Supplement to "Coverage Error Optimal Confidence Intervals for Local Polynomial Regression"

Sebastian Calonico[*] Matias D. Cattaneo[†] Max H. Farrell[‡]

May 28, 2020

This supplement contains proofs of all results, other technical details, and complete simulation results. Notation is kept mostly consistent with the main text, but this document is self-contained as all notation is redefined and all necessary constructions, assumptions, and so forth, are restated. Throughout, clarity is prized over brevity, and repetition is not avoided.

The main result is Theorem S.1, stated in Section S.3. Before that: a complete formalization of the set up and inference procedures is given in Section S.1 and a bias and smoothness issues are discussed thoroughly in S.2, including detail omitted from the main paper. Section S.3 contains theoretical results, which proven in subsequent sections. A few other issues are then discussed. Section S.9 presents complete simulations results and computations. For reference a complete list of notation is given in Section S.10.

Contents

[*]Department of Health Policy and Management, Columbia University.
[†]Department of Operations Research and Financial Engineering, Princeton University.
[‡]Booth School of Business, University of Chicago.

S.1 Setup

We observe a random sample $\{(Y_1, X_1), \ldots, (Y_n, X_n)\}$ from the pair (Y, X), which are distributed according to F, the data-generating process. F is assumed to belong to a class $\mathscr{F}_S$, as defined by Assumption S.1 below, and in particular the pair (Y, X) obeys the heteroskedastic nonparametric regression model

$$Y = \mu_F(X) + \varepsilon, \qquad \mathbb{E}[\varepsilon|X] = 0, \qquad \mathbb{E}[\varepsilon^2|X = x] = v(x). \tag{S.1}$$

The parameter of interest is a derivative of the regression function, defined as

$$\mu^{(\nu)} = \mu_F^{(\nu)}(\mathsf{x}) := \frac{\partial^\nu}{\partial x^\nu}\mathbb{E}_F\left[Y \mid X = x\right]\bigg|_{x=\mathsf{x}}, \tag{S.2}$$

for a point x in the support of X and an nonnegative integer $\nu \leq S$, the latter defined in Assumption S.1, and indexing the class $\mathscr{F}_S$. As usual, we use the notation $\mu_F(\mathsf{x}) = \mu_F^{(0)}(\mathsf{x}) = \mathbb{E}_F[Y \mid X = \mathsf{x}]$.

Expectations and probability statements, as well as parameters and functions, are always understood to depend on F, though for simplicity this will often be omitted when doing so causes no confusion. Similarly, unless it is explicitly required, we will omit the point of evaluation x as an argument. For example,

$$\mu_F^{(\nu)}(\mathsf{x}) = \mu^{(\nu)}(\mathsf{x}) = \mu^{(\nu)}.$$

We study the coverage error of commonly-used Wald-type confidence interval estimators given generically by

$$I = \left[\hat{\theta} - z_u\, \hat{\vartheta}\ ,\ \hat{\theta} - z_l\, \hat{\vartheta}\right], \tag{S.3}$$

for a centering estimator $\hat{\theta}$, scale estimator $\hat{\vartheta}$, and a pair of quantiles z_l and z_u. Our main theoretical result, Theorem 3 of the main text and Theorem S.1 herein, is an Edgeworth expansion for the dual t-statistics of such I that hold uniformly over the class of data-generating processes $\mathscr{F}_S$, defined by Assumption S.1 below.

S.1.1 Centering Estimators

We now define the centering estimators $\hat{\theta}$. These are based on local polynomial regressions. The standard local polynomial (of degree p) point estimator is defined via the local regression

$$\hat{\mu}_p^{(\nu)} = \nu! \boldsymbol{e}_\nu' \hat{\boldsymbol{\beta}}_p = \frac{1}{nh^\nu}\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\boldsymbol{Y}, \qquad \hat{\boldsymbol{\beta}}_p = \underset{\boldsymbol{\beta}\in\mathbb{R}^{p+1}}{\arg\min}\sum_{i=1}^n (Y_i - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta})^2 K\left(X_{h,i}\right), \tag{S.4}$$

where

- $\boldsymbol{e}_k$ 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,
- h is a positive bandwidth sequence that vanishes as n diverges,

- p is an integer greater at least ν, sometimes restricted such that $p-\nu$ odd,
- $\boldsymbol{r}_p(u) = (1, u, u^2, \dots, u^p)'$,
- $X_{h,i} = (X_i - \mathsf{x})/h$, for a bandwidth h and point of interest x,
- to save space, products of functions will often be written together, with only one argument, for example,

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

- $\boldsymbol{W} = \operatorname{diag}\left(h^{-1}K(X_{h,i}) : i = 1, \dots, n\right)$,
- $\boldsymbol{H} = \operatorname{diag}\left(1, h, h^2, \dots, h^p\right)$, where
- $\operatorname{diag}(a_i : i = 1, \dots, k)$ denote the $k \times k$ diagonal matrix constructed using the elements $a_1, a_2, \cdots, a_k$,
- $\boldsymbol{R} = [\boldsymbol{r}_p(X_1 - \mathsf{x}), \cdots, \boldsymbol{r}_p(X_n - \mathsf{x})]'$,
- $\check{\boldsymbol{R}} = \boldsymbol{R}\boldsymbol{H}^{-1} = [\boldsymbol{r}_p(X_{h,1}), \cdots, \boldsymbol{r}_p(X_{h,n})]'$,
- $\boldsymbol{\Gamma} = \frac{1}{nh}\sum_{i=1}^{n}(K\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i}) = (\check{\boldsymbol{R}}'\boldsymbol{W}\check{\boldsymbol{R}})/n$,
- $\boldsymbol{\Omega} = h^{-1}[(K\boldsymbol{r}_p)(X_{h,1}), (K\boldsymbol{r}_p)(X_{h,2}), \dots, (K\boldsymbol{r}_p)(X_{h,n})] = \check{\boldsymbol{R}}'\boldsymbol{W}$, and
- $\boldsymbol{Y} = (Y_1, \dots, Y_n)'$.

We will also use, for bias correction,

- $\hat{\boldsymbol{\beta}}_{p+1}$ which is defined exactly as in Equation (S.4) but with $p+1$ in place of p and b in place of h in all instances.

For more details on local polynomial methods and related theoretical results, see Fan and Gijbels (1996).

For computing the rate of convergence, and clarifying the appearance of $(nh^\nu)^{-1}$ in Equation (S.4), it is useful to spell out the form of $\hat{\boldsymbol{\beta}}_p$, the solution to the minimization in Equation (S.4). Standard least squares algebra yields

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_p &= \left(\boldsymbol{R}'\boldsymbol{W}\boldsymbol{R}\right)^{-1}\boldsymbol{R}'\boldsymbol{W}\boldsymbol{Y} \\
&= \left(\left[\boldsymbol{R}\boldsymbol{H}^{-1}\boldsymbol{H}\right]'\boldsymbol{W}\left[\boldsymbol{R}\boldsymbol{H}^{-1}\boldsymbol{H}\right]\right)^{-1}\left[\boldsymbol{R}\boldsymbol{H}^{-1}\boldsymbol{H}\right]'\boldsymbol{W}\boldsymbol{Y} \\
&= \boldsymbol{H}^{-1}\left(\check{\boldsymbol{R}}'\boldsymbol{W}\check{\boldsymbol{R}}\right)^{-1}\boldsymbol{H}^{-1}\boldsymbol{H}\check{\boldsymbol{R}}'\boldsymbol{W}\boldsymbol{Y} \\
&= \boldsymbol{H}^{-1}\left(\check{\boldsymbol{R}}'\boldsymbol{W}\check{\boldsymbol{R}}\right)^{-1}\check{\boldsymbol{R}}'\boldsymbol{W}\boldsymbol{Y}, \\
&= \boldsymbol{H}^{-1}\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\boldsymbol{Y}/n,
\end{aligned} \tag{S.5}$$

and therefore, because $\boldsymbol{e}_\nu'\boldsymbol{H}^{-1} = \boldsymbol{e}_\nu' h^{-\nu}$,

$$\nu!\boldsymbol{e}_\nu'\hat{\boldsymbol{\beta}}_p = \frac{1}{nh^\nu}\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\boldsymbol{Y}. \tag{S.6}$$

The same applies to $\hat{\boldsymbol{\beta}}_{p+1}$ with the necessary changes to the bandwidth and dimensions.

To conduct valid inference on $\mu^{(\nu)}$ the bias of the nonparametric estimator must be removed. Assuming that the true $\mu^{(\nu)}(\cdot)$ is smooth enough at x (formally, $p+1 \leq S$, such as is required for computing the mean square error optimal bandwidth), we find that the (conditional) bias of $\hat{\mu}_p^{(\nu)}$ is

$$\mathbb{E}\left[\hat{\mu}_p^{(\nu)} \big| X_1, \ldots, X_n\right] - \mu^{(\nu)} = h^{p+1-\nu} \nu! \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1} \boldsymbol{\Lambda}_1 \frac{\mu^{(p+1)}}{(p+1)!} + o_{\mathbb{P}}(h^{p+1-\nu}), \tag{S.7}$$

where

- $\boldsymbol{\Lambda}_k = \boldsymbol{\Omega} \left[X_{h,1}^{p+k}, \ldots, X_{h,n}^{p+k}\right]' / n$, where, in particular $\boldsymbol{\Lambda}_1$ was denoted $\boldsymbol{\Lambda}$ in the main text.

Throughout, asymptotic orders and their in-probability versions hold uniformly in $\mathscr{F}_S$, as required by our framework; e.g., $A_n = o_{\mathbb{P}}(a_n)$ means $\sup_{F \in \mathscr{F}_S} \mathbb{P}_F[|A_n/a_n| > \epsilon] \to 0$ for every $\epsilon > 0$. This expression is valid for $p - \nu$ odd or even, though in the latter case the leading term of will be zero due to symmetry for interior points, i.e. $\boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1} \boldsymbol{\Lambda}_1 = O(h)$, and thus the rate will actually be faster (see Fan and Gijbels, 1996).

Sufficient smoothness for the validity of this calculation need not be available for many of the results herein to apply, and the amount of smoothness assumed to exist is a key factor in determining coverage error rates and optimality. See Section S.2 below for details and derivations in all cases, in addition to the discussion in the main paper. For the present, Equation (S.7) serves to motivate explicit bias correction by subtracting from $\hat{\mu}_p^{(\nu)}$ an estimate of the leading bias term. This estimate is formed as

$$h^{p+1-\nu} \nu! \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1} \boldsymbol{\Lambda}_1 \boldsymbol{e}_{p+1}' \hat{\boldsymbol{\beta}}_{p+1}, \qquad \text{with} \qquad \hat{\boldsymbol{\beta}}_{p+1} = \frac{1}{nb^{p+1}} \bar{\boldsymbol{\Gamma}}^{-1} \bar{\boldsymbol{\Omega}} \boldsymbol{Y},$$

where $\hat{\boldsymbol{\beta}}_{p+1}$ is exactly as in Equation (S.4), but with $p+1$ and b in place of p and h, respectively. Calonico et al. (2018a,b) discuss more general methods of bias correction. It is sometimes convenient to use the form above, but we will also use the more explicit notation for what this approach does: estimating the unknown derivative $\mu^{(p+1)}$ and plugging it in directly

$$h^{p+1-\nu} \nu! \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1} \boldsymbol{\Lambda}_1 \frac{\hat{\mu}_{p+1}^{(p+1)}}{(p+1)!}, \qquad \hat{\mu}_{p+1}^{(p+1)} = (p+1)! \boldsymbol{e}_{p+1}' \hat{\boldsymbol{\beta}}_{p+1} = \frac{1}{nb^{p+1}} (p+1)! \boldsymbol{e}_{p+1}' \bar{\boldsymbol{\Gamma}}^{-1} \bar{\boldsymbol{\Omega}} \boldsymbol{Y},$$

again matching (S.4), but with $p+1$ in place of p and ν and b in place of h. In particular, we have defined the exact analogues for this new local regression:

- $X_{b,i} = (X_i - \mathsf{x})/b$, for a bandwidth b and point of interest x, exactly like $X_{h,i}$ but with b in place of h,
- $\bar{\boldsymbol{\Omega}} = b^{-1}[(K\boldsymbol{r}_{p+1})(X_{b,1}), (K\boldsymbol{r}_{p+1})(X_{b,2}), \ldots, (K\boldsymbol{r}_{p+1})(X_{b,n})]$, exactly like $\boldsymbol{\Omega}$ but with b in place of h and $p+1$ in place of p,
- $\bar{\boldsymbol{\Gamma}} = \frac{1}{nb} \sum_{i=1}^n (K\boldsymbol{r}_{p+1}\boldsymbol{r}_{p+1}')(X_{b,i})$, exactly like $\boldsymbol{\Gamma}$ but with b in place of h and $p+1$ in place of p, and

- $\bar{\boldsymbol{\Lambda}}_k = \bar{\boldsymbol{\Omega}} \left[X_{b,1}^{p+1+k}, \ldots, X_{b,n}^{p+1+k}\right]'/n$, exactly like $\boldsymbol{\Lambda}_k$ but with b in place of h and $p+1$ in place of p (implying $\bar{\boldsymbol{\Omega}}$ in place of $\boldsymbol{\Omega}$).

