Supplement to “Optimal Bandwidth Choice for Robust Bias Corrected Inference in Regression Discontinuity Designs”∗

Sebastian Calonico† Matias D. Cattaneo‡ Max H. Farrell§

September 14, 2018

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.

Contents

S.1 Setup, Notation, and Assumptions 2
  S.1.1 Local Polynomial Point Estimation . . . . . . . . . . . . . . . . . . . . . . . . 2
  S.1.2 Assumptions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3
S.2 Technical Details and Formulas Omitted from the Main Text 4
  S.2.1 Bias and Bias Correction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4
  S.2.2 Variance and Variance Estimators . . . . . . . . . . . . . . . . . . . . . . . . . 7
  S.2.3 Coverage Error Expansion Terms . . . . . . . . . . . . . . . . . . . . . . . . . . 8
S.3 Main Results: Coverage Error and Edgeworth Expansions 10
S.4 Proofs for Main Results 12
  S.4.1 Computing the Terms of the Expansion . . . . . . . . . . . . . . . . . . . . . . 16
S.5 Details of Practical Implementation 19
  S.5.1 Bandwidth Choice: Direct Plug-In (DPI) . . . . . . . . . . . . . . . . . . . . . . 20
  S.5.2 Alternative Standard Errors . . . . . . . . . . . . . . . . . . . . . . . . . . . . 21
  S.5.3 Equivalent Kernels . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 21
S.6 Supplement References 24

∗The second author gratefully acknowledges financial support from the National Science Foundation (SES 1357561 and SES 1459931).
†Department of Economics, University of Miami.
‡Department of Economics and Department of Statistics, University of Michigan.
§Booth School of Business, University of Chicago.
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 \(\nu \geq 0\):

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

Here and elsewhere we drop evaluation points of functions when it causes no confusion. With this notation, \(\tau_0\) corresponds to the standard sharp RD estimand, while \(\tau_1\) denotes the sharp kink RD estimand (up to scale).

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 (2018b,a) for further technical details particularly in the context of Edgeworth expansions.

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

\[
\hat{\beta}_- = \hat{\beta}_-\nu(h) = \arg \min_{b \in \mathbb{R}^{p+1}} \sum_{i=1}^{n} (Y_i - r_p(X_i - c)'b)^2 K_-(X_{h,i}) = \frac{1}{nh^\nu} \Gamma_{-\nu}^{-1} \Omega_{-\nu} Y,
\]

\[
\hat{\beta}_+ = \hat{\beta}_+\nu(h) = \arg \min_{b \in \mathbb{R}^{p+1}} \sum_{i=1}^{n} (Y_i - r_p(X_i - c)'b)^2 K_+(X_{h,i}) = \frac{1}{nh^\nu} \Gamma_{+\nu}^{-1} \Omega_{+\nu} Y,
\]

where:

- \(p\) is an integer greater than \(\min\{1, \nu\}\),
- \(\mathbf{e}_k\) is a conformable zero vector with a one in the \((k+1)\) position, for example \(\mathbf{e}_\nu\) is the \((p+1)\)-vector with a one in the \(\nu^{th}\) position and zeros in the rest,
- \(r_p(u) = (1, u, u^2, \ldots, u^p)'\),

2
• $h$ 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 $c$,

• $K_-(u) = \{u < 0\}K(u)$ and $K_+(u) = \{u \geq 0\}K(u)$ for a kernel function $K(u)$, with in particular $K_-(X_{h,i}) = 1(X_i < c)K(X_{h,i})$ and $K_+(X_{h,i}) = 1(c \leq X_i)K(X_{h,i})$,

• to save space, products of functions will often be written together, with only one argument, for example

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

• $\Gamma_{p,-} = \frac{1}{nh} \sum_{i=1}^{n} (K_-r_p r_p')(X_{h,i})$ and $\Gamma_{p,+} = \frac{1}{nh} \sum_{i=1}^{n} (K_+r_p r_p')(X_{h,i})$,

• $\Omega_{p,-} = h^{-1}[(K_-r_p)(X_{h,1}), \ldots, (K_-r_p)(X_{h,n})]$ and $\Omega_{p,+} = h^{-1}[(K_+r_p)(X_{h,1}), \ldots, (K_+r_p)(X_{h,n})]$,

• $Y = (Y_1, \ldots, Y_n)'$.

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 $\tau_\nu$ of Equation (S.1.1) is then the difference of the appropriate two entries from the one-sided coefficient vectors:

$$\hat{\tau}_\nu = \hat{\tau}_{\nu,\text{US}} = \nu! e'_\nu \hat{\beta}_+ - \nu! e'_\nu \hat{\beta}_- = \frac{1}{nh'} \nu! e'_\nu (\Gamma_{+p}^{-1}\Omega_{+p} - \Gamma_{-p}^{-1}\Omega_{-p}) Y,$$

(S.1.3)

which is also denoted $\hat{\tau}_{\nu,\text{US}}$ 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 $g^{(s)}(x) = \partial^s g(x)/\partial x^s$ for any sufficiently smooth function $g(\cdot)$, with $g(x) = g^{(0)}(x)$ to save notation.

**Assumption 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 $\mu_+^{(S)}(x)$ H"{o}lder continuous with exponent $a \in (0, 1]$, and
\( E[Y_i(t)|X_i = x] \) continuous, for \( t \in \{0,1\} \) and \( \delta > 8 \), with \( \sigma_2^2(x) = \mathbb{V}[Y_i(0)|X_i = x] \) and \( \sigma_4^2(x) = \mathbb{V}[Y_i(1)|X_i = x] \) positive and continuous.

The only difference between this assumption and its counterpart in the main text is that we have defined the function \( \mu_+, \mu_-, \sigma_2^2(x) \), and \( \sigma_4^2(x) \), which we will need later. With this notation the parameter of interest is (cf. (S.1.1))

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

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

\[
\hat{\tau}_\nu = \hat{\tau}_{\nu,US} = \hat{\tau}_\nu(h) = \nu!e_\nu'\hat{\beta}_+ - \nu! e_\nu'\hat{\beta}_- = \hat{\mu}_+^{(\nu)}(c) - \hat{\mu}_-^{(\nu)}(c).
\]

The conditions on the kernel function are as follows.

**Assumption 2 (Kernel).** \( K(u) = \mathbb{1}(u < 0)k(-u) + \mathbb{1}(u \geq 0)k(u) \), where \( k(\cdot) : [0,1] \rightarrow \mathbb{R} \) is bounded and continuous on its support, positive \( \mathbb{0},1 \), zero outside its support, and either is constant or \( (1,K(u)r_3(p+1)(u)) \) is linearly independent on \( \mathbb{0},1 \).

**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 1 and 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 (2018b,a) apply.

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

\[
\hat{\tau}_\nu = \hat{\tau}_\nu(h) = \nu!e_\nu'\hat{\beta}_+ - \nu! e_\nu'\hat{\beta}_- = \frac{1}{nh^p}\nu!e_\nu'\left(\Gamma_{+p,\Omega_{+p}}^{-1} - \Gamma_{-p,\Omega_{-p}}^{-1}\right)Y,
\]

which will also be denoted \( \hat{\tau}_{\nu,US} \) 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

\[
E\left[\hat{\tau}_\nu|X_1, \ldots, X_n\right] - \tau_\nu = h^{p+1-\nu}\mathcal{B} + o_P(h^{p+1-\nu}),
\]

where

\[
\mathcal{B} = \frac{\nu!}{(p+1)!}e_\nu'\left(\Gamma_{+p,\Omega_{+p}}^{-1} - \Gamma_{-p,\Omega_{-p}}^{-1}\right),
\]

\[\text{(S.2.1)}\]
with

- \( \Lambda_{+,p} = \Omega_{+,p} \left[ X_{h,1}^{p+1}, \ldots, X_{h,n}^{p+1} \right] / n \) and similarly for \( \Lambda_{-,p} \), and
- \( \mu^{(p+1)}_+ = \frac{\partial}{\partial x} \mathbb{E}[Y(1)|X_i = x] \big|_{x=c} \), and similarly for \( \mu^{(p+1)}_- \), see Assumption 1.

