - \rho^{p+1}\boldsymbol{\Lambda}_{+,p}\boldsymbol{e}_{p+1}'\boldsymbol{\Gamma}_{+,q}^{-1}\boldsymbol{\Omega}_{+,q}\right)\boldsymbol{Y}$$
$$=: \frac{1}{nh}\sum_{i=1}^{n}\mathcal{K}_{+,p}^{\mathtt{BC}}\big(X_{h,i}; K, \rho\big)Y_i,$$

where the last equality defines the weights (recall the definitions of $\boldsymbol{\Omega}_{+,p}$ and $\boldsymbol{\Omega}_{+,q}$)

$$\mathcal{K}_{+,p}^{\mathtt{BC}}\big(x; K, \rho\big) = \boldsymbol{e}_0'\boldsymbol{\Gamma}_{+,p}^{-1}\Big[(K_+\boldsymbol{r}_p)(x) - \rho^{p+2}\boldsymbol{\Lambda}_{+,p}\boldsymbol{e}_{p+1}'\boldsymbol{\Gamma}_{+,q}^{-1}(K_+\boldsymbol{r}_q)(\rho x)\Big].$$

This function depends on the sample through $\boldsymbol{\Gamma}_{+,p}$, $\boldsymbol{\Lambda}_{+,p}$, and $\boldsymbol{\Gamma}_{+,q}$. To find the equivalent kernel, we replace these with their limiting versions. Note that here, as opposed to elsewhere in the paper, we use the population limiting versions, not fixed-n expectations, i.e. we need the *limit* of $\tilde{\boldsymbol{\Gamma}}_{+,p} = \mathbb{E}[\boldsymbol{\Gamma}_{+,p}]$. Under our assumptions, $\boldsymbol{\Gamma}_{+,p} \to_{\mathbb{P}} f(c)\bar{\boldsymbol{\Gamma}}_{+,p}$, $\boldsymbol{\Lambda}_{+,p} \to_{\mathbb{P}} f(c)\bar{\boldsymbol{\Lambda}}_{+,p}$, and $\boldsymbol{\Gamma}_{+,q}^{-1} \to_{\mathbb{P}} f(c)\bar{\boldsymbol{\Gamma}}_{+,q}$, at sufficient fast rates, such that

$$\hat{\mu}_{+,\mathtt{BC}} = \frac{1}{nh}\sum_{i=1}^{n}\bar{\mathcal{K}}_{+,p}^{\mathtt{BC}}\big(X_{h,i}; K, \rho\big)Y_i\ \{1 + o_{\mathbb{P}}(1)\},$$

where the equivalent kernel is

$$\bar{\mathcal{K}}_{+,p}^{\mathtt{BC}}\big(x; K, \rho\big) = \frac{1}{f(c)}\boldsymbol{e}_0'\bar{\boldsymbol{\Gamma}}_{+,p}^{-1}\Big[(K_+\boldsymbol{r}_p)(x) - \rho^{p+2}\bar{\boldsymbol{\Lambda}}_{+,p}\boldsymbol{e}_{p+1}'\bar{\boldsymbol{\Gamma}}_{+,q}^{-1}(K_+\boldsymbol{r}_q)(\rho x)\Big],$$

with

$$\bar{\mathbf{\Gamma}}_{+,p} = \int (K_+ \boldsymbol{r}_p \boldsymbol{r}_p')(u) du, \qquad \bar{\mathbf{\Lambda}}_{+,p} = \int K_+(u) \boldsymbol{r}_p(u) u^{p+1} du \qquad \text{and} \qquad \bar{\mathbf{\Gamma}}_{+,q} = \int (K_+ \boldsymbol{r}_q \boldsymbol{r}_q')(u) du.$$

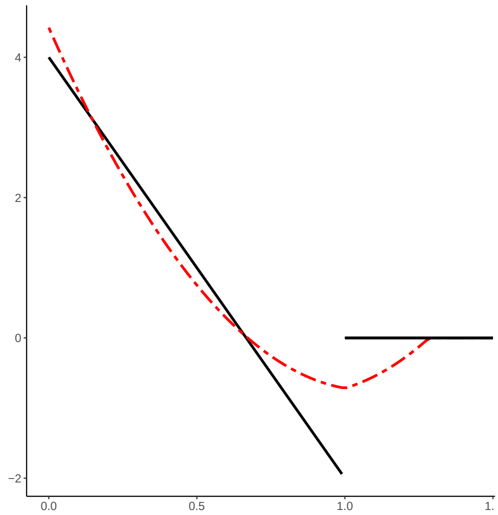
The shape of this equivalent kernel depends on the initial kernel chosen, $K(\cdot)$, and ρ. Cheng, Fan, and Marron (1997) show that the asymptotic variance of a local polynomial point estimator at a boundary point is minimized by employing the uniform kernel $K(u) = \mathbb{1}(|u| \leq 1)$. The resultant equivalent kernel (the "optimal" equivalent kernel) will be denoted $\mathcal{K}^*_{+,p}(x)$ for any p. If the uniform kernel is used when forming $I_{\texttt{RBC}}(h)$, then $\rho = 1$ is optimal in terms of minimizing the asymptotic constant featuring the interval length: that is, $\rho = 1$ makes the induced equivalent kernel, $\bar{\mathcal{K}}^{\texttt{BC}}_{+,p}\big(x; K, \rho\big)$, pointwise equal to the optimal equivalent kernel, $\mathcal{K}^*_{+,p+1}(x)$.