We thus consider two types of centering estimators. Conventional nonparametric local polynomial inference sets $\hat{\theta} = \hat{\mu}_p^{(\nu)}$, which typically requires undersmoothing for valid inference, and robust bias corrected centering, which incorporates the explicit bias correction. In sum, $\hat{\theta}$ of (S.3) is one of

$$
\begin{aligned}
\hat{\mu}_p^{(\nu)} &= \frac{1}{nh^{\nu}} \nu! \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1} \boldsymbol{\Omega} \boldsymbol{Y}; \\
\hat{\theta}_{\mathtt{rbc}} &= \hat{\mu}_p^{(\nu)} - h^{p+1-\nu} \nu! \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1} \boldsymbol{\Lambda}_1 \frac{\hat{\mu}_{p+1}^{(p+1)}}{(p+1)!} = \frac{1}{nh^{\nu}} \nu! \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1} \boldsymbol{\Omega}_{\mathtt{rbc}} \boldsymbol{Y}.
\end{aligned}
\tag{S.8}
$$

where in the latter form of $\hat{\theta}_{\mathtt{rbc}}$, which is useful for defining the scale estimators below, we define

- $\boldsymbol{\Omega}_{\mathtt{rbc}} = \boldsymbol{\Omega} - \rho^{p+1} \boldsymbol{\Lambda}_1 \boldsymbol{e}_{p+1}' \bar{\boldsymbol{\Gamma}}^{-1} \bar{\boldsymbol{\Omega}}$ and
- $\rho = h/b$, the ratio of the two bandwidth sequences.

Comparing the two we see that only the matrix $\boldsymbol{\Omega}$ premultiplying $\boldsymbol{Y}$ changes.

S.1.2 Scale Estimators

The next piece we define are the scaling estimators. As discussed in the paper, it is crucial for coverage error to use fixed-n variance calculations, conditional in this case, to develop the Studentization, and we will focus most of our attention on these. Discussion of other options can be found in Section S.7, with some mention in Section S.3. The fixed-n variance of the centering is defined as

$$
\vartheta^2 = \mathbb{V}\left[\hat{\theta} \middle| X_1, \ldots, X_n\right] = \frac{1}{nh^{1+2\nu}} \nu!^2 \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1} (h \boldsymbol{\Omega}_\bullet \boldsymbol{\Sigma} \boldsymbol{\Omega}_\bullet'/n) \boldsymbol{\Gamma}^{-1} \boldsymbol{e}_\nu,
$$

where either $\boldsymbol{\Omega}_\bullet = \boldsymbol{\Omega}$ or $\boldsymbol{\Omega}_{\mathtt{rbc}}$ depending on the centering and

- $\Sigma = \mathrm{diag}(v(X_i) : i = 1, \ldots, n)$, with $v(x) = \mathbb{V}[Y|X = x]$.

The rateless portions of the variance is defined by $\sigma^2 := (nh^{1+2\nu}) \mathbb{V}\left[\hat{\theta} \middle| X_1, \ldots, X_n\right] = (nh^{1+2\nu}) \vartheta^2$, with, in particular

$$
\begin{aligned}
\sigma_p^2 &= \nu!^2 \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1} (h \boldsymbol{\Omega} \boldsymbol{\Sigma} \boldsymbol{\Omega}'/n) \boldsymbol{\Gamma}^{-1} \boldsymbol{e}_\nu, \qquad \text{and} \\
\sigma_{\mathtt{rbc}}^2 &= \nu!^2 \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1} (h \boldsymbol{\Omega}_{\mathtt{rbc}} \boldsymbol{\Sigma} \boldsymbol{\Omega}_{\mathtt{rbc}}'/n) \boldsymbol{\Gamma}^{-1} \boldsymbol{e}_\nu,
\end{aligned}
\tag{S.9}
$$

The only unknown piece of these is the conditional variance matrix $\boldsymbol{\Sigma}$, which we estimate using either

- $\hat{\boldsymbol{\Sigma}}_p = \mathrm{diag}(\hat{v}(X_i) : i = 1, \ldots, n)$, with $\hat{v}(X_i) = (Y_i - \boldsymbol{r}_p(X_i - \mathsf{x})' \hat{\boldsymbol{\beta}}_p)^2$ for $\hat{\boldsymbol{\beta}}_p$ defined in Equation (S.4), or

- $\hat{\boldsymbol{\Sigma}}_{\texttt{rbc}} = \operatorname{diag}(\hat{v}(X_i) : i = 1, \ldots, n)$, with $\hat{v}(X_i) = (Y_i - \boldsymbol{r}_{p+1}(X_i - \mathsf{x})'\hat{\boldsymbol{\beta}}_{p+1})^2$ for $\hat{\boldsymbol{\beta}}_{p+1}$ defined exactly as in Equation (S.4) but with $p+1$ in place of p and b in place of h.

The estimators $\hat{v}(X_i)$, using either p or $p+1$, are not estimators of the function $v(\cdot)$ of (S.1) per se, but rather are a convenient notation for predicted residuals.

The scale estimator $\hat{\vartheta}$ of I of (S.3) is thus one of

$$\begin{aligned} \hat{\vartheta}^2 &= \frac{\hat{\sigma}_p^2}{nh^{1+2\nu}}, & \hat{\sigma}_p^2 &:= \nu!^2 \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1}(h\boldsymbol{\Omega}\hat{\boldsymbol{\Sigma}}_p\boldsymbol{\Omega}'/n)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu, \qquad \text{or} \\ \hat{\vartheta}^2 &= \hat{\vartheta}^2_{\texttt{rbc}} := \frac{\hat{\sigma}^2_{\texttt{rbc}}}{nh^{1+2\nu}}, & \hat{\sigma}^2_{\texttt{rbc}} &:= \nu!^2 \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1}(h\boldsymbol{\Omega}_{\texttt{rbc}}\hat{\boldsymbol{\Sigma}}_{\texttt{rbc}}\boldsymbol{\Omega}_{\texttt{rbc}}'/n)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu, \end{aligned} \tag{S.10}$$

Remark S.1. For notational, and more importantly, practical/computational simplicity, the standard errors use the same local polynomial regressions (same kernel, bandwidth, and order) as the point estimates. Changing this results in changes to the constants and potentially (depending on the choices of h, b, and p) the rates for the coverage error expansions. Further, the procedure as defined here is simple to implement because the bases $\boldsymbol{r}_p(X_i - \mathsf{x})$ and $\boldsymbol{r}_{p+1}(X_i - \mathsf{x})$ and vectors $\hat{\boldsymbol{\beta}}_p$ and $\hat{\boldsymbol{\beta}}_{p+1}$ are already available. Other standard errors are discussed in Section S.7 and, for asymptotic versions, briefly in the main paper. ⌟

S.1.3 t-Statistics and Confidence Intervals

With the center and scale defined, the dual t-statistics of the confidence intervals in (S.3) are

$$T_p = \frac{\sqrt{nh^{1+2\nu}}(\hat{\mu}_p^{(\nu)} - \mu^{(\nu)})}{\hat{\sigma}_p} \qquad \text{and} \qquad T_{\texttt{rbc}} = \frac{(\hat{\theta}_{\texttt{rbc}} - \mu^{(\nu)})}{\hat{\vartheta}_{\texttt{rbc}}} = \frac{\sqrt{nh^{1+2\nu}}(\hat{\theta}_{\texttt{rbc}} - \mu^{(\nu)})}{\hat{\sigma}_{\texttt{rbc}}}. \tag{S.11}$$

Our main results, Theorem 3 of the paper and Theorem S.1 herein, are uniformly valid Edgeworth expansions of the distribution functions of these statistics, and more general local polynomial based t-statistics using other valid standard errors. These are given in Section S.3 below. From these, for any fixed quantiles z_l and z_u the coverage error can be computed for the confidence intervals (S.3). The class of intervals $\mathscr{I}$ is indexed by the choices of centering, scaling, bandwidths, and quantiles. All of these represent choices made by the researcher, and each choice impacts the coverage error, as made precise below. When discussing specific choices it will be useful notationally to write the intervals as functions of these choices, such as $I(h)$ for an interval based on a bandwidth h or $I(\hat{\theta}, \hat{\vartheta})$ for specific choices of centering and scaling. In particular, let $I_p = I(\hat{\mu}_p^{(\nu)}, \hat{\sigma}_p)$ and $I_{\texttt{rbc}} = I(\hat{\theta}_{\texttt{rbc}}, \hat{\sigma}_{\texttt{rbc}})$, following the convention above.

S.1.4 Assumptions

The two following assumptions are sufficient for our results, both directly copied from the main text. See discuss there. The first defines the class of distributions of the data, denoted $\mathscr{F}_S$.

Assumption S.1. *Let $\mathscr{F}_S$ be the set of distributions F for the pair (Y, X) which obey model (S.1) and the following. There exist constants $S \geq \nu$, $s \in (0, 1]$, $0 < c < C < \infty$, and a neighborhood of x on the support of X, none of which depend on F, such that for all x, x' in the neighborhood:*

- **(a)** *the Lebesgue density of (Y, X), $f_{yx}(\cdot)$, is continuous and $c \leq f_{yx}(\cdot) \leq C$; the Lebesgue density of X, $f(\cdot)$, is continuous and $c \leq f(x) \leq C$; $v(x) := \mathbb{V}[Y|X = x] \geq c$ and continuous; and $\mathbb{E}[|Y|^{8+c}|X = x] \leq C$, and*
- **(b)** *$\mu(\cdot)$ is S-times continuously differentiable and $|\mu^{(S)}(x) - \mu^{(S)}(x')| \leq C|x - x'|^s$.*

Throughout, $\{(Y_1, X_1), \ldots, (Y_n, X_n)\}$ is a random sample from (Y, X).

Second, the class of confidence intervals is governed by the following condition on the kernel function $K(\cdot)$ and polynomial degree p. We impose the following throughout.

Assumption S.2. *The kernel K is supported on $[-1, 1]$, positive, bounded, and even. Further, $K(u)$ is either constant (the uniform kernel) or $(1, K(u)\boldsymbol{r}_{3(k+1)}(u))'$ is linearly independent on $[-1, 0]$ and $[0, 1]$, where $k = p$ for T_p or $k = p + 1$ for $T_{\mathtt{rbc}}$. The order p is at least ν.*

We define

- $\mathscr{I}_p$ as the set of intervals I of the form (S.3) governed by Assumption S.2 for a given, fixed p, centering and scaling from (S.8) and (S.10) respectively, and possibly, as needed, restrictions on the bandwidths and other user choices defined below in the main results. Other choices may be accommodated with $\mathscr{I}_p$, such as the different standard errors discussed in Section S.7. These will require tweaks to the proofs below, though in many cases no conceptual alteration to the proof strategy. Other bias correction methods could be accommodated as well, see Calonico et al. (2018a) for some discussion.

S.2 Bias and the Role of Smoothness

In this section we derive (and list) all the necessary bias terms, both in generic form and for special cases. We will cover different centerings, different smoothness cases, as well as interior and boundary points. We first give a generic derivation, followed by discussion of the bias of $\hat{\theta} = \hat{\mu}^{(\nu)}_{p+1}$ and then $\hat{\theta}_{\mathtt{rbc}}$, and in the final subsection, a complete list of all results and formulae.

The conditional bias defined above in Equation (S.7), and the similarly computed $\mathbb{E}[\hat{\theta}_{\mathtt{rbc}} \big| X_1, \ldots, X_n]$, are useful for describing bias correction, first order asymptotics, and computing and implementing optimal bandwidths. However, these can not be present in the Edgeworth and coverage error expansions because they are random quantities. Further, the leading term isolated in Equation (S.7) presumes sufficient smoothness, which we avoid for general results. (The analogous calculation for $\hat{\theta}_{\mathtt{rbc}}$ is shown below.)

The bias terms in the expansions are generic and nonrandom. In Theorem S.1 we denote the bias contribution by $\Psi_{T,F}$. This term, and its particular cases $\Psi_{T_p,F}$ and $\Psi_{\mathtt{rbc},F} = \Psi_{T_{\mathtt{rbc}},F}$ in particular, capture the entire bias, that is both the rate and the constant. These terms are defined both (i) before a Taylor approximation is performed, and (ii) with $\mathbf{\Gamma}$, $\bar{\mathbf{\Gamma}}$, and $\mathbf{\Lambda}_1$ replaced with their fixed-n expectations, denoted $\tilde{\mathbf{\Gamma}}$, $\tilde{\bar{\mathbf{\Gamma}}}$, and $\tilde{\mathbf{\Lambda}}_1$. In both sense, these bias terms reflect the "fixed-n" approach. (A tilde always denotes a fixed-n expectation, and all expectations are fixed-n calculations unless explicitly denoted otherwise.)

For notation, we maintain the dependence on F if it is useful to emphasize that for certain $F \in \mathscr{F}_S$ the bias may be lower or higher. For example, if it happens that $\mu_F^{(p+1)}(\mathsf{x}) = 0$, the leading term of Equation (S.7) will be zero even if $p - \nu$ is odd. Further, at present we explicitly write these as functions of the t-statistic, as the expansions in Section S.3 are for the t-statistics, but it would be equivalent to write them as functions of the corresponding interval: that is $\Psi_{I,F} \equiv \Psi_{T,F}$, in terms of I and F. For example, $\Psi_{\mathtt{rbc},F} = \Psi_{T_{\mathtt{rbc}},F} = \Psi_{I_{\mathtt{rbc}},F}$.