The bias of (S.2.1) 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 \( \mathcal{B} \) and subtracting this estimate from the point estimator \( \hat{\tau}_\nu \). The estimator of \( \mathcal{B} \) will also be based on one-sided local polynomial regressions, of exactly the same form as (S.1.2) but with one degree higher order polynomial, \( q = p + 1 \) (see Remark 1), and a bandwidth \( b \) defined as \( b = \rho^{-1} h \). Specifically,

\[
\hat{\tau}_{\nu,BC} = \hat{\tau}_\nu - h^{p+1-\nu} \hat{\mathcal{B}} = \frac{1}{nh^{p+1}} e'_\nu \left( \Gamma_{+,p}^{-1} \Omega_{+,BC} - \Gamma_{-,p}^{-1} \Omega_{-,BC} \right) Y,
\]

where

\[
\hat{\mathcal{B}} = \frac{\nu!}{(p+1)!} e'_\nu \left( \Gamma_{+,p}^{-1} \Lambda_{+,p} \hat{\mu}^{(p+1)}_+ - \Gamma_{-,p}^{-1} \Lambda_{-,p} \hat{\mu}^{(p+1)}_- \right),
\]

and

\[
\Omega_{+,BC} = \Omega_{+,p} - \rho^{p+1} \Lambda_{+,p} e'_{p+1} \Gamma_{+,q}^{-1} \Omega_{+,q} \quad \text{and} \quad \Omega_{-,BC} = \Omega_{-,p} - \rho^{p+1} \Lambda_{-,p} e'_{p+1} \Gamma_{-,q}^{-1} \Omega_{-,q}
\]

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

\[
\hat{\mu}^{(p+1)}_+ = \frac{1}{nb^{p+1}} (p+1)! e'_{p+1} \Gamma_{+,q}^{-1} \Omega_{+,q} Y \quad \text{and} \quad \hat{\mu}^{(p+1)}_- = \frac{1}{nb^{p+1}} (p+1)! e'_{p+1} \Gamma_{-,q}^{-1} \Omega_{-,q} Y,
\]

with

- an integer \( q \geq p \) taken throughout to be \( q = p + 1 \) (Calonico, Cattaneo and Farrell (2018b) show why \( q = p + 1 \) is the optimal choice for coverage considerations. See also Remark 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:

- \( r_q(u) = (1, u, u^2, \ldots, u^q)' \),
- \( X_{b,i} = (X_i - c)/b \), for a bandwidth \( b \) and point of interest \( c \),
- \( \Gamma_{-,q} = \frac{1}{nb} \sum_{i=1}^n (K_r r_q')(X_{b,i}) \) and \( \Omega_{-,q} = b^{-1} [(K_r r_q)(X_{b,1}), \ldots, (K_r r_q)(X_{b,n})] \) and similarly for \( \Gamma_{+,q} \) and \( \Omega_{+,q} \).
The bias of $\hat{\tau}_{\nu,BC}$ itself is an important quantity for the coverage error expansions and feasible inference-optimal bandwidth selectors. This is given by

$$
\mathbb{E} [\hat{\tau}_{\nu,BC} | X_1, \ldots, X_n] - \tau_\nu = \begin{cases} 
O(h^{S+a-\nu}) & \text{if } S \leq p + 1 \\
h^{p+2-\nu} \mathcal{B}_{BC} + o_F(h^{p+2-\nu}) & \text{if } S \geq p + 2,
\end{cases}
$$

(5.2.4)

where

$$
\mathcal{B}_{BC} = \frac{\mu_+^{(p+2)}}{(p+2)!} \nu^q e'^q \Gamma_{+,p}^{-1}(\Lambda_{+,p,2} - \rho^{-1} \Lambda_{+,p} e'_{p+1} \Gamma_{+,q}^{-1} \Lambda_{+,q})
$$

$$
- \frac{\mu_-^{(p+2)}}{(p+2)!} \nu^q e'^q \Gamma_{-,p}^{-1}(\Lambda_{-,p,2} - \rho^{-1} \Lambda_{-,p} e'_{p+1} \Gamma_{-,q}^{-1} \Lambda_{-,q})
$$

using the notation

- $\rho = h/b$,
- $\Lambda_{+,p,k} = \Omega_{+,p} \left[ X_{h,1}^{p+k}, \ldots, X_{b,n}^{p+k} \right] / n$, with $\Lambda_{+,p,1} = \Lambda_{+,p}$ in particular, and similarly for $\Lambda_{-,p,k}$, and
- $\Lambda_{+,q,k} = \Omega_{+,q} \left[ X_{b,1}^{q+k}, \ldots, X_{b,n}^{q+k} \right] / n$, with $\Lambda_{+,q,1} = \Lambda_{+,q}$ in particular, and similarly for $\Lambda_{-,q,k}$.

**Remark 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 not improve coverage. The former will tend to inflate variance constants and (to be made feasible) require estimation of higher derivatives, while the latter leads to a slower variance rate. To see why, first, the general form of $\mathcal{B}_{BC}$, provided all derivatives exist (and if they do not, there is even less point to higher $q$) will be

$$
\mathcal{B}_{BC} = \frac{\mu_+^{(p+2)}}{(p+2)!} \nu^q e'^q \Gamma_{+,p}^{-1}(\Lambda_{+,p,2} - \rho^{-1} \Lambda_{+,p} e'_{p+1} \Gamma_{+,q}^{-1} \Lambda_{+,q})
$$

$$
- \frac{\mu_-^{(p+2)}}{(p+2)!} \nu^q e'^q \Gamma_{-,p}^{-1}(\Lambda_{-,p,2} - \rho^{-1} \Lambda_{-,p} e'_{p+1} \Gamma_{-,q}^{-1} \Lambda_{-,q})
$$

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 $\rho$ balances these two.

See Calonico, Cattaneo and Farrell (2018b) 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 matrixes

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

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

$$\mathbb{V} \left[ \hat{\tau}_\nu | X_1, \ldots, X_n \right] = \frac{1}{nh^{1+2\nu}} \nu^\prime,$$

(S.2.5)

$$\nu^\prime = v_{0S}^2 = \frac{h}{n} \nu^2 \epsilon^\prime \left( \Gamma^{-1}_{+,p} \Omega_{+,p} \Sigma_+ \Omega'_{-,p} \Gamma^{-1}_{+,p} + \Gamma^{-1}_{-,p} \Omega_{-,p} \Sigma_- \Omega'_{-,p} \Gamma^{-1}_{-,p} \right) \epsilon^\prime,$$

where we simultaneously define $\nu^\prime$ and $v_{0S}^2$. These are identical, but it will frequently be convenient to write $v_{0S}$ rather than $\nu^\prime$. 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 $\nu^\prime$ instead of $\nu^\prime(h)$.

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

$$\mathbb{V} \left[ \hat{\tau}_{\nu,BC} | X_1, \ldots, X_n \right] = \frac{1}{nh^{1+2\nu}} \nu_{BC}^\prime,$$

(S.2.6)

$$\nu_{BC}^\prime = v_{BC}^2 = \frac{h}{n} \nu^2 \epsilon^\prime \left( \Gamma^{-1}_{+,p} \Omega_{+,BC} \Sigma_+ \Omega'_{+,BC} \Gamma^{-1}_{+,p} \right) \epsilon^\prime,$$

where we simultaneously define $\nu_{BC}$ and $v_{BC}^2$. These are identical, but it will frequently be convenient to write $v_{BC}$ rather than $\nu_{BC}^\prime$. Notice that the only change is replacing $\Omega_{+,BC}$ and $\Omega_{-,BC}$ for $\Omega_{+,p}$ and $\Omega_{-,p}$, as expected from comparing (S.2.2) and (S.1.3).

To estimate these variances we must only estimate the diagonal matrixes $\Sigma_+$ and $\Sigma_-.$ Define

$$\hat{\Sigma}_{+,p} = \text{diag} \left( (Y_i - r_p(X_i - c)'\hat{\beta}_{+,p})^2 : i = 1, \ldots, n \right),$$

$$\hat{\Sigma}_{-,p} = \text{diag} \left( (Y_i - r_p(X_i - c)'\hat{\beta}_{-,p})^2 : i = 1, \ldots, n \right),$$

$$\hat{\Sigma}_{+,BC} = \text{diag} \left( (Y_i - r_q(X_i - c)'\hat{\beta}_{+,q})^2 : i = 1, \ldots, n \right),$$

and

$$\hat{\Sigma}_{-,BC} = \text{diag} \left( (Y_i - r_q(X_i - c)'\hat{\beta}_{-,q})^2 : i = 1, \ldots, n \right),$$

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

\[
\begin{align*}
\hat{\psi} &= \varphi_{\text{US}}^2 = \frac{h}{n} \nu !^2 \nu' \left( \Gamma^-_+ \Omega^-_+ \hat{\Sigma}_+ \Gamma^-_+ + \Gamma^-_\cdot \Omega^-_\cdot \hat{\Sigma}_\cdot \Gamma^-_\cdot \right) e_\nu \\
\hat{\psi}_{\text{BC}} &= \varphi_{\text{BC}}^2 = \frac{h}{n} \nu !^2 \nu' \left( \Gamma^-_+ \Omega^-_+ \Omega^-_\cdot \hat{\Sigma}_+ \Omega^-_\cdot \Gamma^-_+ + \Gamma^-_\cdot \Omega^-_\cdot \hat{\Sigma}_\cdot \Omega^-_\cdot \Gamma^-_\cdot \right) e_\nu 
\end{align*}
\]

(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 (2018b,a).