However, if a kernel other than uniform is used, we can find the optimal choice of ρ in terms of minimizing the L_2 distance between the induced equivalent kernel, $\bar{\mathcal{K}}^{\texttt{BC}}_{+,p}\big(x; K, \rho\big)$, and the optimal variance-minimizing equivalent kernel, $\mathcal{K}^*_{+,p+1}(x)$. To be precise, we compute

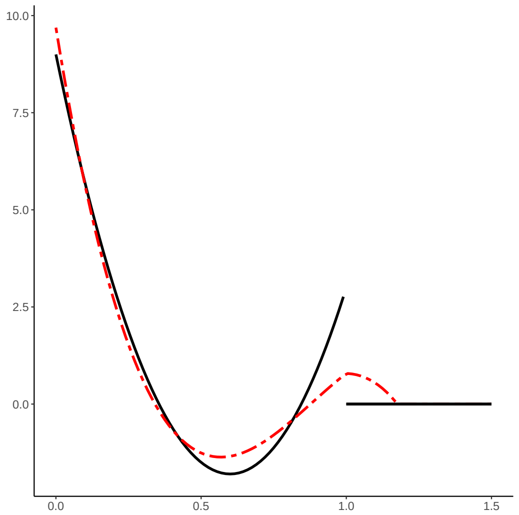
$$\rho^* = \operatorname*{arg\,min}_{\rho > 0} \int \left| \bar{\mathcal{K}}^{\texttt{BC}}_{+,p}\big(x; K, \rho\big) - \mathcal{K}^*_{+,p+1}(x) \right|^2 dx.$$

A common choice is the triangular kernel $K(u) = (1 - |u|)\mathbb{1}(|u| \leq 1)$, which Cheng, Fan, and Marron (1997) show is MSE-optimal (i.e., optimal from a point estimation perspective). We illustrate the shape of the resulting equivalent kernel under the L_2-optimal choice of ρ in Figure S.1 for the triangular bias-corrected equivalent kernel and different choices of p. The corresponding values of ρ^* were given in Table 1 of the paper.

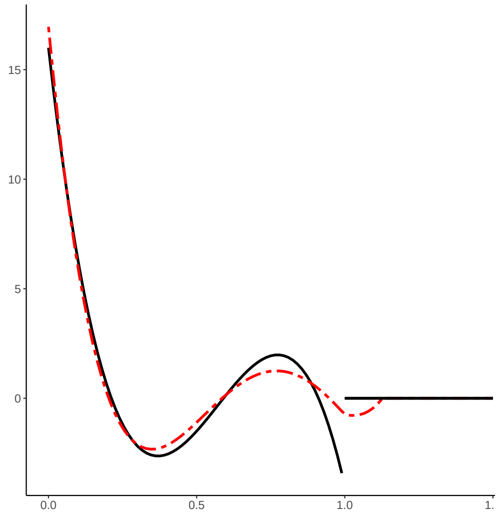
Figure S.1: $\mathcal{K}^*_{+,p+1}(x)$ vs. $\bar{\mathcal{K}}^{\mathtt{BC}}_{+,p}(x; K, \rho^*)$



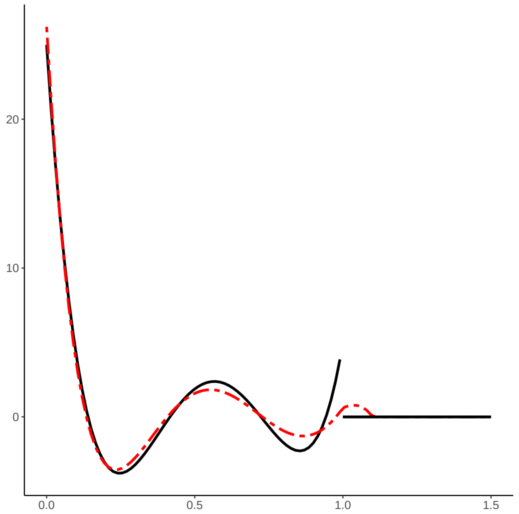
(a) $p = 0$



(b) $p = 1$



(c) $p = 2$



(d) $p = 3$

REFERENCES

Abadie, A. and G. W. Imbens (2008). Estimation of the conditional variance in paired experiments. *Annales d'Economie et de Statistique* (91/92), 175–187.

Bhattacharya, R. N. and R. R. Rao (1976). *Normal Approximation and Asymptotic Expansions*. John Wiley and Sons.

Calonico, S., M. D. Cattaneo, and M. H. Farrell (2018). On the effect of bias estimation on coverage accuracy in nonparametric inference. *Journal of the American Statistical Association 113*(522), 767–779.

Calonico, S., M. D. Cattaneo, and M. H. Farrell (2019). Coverage error optimal confidence intervals for local polynomial regression. arXiv:1808.01398.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2019). Regression discontinuity designs using covariates. *Review of Economics and Statistics 101*(3), 442–451.

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

Cheng, M.-Y., J. Fan, and J. S. Marron (1997). On automatic boundary corrections. *Annals of Statistics 25*(4), 1691–1708.

Fan, J. and I. Gijbels (1996). *Local Polynomial Modelling and Its Applications*. New York: Chapman & Hall/CRC.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. New York: Springer-Verlag.

MacKinnon, J. G. (2013). *Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis*, Chapter Thirty Years of Heteroskedasticity-Robust Inference, pp. 437–461. Springer.

Muller, H.-G. and U. Stadtmuller (1987). Estimation of heteroscedasticity in regression analysis. *Annals of Statistics 15*(2), 610–625.

Skovgaard, I. M. (1986). On multivariate edgeworth expansions. *International Statistical Review 54*(2), 169–186.