S.2.1 Generic Bias Formulas

Define

- $\boldsymbol{\beta}_k$ (usually $k = p$ or $k = p + 1$) as the $k + 1$ vector with $(j + 1)$ element equal to $\mu^{(j)}(\mathsf{x})/j!$ for $j = 0, 1, \ldots, k$ as long as $j \leq S$, and zero otherwise,
- $\boldsymbol{M} = [\mu(X_1), \ldots, \mu(X_n)]'$,
- $\boldsymbol{B}_k$ as the n-vector with i^{th} entry $[\mu(X_i) - \boldsymbol{r}_k(X_i - \mathsf{x})'\boldsymbol{\beta}_k]$,
- $\rho = h/b$, the ratio of the two bandwidth sequences, and
- $\tilde{\mathbf{\Gamma}} = \mathbb{E}[\mathbf{\Gamma}]$, $\tilde{\bar{\mathbf{\Gamma}}} = \mathbb{E}[\bar{\mathbf{\Gamma}}]$, $\tilde{\mathbf{\Lambda}}_1 = \mathbb{E}[\mathbf{\Lambda}_1]$, and so forth. A tilde always denotes a fixed-n expectation, and all expectations are fixed-n calculations unless explicitly denoted otherwise. The dependence on F and $\mathscr{F}_S$ is suppressed. As a concrete example:

$$\mathbf{\Lambda}_k = \mathbf{\Omega} \left[X_{h,1}^{p+k}, \ldots, X_{h,n}^{p+k} \right]' /n = \frac{1}{nh} \sum_{i=1}^{n} (K\boldsymbol{r}_p)(X_{h,i}) X_{h,i}^{p+k},$$

  and so

$$\begin{aligned}
\tilde{\mathbf{\Lambda}}_k = \mathbb{E}[\mathbf{\Lambda}_k] &= h^{-1}\mathbb{E}\left[(K\boldsymbol{r}_p)(X_{h,i}) X_{h,i}^{p+k} \right] \\
&= h^{-1} \int_{\text{supp}\{X\}} K\left(\frac{X_i - \mathsf{x}}{h}\right) \boldsymbol{r}_p\left(\frac{X_i - \mathsf{x}}{h}\right) \left(\frac{X_i - \mathsf{x}}{h}\right)^{p+k} f(X_i) dX_i \\
&= \int_{-1}^{1} K(u) \boldsymbol{r}_p(u) u^{p+k} f(\mathsf{x} + uh) du.
\end{aligned}$$

  The range of integration for integrals will generally be left implicit. The range will change when the point of interest is on a boundary, but the notation will remain the same and it is to be understood that moments and moments of the kernel be replaced by the appropriate truncated version. For example, if $\text{supp}\{X\} = [0, \infty)$ and the point of interest is $\mathsf{x} = 0$, then by a change of variables

$$\tilde{\mathbf{\Lambda}}_k = h^{-1}\int_{\text{supp}\{X\}} (Kr_p)(X_{h,i})X_{h,i}^{p+k} f(X_i)dX_i = \int_0^\infty (Kr_p)(u)u^{p+k}f(uh)du,$$

whereas if $\text{supp}\{X\} = (-\infty, 0]$ and $\mathsf{x} = 0$, then

$$\tilde{\mathbf{\Lambda}}_k = \int_{-\infty}^0 (Kr_p)(u)u^{p+k}f(-uh)du.$$

For the remainder of this section, the notation is left generic.

To compute the terms $\Psi_{T_p,F}$ and $\Psi_{\texttt{rbc},F}$, begin with the conditional mean of $\hat{\mu}_p^{(\nu)}$:

$$\begin{aligned}\mathbb{E}\left[\hat{\mu}_p^{(\nu)}\middle|X_1,\ldots,X_n\right] &= \nu!\boldsymbol{e}_\nu'\mathbb{E}\left[\hat{\boldsymbol{\beta}}_p\middle|X_1,\ldots,X_n\right] = \frac{1}{nh^\nu}\nu!\boldsymbol{e}_\nu'\mathbf{\Gamma}^{-1}\mathbf{\Omega}\boldsymbol{M} \\ &= \frac{1}{nh^\nu}\nu!\boldsymbol{e}_\nu'\mathbf{\Gamma}^{-1}\mathbf{\Omega}(\boldsymbol{M}-\boldsymbol{R}\boldsymbol{\beta}_p) + \frac{1}{nh^\nu}\nu!\boldsymbol{e}_\nu'\mathbf{\Gamma}^{-1}\mathbf{\Omega}\boldsymbol{R}\boldsymbol{\beta}_p \\ &= \frac{1}{nh^\nu}\nu!\boldsymbol{e}_\nu'\mathbf{\Gamma}^{-1}\mathbf{\Omega}\boldsymbol{B}_p + \frac{1}{nh^\nu}\nu!\boldsymbol{e}_\nu'\mathbf{\Gamma}^{-1}\mathbf{\Omega}\boldsymbol{R}\boldsymbol{\beta}_p.\end{aligned}$$

Because $h^{-\nu}\boldsymbol{e}_\nu' = \boldsymbol{e}_\nu'\boldsymbol{H}^{-1}$, $\check{\boldsymbol{R}} = \boldsymbol{R}\boldsymbol{H}^{-1}$, $\mathbf{\Omega} = \check{\boldsymbol{R}}'\boldsymbol{W}$, and $\mathbf{\Gamma} = \check{\boldsymbol{R}}'\boldsymbol{W}\check{\boldsymbol{R}}/n = \mathbf{\Omega}\check{\boldsymbol{R}}/n$, (the same calculations used for (S.5) and (S.6)) the second term above is

$$\nu!\left(\boldsymbol{e}_\nu'\boldsymbol{H}^{-1}\right)\mathbf{\Gamma}^{-1}\left(\mathbf{\Omega}\check{\boldsymbol{R}}/n\right)\boldsymbol{H}\boldsymbol{\beta}_p = \nu!\boldsymbol{e}_\nu'\boldsymbol{\beta}_p = \mu^{(\nu)}(\mathsf{x}), \tag{S.12}$$

using the definition of $\boldsymbol{\beta}_p$ (the $\nu+1$ element of the vector $\boldsymbol{\beta}_p$ will not be zero, as $\nu \leq S$ holds by Assumption S.1). Therefore

$$\begin{aligned}\mathbb{E}\left[\hat{\mu}_p^{(\nu)}\middle|X_1,\ldots,X_n\right] - \mu^{(\nu)} &= \frac{1}{nh^\nu}\nu!\boldsymbol{e}_\nu'\mathbf{\Gamma}^{-1}\mathbf{\Omega}\boldsymbol{B}_p \\ &= h^{-\nu}\nu!\boldsymbol{e}_\nu'\mathbf{\Gamma}^{-1}\frac{1}{nh}\sum_{i=1}^n (K\boldsymbol{r}_p)(X_{h,i})\left(\mu(X_i) - \boldsymbol{r}_p(X_i-\mathsf{x})'\boldsymbol{\beta}_p\right).\end{aligned} \tag{S.13}$$

From here, a Taylor expansion of $\mu(X_i)$ around $X = \mathsf{x}$ immediately gives Equation (S.7), provided that $S \geq p+1$. Instead, the bias terms of the Edgeworth expansions use this form directly, replacing the sample averages with population averages. The biases, $\Psi_{T,F}$ in general and $\Psi_{T_p,F}$ and $\Psi_{\texttt{rbc},F}$ in particular, must explicitly account for the rate scaling of $\sqrt{nh^{1+2\nu}}$, because the Edgeworth expansions are proven directly for the t-statistics.

For $\hat{\theta} = \hat{\mu}^{(\nu)}$, for T_p or I_p, we apply the rate scaling to the above display and then define

$$\Psi_{T_p,F} = \sqrt{nh^{1+2\nu}}h^{-\nu}\nu!\boldsymbol{e}_\nu'\tilde{\mathbf{\Gamma}}^{-1}\mathbb{E}\left[h^{-1}(K\boldsymbol{r}_p)(X_{h,i})\left(\mu(X_i) - \boldsymbol{r}_p(X_i-\mathsf{x})'\boldsymbol{\beta}_p\right)\right].$$

Note that the $h^{-\nu}$ cancels, and thus the rate of decay of the scaled bias does not depend on the level of derivative of interest. Because of the fixed-n nature of this calculation, the parity of $p-\nu$ does not matter. If a Taylor series were performed *and* the matrixes were allowed to converge to their limit, the well-known symmetry cancellation would occur for $p-\nu$ even at interior x (Fan and Gijbels, 1996). The generic expansions are stated without being explicit on this, but for certain derivations and specific cases the symmetry will be exploited. It holds that $\Psi_{T_p,F} = O(\sqrt{nh}h^{\zeta})$ uniformly in $\mathscr{F}_S$ where ζ varies depending on smoothness, parity of $p-\nu$, and location of x. If p is small relative to S, depending again on parity and location, we can isolate the leading term $\psi_{T_p,F}$ such that $\Psi_{T_p,F} = \sqrt{nh}h^{\zeta}\psi_{T_p,F}[1+o(1)]$ where $\psi_{T_p,F} = O(1)$ uniformly in $\mathscr{F}_S$ and is nonzero for some $F \in \mathscr{F}_S$. Results for every case are given in Section S.2.2 and summarized in Table S.1.

For $\hat{\theta}_{\texttt{rbc}}$ (i.e. for $T_{\texttt{rbc}}$ and $I_{\texttt{rbc}}$),

$$\mathbb{E}\left[\hat{\theta}_{\texttt{rbc}}\middle|X_1,\ldots,X_n\right] - \mu^{(\nu)} = \left\{\mathbb{E}\left[\hat{\mu}^{(\nu)}\middle|X_1,\ldots,X_n\right] - \mu^{(\nu)}\right\} \\ - \left\{h^{p+1-\nu}\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Lambda}_1\frac{1}{(p+1)!}\mathbb{E}\left[\hat{\mu}^{(p+1)}\middle|X_1,\ldots,X_n\right]\right\}.$$

The first term is given exactly in (S.13). For the second term, following exactly the same steps that we used to arrive at (S.13), but with $(p+1)$ in place of v and p and b in place of h, we find that

$$\mathbb{E}\left[\hat{\mu}^{(p+1)}\middle|X_1,\ldots,X_n\right] = (p+1)!\boldsymbol{e}_{p+1}'\boldsymbol{\beta}_{p+1} \\ + b^{-p-1}(p+1)!\boldsymbol{e}_{p+1}'\bar{\boldsymbol{\Gamma}}^{-1}\frac{1}{nb}\sum_{i=1}^{n}(K\boldsymbol{r}_{p+1})(X_{b,i})\left(\mu(X_i) - \boldsymbol{r}_{p+1}(X_i-\mathsf{x})'\boldsymbol{\beta}_{p+1}\right)$$

Inserting this result and (S.13) into $\mathbb{E}\left[\hat{\theta}_{\texttt{rbc}}\middle|X_1,\ldots,X_n\right] - \mu^{(\nu)}$, we find that

$$\begin{aligned}
&\mathbb{E}\left[\hat{\theta}_{\texttt{rbc}}\middle|X_1,\ldots,X_n\right] - \mu^{(\nu)} \\
&= h^{-\nu}\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\frac{1}{nh}\sum_{i=1}^{n}(K\boldsymbol{r}_p)(X_{h,i})\left(\mu(X_i) - \boldsymbol{r}_p(X_i-\mathsf{x})'\boldsymbol{\beta}_p\right) - h^{p+1-\nu}\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Lambda}_1\frac{1}{(p+1)!}(p+1)!\boldsymbol{e}_{p+1}'\boldsymbol{\beta}_{p+1} \\
&\quad - h^{p+1-\nu}\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Lambda}_1\frac{1}{(p+1)!}b^{-p-1}(p+1)!\boldsymbol{e}_{p+1}'\bar{\boldsymbol{\Gamma}}^{-1}\times\frac{1}{nb}\sum_{i=1}^{n}(K\boldsymbol{r}_{p+1})(X_{b,i})\left(\mu(X_i) - \boldsymbol{r}_{p+1}(X_i-\mathsf{x})'\boldsymbol{\beta}_{p+1}\right) \\
&= h^{-\nu}\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\frac{1}{nh}\sum_{i=1}^{n}(K\boldsymbol{r}_p)(X_{h,i})\left(\mu(X_i) - \boldsymbol{r}_p(X_i-\mathsf{x})'\boldsymbol{\beta}_p\right) - h^{p+1-\nu}\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Lambda}_1\boldsymbol{e}_{p+1}'\boldsymbol{\beta}_{p+1} \\
&\quad - h^{-\nu}\rho^{p+1}\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Lambda}_1\boldsymbol{e}_{p+1}'\bar{\boldsymbol{\Gamma}}^{-1}\times\frac{1}{nb}\sum_{i=1}^{n}(K\boldsymbol{r}_{p+1})(X_{b,i})\left(\mu(X_i) - \boldsymbol{r}_{p+1}(X_i-\mathsf{x})'\boldsymbol{\beta}_{p+1}\right) \\
&= h^{-\nu}\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\frac{1}{nh}\sum_{i=1}^{n}(K\boldsymbol{r}_p)(X_{h,i})\left(\mu(X_i) - \boldsymbol{r}_{p+1}(X_i-\mathsf{x})'\boldsymbol{\beta}_{p+1}\right) \\
&\quad - h^{-\nu}\rho^{p+1}\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Lambda}_1\boldsymbol{e}_{p+1}'\bar{\boldsymbol{\Gamma}}^{-1}\times\frac{1}{nb}\sum_{i=1}^{n}(K\boldsymbol{r}_{p+1})(X_{b,i})\left(\mu(X_i) - \boldsymbol{r}_{p+1}(X_i-\mathsf{x})'\boldsymbol{\beta}_{p+1}\right).
\end{aligned} \tag{S.14}$$

where the last equality combines the first two terms (in the penultimate line), by noticing that

$$\begin{aligned} h^{p+1-\nu}\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Lambda}_1\boldsymbol{e}_{p+1}'\boldsymbol{\beta}_{p+1} &= h^{p+1-\nu}\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\frac{1}{nh}\sum_{i=1}^n (K\boldsymbol{r}_p)(X_{h,i})(X_{h,i})^{p+1}\boldsymbol{e}_{p+1}'\boldsymbol{\beta}_{p+1} \\ &= h^{p+1-\nu}\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\frac{1}{nh}\sum_{i=1}^n (K\boldsymbol{r}_p)(X_{h,i})h^{-p-1}(X_i-\mathsf{x})^{p+1}\boldsymbol{e}_{p+1}'\boldsymbol{\beta}_{p+1}, \end{aligned}$$

and that $(X_i-\mathsf{x})^{p+1}\boldsymbol{e}_{p+1}'\boldsymbol{\beta}_{p+1}$ is exactly the difference between $\boldsymbol{r}_p(X_i-\mathsf{x})'\boldsymbol{\beta}_p$ and $\boldsymbol{r}_{p+1}(X_i-\mathsf{x})'\boldsymbol{\beta}_{p+1}$.