### S.2.3 Coverage Error Expansion Terms

We now give the precise definition of the terms \(Q_{\text{US},k}\) and \(Q_{\text{BC},k}\), \(k = 1, 2, 3\), 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 \(n, h, b, \nu, p, \) and \(K\), though this is mostly suppressed notionally. These functions are all calculated in a fixed-\(n\) sense and are all bounded and rateless.

\[
\begin{align*}
L^0_{\text{US}}(X_i) &= \nu e_\nu \left\{ \tilde{\Gamma}^-_+ (K_+ r_p)(X_{h,i}) - \tilde{\Gamma}^-_\cdot (K_- r_p)(X_{h,i}) \right\} \\
L^0_{\text{BC}}(X_i) &= \nu e_\nu \left\{ \tilde{\Gamma}^-_+ (K_+ r_p)(X_{h,i}) \right\} + \nu e_\nu \left\{ \tilde{\Gamma}^-_\cdot (K_- r_p)(X_{h,i}) \right\} \\
L^1_{\text{US}}(X_i, X_j) &= \nu e_\nu \left\{ \tilde{\Gamma}^-_+ (K_+ r_p')(X_{h,j}) - \tilde{\Gamma}^-_\cdot (K_- r_p')(X_{h,j}) \right\} \\
L^1_{\text{BC}}(X_i, X_j) &= \nu e_\nu \left\{ \tilde{\Gamma}^-_+ (K_+ r_p')(X_{h,j}) \right\} + \nu e_\nu \left\{ \tilde{\Gamma}^-_\cdot (K_- r_p')(X_{h,j}) \right\} \\
\end{align*}
\]

Further, define

\[
\begin{align*}
\varepsilon_i &= \mathbb{1}\{X_i < c\} \varepsilon_{-,i} + \mathbb{1}\{X_i \geq c\} \varepsilon_{+,i} \\
v(X_i) &= \mathbb{1}\{X_i < c\} \sigma^2(X_i) + \mathbb{1}\{X_i \geq c\} \sigma^2(X_i) 
\end{align*}
\]
Let $I$ ("$I$" for Interval) stand in for either US or RBC.\footnote{More precisely, with this generic "$I$" notation, $I = \text{RBC}$ refers to quantities appearing in $\mathcal{Q}_{\text{RBC},k}$, $k = 1, 2, 3$, i.e. those relevant for $I_{\text{RBC}}$, which include notations with a subscript BC, such as $v_{\text{BC}}$.} Define $\tilde{v}_{I}^2 = \mathbb{E}[h^{-1}\mathcal{L}_{I}^0(X)^2v(X)]$.

Now we define three functions $Q_{\text{US},k}$ and $Q_{\text{RBC},k}$, $k = 1, 2, 3$ which serve as the main building blocks of the terms of the expansion, capturing in particular all dependence on the DGP other than the bias. $Q_{I,1}$ is the most cumbersome notationally. Begin with the others. Define

$$Q_{I,2}(z) = -\tilde{v}_{I}^{-2}\{z/2\}$$

and

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

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

$$Q_{I,1}(z) = \tilde{v}_{I}^{-6}\mathbb{E}\left[h^{-1}\mathcal{L}_{I}^0(X)^3\varepsilon_i^3\right]^2\left\{z^3/3 + 7z/4 + \tilde{v}_{I}^2z(z^2 - 3)/4\right\}$$

$$+ \tilde{v}_{I}^{-2}\mathbb{E}\left[h^{-1}\mathcal{L}_{I}^0(X)\mathcal{L}_{I}^0(X_i)\varepsilon_i^2\right]\{-z(z^2 - 3)/2\}$$

$$+ \tilde{v}_{I}^{-4}\mathbb{E}\left[h^{-1}\mathcal{L}_{I}^0(X)^4(\varepsilon_i^4 - v(X_i)^2)\right]\{z(z^2 - 3)/8\}$$

$$- \tilde{v}_{I}^{-2}\mathbb{E}\left[h^{-1}\mathcal{L}_{I}^0(X)^2\mathcal{G}_{+,1}(K \mathcal{p}_X)(X_{d_n,i})\varepsilon_i^2\right]\{z(z^2 - 1)/2\}$$

$$- \tilde{v}_{I}^{-2}\mathbb{E}\left[h^{-1}\mathcal{L}_{I}^0(X)^2\mathcal{G}_{-,1}(K \mathcal{p}_X)(X_{d_n,i})\varepsilon_i^2\right]\{z(z^2 - 1)/2\}$$

$$- \tilde{v}_{I}^{-4}\mathbb{E}\left[h^{-1}\mathcal{L}_{I}^0(X)^3\mathcal{p}_X(X_{d_n,i})\mathcal{G}_{+,1}(K \mathcal{p}_X)(X_{d_n,i})\varepsilon_i^2\right]\{z(z^2 - 1)\}$$

$$- \tilde{v}_{I}^{-4}\mathbb{E}\left[h^{-1}\mathcal{L}_{I}^0(X)^3\mathcal{p}_X(X_{d_n,i})\mathcal{G}_{-,1}(K \mathcal{p}_X)(X_{d_n,i})\varepsilon_i^2\right]\{z(z^2 - 1)\}$$

$$+ \tilde{v}_{I}^{-2}\mathbb{E}\left[h^{-2}\mathcal{L}_{I}^0(X)^2(\mathcal{p}_X(X_{d_n,i})\mathcal{G}_{+,1}(K \mathcal{p}_X)(X_{d_n,i}))\varepsilon_i^2\right]\{z(z^2 - 1)/4\}$$

$$+ \tilde{v}_{I}^{-2}\mathbb{E}\left[h^{-2}\mathcal{L}_{I}^0(X)^2(\mathcal{p}_X(X_{d_n,i})\mathcal{G}_{-,1}(K \mathcal{p}_X)(X_{d_n,i}))\varepsilon_i^2\right]\{z(z^2 - 1)/4\}$$

$$+ \tilde{v}_{I}^{-4}\mathbb{E}\left[h^{-3}\mathcal{L}_{I}^0(X)^2\mathcal{p}_X(X_{d_n,i})\mathcal{G}_{+,1}(K \mathcal{p}_X)(X_{d_n,i})\mathcal{L}_{I}^0(X)\mathcal{L}_{I}^0(X)\varepsilon_i^2\varepsilon_i^2\right]\{z(z^2 - 1)/2\}$$

$$+ \tilde{v}_{I}^{-4}\mathbb{E}\left[h^{-3}\mathcal{L}_{I}^0(X)^2\mathcal{p}_X(X_{d_n,i})\mathcal{G}_{-,1}(K \mathcal{p}_X)(X_{d_n,i})\mathcal{L}_{I}^0(X)\mathcal{L}_{I}^0(X)\varepsilon_i^2\varepsilon_i^2\right]\{z(z^2 - 1)/2\}$$

$$+ \tilde{v}_{I}^{-4}\mathbb{E}\left[h^{-1}\mathcal{L}_{I}^0(X)^4\varepsilon_i^4\right]\{-z(z^2 - 3)/24\}$$

$$+ \tilde{v}_{I}^{-4}\mathbb{E}\left[h^{-1}(\mathcal{L}_{I}^0(X)^2v(X_i) - \mathbb{E}(\mathcal{L}_{I}^0(X)^2v(X_i))\mathcal{L}_{I}^0(X)^2\varepsilon_i^2\right]\{z(z^2 - 1)/4\}$$

$$+ \tilde{v}_{I}^{-4}\mathbb{E}\left[h^{-2}\mathcal{L}_{I}^0(X)\mathcal{L}_{I}^0(X)\mathcal{L}_{I}^0(X)\mathcal{L}_{I}^0(X)\varepsilon_i^2\varepsilon_i^2\right]\{z(z^2 - 3)\}$$

$$+ \tilde{v}_{I}^{-4}\mathbb{E}\left[h^{-2}\mathcal{L}_{I}^0(X)\mathcal{L}_{I}^0(X)\mathcal{L}_{I}^0(X)\mathcal{L}_{I}^0(X)\varepsilon_i^2\varepsilon_i^2\right]\{-z\}$$
We now state the main theoretical results: Edgeworth expansion for the distributions of the 

t-statistics

\[ T_{US} = \frac{\sqrt{n} h^{1+2\nu} (\bar{\tau}_{\nu,US} - \tau_{\nu})}{\bar{\nu}_{US}} \quad \text{and} \quad T_{RBC} = \frac{\sqrt{n} h^{1+2\nu} (\bar{\tau}_{\nu,BC} - \tau_{\nu})}{\bar{\nu}_{BC}}. \]  

(S.3.1)

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

\[
\mathbb{E} \left[ h^{-3} L_{ij}^0(X_j) r_{\tilde{p}}(X_{d_a,j})^1 \tilde{G}^{-1} (K \tilde{r}_{\tilde{p}})(X_{d_a,i}) L_{ij}^0(X_i) r_{\tilde{p}}(X_{d_a,j})^1 \tilde{G}^{-1} (K \tilde{r}_{\tilde{p}})(X_{d_a,k}) L_{ik}^0(X_k) \varepsilon_i \varepsilon_k \right] \\
= \mathbb{E} \left[ h^{-1} L_{ij}^0(X_j) \varepsilon_i^2 (K \tilde{r}_{\tilde{p}})(X_{d_a,i}) \tilde{G}^{-1} \right] \mathbb{E} \left[ h^{-1} r_{\tilde{p}}(X_{d_a,j}) L_{ij}^0(X_j) r_{\tilde{p}}(X_{d_a,j})^1 \tilde{G}^{-1} \right] \\
\times \mathbb{E} \left[ h^{-1} (K \tilde{r}_{\tilde{p}})(X_{d_a,j}) L_{ik}^0(X_k) \varepsilon_k^2 \right].
\]

This will greatly ease implementation.