As before, $\Psi_{\mathtt{rbc},F}$ is now defined replacing sample averages with population averages and applying the scaling of $\sqrt{nh^{1+2\nu}}$ from the t-statistic. Again the $h^{-\nu}$ cancels, and thus the rate of decay of the scaled bias does not depend on the level of derivative of interest.

In sum, the generic formulas are

$$\begin{aligned} \Psi_{T_p,F} &= \sqrt{nh}\;\nu!\boldsymbol{e}_\nu'\tilde{\boldsymbol{\Gamma}}^{-1}\mathbb{E}\left[h^{-1}(K\boldsymbol{r}_p)(X_{h,i})\left(\mu(X_i)-\boldsymbol{r}_p(X_i-\mathsf{x})'\boldsymbol{\beta}_p\right)\right], \\ \Psi_{\mathtt{rbc},F} &= \sqrt{nh}\;\nu!\boldsymbol{e}_\nu'\tilde{\boldsymbol{\Gamma}}^{-1}\mathbb{E}\Bigg[\left\{h^{-1}(K\boldsymbol{r}_p)(X_{h,i})-\rho^{p+1}\tilde{\boldsymbol{\Lambda}}_1\boldsymbol{e}_{p+1}'\tilde{\bar{\boldsymbol{\Gamma}}}^{-1}b^{-1}(K\boldsymbol{r}_{p+1})(X_{b,i})\right\} \\ &\qquad\qquad \times\left(\mu(X_i)-\boldsymbol{r}_{p+1}(X_i-\mathsf{x})'\boldsymbol{\beta}_{p+1}\right)\Bigg] \end{aligned} \tag{S.15}$$

or using $\boldsymbol{\Omega}$ and $\boldsymbol{\Omega}_{\mathtt{rbc}}$ as in Eqn. (S.8), and $\boldsymbol{B}_k$,

$$\Psi_{T_p,F} = \sqrt{nh}\;\nu!\boldsymbol{e}_\nu'\tilde{\boldsymbol{\Gamma}}^{-1}\mathbb{E}[\boldsymbol{\Omega}\boldsymbol{B}_p]$$

and

$$\Psi_{\mathtt{rbc},F} = \sqrt{nh}\;\nu!\boldsymbol{e}_\nu'\tilde{\boldsymbol{\Gamma}}^{-1}\left(\mathbb{E}[\boldsymbol{\Omega}\boldsymbol{B}_{p+1}]-\rho^{p+1}\tilde{\boldsymbol{\Lambda}}\boldsymbol{e}_{p+1}'\tilde{\bar{\boldsymbol{\Gamma}}}^{-1}\mathbb{E}[\bar{\boldsymbol{\Omega}}\boldsymbol{B}_{p+1}]\right).$$

For the generic results of coverage error or the generic Edgeworth expansions of Theorem S.1 below, these definitions are suitable and the $\Psi_{T_p,F}$ and $\Psi_{\mathtt{rbc},F}$ may appear directly. For T_p, parity of $p-\nu$ is not used, but can matter: the rate at which $\Psi_{T_p,F}$ vanishes is faster by one factor of h at interior points (Fan and Gijbels, 1996). The validity of the Edgeworth expansions is not affected by this; the statements are seamless.

However, it is also useful to separate the rate and leading constant term of these biases when possible. When it is possible we will isolate both the rate and the constant term of the bias. It holds that $\Psi_{T_p,F} = O(\sqrt{nh}h^\zeta)$ uniformly in $\mathscr{F}_S$ and if p is small relative to S, depending again on parity and location, we can isolate the leading term $\psi_{T_p,F}$ such that $\Psi_{T_p,F} = \sqrt{nh}h^\zeta\psi_{T_p,F}[1+o(1)]$ where $\psi_{T_p,F} = O(1)$ uniformly in $\mathscr{F}_S$ and is nonzero for some $F\in\mathscr{F}_S$. Similarly, it is always possible to show that $\Psi_{\mathtt{rbc},F} = O(\sqrt{nh}\;t(h,b))$ for a function $t(\cdot,\cdot)$ and further, if $\rho = h/b$ is bounded and bounded away from zero then $t(\cdot,\cdot)$ can be simplified to h^ζ. If p is small relative to S we can isolate the leading terms via a Taylor expansion. If p is small and ρ is bounded and bounded away from zero, we can write $\Psi_{\mathtt{rbc},F} = \sqrt{nh}h^\zeta\psi_{\mathtt{rbc},F}[1+o(1)]$.

For both $\Psi_{T_p,F}$ and $\Psi_{\mathtt{rbc},F}$, ζ, $t(h,b)$, $\psi_{T_p,F}$ and $\psi_{\mathtt{rbc},F}$ depend on smoothness, parity of $p-\nu$, and location of x. Complete derivations for $\Psi_{T_p,F}$ and $\Psi_{\texttt{rbc},F}$ are given in Sections S.2.2 and S.2.3 below and both are summarized in Tables S.1 and S.2 for lists of all cases.

The starting point of the derivations is a Taylor approximation. Recall the definitions of $\boldsymbol{r}_p(u)$ and $\boldsymbol{\beta}_p$, where in particular elements of the latter beyond $S+1$ are zero. A Taylor approximation, for some $\bar{x}$, gives

$$\begin{aligned}
\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p &= \sum_{k=0}^{S} \frac{1}{k!}(X_i-\mathsf{x})^k \mu^{(k)}(\mathsf{x}) + \frac{1}{S}(X_i - \mathsf{x})^S \left(\mu^{(S)}(\bar{x}) - \mu^{(S)}(\mathsf{x})\right) \\
&\quad - \sum_{k=0}^{S\wedge p} \frac{1}{k!}(X_i-\mathsf{x})^k \mu^{(k)}(\mathsf{x}) \\
&= \sum_{k=S\wedge p+1}^{S} \frac{1}{k!}(X_i-\mathsf{x})^k \mu^{(k)}(\mathsf{x}) + \frac{1}{S!}(X_i - \mathsf{x})^S \left(\mu^{(S)}(\bar{x}) - \mu^{(S)}(\mathsf{x})\right) \\
&= \sum_{k=S\wedge p+1}^{S} \frac{h^k}{k!}(X_{h,i})^k \mu^{(k)}(\mathsf{x}) + O(h^{S+s}), \qquad \text{(S.16)}
\end{aligned}$$

where the first summation in the last two lines is taken to be zero if $p \geq S$, and we have applied Assumption S.1 and restricted to $X_i \in [\mathsf{x} \pm h]$ (i.e. $K(X_{h,i}) > 0$). Note that by assumption the order of the remainder, $O(h^{S+s})$, holds uniformly in $\mathscr{F}_S$. We will use this expansion repeatedly below, or analogous results for other bandwidths and polynomial degrees.

S.2.2 No Bias Correction: Specific Cases and Leading Terms

We now turn to specific cases for $\Psi_{T_p,F}$. We will characterize the rate and leading constant terms in all cases, depending on depending on the relationship of p and S, the parity of $p-\nu$, and whether x is an interior point or on the boundary. Note that here, unlike Equation (S.7), we are working with nonrandom quantities. The general case, from Equation (S.15), which appears in the Edgeworth expansion is

$$\Psi_{T_p,F} = \sqrt{nh}\; \nu! \boldsymbol{e}_\nu' \tilde{\boldsymbol{\Gamma}}^{-1} \mathbb{E}[\boldsymbol{\Omega} \boldsymbol{B}_p] = \sqrt{nh}\; \nu! \boldsymbol{e}_\nu' \tilde{\boldsymbol{\Gamma}}^{-1} \mathbb{E}\left[h^{-1}(K\boldsymbol{r}_p)(X_{h,i})\left(\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p\right)\right].$$

It is always true that the rate is captured by the exponent ζ in the form

$$\Psi_{T_p,F} = O(\sqrt{nh}h^{\zeta}).$$

If p is small enough relative to S, then we write

$$\Psi_{T_p,F} = \sqrt{nh}h^{\zeta}\psi_{T_p,F}[1 + o(1)]$$

and call $\psi_{T_p,F}$ the leading constant. Recall that $\psi_{T_p,F}$ is not truly constant, but rather a nonrandom sequence that is $O(1)$ uniformly in $\mathscr{F}_S$ and is nonzero for some $F \in \mathscr{F}_S$. Table S.1 is complete list

<table>
<tr><th>Location of x</th><th>Parity of $p-\nu$</th><th>Smoothness</th><th>Rate Exponent ζ</th><th>$\psi_{T_p,F}$</th></tr>
<tr><td rowspan="2">Boundary</td><td rowspan="2">odd or even</td><td>$p < S$</td><td>$p+1$</td><td>$\nu! \boldsymbol{e}_\nu' \tilde{\boldsymbol{\Gamma}}^{-1} \tilde{\boldsymbol{\Lambda}}_1 \frac{\mu^{(p+1)}}{(p+1)!}$</td></tr>
<tr><td>$p \geq S$</td><td>$S+s$</td><td>N/A</td></tr>
<tr><td rowspan="4">Interior</td><td rowspan="2">odd</td><td>$p < S$</td><td>$p+1$</td><td>$\nu! \boldsymbol{e}_\nu' \tilde{\boldsymbol{\Gamma}}^{-1} \tilde{\boldsymbol{\Lambda}}_1 \frac{\mu^{(p+1)}}{(p+1)!}$</td></tr>
<tr><td>$p \geq S$</td><td>$S+s$</td><td>N/A</td></tr>
<tr><td rowspan="2">even</td><td>$p+2 \leq S$</td><td>$p+2$</td><td>$\nu! \boldsymbol{e}_\nu' \tilde{\boldsymbol{\Gamma}}^{-1} \left(h^{-1} \tilde{\boldsymbol{\Lambda}}_1 \frac{\mu^{(p+1)}}{(p+1)!} + \tilde{\boldsymbol{\Lambda}}_2 \frac{\mu^{(p+2)}}{(p+2)!} \right)$</td></tr>
<tr><td>$p+2 > S$</td><td>$S+s$</td><td>N/A</td></tr>
</table>

Table S.1: Summary of Bias Terms in All Cases For Uncorrected Centering $\hat{\mu}_p^{(\nu)}$. Rate exponent ζ is such that $\Psi_{T_p,F} = O(\sqrt{nh} h^\zeta)$. When possible, $\psi_{T_p,F}$ is such that $\Psi_{T_p,F} = \sqrt{nh} h^\zeta \psi_{T_p,F}[1 + o(1)]$.

of the results, including ζ and $\psi_{T_p,F}$. These cases are derived in the rest of this section.

As an aside, it is technically possible to obtain the representation $\Psi_{T_p,F} = \sqrt{nh} h^\zeta \psi_{T_p,F}[1+o(1)]$ in general, that is for any p, by letting $\psi_{T_p,F}$ to capture the final term in the Taylor expansion, $(X_i - \mathsf{x})^S [\mu^{(S)}(\bar{x}) - \mu^{(S)}(\mathsf{x})]/S!$, see the penultimate step of Equation (S.16), and taking the $o(1)$ term to be exactly zero. However, we do not use $\psi_{T_p,F}$ in this case because the representation is not useful for practice nor is it more concrete than simply using $\Psi_{T_p,F}$, since in this case $\Psi_{T_p,F} = \sqrt{nh} h^\zeta \psi_{T_p,F}$ amounts to little more than a redefinition of notation.

S.2.2.1 Boundary Point

Here parity plays no role.

Case 1: $\boldsymbol{p < S}$. The leading bias term can be characterized, and we find (cf. Equation (S.7))

$$\Psi_{T_p,F} = \sqrt{nh^{1+2\nu}} h^{-\nu} h^{p+1} \frac{\mu^{(p+1)}}{(p+1)!} \nu! \boldsymbol{e}_\nu' \tilde{\boldsymbol{\Gamma}}^{-1} \tilde{\boldsymbol{\Lambda}}_1 \left[1 + o(1)\right].$$

Note that this holds regardless of whether x is an interior or boundary point, with suitable changes to the ranges of integration in $\tilde{\boldsymbol{\Gamma}}$ and $\tilde{\boldsymbol{\Lambda}}_1$.

Case 2: $\boldsymbol{p \geq S}$. All that is left in Equation (S.16) is this remainder term, and we therefore have

$$\Psi_{T_p,F} = \sqrt{nh^{1+2\nu}} h^{-\nu} O(h^{S+s}) = O(\sqrt{nh} h^{S+s}),$$

and cannot say anything further regarding constants. This result applies any time $p \geq S$, regardless of ν, parity of $p - \nu$, and at interior and boundary points.

S.2.2.2 Interior Point: $p-\nu$ odd

The results for $p-\nu$ odd are identical to the boundary point case. This automatic boundary carpentry is discussed briefly in the main text. It is one of the celebrated features of local polynomial regression, known for point estimation since their inception, see Fan and Gijbels (1996) for review, and proven for inference for the first time in Calonico et al. (2018a).

Case 1: $\boldsymbol{p < S}$. The leading bias term can be characterized, and we find (cf. Equation (S.7))

$$\Psi_{T_p,F} = \sqrt{nh^{1+2\nu}}h^{-\nu}h^{p+1}\frac{\mu^{(p+1)}}{(p+1)!}\nu!\boldsymbol{e}_\nu'\tilde{\boldsymbol{\Gamma}}^{-1}\tilde{\boldsymbol{\Lambda}}_1\left[1+o(1)\right].$$

Note that this holds regardless of whether x is an interior or boundary point, with suitable changes to the ranges of integration in $\tilde{\boldsymbol{\Gamma}}$ and $\tilde{\boldsymbol{\Lambda}}_1$.