The final ingredient require to define the \( \mathcal{D}_{US, k} \) and \( \mathcal{D}_{RBC, k} \) 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

\[
\bar{\mathcal{B}}_{US} = \frac{\nu!}{(p + 1)!} \nu e_{\nu} \left( \tilde{\Gamma}_{+,p} \tilde{\Lambda}_{+,p}^{(p+1)} - \tilde{\Gamma}_{-,p} \tilde{\Lambda}_{-,p}^{(p+1)} \right)
\]

and

\[
\bar{\mathcal{B}}_{BC} = \frac{\mu_{+}^{(p+2)}}{(p + 2)!} \nu e_{\nu} \tilde{\Gamma}_{+,p} \left\{ \tilde{\Lambda}_{+,p,2} - \rho^{-1} \tilde{\Lambda}_{+,p} e_{p+1} \tilde{\Gamma}_{+,q} \tilde{\Lambda}_{+,q} \right\}
\]

\[
- \frac{\mu_{-}^{(p+2)}}{(p + 2)!} \nu e_{\nu} \tilde{\Gamma}_{-,p} \left\{ \tilde{\Lambda}_{-,p,2} - \rho^{-1} \tilde{\Lambda}_{-,p} e_{p+1} \tilde{\Gamma}_{-,q} \tilde{\Lambda}_{-,q} \right\},
\]

where

- \( \tilde{\Gamma}_{+,p} = \mathbb{E}[\Gamma_{+,p}], \tilde{\Lambda}_{+,p} = \mathbb{E}[\Lambda_{+,p}] \), and so forth.

Finally, the \( \mathcal{D}_{US, k} \) and \( \mathcal{D}_{RBC, k} \) terms are defined as follows, where as usual \( I \) stands in for either \( I_{US} \) or \( I_{RBC} \),

\[
\mathcal{D}_{I,1} = 2 \phi(z_{\alpha/2}) Q_{I,1}(z_{\alpha/2})
\]
\[
\mathcal{D}_{I,2} = 2 \phi(z_{\alpha/2}) Q_{I,2}(z_{\alpha/2}) \bar{\mathcal{B}}_{I}^2
\]
\[
\mathcal{D}_{I,3} = 2 \phi(z_{\alpha/2}) Q_{I,3}(z_{\alpha/2}) \bar{\mathcal{B}}_{I}
\]

(S.2.8)

(S.3.1)

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

\[
T_{US} = \frac{\sqrt{n} h^{1+2\nu} (\bar{\tau}_{\nu,US} - \tau_{\nu})}{\bar{\nu}_{US}} \quad \text{and} \quad T_{RBC} = \frac{\sqrt{n} h^{1+2\nu} (\bar{\tau}_{\nu,BC} - \tau_{\nu})}{\bar{\nu}_{BC}}.
\]
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 $Q_{1,k}$, $k = 1, 2, 3$, two other terms appear in the Edgeworth expansion for the $t$-statistic, which then cancel upon computing coverage error due to symmetry. These are

$$Q_{1,4}(z) = \tilde{\nu}_I^{-3}\mathbb{E} \left[h^{-1}L_I^0(X_i)^3\varepsilon_i^3\right] \{2z^2 - 1\} / 6 \quad \text{and} \quad Q_{1,5}(z) = -\tilde{\nu}_I^{-1}.$$ 

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 $T_{US}$ and $T_{RBC}$. For the standard, or undersmoothing, approach, we have the following result.

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

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

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

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

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

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

$$h_{US} = H_{US}n^{-1/(p+2)}, \quad \text{with} \quad H_{US} = \arg\min_{H > 0} \left|\frac{1}{H}Q_{I_{US,1}} + H^{3+2p}Q_{I_{US,2}} + H^{1+p}Q_{I_{US,3}}\right|.$$ 

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 2** (Edgeworth Expansions for $T_{RBC}$). Suppose Assumptions 1 and 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 $\sqrt{nh^{p+2}}(1 + \rho^{-1}) \log(nh)^{1+\gamma} \to 0$ for any $\gamma > 0$ then

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

11
where $\epsilon_{\text{RBC}} = o((n\hbar)^{-1}) + O(n\hbar^{5+2p+2a} + h^{p+2+a})$ and

$$
\mathcal{E}_{\text{RBC}}(z) = \frac{1}{n\hbar} \phi(z)Q_{I_\text{asc},1} + n\hbar^{5+2p}\phi(z)Q_{I_\text{asc},2}\tilde{\mathcal{R}}_{\text{BC}}^2 + h^{p+2}\phi(z)Q_{I_\text{asc},3}\tilde{\mathcal{R}}_{\text{BC}} + \frac{1}{\sqrt{n\hbar}}Q_{I_\text{asc},4}(z) + \sqrt{n\hbar h^{p+1}}\tilde{\mathcal{R}}_{\text{BC}} Q_{5,I_\text{asc}}(z).
$$

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

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

where $\epsilon_{\text{RBC}} = o((n\hbar)^{-1}) + O(n\hbar^{3+2p+2a} + h^{p+1+a})$ and

$$
\Psi_{I_\text{asc}} = \sqrt{n\hbar} \nu! e_{\nu}^p \tilde{\mathcal{R}}_{-p}^{-1} \mathbb{E} \left\{ h^{-1}(K_+ r_p)(X_{h,i}) - \rho^{p+1} \tilde{\Lambda}_{+,p} e_p^{r_{p+1}} \tilde{\mathcal{R}}_{-q}^{-1} h^{-1}(K_+ r_{p+1})(X_{b,i}) \right\}
$$

$$
\times \left( \mu_+(X_i) - r_{p+1}(X_i - c)\beta_{+,p+1} \right)
$$

$$
- \sqrt{n\hbar} \nu! e_{\nu}^p \tilde{\mathcal{R}}_{-p}^{-1} \mathbb{E} \left\{ h^{-1}(K_- r_p)(X_{h,i}) - \rho^{p+1} \tilde{\Lambda}_{-,p} e_p^{r_{p+1}} \tilde{\mathcal{R}}_{-q}^{-1} h^{-1}(K_- r_{p+1})(X_{b,i}) \right\}
$$

$$
\times \left( \mu_-(X_i) - r_{p+1}(X_i - c)\beta_{-,p+1} \right),
$$

with $\beta_{+,k}$ 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 $\beta_{-,k}$.

### S.4 Proofs for Main Results

We now present proofs for the main theoretical results. We present details for Theorem 1, as the proof for Theorem 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 (2018a), 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

$$
\mathbb{P}[T_{\text{us}} < z] = \mathbb{P}[\hat{T} < z] + o\left((n\hbar)^{-1} + h^{p+1} + n\hbar^{3+2p}\right),
$$

(S.4.1)

for a smooth function $\hat{T} := \hat{T}(\mathbf{Z}) = \sum_{i=1}^{n} Z_i$, where $\mathbf{Z}_i$ a random vector consisting of functions of the data, that, among other requirements, obeys Cramér’s condition under our assumptions.