Case 2: $\boldsymbol{p \geq S}$. All that is left in Equation (S.16) is this remainder term, and we therefore have

$$\Psi_{T_p,F} = \sqrt{nh^{1+2\nu}}h^{-\nu}O(h^{S+s}) = O(\sqrt{nh}h^{S+s}),$$

and cannot say anything further regarding constants. This result applies any time $p \geq S$, regardless of ν, parity of $p-\nu$, and at interior and boundary points.

S.2.2.3 Interior Point: $p-\nu$ even

Here the parity of p will matter. It is worth spelling out three smoothness cases, though we will find the same result for the latter two.

Case 1: $\boldsymbol{p+2 \leq S}$. We begin by retaining *two* terms of Equation (S.16):

$$\Psi_{T_p,F} = \sqrt{nh^{1+2\nu}}h^{-\nu}h^{p+1}\nu!\boldsymbol{e}_\nu'\tilde{\boldsymbol{\Gamma}}^{-1}\left(\tilde{\boldsymbol{\Lambda}}_1\frac{\mu^{(p+1)}}{(p+1)!}+h\tilde{\boldsymbol{\Lambda}}_2\frac{\mu^{(p+2)}}{(p+2)!}\right)\left[1+o(1)\right].$$

To find the leading term, we must appeal to the limits of (the fixed-n) expectations $\tilde{\boldsymbol{\Gamma}}^{-1}$ and $\tilde{\boldsymbol{\Lambda}}_k$ where it holds that

$$\boldsymbol{e}_\nu'\tilde{\boldsymbol{\Gamma}}^{-1}\tilde{\boldsymbol{\Lambda}}_k = A + hB + o(h), \text{ with } A=0 \text{ if } (p+k-\nu) \text{ is odd and } \mathsf{x} \text{ is in the interior.} \tag{S.17}$$

Note that at present we use this fact with $k=1$, and hence $(p+k-\nu)$ is odd if $p-\nu$ is even, the more common way of referring to this cancellation. Rather than derive the precise form of A and B in (S.17), we maintain the fixed-n approach by stabilizing $\boldsymbol{e}_\nu'\tilde{\boldsymbol{\Gamma}}^{-1}\tilde{\boldsymbol{\Lambda}}_k$ for interior points when needed. This has the dual the advantages of easy implementability (using the sample, non-tilde versions) and capturing all terms. We will thus write

$$\Psi_{T_p,F} = \sqrt{nh^{1+2\nu}}h^{-\nu}h^{p+2}\nu!\boldsymbol{e}_\nu'\tilde{\boldsymbol{\Gamma}}^{-1}\left(h^{-1}\tilde{\boldsymbol{\Lambda}}_1\frac{\mu^{(p+1)}}{(p+1)!}+\tilde{\boldsymbol{\Lambda}}_2\frac{\mu^{(p+2)}}{(p+2)!}\right)\left[1+o(1)\right].$$

Case 2: $\boldsymbol{p+1=S}$. We can no longer retain the second term above, because $\mu^{(p+2)}$ does not exist. Instead we find that

$$\Psi_{T_p,F} = \sqrt{nh^{1+2\nu}} h^{-\nu} h^{p+1} \nu! \boldsymbol{e}_\nu' \tilde{\boldsymbol{\Gamma}}^{-1} \left(\tilde{\boldsymbol{\Lambda}}_1 \frac{\mu^{(p+1)}}{(p+1)!} + O(h^s) \right) [1 + o(1)] .$$

The same symmetry still applies to the first term however, and thus we have

$$\Psi_{T_p,F} = \sqrt{nh^{1+2\nu}} h^{-\nu} h^{p+1+s} \nu! \boldsymbol{e}_\nu' \tilde{\boldsymbol{\Gamma}}^{-1} \left(h^{1-s} h^{-1} \tilde{\boldsymbol{\Lambda}}_1 \frac{\mu^{(p+1)}}{(p+1)!} + O(1) \right) [1 + o(1)] ,$$

but since $s \leq 1$, the second term is (part of) the leading form, and we therefore write

$$\Psi_{T_p,F} = \sqrt{nh^{1+2\nu}} h^{-\nu} O(h^{p+1+s}) = \sqrt{nh^{1+2\nu}} h^{-\nu} O(h^{S+s}),$$

with the final equality holding because, by assumption, $p+1=S$ in this case.

Case 3: $\boldsymbol{p \geq S}$. All that is left in Equation (S.16) is this remainder term, and we therefore have

$$\Psi_{T_p,F} = \sqrt{nh^{1+2\nu}} h^{-\nu} O(h^{S+s}) = O(\sqrt{nh} h^{S+s}),$$

and cannot say anything further regarding constants. This result applies any time $p \geq S$, regardless of ν, parity of $p-\nu$, and at interior and boundary points.

S.2.3 Post Bias Correction: Specific Cases and Leading Terms

The general case, from Equation (S.15), which appears in the Edgeworth expansion is

$$\begin{aligned}
\Psi_{\mathtt{rbc},F} &= \sqrt{nh}\ \nu! \boldsymbol{e}_\nu' \tilde{\boldsymbol{\Gamma}}^{-1} \Big(\mathbb{E}\left[\boldsymbol{\Omega} \boldsymbol{B}_{p+1}\right] - \rho^{p+1} \tilde{\boldsymbol{\Lambda}} \boldsymbol{e}_{p+1}' \tilde{\bar{\boldsymbol{\Gamma}}}^{-1} \mathbb{E}\left[\bar{\boldsymbol{\Omega}} \boldsymbol{B}_{p+1}\right] \Big) \\
&= \sqrt{nh}\ \nu! \boldsymbol{e}_\nu' \tilde{\boldsymbol{\Gamma}}^{-1} \mathbb{E}\Big[\Big\{ h^{-1} (K \boldsymbol{r}_p)(X_{h,i}) - \rho^{p+1} \tilde{\boldsymbol{\Lambda}}_1 \boldsymbol{e}_{p+1}' \tilde{\bar{\boldsymbol{\Gamma}}}^{-1} b^{-1} (K \boldsymbol{r}_{p+1})(X_{b,i}) \Big\} \\
&\qquad\qquad \times \big(\mu(X_i) - \boldsymbol{r}_{p+1}(X_i - \mathsf{x})' \boldsymbol{\beta}_{p+1} \big) \Big].
\end{aligned}$$

It is always true that the rate is captured by a function $t(\cdot,\cdot)$ such that

$$\Psi_{\mathtt{rbc},F} = O(\sqrt{nh}\ t(h,b)),$$

or if ρ is bounded and bounded away from zero, the rate is captured by the exponent ζ such that

$$\Psi_{\mathtt{rbc},F} = O(\sqrt{nh} h^{\zeta}).$$

Location of x	Parity of $p-\nu$	Smoothness	Rate $t(h,b)$	ρ bounded above 0, below ∞	
				ζ	$\psi_{\texttt{rbc},F}$
Boundary	odd or even	$p+2 \leq S$	$h^{p+2}(1+\rho^{-1})$	$p+2$	(S.18a)
		$p+2 > S$	$h^{S+s}[1+\rho^{p+1-S-s}]$	$S+s$	N/A
Interior	even	$p+2 \leq S$	h^{p+2}	$p+2$	(S.18b)
		$p+2 > S$	$h^{S+s}\left[1+\rho^{p+1-S-s}\right]$	$S+s$	N/A
	odd	$p+3 \leq S$	$h^{p+3}(1+\rho^{-2})$	$p+3$	(S.18c)
		$p+2 = S$	$h^{p+2+s}[1+\rho^{-1-s}]$	$p+2+s = S+s$	N/A
		$p+2 > S$	$h^{S+s}\left[1+\rho^{p+1-S-s}\right]$	$S+s$	N/A

Table S.2: Summary of Bias Terms in All Cases For Bias-Corrected Centering $\hat{\theta}_{\texttt{rbc}}$. Rate function $t(h,b)$ is such that $\Psi_{\texttt{rbc},F} = O(\sqrt{nh}\, t(h,b))$. If ρ is bounded and bounded away from zero then we can take $t(h,b) = h^{\zeta}$. When possible, $\psi_{T_p,F}$ is such that $\Psi_{\texttt{rbc},F} = \sqrt{nh}h^{\zeta}\psi_{\texttt{rbc},F}[1+o(1)]$.

Additionally, if p is small enough relative to S, then we write

$$\Psi_{\texttt{rbc},F} = \sqrt{nh}h^{\zeta}\psi_{\texttt{rbc},F}[1+o(1)],$$

and call $\psi_{\texttt{rbc},F}$ the leading constant. Recall that $\psi_{\texttt{rbc},F}$ is not truly constant, but rather a nonrandom sequence that is $O(1)$ uniformly in $\mathscr{F}_S$ and is nonzero for some $F \in \mathscr{F}_S$. Table S.2 is complete list of the results, including $t(h,b)$), and where possible, ζ and $\psi_{T_p,F}$. These cases are derived in the rest of this section.

$$\psi_{\texttt{rbc},F} \text{ in Table S.2 can be } \begin{cases} \dfrac{\mu^{(p+2)}}{(p+2)!}\nu! \boldsymbol{e}_\nu' \tilde{\boldsymbol{\Gamma}}^{-1}\left\{\tilde{\boldsymbol{\Lambda}}_2 - \rho^{-1}\tilde{\boldsymbol{\Lambda}}_1 \boldsymbol{e}_{p+1}' \bar{\tilde{\boldsymbol{\Gamma}}}^{-1}\bar{\tilde{\boldsymbol{\Lambda}}}_1\right\}, & \text{(S.18a)} \\ \dfrac{\mu^{(p+2)}}{(p+2)!}\nu! \boldsymbol{e}_\nu' \tilde{\boldsymbol{\Gamma}}^{-1}\tilde{\boldsymbol{\Lambda}}_2, \quad \text{or} & \text{(S.18b)} \\ \nu! \boldsymbol{e}_\nu' \tilde{\boldsymbol{\Gamma}}^{-1}\left\{ \dfrac{\mu^{(p+2)}}{(p+2)!}\left[h^{-1}\tilde{\boldsymbol{\Lambda}}_2 - \rho^{-2}b^{-1}\tilde{\boldsymbol{\Lambda}}_1 \boldsymbol{e}_{p+1}' \bar{\tilde{\boldsymbol{\Gamma}}}^{-1}\bar{\tilde{\boldsymbol{\Lambda}}}_1\right] \right. \\ \qquad \left. + \dfrac{\mu^{(p+3)}}{(p+3)!}\left[\tilde{\boldsymbol{\Lambda}}_3 - \rho^{-2}\tilde{\boldsymbol{\Lambda}}_1 \boldsymbol{e}_{p+1}' \bar{\tilde{\boldsymbol{\Gamma}}}^{-1}\bar{\tilde{\boldsymbol{\Lambda}}}_2\right]\right\}, & \text{(S.18c)} \end{cases}$$

The starting point of all the derivations is again a Taylor approximation. We use Equation (S.16) with different choices for the bandwidth and polynomial degree. It will be useful at times to consider the two terms of $\psi_{\texttt{rbc},F}$ in Equation (S.15) separately, as the bandwidths h and b may be different and even vanish at different rates. The two terms represent (i) the second bias term of $\hat{\mu}_p^{(\nu)}$, not targeted by bias correction, and (ii) the bias of the bias estimator. For discussion in the context of kernel-based density estimation, see Hall (1992b) and Calonico et al. (2018a,b). See the latter also for bias correction using a generic polynomial of degree $q \geq p+1$; here we maintain degree $p+1$ for bias correction throughout.

The two terms of $\psi_{\mathtt{rbc},F}$ in Equation (S.15) are separated appropriately in Equation (S.14). We will resume there and apply the Taylor expansion Equation (S.16) with $p+1$ in place of p and, for the second term of (S.14), also with b in place of h. Doing this, assuming for the present sufficient smoothness, and applying the definitions of $\mathbf{\Lambda}_k$ and $\bar{\mathbf{\Lambda}}_k$ and their respective fixed-n expectations, we have,

$$\begin{aligned}
&\mathbb{E}\left[\hat{\theta}_{\mathtt{rbc}} \middle| X_1, \ldots, X_n\right] - \mu^{(\nu)} \\
&= h^{-\nu}\nu! \boldsymbol{e}_\nu' \mathbf{\Gamma}^{-1} \frac{1}{nh}\sum_{i=1}^{n} (K\boldsymbol{r}_p)(X_{h,i}) \left(\mu(X_i) - \boldsymbol{r}_{p+1}(X_i - \mathsf{x})'\boldsymbol{\beta}_{p+1}\right) \\
&\quad - h^{-\nu}\rho^{p+1}\nu!\boldsymbol{e}_\nu'\mathbf{\Gamma}^{-1}\mathbf{\Lambda}_1\boldsymbol{e}_{p+1}'\bar{\mathbf{\Gamma}}^{-1} \times \frac{1}{nb}\sum_{i=1}^{n}(K\boldsymbol{r}_{p+1})(X_{b,i})\left(\mu(X_i) - \boldsymbol{r}_{p+1}(X_i - \mathsf{x})'\boldsymbol{\beta}_{p+1}\right) \\
&= h^{-\nu}\nu!\boldsymbol{e}_\nu'\mathbf{\Gamma}^{-1}\left(h^{p+2}\mathbf{\Lambda}_2\frac{\mu^{(p+2)}}{(p+2)!} + h^{p+3}\mathbf{\Lambda}_3\frac{\mu^{(p+3)}}{(p+3)!}\right)[1 + o_{\mathbb{P}}(1)] \\
&\quad - h^{-\nu}\rho^{p+1}\nu!\boldsymbol{e}_\nu'\mathbf{\Gamma}^{-1}\mathbf{\Lambda}_1\boldsymbol{e}_{p+1}'\bar{\mathbf{\Gamma}}^{-1}\left(b^{p+2}\bar{\mathbf{\Lambda}}_1\frac{\mu^{(p+2)}}{(p+2)!} + b^{p+3}\bar{\mathbf{\Lambda}}_2\frac{\mu^{(p+3)}}{(p+3)!}\right)[1 + o_{\mathbb{P}}(1)].
\end{aligned}$$

Collecting terms and replacing sample averages with expectations, we arrive at

$$\begin{aligned}
&= h^{p+2-\nu}\nu!\boldsymbol{e}_\nu'\tilde{\mathbf{\Gamma}}^{-1}\Bigg\{\frac{\mu^{(p+2)}}{(p+2)!}\left(\tilde{\mathbf{\Lambda}}_2 - \rho^{-1}\tilde{\mathbf{\Lambda}}_1\boldsymbol{e}_{p+1}'\tilde{\bar{\mathbf{\Gamma}}}^{-1}\tilde{\bar{\mathbf{\Lambda}}}_1\right) \\
&\qquad + \frac{\mu^{(p+3)}}{(p+3)!}\left(h\tilde{\mathbf{\Lambda}}_3 - \rho^{-1}b\tilde{\mathbf{\Lambda}}_1\boldsymbol{e}_{p+1}'\tilde{\bar{\mathbf{\Gamma}}}^{-1}\tilde{\bar{\mathbf{\Lambda}}}_2\right)\Bigg\}[1 + o_{\mathbb{P}}(1)]
\end{aligned} \tag{S.19}$$

This final form will serve as the starting point for the special cases that follow.

S.2.3.1 Boundary Point

Here parity does not matter. Therefore we need only the first term of (S.19), containing $\mu^{(p+2)}$. It matters only if there is sufficient smoothness.

Case 1: $p+2 \leq S$. The first term of (S.19) exists and dominates others if they exist, and so

$$\Psi_{\mathtt{rbc},F} = \sqrt{nh^{1+2\nu}}h^{-\nu}h^{p+2}\frac{\mu^{(p+2)}}{(p+2)!}\nu!\boldsymbol{e}_\nu'\tilde{\mathbf{\Gamma}}^{-1}\left\{\tilde{\mathbf{\Lambda}}_2 - \rho^{-1}\tilde{\mathbf{\Lambda}}_1\boldsymbol{e}_{p+1}'\tilde{\bar{\mathbf{\Gamma}}}^{-1}\tilde{\bar{\mathbf{\Lambda}}}_1\right\}[1 + o(1)].$$

Case 2: $p+2 > S$. In this case $\mu^{(p+2)}$ does not exist, and therefore

$$\Psi_{\mathtt{rbc},F} = \sqrt{nh^{1+2\nu}}h^{-\nu}\left(O(h^{S+s}) + \rho^{p+1}O(b^{S+s})\right)$$

$$= O\left(\sqrt{nh}h^{S+s}[1 + \rho^{p+1-S-s}]\right).$$

The final rate depends on p and ρ in three cases: (i) if ρ is bounded and bounded away from zero, then $\rho^{p+1-S-s} \asymp 1$ and $\Psi_{\mathtt{rbc},F} = O\left(\sqrt{nh}h^{S+s}\right)$; (ii) the same rate is obtained if $\rho \to 0$ and $p+1 > S$, because, since $p \geq S$ and $1 \geq s$, the exponent on ρ is positive and, with ρ bounded, $\Psi_{\mathtt{rbc},F} = O\left(\sqrt{nh}h^{S+s}\right)$; (iii) if $\rho \to 0$ and $p+1 = S$, then the second term is $\rho^{-s} \to \infty$, thus $\Psi_{\mathtt{rbc},F} = O\left(\sqrt{nh}h^{S+s}\rho^{-s}\right)$.

S.2.3.2 Interior Point: $p - \nu$ odd

Cancellations due to symmetry will occur here as well, even though the initial centering uses $p - \nu$ odd, because bias correction involves $p+1-\nu$, which is even. Again we will have three smoothness cases, though we will find the same result for the latter two.

The analogue of Equation (S.17) for the bias correction is

$$\boldsymbol{e}_\nu' \tilde{\bar{\boldsymbol{\Gamma}}}^{-1} \tilde{\bar{\boldsymbol{\Lambda}}}_k = \bar{A} + b\bar{B} + o(b), \text{ with } \bar{A} = 0 \text{ if } (p+1+k-\nu) \text{ is odd and } \mathsf{x} \text{ is in the interior.} \qquad \text{(S.20)}$$

We will use this along with (S.17); both matter here because $\hat{\theta}_{\mathtt{rbc}}$ involves both $\hat{\mu}_p^{(\nu)}$ and $\hat{\mu}_{p+1}^{(p+1)}$.

Case 1: $\boldsymbol{p+3 \leq S}$. Starting with the formula for $\Psi_{\mathtt{rbc},F}$ at the boundary given above, Equations (S.17) and (S.20) yield $\boldsymbol{e}_\nu' \tilde{\boldsymbol{\Gamma}}^{-1} \tilde{\boldsymbol{\Lambda}}_2 = O(h)$ and $\boldsymbol{e}_{p+1}' \tilde{\bar{\boldsymbol{\Gamma}}}^{-1} \tilde{\bar{\boldsymbol{\Lambda}}}_1 = O(b)$. Therefore, these are the same order as the appropriate "next" term in the expansion (S.19), i.e. one further derivative must be retained. This is possible with $p+3 \leq S$.

Applying this to $\Psi_{\mathtt{rbc},F}$, we find that

$$\begin{aligned}\Psi_{\mathtt{rbc},F} = \sqrt{nh}h^{p+3}\ \nu! \boldsymbol{e}_\nu' \tilde{\boldsymbol{\Gamma}}^{-1} \Bigg\{ & \frac{\mu^{(p+2)}}{(p+2)!} \Big[h^{-1} \tilde{\boldsymbol{\Lambda}}_2 - \rho^{-2} b^{-1} \tilde{\boldsymbol{\Lambda}}_1 \boldsymbol{e}_{p+1}' \tilde{\bar{\boldsymbol{\Gamma}}}^{-1} \tilde{\bar{\boldsymbol{\Lambda}}}_1 \Big] \\ & + \frac{\mu^{(p+3)}}{(p+3)!} \Big[\tilde{\boldsymbol{\Lambda}}_3 - \rho^{-2} \tilde{\boldsymbol{\Lambda}}_1 \boldsymbol{e}_{p+1}' \tilde{\bar{\boldsymbol{\Gamma}}}^{-1} \tilde{\bar{\boldsymbol{\Lambda}}}_2 \Big] \Bigg\} \ [1 + o(1)].\end{aligned}$$

Notice that rather than spell out the limiting form of $\boldsymbol{e}_\nu' \tilde{\boldsymbol{\Gamma}}^{-1} \tilde{\boldsymbol{\Lambda}}_2$ and $\boldsymbol{e}_{p+1}' \tilde{\bar{\boldsymbol{\Gamma}}}^{-1} \tilde{\bar{\boldsymbol{\Lambda}}}_1$, that is, the $\boldsymbol{C}_2$ and $\bar{\boldsymbol{C}}_2$ above, we keep with the fixed-n spirit and write $h^{-1} \boldsymbol{e}_\nu' \tilde{\boldsymbol{\Gamma}}^{-1} \tilde{\boldsymbol{\Lambda}}_2$ and $b^{-1} \boldsymbol{e}_{p+1}' \tilde{\bar{\boldsymbol{\Gamma}}}^{-1} \tilde{\bar{\boldsymbol{\Lambda}}}_1$, which dual the advantages of easy implementability (using the sample, non-tilde versions) and capturing all terms.

Case 2: $\boldsymbol{p+2 = S}$. The terms above involving $\mu^{(p+3)}$ must be replaced by the $O(h^{S+s})$ (or b^{S+s}) term of (S.16), which if $p+2 = S$, leaves the exponent as $p+2+s$. This gives

$$\begin{aligned}\Psi_{\mathtt{rbc},F} = \sqrt{nh}h^{p+3}\ \nu! \boldsymbol{e}_\nu' \tilde{\boldsymbol{\Gamma}}^{-1} \Bigg\{ & \frac{\mu^{(p+2)}}{(p+2)!} \Big[h^{-1} \tilde{\boldsymbol{\Lambda}}_2 - \rho^{-2} b^{-1} \tilde{\boldsymbol{\Lambda}}_1 \boldsymbol{e}_{p+1}' \tilde{\bar{\boldsymbol{\Gamma}}}^{-1} \tilde{\bar{\boldsymbol{\Lambda}}}_1 \Big] \Bigg\} \\ & + O\left(\sqrt{nh}h^{p+2+s}\right) + O\left(\sqrt{nh}\rho^{p+1} b^{p+2+s}\right)\end{aligned}$$

$$
\begin{aligned}
&= \sqrt{nh} h^{p+3}\ \nu! \boldsymbol{e}_\nu' \tilde{\boldsymbol{\Gamma}}^{-1} \left\{ \frac{\mu^{(p+2)}}{(p+2)!} \left[h^{-1} \tilde{\boldsymbol{\Lambda}}_2 - \rho^{-2} b^{-1} \tilde{\boldsymbol{\Lambda}}_1 \boldsymbol{e}_{p+1}' \tilde{\tilde{\boldsymbol{\Gamma}}}^{-1} \tilde{\tilde{\boldsymbol{\Lambda}}}_1 \right] \right\} \\
&\quad + O\left(\sqrt{nh} h^{p+2+s} [1 + \rho^{-1-s}] \right).
\end{aligned}
$$

(Note that the order of second term is equivalently $\sqrt{nh} h^{S+s}[1+\rho^{-1-s}]$.) Recall that $s \in (0,1]$. Therefore the first term above is higher order unless $s = 1$ (which is not known) and $\rho \to \bar{\rho} \in (0,\infty)$, in which case the two are of the same order. Otherwise, the second term dominates, and further, if the ρ^{-1-s} portion is the dominant rate if $\rho = h/b \to 0$ regardless of s. Therefore in this case it is more clear to suppress the constants of the higher order term and write

$$
\Psi_{\texttt{rbc},F} = O\left(\sqrt{nh} h^{p+2+s} [1 + \rho^{-1-s}] \right).
$$

Case 3: $\boldsymbol{p+2 > S}$. Now the symmetry does not apply (because only when the derivatives exist do the Taylor series terms collapse to $\boldsymbol{\Lambda}_k$ and $\bar{\boldsymbol{\Lambda}}_k$) and so we find that $\Psi_{\texttt{rbc},F} = O\left(\sqrt{nh} \left[h^{S+s} + \rho^{p+1} b^{S+s} \right] \right) = O\left(\sqrt{nh} h^{S+s} \left[1 + \rho^{p+1-S-s} \right] \right)$.

S.2.3.3 Interior Point: $p - \nu$ even

Case 1: $\boldsymbol{p+3 \leq S}$. The conditions for $A = 0$ and $\bar{A} = 0$ in Equations (S.17) and (S.20) reduce to whether or not k is odd, because $p - \nu$ is even for the former and the latter is always applied with $\nu = p+1$. Using this to add the stabilization needed to Equation (S.19) yields

$$
\begin{aligned}
&\mathbb{E}\left[\hat{\theta}_{\texttt{rbc}} \middle| X_1, \ldots, X_n \right] - \mu^{(\nu)} \\
&= h^{p+2-\nu} \nu! \boldsymbol{e}_\nu' \tilde{\boldsymbol{\Gamma}}^{-1} \left\{ \frac{\mu^{(p+2)}}{(p+2)!} \left(\tilde{\boldsymbol{\Lambda}}_2 - \rho^{-1} h h^{-1} \tilde{\boldsymbol{\Lambda}}_1 b b^{-1} \boldsymbol{e}_{p+1}' \tilde{\tilde{\boldsymbol{\Gamma}}}^{-1} \tilde{\tilde{\boldsymbol{\Lambda}}}_1 \right) \right. \\
&\qquad\qquad \left. + \frac{\mu^{(p+3)}}{(p+3)!} \left(h^2 h^{-1} \tilde{\boldsymbol{\Lambda}}_3 - \rho^{-1} h h^{-1} b \tilde{\boldsymbol{\Lambda}}_1 \boldsymbol{e}_{p+1}' \tilde{\tilde{\boldsymbol{\Gamma}}}^{-1} \tilde{\tilde{\boldsymbol{\Lambda}}}_2 \right) \right\} [1 + o_{\mathbb{P}}(1)] \\
&= h^{p+2-\nu} \nu! \boldsymbol{e}_\nu' \tilde{\boldsymbol{\Gamma}}^{-1} \left\{ \frac{\mu^{(p+2)}}{(p+2)!} \left(\tilde{\boldsymbol{\Lambda}}_2 - b^2 \left[h^{-1} \tilde{\boldsymbol{\Lambda}}_1 b^{-1} \boldsymbol{e}_{p+1}' \tilde{\tilde{\boldsymbol{\Gamma}}}^{-1} \tilde{\tilde{\boldsymbol{\Lambda}}}_1 \right] \right) \right. \\
&\qquad\qquad \left. + \frac{\mu^{(p+3)}}{(p+3)!} \left(h^2 \left[h^{-1} \tilde{\boldsymbol{\Lambda}}_3 \right] - b^2 \left[h^{-1} \tilde{\boldsymbol{\Lambda}}_1 \boldsymbol{e}_{p+1}' \tilde{\tilde{\boldsymbol{\Gamma}}}^{-1} \tilde{\tilde{\boldsymbol{\Lambda}}}_2 \right] \right) \right\} [1 + o_{\mathbb{P}}(1)].
\end{aligned}
$$

Therefore

$$
\Psi_{\texttt{rbc},F} = \sqrt{nh} h^{p+2} \frac{\mu^{(p+2)}}{(p+2)!} \nu! \boldsymbol{e}_\nu' \tilde{\boldsymbol{\Gamma}}^{-1} \tilde{\boldsymbol{\Lambda}}_2 [1 + o(1)].
$$

To build intuition for why this result is correct, recall that if $\rho = 1$ then $\hat{\theta}_{\texttt{rbc}} = \hat{\mu}_p^{(\nu)} - h^{p+1-\nu} \nu! \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1} \boldsymbol{\Lambda}_1 \frac{\hat{\mu}_{p+1}^{(p+1)}}{(p+1)!} = \hat{\mu}_{p+1}^{(\nu)}$, that is, $\hat{\theta}_{\texttt{rbc}}$ is equivalent to fitting a $p+1$ degree local polynomial rather than p. Here we are working under $p - \nu$ even and so $p + 1 - \nu$ is odd, and so naturally we recover the standard result for odd degree local polynomials.