Define

- $s_n = \sqrt{n\hbar}$.
The $t$-statistic at hand is

$$T_{US} = \frac{\sqrt{nh^{1+2v} (\tau_{i,US} - \tau_v)}}{\nu_{US}} = \frac{s_n \nu' e'_{\nu} \left( \Gamma_{+p}^{-1} \Omega_{+p} (Y - R \beta_{+p}) - \Gamma_{-p}^{-1} \Omega_{-p} (Y - R \beta_{-p}) \right)}{\nu_{US}}.$$

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

$$\hat{v} = v_{US}^2 = \frac{h}{n} \nu^2 e'_{\nu} \left( \Gamma_{+p}^{-1} \Omega_{+p} \hat{\Sigma}_{+p} \Omega_{+p}' + \Gamma_{-p}^{-1} \Omega_{-p} \hat{\Sigma}_{-p} \Omega_{-p}' \Gamma_{-p}^{-1} \right) e_{\nu}.$$

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

$$\frac{h}{n} \Omega_{+p} \hat{\Sigma}_{+p} \Omega_{+p}' = \frac{1}{nh} \sum_{i=1}^n (K^2_{+p} r_p r_p') (X_{h, i}) \left( Y_i - r_p (X_i - c)' \hat{\beta}_+ \right)^2$$

$$= \frac{1}{nh} \sum_{i=1}^n (K^2_{+p} r_p r_p') (X_{h, i}) \left( \varepsilon_i + [\mu_+ (X_i) - r_p (X_i - c)' \beta_{+p}] + r_p (X_i - c)' \left[ \beta_{+p} - \hat{\beta}_+ \right] \right)^2$$

$$= \frac{1}{nh} \sum_{i=1}^n (K^2_{+p} r_p r_p') (X_{h, i}) \left( \varepsilon_i + [\mu_+ (X_i) - r_p (X_i - c)' \beta_{+p}] - r_p (X_i - c)' \Omega_{+p}^{-1} \Omega_{+p} [Y - R \beta_{+p}] / n \right)^2.$$

Define

$$V_1^+ = \frac{1}{nh} \sum_{i=1}^n (K^2_{+p} r_p r_p') (X_{h, i}) \varepsilon_i^2,$$

$$V_2^+ = \frac{1}{nh} \sum_{i=1}^n (K^2_{+p} r_p r_p' r_p') (X_{h, i}) \varepsilon_i \Gamma_{+p}^{-1} \Omega_{+p} [Y - R \beta_{+p}] / n,$$

$$V_3^+ = \frac{1}{nh} \sum_{i=1}^n (K^2_{+p} r_p r_p') (X_{h, i}) [\mu_+ (X_i) - r_p (X_i - c)' \beta_{+p}]^2,$$

$$V_4^+ = \frac{1}{nh} \sum_{i=1}^n (K^2_{+p} r_p r_p') (X_{h, i}) \{ \varepsilon_i [\mu_+ (X_i) - r_p (X_i - c)' \beta_{+p}] \},$$

$$V_5^+ = \frac{1}{nh} \sum_{i=1}^n (K^2_{+p} r_p r_p' r_p') (X_{h, i}) [\mu_+ (X_i) - r_p (X_i - c)' \beta_{+p}] \Gamma_{+p}^{-1} \Omega_{+p} [Y - R \beta_{+p}] / n,$$

$$V_6^+ = \frac{1}{nh} \sum_{i=1}^n (K^2_{+p} r_p r_p') (X_{h, i}) \{ r_p (X_{h, i}) \Gamma_{+p}^{-1} \Omega_{+p} [Y - R \beta_{+p}] / n \}^2.$$
and

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

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

where \([\Gamma^{-1}_{+,p}]_{l_i, l_j}\) the \(\{l_i + 1, l_j + 1\}\) element of \(\Gamma^{-1}_{+,p}\), and the corresponding “−” versions of all these.

With these definitions in hand, rewrite \(\hat{\gamma} = \hat{\nu}_{US}^2\) as

\[
\hat{\nu}_{US}^2 = \nu^2 e'_\nu \Gamma^{-1}_{+,p} \left( V_1^+ + 2V_4^+ - 2V_2^+ + V_3^+ - 2V_5^+ + V_6^+ \right) \Gamma^{-1}_{+,p} e_\nu \\
+ \nu^2 e'_\nu \Gamma^{-1}_{-,p} \left( V_1^- + 2V_4^- - 2V_2^- + V_3^- - 2V_5^- + V_6^- \right) \Gamma^{-1}_{-,p} e_\nu
\]

and let

\[
\hat{\nu}_{US}^2 = \nu^2 e'_\nu \Gamma^{-1}_{+,p} \left( V_1^+ - 2V_2^+ + 2V_4^+ - 2V_5^+ + \hat{V}_5^- + \hat{V}_6^- \right) \Gamma^{-1}_{+,p} e_\nu \\
+ \nu^2 e'_\nu \Gamma^{-1}_{-,p} \left( V_1^- - 2V_2^- + 2V_4^- - 2\hat{V}_5^- + \hat{V}_6^- \right) \Gamma^{-1}_{-,p} e_\nu.
\]

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

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

with

\[
U_n = \left( \hat{\nu}_{US}^{-1} - \hat{\nu}_{US}^{-1} \right) s_n \nu' e'_\nu \left( \Gamma^{-1}_{+,p} \Omega_{+,p} (Y - R\beta_{+,p}) - \Gamma^{-1}_{-,p} \Omega_{-,p} (Y - R\beta_{-,p}) \right) / n
\]

and

\[
\hat{T} = \hat{\nu}_{US}^{-1} s_n \nu' e'_\nu \left( \Gamma^{-1}_{+,p} \Omega_{+,p} (Y - R\beta_{+,p}) - \Gamma^{-1}_{-,p} \Omega_{-,p} (Y - R\beta_{-,p}) \right) / n.
\]

As required, \(\hat{T} := \hat{T} (s_n^{-1} \sum_{i=1}^n Z_i)\) is a smooth function of the sample average of \(Z_i = (Z_i^{+\prime}, Z_i^{-\prime})'\),

14
where \( Z_i^+ \) is defined as

\[
Z_i^+ = \left\{ (K + r_p)(X_{h,i})(Y_i - r_p(X_i - c)'\beta_{+,p}) \right\}',
\]

\[
\text{vech} \left\{ (K + r_p r_p')(X_{h,i}) \right\}',
\]

\[
\text{vech} \left\{ (K^2 r_p r_p')(X_{h,i}) \varepsilon^2_{+,i} \right\}',
\]

\[
\text{vech} \left\{ (K^2 r_p r_p')(X_{h,i}) (X_{h,i})^0 \varepsilon_{+,i} \right\}', \text{vech} \left\{ (K^2 r_p r_p')(X_{h,i}) (X_{h,i})^1 \varepsilon_{+,i} \right\}',
\]

\[
\text{vech} \left\{ (K^2 r_p r_p')(X_{h,i}) (X_{h,i})^2 \varepsilon_{+,i} \right\}', \ldots, \text{vech} \left\{ (K^2 r_p r_p')(X_{h,i}) (X_{h,i})^p \varepsilon_{+,i} \right\}',
\]

\[
\text{vech} \left\{ (K^2 r_p r_p')(X_{h,i}) \{ \varepsilon_{+,i} [\mu(X_i) - r_p(X_i - c)'\beta_{+,p}] \} \right\}'.
\]

and \( Z_i^- \) is analogous. In order of their listing above, these pieces come from (i) the “score” portion of the numerator, (ii) the “Gram” matrix \( \Gamma_{+,p} \), (iii) \( V_1^+ \), (iv) \( V_2^+ \), and (v) \( V_4^+ \). Notice that \( \hat{V}_5^+ \) and \( \hat{V}_6^+ \) do not add any additional elements to \( Z_i \).

Equation (S.4.1) now follows from the Delta method for Edgeworth expansions (see Calonico et al., 2018a, Lemma S.II.1 and discussion there), if we can show that

\[
r_{l_{us}}^{-1} \mathbb{P} [ |U_n| > r_n ] = o(1),
\]

where \( r_{l_{us}} = \max \{ s^{-2}, nh^{3+2p}, h^{p+1} \} \) and \( r_n = o(r_{l_{us}}) \).

For a point \( \hat{\nu}^2 \in [\hat{v}_{us}^2, \hat{v}_{us}^2] \), a Taylor expansion gives

\[
\hat{v}_{us}^2 - \hat{v}_{us}^{-1} = -\frac{1}{2} \left( \frac{\hat{v}_{us}^2 - \hat{v}_{us}^2}{\hat{v}_{us}^3} \right) + \frac{3}{8} \left( \frac{\hat{v}_{us}^2 - \hat{v}_{us}^2}{\hat{v}^5} \right).
\]

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

\[
r_{l_{us}}^{-1} \mathbb{P} \left[ \left| \left( s \nu^t e'_p \left( \Gamma_{+,p} - \Omega_{-p} (Y - R \beta_{+,p}) \right) / n \right) \right| > r_n \right] = \left[ \right] = \left[ \right] = o(1).
\]

Recall that \( r_{l_{us}} = \max \{ s^{-2}, nh^{3+2p}, h^{p+1} \} \) and \( r_n = o(r_{l_{us}}) \). The result then follows by the same argument as Section S.II.5.1 of Calonico, Cattaneo and Farrell (2018a); cf. their Equation (S.II.23) and notice that all products of “+” and “−” are zero because of their respective indicator functions.

15
Thus we have established Equation (S.4.1). Section S.II.5.2 of Calonico, Cattaneo and Farrell (2018a) shows that \( \sum_{i=1}^{n} \mathbb{V}[Z_i]^{-1/2}(Z_i - \mathbb{E}[Z_i])/\sqrt{n} \) obeys an Edgeworth expansion. From this, we deduce that \( \bar{T} = \bar{T} (\mathbb{V}[Z_i]^{-1/2} S_{n} + n \mathbb{E}[Z_i]/s_{n}) \) has its own expansion by Skovgaard (1986), and the result for \( T_{\text{US}} \) holds by combining the expansion for \( \bar{T} \) with Equation (S.4.1). This completes the proof of Theorem 1.

Let us turn to Theorem 2. The starting point of the proof is the same as that of Theorem 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 \( \Omega_{\pm,*} \). 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 \( \Omega_{\pm,BC} \) in place of \( \Omega_{\pm,p} \). The latter lead to the weights \( (K_{+} r_{p} r'_{p})(X_{h,i}) \), and these are simply replaced by

\[
\left( (K_{+} r_{p})(X_{h,i}) - \rho^{p+1} A_{+} e'_{p+1} \Gamma^{-1}_{+q} (K_{+} r_{p+1})(X_{b,i}) \right) \left( (K_{+} r_{p})(X_{h,i}) - \rho^{p+1} A_{+} e'_{p+1} \Gamma^{-1}_{+q} (K_{+} r_{p+1})(X_{b,i}) \right)'
\]

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 (2018a).

\[ \square \]

S.4.1 Computing the Terms of the Expansion

Computing the terms of the Edgeworth expansions of Theorems 1 and 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 (2018a). In what follows, will always discard higher order terms and write \( A \overset{\text{def}}{=} B \) 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 \( \tilde{G}_{+} \) stand in for \( \tilde{G}_{+}^{p} \) or \( \tilde{G}_{+,q} \) and similarly for \( \tilde{G}_{-} \). \( \bar{p} \) stand in for \( p \) or \( p + 1 \), and \( d_{n} \) stand in for \( h \) or \( b \), all depending on if \( T = T_{\text{US}} \) or \( T_{\text{BC}} \). Note however, that \( h \) is still used in many places, in particular for stabilizing fixed- \( n \) expectations, for \( T_{\text{BC}} \). 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 \( \hat{\nu}^{-1} \) and \( \tilde{G}^{-1} \). The cumulants of this linearized version are the approximate cumulants of \( T \) itself, which determine the terms of the expansion (Bhattacharya and Rao, 1976; Hall, 1992).

It is important to note that the functions \( \mathcal{L}^{0}_{i}(X_{i}) \) and \( \mathcal{L}^{1}_{i}(X_{i},X_{j}) \) already include terms to the left and right of the cutoff. The same is true of

\[
\varepsilon_{i} = 1\{X_{i} < c\} \varepsilon_{-,i} + 1\{X_{i} \geq c\} \varepsilon_{+,i}
\]

\[
v(X_{i}) = 1\{X_{i} < c\} \sigma^{2}(X_{i}) + 1\{X_{i} \geq c\} \sigma^{2}_{+}(X_{i}).
\]

Notice that, because of the indicator functions for each side, products such as \( \mathcal{L}^{0}_{i}(X_{i})^{2} \) or \( \mathcal{L}^{0}_{i}(X_{i})\mathcal{L}^{1}_{i}(X_{i}, X_{j}) \)
or $L_0^0(X_i)e_i^2$, etc., are always correct.

The Taylor expansion is

$$T = \frac{1}{2\hat{v}_I^2} (W_{I,1} + W_{I,2} + W_{I,3}) + \frac{3}{8\hat{v}_I} (W_{I,1} + W_{I,2} + W_{I,3})^2$$

$$\times \hat{v}_I^{-1} \{ E_{I,1} + E_{I,2} + E_{I,3} + B_{I,1} \},$$

where

$$W_{I,1} = \frac{1}{nh} \sum_{i=1}^n \left\{ L_0^0(X_i)^2 (e_i^2 - v(X_i)) \right\}$$

$$- 2 \frac{1}{n^2h^2} \sum_{i=1}^n \sum_{j=1}^n \left\{ L_0^0(X_i)^2 r_p(X_{d_n,i}) \left( \hat{G}_0^0 + \hat{G}_0^- \right) \left( (K_+ + K_-) r_p \right)(X_{d_n,i})e_i e_j \right\}$$

$$+ \frac{1}{n^3h^3} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \left\{ L_0^0(X_i)^2 r_p(X_{d_n,i}) \left( \hat{G}_0^0 + \hat{G}_0^- \right) \left( (K_+ + K_-) r_p \right)(X_{d_n,i})e_j e_k \right\},$$

$$W_{I,2} = \frac{1}{nh} \sum_{i=1}^n \left\{ L_0^0(X_i)^2 v(X_i)^2 - \mathbb{E}[L_0^0(X_i)^2 v(X_i)^2] \right\} + 2 \frac{1}{n^2h^2} \sum_{i=1}^n \sum_{j=1}^n L_1^0(X_i, X_j) L_0^0(X_i) v(X_i),$$

$$W_{I,3} = \frac{1}{n^3h^3} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n L_1^0(X_i, X_j) L_1^0(X_i, X_k) v(X_i) + 2 \frac{1}{n^3h^3} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n L_2^0(X_i, X_j, X_k) L_0^0(X_i) v(X_i),$$

$$B_{I,1} = s_n \frac{1}{nh} \sum_{i=1}^n L_0^0(X_i) \left( [\mu_+(X_i) - r_\hat{p}(X_i - x)\beta_{+,\hat{p}}] - [\mu_-(X_i) - r_\hat{p}(X_i - x)\beta_{-,\hat{p}}] \right),$$

$$E_{I,1} = s_n \frac{1}{nh} \sum_{i=1}^n L_0^0(X_i) e_i,$$

$$E_{I,2} = s_n \frac{1}{(nh)^2} \sum_{i=1}^n \sum_{j=1}^n L_1^0(X_i, X_j) e_i,$$

$$E_{I,3} = s_n \frac{1}{(nh)^3} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n L_2^0(X_i, X_j, X_k) e_i,$$

with the final line defining $L_2^0(X_i, X_j, X_k)$ in the obvious way following $L_1^0$, i.e. taking account of the next set of remainders. Terms involving $L_2^0(X_i, X_j, X_k)$ are higher-order, which is why it is not needed in Section S.2.3.

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

$$\mathbb{E}[T] = \hat{v}_I^{-1} \mathbb{E}[B_{I,1}] - \frac{1}{2\hat{v}_I^2} \mathbb{E}[W_{I,1} E_{I,1}],$$

$$\mathbb{E}[T^2] = \frac{1}{\hat{v}_I^2} \mathbb{E}[E_{I,1}^2 + E_{I,2}^2 + 2E_{I,1}E_{I,2} + 2E_{I,1}E_{I,3}]$$

17
\[- \frac{1}{\tilde{V}_f} \mathbb{E} \left[ W_{I,1} E_{I,1}^2 + W_{I,2} E_{I,1}^2 + W_{I,3} E_{I,1}^2 + 2W_{I,2} E_{I,1} E_{I,2} \right] \\
+ \frac{1}{\tilde{V}_f} \mathbb{E} \left[ W_{I,2} E_{I,1}^2 + W_{I,2} E_{I,1}^2 \right] + \frac{1}{\tilde{V}_f} \mathbb{E} \left[ B_{I,1}^2 \right] - \frac{1}{\tilde{V}_f} \mathbb{E} \left[ W_{I,1} E_{I,1} B_{I,1} \right], \]

\[\mathbb{E}[T^3] \overset{\circ}{=} \frac{1}{\tilde{V}_f} \mathbb{E} \left[ E_{I,1}^3 \right] - \frac{3}{2\tilde{V}_f^2} \mathbb{E} \left[ W_{I,1} E_{I,1}^3 \right] + \frac{3}{\tilde{V}_f} \mathbb{E} \left[ E_{I,1}^2 B_{I,1} \right], \]

and

\[\mathbb{E}[T^4] \overset{\circ}{=} \frac{1}{\tilde{V}_f} \mathbb{E} \left[ E_{I,1}^4 + 4E_{I,1}^3 E_{I,2} + 4E_{I,1}^3 E_{I,3} + 6E_{I,1}^2 E_{I,1}^2 \right] \\
- \frac{2}{\tilde{V}_f^2} \mathbb{E} \left[ W_{I,1} E_{I,1}^4 + W_{I,2} E_{I,1}^4 + 4W_{I,2} E_{I,1}^3 E_{I,2} + W_{I,3} E_{I,1} \right] \\
+ \frac{3}{\tilde{V}_f^3} \mathbb{E} \left[ W_{I,2} E_{I,1}^4 + W_{I,2} E_{I,1}^4 \right] \\
+ \frac{4}{\tilde{V}_f} \mathbb{E} \left[ E_{I,1}^2 B_{I,1} \right] - \frac{8}{\tilde{V}_f^2} \mathbb{E} \left[ W_{I,1} E_{I,1}^3 B_{I,1} \right] + \frac{6}{\tilde{V}_f} \mathbb{E} \left[ E_{I,1}^2 B_{I,1}^2 \right]. \]