Case 2: $\boldsymbol{p+2=S}$. The terms above involving $\mu^{(p+3)}$ must be replaced by the $O(h^{S+s})$ (or b^{S+s}) term of (S.16), which if $p+2=S$, leaves the exponent as $p+2+s$. Thus the leading term on the right of Equation (S.19) becomes

$$h^{p+2-\nu}\nu!\boldsymbol{e}_\nu'\tilde{\boldsymbol{\Gamma}}^{-1}\left\{\frac{\mu^{(p+2)}}{(p+2)!}\left(\tilde{\boldsymbol{\Lambda}}_2-\rho^{-1}\tilde{\boldsymbol{\Lambda}}_1\boldsymbol{e}_{p+1}'\tilde{\bar{\boldsymbol{\Gamma}}}^{-1}\tilde{\bar{\boldsymbol{\Lambda}}}_1\right)+O(h^s+\rho^{-1}b^s)\right\}.$$

The same symmetry applies as in the previous case, and therefore we still have

$$\Psi_{\texttt{rbc},F}=\sqrt{nh}h^{p+2}\frac{\mu^{(p+2)}}{(p+2)!}\nu!\boldsymbol{e}_\nu'\tilde{\boldsymbol{\Gamma}}^{-1}\tilde{\boldsymbol{\Lambda}}_2[1+o(1)].$$

Case 3: $\boldsymbol{p+2>S}$. Now the symmetry does not apply (because only when the derivatives exist do the Taylor series terms collapse to $\boldsymbol{\Lambda}_k$ and $\bar{\boldsymbol{\Lambda}}_k$) and so we find that $\Psi_{\texttt{rbc},F}=O\left(\sqrt{nh}\left[h^{S+s}+\rho^{p+1}b^{S+s}\right]\right)=O\left(\sqrt{nh}h^{S+s}\left[1+\rho^{p+1-S-s}\right]\right)$.

S.3 Main Theoretical Results

We now give the main technical result of the paper, a uniformly (in $F\in\mathscr{F}_S$) valid Edgeworth expansion of the distribution function of a generic local polynomial based t-statistic, from which coverage error follows for any I. This result is the same as Theorem 3 in the main text. Recall that the expansions are for the $t-$ statistics defined in Section S.1, of which we are mainly interested in those from Equation (S.11):

$$T_p=\frac{\sqrt{nh^{1+2\nu}}(\hat{\mu}_p^{(\nu)}-\mu^{(\nu)})}{\hat{\sigma}_p}\qquad\text{and}\qquad T_{\texttt{rbc}}=\frac{(\hat{\theta}_{\texttt{rbc}}-\mu^{(\nu)})}{\hat{\vartheta}_{\texttt{rbc}}}=\frac{\sqrt{nh^{1+2\nu}}(\hat{\theta}_{\texttt{rbc}}-\mu^{(\nu)})}{\hat{\sigma}_{\texttt{rbc}}}.$$

Appearing the expansion are the six functions $\omega_{k,T,F}(z)$, $k=1,2,\ldots,6$, whose exact forms are computed in Section S.3.1. All six are known for all I and $F\in\mathscr{F}_S$, bounded, and bounded away from zero for at least some $F\in\mathscr{F}_S$, and most crucially, that ω_1, ω_2, and ω_3 are *even* functions of z, while ω_4, ω_5, and ω_6 are *odd*.

Also appearing is $\lambda_{T,F}$, a generic placeholder capturing the mismatch between the variance of the numerator of the t-statistic and the population standardization chosen (i.e. the quantity estimated by $\hat{\sigma}$ of T). We can not make this error precise for all choices, but give two important special cases. First, employing an estimate of the asymptotic variance renders $\lambda_{T,F}=O(h)$ at boundary points. Second, the fixed-n Studentizations (S.10) yield $\lambda_{T,F}\equiv 0$. For other choices, the rates and constants may change, but it is important to point out that the coverage error rate cannot be improved beyond the others shown through the choice of Studentization alone (see discussion in Section S.7 and Calonico et al. (2018a,b)). Let λ_T such that $\sup_{F\in\mathscr{F}_S}\lambda_{T,F}=O(\lambda_T)=o(1)$ (if this were not $o(1)$, the expansion is not valid as the t-statistic does not converge to standard Normal). Like the bias terms, $\lambda_{T,F}\equiv\lambda_{I,F}$, and here we maintain the former only because the theorem is directly for T.

Theorem S.1. *Let Assumptions S.1 and S.2 hold, and suppose*

$$nh/\log(nh)^{2+\gamma} \to \infty \qquad \text{and} \qquad \Psi_{T,F}\log(nh)^{1+\gamma} \to 0,$$

for any γ bounded away from zero uniformly in $\mathscr{F}_S$, and if bias correction is used, $p-\nu$ is odd and ρ is bounded uniformly in $\mathscr{F}_S$. Then

$$\sup_{F\in\mathscr{F}_S}\sup_{z\in\mathbb{R}} r_T^{-1}\Big|\mathbb{P}_F\left[T<z\right]-\Phi(z)-E_{T,F}(z)\Big| = o\left(1\right),$$

where $\Phi(z)$ is the standard Normal distribution function, $r_T = \max\{(nh)^{-1}, \Psi_{T,F}^2, (nh)^{-1/2}\Psi_{T,F}, \lambda_T\}$, i.e. the slowest vanishing of the rates, and

$$\begin{aligned} E_{T,F}(z) = &\frac{1}{\sqrt{nh}}\omega_{1,T,F}(z) + \Psi_{T,F}\omega_{2,T,F}(z) + \lambda_{T,F}\omega_{3,T,F}(z) \\ &+ \frac{1}{nh}\omega_{4,T,F}(z) + \Psi_{T,F}^2\omega_{5,T,F}(z) + (nh)^{-1/2}\Psi_{T,F}\omega_{6,T,F}(z). \end{aligned} \tag{S.21}$$

with $\omega_{k,T,F}(z)$, $k=1,2,\ldots,6$, bounded uniformly in $\mathscr{F}_S$, and bounded away from zero for at least some F.

This Theorem is the main technical result which allows the deduction of all others. At this level of generality, $p-\nu$ may be odd or even for $T=T_p$. The terms of the expansion are computed in the following subsection. We then turn our attention over to the proof of this result.

The requirement that $\Psi_{T,F}\log(nh)^{1+\gamma}\to 0$ is a high-level condition on the bandwidth(s) h (and b). If ρ is additionally bounded away from zero, it can be simplified to the more transparent $\sqrt{nh}h^{\zeta}\log(nh)^{1+\gamma}$, where ζ can be read off Tables S.1 or S.2, as discussed there.

S.3.1 Terms of the Edgeworth Expansion

We now give the precise forms of the terms in the Edgeworth expansion, $E_{T,F}(z)$. This amounts to defining the terms ω_k, $k=1,2,\ldots,6$, $\Psi_{T,F}$, and $\lambda_{T,F}$. For all T (or I), $\Psi_{T,F}$ is given in Section S.2 and explicitly given in Equation (S.15). For the expansion, the special cases are not needed. For the variance errors $\lambda_{T,F}$, we mention a few examples. First, as already discussed, the fixed-n standard errors of Equation (S.10) yield $\lambda_{T,F}\equiv 0$. When it is nonzero, typically $\lambda_{T,F}$ has the form $\lambda_{T,F}=l_nL$, for a rate $l_n\to 0$ and a constant (or at least, a sequence bounded and bounded away from zero) L. The term L is exactly the difference between the variance of the numerator of the t-statistic and the population standardization chosen. This has nothing to do with estimation error. Loosely speaking,

$$L = \frac{\mathbb{V}\left[\sqrt{nh^{1+2\nu}}(\hat{\theta}-\mu^{(\nu)})\right]}{\sigma^2} - 1,$$

where σ^2 is the limit of Studentization whatever $\hat{\sigma}^2$ has been chosen (c.f. Equation (S.11)). As an example, consider traditional explicit bias correction, where the point estimate (or numerator of T) is bias-corrected but it is assumed that σ_p provides valid standardization (this requires $\rho \to 0$), we find that $\lambda_{T,F} = \rho^{p+2}(L_1 + \rho^{p+2}L_2)$, where L_1 captures the (scaled) covariance between $\hat{\mu}^{(\nu)}$ and $\hat{\mu}^{(p+1)}$ and L_2 the variance of $\hat{\mu}^{(p+1)}$; see Calonico et al. (2018a,b) for the exact expressions. For another example, for inference at the boundary when using the asymptotic variance for standardization (i.e. the probability limit of the conditional variance of the numerator), one finds $l_n = h$ and L capturing the difference between the conditional variance and its limit, based on the localization of the kernel; see Chen and Qin (2002) for the exact expression.

It remains to define ω_k, $k = 1, 2, \ldots, 6$. More notation is required. As with the bias, all terms must be nonrandom. We will maintain, as far as possible, fixed-n calculations. First, define the following functions, which depend on F, 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.

$$
\begin{aligned}
\ell^0_{T_p}(X_i) &= \nu! \boldsymbol{e}_\nu' \tilde{\boldsymbol{\Gamma}}^{-1} (K\boldsymbol{r}_p)(X_{h,i}); \\
\ell^0_{T_{\mathtt{rbc}}}(X_i) &= \ell^0_{T_p}(X_i) - \rho^{p+1}\nu! \boldsymbol{e}_\nu' \tilde{\boldsymbol{\Gamma}}^{-1} \tilde{\boldsymbol{\Lambda}}_1 \boldsymbol{e}_{p+1}' \tilde{\bar{\boldsymbol{\Gamma}}}^{-1} (K\boldsymbol{r}_{p+1})(X_{b,i}); \\
\ell^1_{T_p}(X_i, X_j) &= \nu! \boldsymbol{e}_\nu' \tilde{\boldsymbol{\Gamma}}^{-1} \left(\mathbb{E}[(K\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,j})] - (K\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,j}) \right) \tilde{\boldsymbol{\Gamma}}^{-1} (K\boldsymbol{r}_p)(X_{h,i}); \\
\ell^1_{T_{\mathtt{rbc}}}(X_i, X_j) &= \ell^1_{T_p}(X_i, X_j) - \rho^{p+1}\nu! \boldsymbol{e}_\nu' \tilde{\boldsymbol{\Gamma}}^{-1} \Big\{ \left(\mathbb{E}[(K\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,j})] - (K\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,j}) \right) \tilde{\boldsymbol{\Gamma}}^{-1} \tilde{\boldsymbol{\Lambda}}_1 \boldsymbol{e}_{p+1}' \\
&\qquad + \Big((K\boldsymbol{r}_p)(X_{h,j}) X_{h,i}^{p+1} - \mathbb{E}[(K\boldsymbol{r}_p)(X_{h,j}) X_{h,i}^{p+1}] \Big) \boldsymbol{e}_{p+1}' \\
&\qquad + \tilde{\boldsymbol{\Lambda}}_1 \boldsymbol{e}_{p+1}' \tilde{\bar{\boldsymbol{\Gamma}}}^{-1} \left(\mathbb{E}[(K\boldsymbol{r}_{p+1}\boldsymbol{r}_{p+1}')(X_{b,j})] - (K\boldsymbol{r}_{p+1}\boldsymbol{r}_{p+1}')(X_{b,j}) \right) \Big\} \tilde{\bar{\boldsymbol{\Gamma}}}^{-1} (K\boldsymbol{r}_{p+1})(X_{b,i}).
\end{aligned}
$$

With this notation, define

$$\tilde{\sigma}_T^2 = \mathbb{E}[h^{-1}\ell^0_T(X)^2 v(X)].$$

We can also rewrite the bias terms using this notation as

$$\Psi_{T_p,F} = \sqrt{nh}\,\mathbb{E}\left[h^{-1}\ell^0_{T_p}(X_i)[\mu(X_i) - \boldsymbol{r}_p(X_i - x)'\boldsymbol{\beta}_p] \right]$$

and

$$\Psi_{\mathtt{rbc},F} = \sqrt{nh}\,\mathbb{E}\Big[h^{-1}\ell^0_{T_{\mathtt{rbc}}}(X_i)[\mu(X_i) - \boldsymbol{r}_{p+1}(X_i - x)'\boldsymbol{\beta}_{p+1}] \Big].$$