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, \(\mathbb{E}[B_{I,1}]\) is simply the fixed-\(n\) version of the bias terms.

\[\mathbb{E} \left[ W_{I,1} E_{I,1} \right] \overset{\circ}{=} s_n^{-1} \mathbb{E} \left[ h^{-1} L_0^0(X_i) \varepsilon_i \right], \]

\[\mathbb{E} \left[ E_{I,1}^2 \right] \overset{\circ}{=} \tilde{V}_f, \]

\[\mathbb{E} \left[ E_{I,1}^3 \right] \overset{\circ}{=} s_n^{-2} \mathbb{E} \left[ h^{-1} L_1^0(X, X_1) L_0^0(X_i) \varepsilon_i \right], \]

\[\mathbb{E} \left[ E_{I,2}^2 \right] \overset{\circ}{=} s_n^{-1} \mathbb{E} \left[ h^{-1} L_1^1(X, X_1, X_2) \varepsilon_i \right], \]

\[\mathbb{E} \left[ E_{I,2}^3 \right] \overset{\circ}{=} s_n^{-2} \mathbb{E} \left[ h^{-2} L_0^0(X, X_j, X_j) L_0^0(X_i) \varepsilon_i \right], \]

\[\mathbb{E} \left[ W_{I,1}^2 E_{I,1}^2 \right] \overset{\circ}{=} s_n^{-2} \left\{ \mathbb{E} \left[ h^{-1} L_0^0(X_i) \varepsilon_i \right] - \mathbb{E} \left[ h^{-1} L_0^0(X_i)^2 \varepsilon_i \right] \right\}, \]

\[\mathbb{E} \left[ W_{I,2} E_{I,1}^2 \right] \overset{\circ}{=} s_n^{-2} \left\{ \mathbb{E} \left[ h^{-1} L_0^0(X_i)^2 v(X_i) - \mathbb{E}(L_0^0(X_i)^2 v(X_i)) \right] \right\} L_0^0(X_i)^2 \varepsilon_i^2 \]
\[ + 2\sqrt{2}\mathbb{E}\left[h^{-1}\mathcal{L}_1(X_i, X_i)\mathcal{L}_0^3(X_i)v(X_i)\right], \]
\[ \mathbb{E}[W_{I,2}E_{I,1}E_{I,2}] \triangleq s_n^{-2}\left\{ \mathbb{E}\left[h^{-2}\left(\mathcal{L}_0^2(X_j)^2v(X_j) - \mathbb{E}[\mathcal{L}_0^2(X_j)^2v(X_j)]\right)\mathcal{L}_1(X_i, X_i)\mathcal{L}_0^3(X_i)v(X_i)\right] \right. \]
\[ \left. + 2\mathbb{E}\left[h^{-3}\mathcal{L}_1(X_i, X_i)\mathcal{L}_1(X_k, X_j)\mathcal{L}_0^2(X_i)\mathcal{L}_1(X_k)v(X_i)v(X_k)\right] \right\}, \]
\[ \mathbb{E}[W_{I,3}E_{I,1}^2] \triangleq s_n^{-2}\left\{ \sqrt{2}\mathbb{E}\left[h^{-2}\left(\mathcal{L}_1(X_i, X_j)^2 + 2\mathcal{L}_1(X_i, X_j, X_j) v(X_i)\right)\right] \right\}, \]
\[ \mathbb{E}[W_{I,1}E_{I,1}^2] \triangleq s_n^{-2}\left\{ \mathbb{E}\left[h^{-1}\mathcal{L}_0^3(X_i)^4(\varepsilon_i^4 - v(X_i)^2)\right] + 2\mathbb{E}\left[h^{-1}\mathcal{L}_0^3(X_i)^3(\varepsilon_i^3)^2\right] \right\}, \]
\[ \mathbb{E}[W_{I,2}E_{I,1}^2] \triangleq s_n^{-2}\sqrt{2}\mathbb{E}\left[h^{-1}\mathcal{L}_0^3(X_i)^2v(X_i) - \mathbb{E}[\mathcal{L}_0^3(X_i)^2v(X_i)]\right] \left. \right\] \[ + 4\mathbb{E}\left[h^{-2}\left(\mathcal{L}_0^2(X_i)^2v(X_i) - \mathbb{E}[\mathcal{L}_0^2(X_i)^2v(X_i)]\right)\mathcal{L}_1(X_i, X_j)\mathcal{L}_0^3(X_j)v(X_j)\right] \]
\[ + 4\mathbb{E}\left[h^{-3}\mathcal{L}_1(X_i, X_j)\mathcal{L}_0^2(X_i)v(X_i)\mathcal{L}_1(X_k, X_j)\mathcal{L}_0^3(X_k)v(X_k)\right] \right\}, \]
\[ \mathbb{E}[W_{I,1}E_{I,1}E_{I,1}^2] \triangleq \mathbb{E}[W_{I,1}E_{I,1}^2] \mathbb{E}[B_{I,1}], \]
\[ \mathbb{E}[E_{I,1}^3] \triangleq s_n^{-1}\mathbb{E}\left[h^{-1}\mathcal{L}_0^3(X_i)^3(\varepsilon_i^3)^2\right], \]
\[ \mathbb{E}[W_{I,1}E_{I,1}^3] \triangleq \mathbb{E}[E_{I,1}^3] \mathbb{E}[W_{I,1}E_{I,1}], \]
\[ \mathbb{E}[E_{I,1}^4] \triangleq 3\sqrt{2} + s_n^{-2}\mathbb{E}\left[h^{-1}\mathcal{L}_0^3(X_i)^4(\varepsilon_i^4)^2\right], \]
\[ \mathbb{E}[E_{I,1}^3E_{I,2}] \triangleq s_n^{-2}6\sqrt{2}\mathbb{E}\left[h^{-1}\mathcal{L}_1(X_i, X_i)\mathcal{L}_0^4(X_i)(\varepsilon_i^2)^2\right], \]
\[ \mathbb{E}[E_{I,1}^3E_{I,3}] \triangleq s_n^{-2}3\sqrt{2}\mathbb{E}\left[h^{-2}\mathcal{L}_1(X_i, X_j, X_j)\mathcal{L}_0^4(X_i)(\varepsilon_i^2)^2\right], \]
\[ \mathbb{E}[E_{I,1}^2E_{I,2}] \triangleq s_n^{-2}\left\{ \sqrt{2}\mathbb{E}\left[h^{-2}\mathcal{L}_1(X_i, X_j)^2(\varepsilon_i^2)^2 + 2\mathbb{E}\left[h^{-3}\mathcal{L}_1(X_i, X_j)\mathcal{L}_1(X_i, X_j)\mathcal{L}_0^2(X_i)\mathcal{L}_0^3(X_k)\varepsilon_i^2(\varepsilon_k^2)^2\right] \right\}, \]
\[ \mathbb{E}[W_{I,1}E_{I,1}^2] \triangleq s_n^{-2}\left\{ \mathbb{E}\left[h^{-1}\mathcal{L}_0^3(X_i)^3(\varepsilon_i^3)^2\right] \mathbb{E}\left[h^{-1}\mathcal{L}_0^3(X_i)^3(\varepsilon_i^3)^2\right] + 6\mathbb{E}[E_{I,1}^2] \mathbb{E}[W_{I,1}E_{I,1}^2] \right\}, \]
\[ \mathbb{E}[W_{I,2}E_{I,1}^4] \triangleq s_n^{-2}\sqrt{2}6\left\{ \mathbb{E}\left[h^{-1}\left(\mathcal{L}_0^3(X_i)^2v(X_i) - \mathbb{E}[\mathcal{L}_0^3(X_i)^2v(X_i)]\right)\mathcal{L}_0^2(X_i)(\varepsilon_i^2)^2\right] \right. \]
\[ \left. + 2\mathbb{E}\left[h^{-2}\mathcal{L}_1(X_i, X_j)\mathcal{L}_0^2(X_i)\mathcal{L}_0^3(X_i)^2(\varepsilon_i^2)^2v(X_i)\right] + \mathbb{E}\left[h^{-1}\mathcal{L}_1(X_i, X_i)\mathcal{L}_0^3(X_i)v(X_i)\right] \right\}, \]
\[ \mathbb{E}[W_{I,2}E_{I,1}E_{I,2}^2] \triangleq 3\mathbb{E}[E_{I,1}^2] \mathbb{E}[W_{I,2}E_{I,1}E_{I,2}], \]
\[ \mathbb{E}[W_{I,3}E_{I,1}^2] \triangleq 3\mathbb{E}[E_{I,1}^3] \mathbb{E}[W_{I,3}E_{I,1}^2], \]
\[ \mathbb{E}[W_{I,1}E_{I,1}^4] \triangleq 3\mathbb{E}[E_{I,1}^2] \mathbb{E}[W_{I,2}E_{I,1}^2], \]
\[ \mathbb{E}[W_{I,2}E_{I,1}^4] \triangleq 3\mathbb{E}[E_{I,1}^2] \mathbb{E}[W_{I,2}E_{I,1}^2]. \]

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 (Hall, 1992, Equation (2.6)) when then become the terms of the expansion. See Hall (1992) for the general case and Calonico, Cattaneo and Farrell (2018b,a) 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 \( \rho^* \). These methods are
implemented in R and STATA via the \texttt{rdrobust} package, available from \url{http://sites.google.com/site/rdpackages/rdrobust}.

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

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

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

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

3. Using the pilot bandwidth and a choice of $p$, estimate the terms $\hat{Q}_{RBC,k}$, $k = 1, 2, 3$. As discussed more just below, from the formulas in Section S.2.3, the estimates are defined by replacing:
   (i) $h$ with $\hat{h}_{\text{MSE}},$
   (ii) population expectations with sample averages,
   (iii) residuals $\varepsilon_i$ with $\hat{\varepsilon}_i$, and
   (iv) limiting matrices with the corresponding sample versions using the pilot bandwidth.

4. To estimate the bias constants $\hat{B}_{BC}$, 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:

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

Consistency of this bandwidth, meaning $\hat{h}_{RBC}/h_{RBC} \to_{P} 1$, will follow under natural conditions. In particular, all that is required is consistent estimates for the constants appearing $\hat{Q}_{RBC,k}$, $k = 1, 2, 3$, as listed in Section S.2.3. The constants involved are fixed-n computations, and so “consistent” we mean $\hat{Q}_{RBC,1}/Q_{RBC,k} \to_{P} 1$. 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 $\Gamma_{-,p}(h) = \mathbb{E}[h^{-1}(K_{p}r_{p}'(X_{h,i}))]$ is, for some pilot bandwidth $\bar{h}$, $\Gamma_{-,p}(\bar{h}) = \sum_{i=1}^{n} (K_{p}r_{p}'((X_{i} - c)/\bar{h}))/n\bar{h}$. If $n\bar{h} \to \infty$, a law of large numbers yields that $\Gamma_{-,p}(\bar{h})$ is consistent for its fixed-n expectation, as in $\mathbb{E}[\Gamma_{-,p}(\bar{h})] \to_{P} 1$. If $h \lor \bar{h} \to 0$ then the limits of both fixed-n expectations agree, $\mathbb{E}[\Gamma_{-,p}(\bar{h})]/\hat{\Gamma}_{-,p}(h) \to 1$. This yields the desired result.
The logic for all the remaining terms is similar, with the possible addition of a consistent estimator for \( \mu_+ \) or \( \mu_- \), 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 \( \mathcal{Q}_{\text{RBC},3} = 2\phi(z_{\alpha/2})Q_{\text{RBC},3}(z_{\alpha/2})\hat{\beta}_{\text{BC}} \). This requires estimates of \( Q_{\text{RBC},3}(z_{\alpha/2}) \) and \( \hat{\beta}_{\text{BC}} \). The former term is \( Q_{\text{RBC},3}(z) = \tilde{v}_I^{-4}\mathbb{E}[h^{-1}L_I^0(X_i)^3\varepsilon_i^3]\{z^3/3\} \). First, \( \tilde{v}_I^{-4} \) can be estimated by employing \( \tilde{v}_I^2 \) following Section S.2.2, all that is required is a pilot bandwidth that consistently estimates \( \mu_+ \) and \( \mu_- \), 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 \( \tilde{v}_I^2 = \mathbb{E}[h^{-1}L_I^0(X)^2v(X)] \), and so if we can estimate this quantity it is obvious that replacing the squaring with cubing estimates the factor \( \mathbb{E}[h^{-1}L_I^0(X_i)^3\varepsilon_i^3] \), and altogether we find that \( Q_{\text{RBC},3}(z_{\alpha/2})/Q_{\text{RBC},3}(z_{\alpha/2})(h) \to \text{P} 1 \). 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 \( \Sigma_+ \) and \( \Sigma_- \) 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 HC\(k \) heteroskedasticity consistent estimators; see MacKinnon (2013) for details and a recent review. Calonico, Cattaneo, Farrell and Titiunik (2018c) discuss how they can be applied in the context of local polynomial estimation to construct \( \tilde{v}_I^2 \)-HC\( k \), \( k = 0,1,2,3 \), where \( \tilde{v}_I^2 \)-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 \( \hat{e}_{+,i} = 1(X_i \geq c)\sqrt{\frac{I}{J+1}}(Y_i - \sum_{j=1}^J Y_{j(i)}/J) \), \( \hat{e}_{-,i} = 1(X_i < c)\sqrt{\frac{I}{J+1}}(Y_i - \sum_{j=1}^J Y_{j(i)}/J) \).

As discussed in Calonico, Cattaneo and Farrell (2018b), 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 (Fan and Gijbels, 1996, Section 3.2.2). Detailed derivations are found there.

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

\[
\tilde{\mu}_{+,\text{BC}}^{(0)}(c) = \tilde{\mu}_{+,\text{BC}} = \frac{1}{n}e_0'\Gamma_{+,p}^{-1}\Omega_{+,\text{BC}} Y = \frac{1}{n}e_0'\Gamma_{+,p}^{-1}(\Omega_{+,p} - \rho^{p+1}\Lambda_{+,p}e_p'\Gamma_{+,q}^{-1}\Omega_{+,q}) Y
\]

\[
= \frac{1}{nh} \sum_{i=1}^n \kappa_{+,p}^{\text{BC}}(X_{h,i}; K, \rho) Y_i,
\]

21
where the last equality defines the weights (recall the definitions of $\Omega_{+,p}$ and $\Omega_{+,q}$)

$$K_{+,p}^{\text{BC}}(x; K, \rho) = e_0' \Gamma_{+,p}^{-1} (K_+ r_p)(x) - \rho^{p+2} A_{+,p} e_{p+1}' \Gamma_{+,q}^{-1} (K_+ r_q)(\rho x).$$

This function depends on the sample through $\Gamma_{+,p}$, $A_{+,p}$, and $\Gamma_{+,q}$. 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 $\bar{\Gamma}_{+,p} = \mathbb{E}[\Gamma_{+,p}]$. Under our assumptions, $\Gamma_{+,p} \to P f(c) \bar{\Gamma}_{+,p}$, $A_{+,p} \to P f(c) \bar{\Lambda}_{+,p}$, and $\Gamma_{+,q}^{-1} \to P f(c) \bar{\Gamma}_{+,q}$, at sufficient fast rates, such that

$$\hat{\mu}_{+,\text{BC}} = \frac{1}{nh} \sum_{i=1}^n \bar{K}_{+,p}^{\text{BC}}(X_{h,i}; K, \rho) Y_i \{1 + o_P(1)\},$$

where the equivalent kernel is

$$\bar{K}_{+,p}^{\text{BC}}(x; K, \rho) = \frac{1}{f(c)} e_0' \bar{\Gamma}_{+,p}^{-1} (K_+ r_p)(x) - \rho^{p+2} \bar{\Lambda}_{+,p} e_{p+1}' \bar{\Gamma}_{+,q}^{-1} (K_+ r_q)(\rho x),$$

with

$$\bar{\Gamma}_{+,p} = \int (K_+ r_p r_p')(u) du, \quad \bar{\Lambda}_{+,p} = \int K_+(u) r_p(u) u^{p+1} du \quad \text{and} \quad \bar{\Gamma}_{+,q} = \int (K_+ r_q r_q')(u) du.$$

The shape of this equivalent kernel depends on the initial kernel chosen, $K(\cdot)$, and $\rho$. 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 $K(u) = \mathbb{I}(|u| \leq 1)$. The resultant equivalent kernel (the “optimal” equivalent kernel) will be denoted $K_{+,p}^{*}(x)$ for any $p$. If the uniform kernel is used when forming $I_{\text{RBC}}(h)$, 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, $\bar{K}_{+,p}^{\text{BC}}(x; K, \rho)$, pointwise equal to the optimal equivalent kernel, $K_{+,p+1}^{*}(x)$.

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

$$\rho^* = \arg \min_{\rho > 0} \int \left| \bar{K}_{+,p}^{\text{BC}}(x; K, \rho) - K_{+,p+1}^{*}(x) \right|^2 dx.$$

A common choice is the triangular kernel $K(u) = (1 - |u|)\mathbb{I}(|u| \leq 1)$ is used for $I_{\text{RBC}}$, 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 $\rho$ in Figure S.1 for the triangular bias-corrected equivalent kernel and different choices of $p$. The corresponding values of $\rho^*$ were given in Table 1 of the paper.
Figure S.1: $K^*_+,p+1(x)$ vs. $\hat{K}^{BC}_+,p(x; K, \rho^*)$
S.6 Supplement References