Now we can define the Edgeworth expansion polynomials ω_k, $k = 1, 2, \ldots, 6$. The standard Normal density is $\phi(z)$. The term ω_4 is the most cumbersome. Beginning with the others:

$$
\begin{aligned}
\omega_{1,T,F}(z) &= \phi(z)\tilde{\sigma}_T^{-3}\mathbb{E}\left[h^{-1}\ell^0_T(X_i)^3\varepsilon_i^3 \right] \left\{ (2z^2 - 1)/6 \right\}, \\
\omega_{2,T,F}(z) &= -\phi(z)\tilde{\sigma}_T^{-1}, \\
\omega_{3,T,F}(z) &= -\phi(z)\left\{ z/2 \right\}, \\
\omega_{5,T,F}(z) &= -\phi(z)\tilde{\sigma}_T^{-2}\left\{ z/2 \right\},
\end{aligned}
$$

$$\omega_{6,T,F}(z) = \phi(z)\tilde{\sigma}_T^{-4}\mathbb{E}[h^{-1}\ell_T^0(X_i)^3\varepsilon_i^3]\left\{z^3/3\right\}.$$

For ω_3, it is not quite as simple to state a generic version. Let $\tilde{\boldsymbol{G}}$ stand in for $\tilde{\boldsymbol{\Gamma}}$ or $\tilde{\bar{\boldsymbol{\Gamma}}}$, $\tilde{p}$ stand in for p or $p+1$, and d_n stand in for h or b, all depending on if $T = T_p$ or $T_{\texttt{rbc}}$. Note however, that h is still used in many places, in particular for stabilizing fixed-n expectations, for $T_{\texttt{rbc}}$. Indexes i, j, and k are always distinct (i.e. $X_{h,i} \neq X_{h,j} \neq X_{h,k}$).

$$\begin{aligned}
\omega_{4,T,F}(z) = {} & \phi(z)\tilde{\sigma}_T^{-6}\mathbb{E}\left[h^{-1}\ell_T^0(X_i)^3\varepsilon_i^3\right]^2\left\{z^3/3 + 7z/4 + \tilde{\sigma}_T^2 z(z^2-3)/4\right\} \\
& + \phi(z)\tilde{\sigma}_T^{-2}\mathbb{E}\left[h^{-1}\ell_T^0(X_i)\ell_T^1(X_i,X_i)\varepsilon_i^2\right]\left\{-z(z^2-3)/2\right\} \\
& + \phi(z)\tilde{\sigma}_T^{-4}\mathbb{E}\left[h^{-1}\ell_T^0(X_i)^4(\varepsilon_i^4 - v(X_i)^2)\right]\left\{z(z^2-3)/8\right\} \\
& - \phi(z)\tilde{\sigma}_T^{-2}\mathbb{E}\left[h^{-1}\ell_T^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\} \\
& - \phi(z)\tilde{\sigma}_T^{-4}\mathbb{E}\left[h^{-1}\ell_T^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})\ell_T^0(X_i)\varepsilon_i^2\right]\left\{z(z^2-1)\right\} \\
& + \phi(z)\tilde{\sigma}_T^{-2}\mathbb{E}\left[h^{-2}\ell_T^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\} \\
& + \phi(z)\tilde{\sigma}_T^{-4}\mathbb{E}\left[h^{-3}\ell_T^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})\ell_T^0(X_i)\boldsymbol{r}_{\tilde{p}}(X_{d_n,j})'\tilde{\boldsymbol{G}}^{-1}(K\boldsymbol{r}_{\tilde{p}})(X_{d_n,k})\ell_T^0(X_k)\varepsilon_i^2\varepsilon_k^2\right] \\
& \qquad\qquad \times \left\{z(z^2-1)/2\right\} \\
& + \phi(z)\tilde{\sigma}_T^{-4}\mathbb{E}\left[h^{-1}\ell_T^0(X_i)^4\varepsilon_i^4\right]\left\{-z(z^2-3)/24\right\} \\
& + \phi(z)\tilde{\sigma}_T^{-4}\mathbb{E}\left[h^{-1}\left(\ell_T^0(X_i)^2v(X_i) - \mathbb{E}[\ell_T^0(X_i)^2v(X_i)]\right)\ell_T^0(X_i)^2\varepsilon_i^2\right]\left\{z(z^2-1)/4\right\} \\
& + \phi(z)\tilde{\sigma}_T^{-4}\mathbb{E}\left[h^{-2}\ell_T^1(X_i,X_j)\ell_T^0(X_i)\ell_T^0(X_j)^2\varepsilon_j^2v(X_i)\right]\left\{z(z^2-3)\right\} \\
& + \phi(z)\tilde{\sigma}_T^{-4}\mathbb{E}\left[h^{-2}\ell_T^1(X_i,X_j)\ell_T^0(X_i)\left(\ell_T^0(X_j)^2v(X_j) - \mathbb{E}[\ell_T^0(X_j)^2v(X_j)]\right)\varepsilon_i^2\right]\left\{-z\right\} \\
& + \phi(z)\tilde{\sigma}_T^{-4}\mathbb{E}\left[h^{-1}\left(\ell_T^0(X_i)^2v(X_i) - \mathbb{E}[\ell_T^0(X_i)^2v(X_i)]\right)^2\right]\left\{-z(z^2+1)/8\right\}.
\end{aligned}$$

For computation, note that the seventh term 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

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

This will greatly ease implementation.

S.4 Proof of Theorem S.1 without Bias Correction

The goal of this section is to prove that the Edgeworth expansion of Theorem S.1 is valid for $T_p = T(\hat{\mu}_{p+1}^{(\nu)}, \hat{\sigma}_p^2/(nh^{1+2\nu}))$. The proof for $T_{\texttt{rbc}}$ is essentially the same from a conceptual and technical point of view, just with more notation and a repetition of the same steps, and so only a sketch is provided. See Section S.5. We also restrict to the fixed-n, HC0 standard errors of (S.10), which, in particular, render $\lambda_{T,F} \equiv 0$. Other possibilities are discussed in Section S.7. The terms of the expansion are computed, in a formal manner, in Section S.3.1.

For notational ease, we sometimes drop subscripts, along with the point of evaluation and/or dependence on F. Also define

- $s_n = \sqrt{nh}$

The proof consists of three main steps, which are tackled in the subsections below.

Step (I) – Section S.4.1

Show that

$$\mathbb{P}_F\left[T_p < z\right] = \mathbb{P}_F\left[\check{T} < z\right] + o\left((nh)^{-1} + (nh)^{-1/2}\Psi_{T_p,F} + \Psi^2_{T_p,F}\right), \tag{S.22}$$

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

Step (II) – Section S.4.1

Prove 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.

Step (III) – Section S.4.2

Prove that the expansion for T_p holds and that it holds uniformly over $F \in \mathscr{F}_S$.

Numerous intermediate results relied upon in the proof are collected as lemmas that are stated and proved in Section S.4.3.

Unless it is important to emphasize the dependence on F, this will be suppressed to save notation; for example $\mathbb{P} = \mathbb{P}_F$. Throughout proofs C shall be a generic conformable constant that may take different values in different places. If more than one constant is needed, C_1, C_2, ..., will be used. Also define

- $r_{T_p} = \max\{s_n^{-2}, \Psi^2_{T_p,F}, s_n^{-1}\Psi_{T_p,F}\}$, i.e. the slowest vanishing of the rates, and
- r_n as a generic sequence that obeys $r_n = o(r_{T_p})$.

We will frequently use the elementary probability bounds that for random A and B and positive fixed scalars a and b, $\mathbb{P}[|A + B| > a] \le \mathbb{P}[|A| > a/2] + \mathbb{P}[|B| > a/2]$ and $\mathbb{P}[|AB| > a] \le \mathbb{P}[|A| > b] + \mathbb{P}[|B| > a/b]$, also relying on the elementary bound $|AB| \le |A||B|$ for conformable vectors or matrixes A and B.

S.4.1 Step (I)

We now prove Equation (S.22) holds for suitable choices of $\check{T}$ and $\boldsymbol{Z}_i$. Notice that the "numerator" portion, $\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{Y} - \boldsymbol{R}\boldsymbol{\beta}_p\right)/n$ is already a smooth function of well-behaved random variables, and will thus be incorporated into $\check{T}$. Our difficulty lies with the Studentization, and in particular, the estimated residuals. We will start by expanding $\hat{\sigma}_p^2$ (see Equation (S.23)). Substituting this expansion into T_p, we will identify the leading terms, collected as appropriate into $\check{T}$ (Equation (S.25)) and $\boldsymbol{Z}_i$ (Equation (S.26)), and the remainder terms, collected in $U_n := T_p - \check{T}$ (Equation (S.24)). **Step (I)** is complete upon showing that U_n can be ignored in the expansion; this occupies the latter half of the present subsection.

To begin, recall that $\hat{\sigma}_p^2 = \nu!^2 \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1}(h\boldsymbol{\Omega}\hat{\boldsymbol{\Sigma}}_p\boldsymbol{\Omega}'/n)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu$. The matrix $\boldsymbol{\Gamma}^{-1}$, present in the numerator as well, enters smoothly and is itself smooth in elements of $s_n^{-1}\sum_{i=1}^n \boldsymbol{Z}_i$. Thus our focus is on the center matrix, $(h\boldsymbol{\Omega}\hat{\boldsymbol{\Sigma}}_p\boldsymbol{\Omega}'/n)$, which contains the estimated residuals. Using $\check{\boldsymbol{R}}\boldsymbol{H} = \boldsymbol{R}$ (and for each observation, $\boldsymbol{r}_p(X_i - \mathsf{x})\boldsymbol{H}^{-1} = \boldsymbol{r}_p(X_{h,i})$) and $\boldsymbol{\Gamma} = \boldsymbol{\Omega}\check{\boldsymbol{R}}/n$ we have

$$\boldsymbol{r}_p(X_i - \mathsf{x})'\hat{\boldsymbol{\beta}}_p = \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{H}^{-1}\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\boldsymbol{Y}/n = \boldsymbol{r}_p(X_{h,i})'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\boldsymbol{Y}/n$$

and

$$\boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p = \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{H}^{-1}\boldsymbol{\Gamma}^{-1}(\boldsymbol{\Omega}\check{\boldsymbol{R}}/n)\boldsymbol{H}\boldsymbol{\beta}_p = \boldsymbol{r}_p(X_{h,i})'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\boldsymbol{R}\boldsymbol{\beta}_p/n.$$

We use these forms to expand as follows:

$$\begin{aligned}
\frac{h}{n}\boldsymbol{\Omega}\hat{\boldsymbol{\Sigma}}_p\boldsymbol{\Omega}' &= \frac{1}{nh}\sum_{i=1}^n (K^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})\hat{v}(X_i) \\
&= \frac{1}{nh}\sum_{i=1}^n (K^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})\left(Y_i - \boldsymbol{r}_p(X_i - \mathsf{x})'\hat{\boldsymbol{\beta}}_p\right)^2 \\
&= \frac{1}{nh}\sum_{i=1}^n (K^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})\left(\varepsilon_i + \left[\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p\right] + \boldsymbol{r}_p(X_i - \mathsf{x})'\left[\boldsymbol{\beta}_p - \hat{\boldsymbol{\beta}}_p\right]\right)^2 \\
&= \frac{1}{nh}\sum_{i=1}^n (K^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})\left(\varepsilon_i + \left[\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p\right] - \boldsymbol{r}_p(X_{h,i})'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left[\boldsymbol{Y} - \boldsymbol{R}\boldsymbol{\beta}_p\right]/n\right)^2.
\end{aligned}$$

The expansion of $\hat{\sigma}_p^2$ is then

$$\hat{\sigma}_p^2 = \nu!^2\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\Big(\boldsymbol{V}_1 + 2\boldsymbol{V}_4 - 2\boldsymbol{V}_2 + \boldsymbol{V}_3 - 2\boldsymbol{V}_5 + \boldsymbol{V}_6\Big)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu \tag{S.23}$$

where

$$\begin{aligned}
\boldsymbol{V}_1 &= \frac{1}{nh}\sum_{i=1}^n (K^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})\varepsilon_i^2, \\
\boldsymbol{V}_2 &= \frac{1}{nh}\sum_{i=1}^n (K^2\boldsymbol{r}_p\boldsymbol{r}_p'\boldsymbol{r}_p')(X_{h,i})\varepsilon_i\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left[\boldsymbol{Y} - \boldsymbol{R}\boldsymbol{\beta}_p\right]/n, \\
\boldsymbol{V}_3 &= \frac{1}{nh}\sum_{i=1}^n (K^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})\left[\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p\right]^2,
\end{aligned}$$

$$
\begin{aligned}
\boldsymbol{V}_4 &= \frac{1}{nh}\sum_{i=1}^{n}(K^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})\left\{\varepsilon_i\left[\mu(X_i)-\boldsymbol{r}_p(X_i-\mathsf{x})'\boldsymbol{\beta}_p\right]\right\},\\
\boldsymbol{V}_5 &= \frac{1}{nh}\sum_{i=1}^{n}(K^2\boldsymbol{r}_p\boldsymbol{r}_p'\boldsymbol{r}_p')(X_{h,i})\left[\mu(X_i)-\boldsymbol{r}_p(X_i-\mathsf{x})'\boldsymbol{\beta}_p\right]\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left[\boldsymbol{Y}-\boldsymbol{R}\boldsymbol{\beta}_p\right]/n,\\
\boldsymbol{V}_6 &= \frac{1}{nh}\sum_{i=1}^{n}(K^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})\big\{\boldsymbol{r}_p(X_{h,i})'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left[\boldsymbol{Y}-\boldsymbol{R}\boldsymbol{\beta}_p\right]/n\big\}^2.
\end{aligned}
$$

With these terms in hand, define

- $s_n = \sqrt{nh}$
- $\breve{\sigma}_p^2 = \nu!^2\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\Big(\boldsymbol{V}_1 - 2\boldsymbol{V}_2 + 2\boldsymbol{V}_4 - 2\breve{\boldsymbol{V}}_5 + \breve{\boldsymbol{V}}_6\Big)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu$, where, with $\left[\boldsymbol{\Gamma}^{-1}\right]_{l_i,l_j}$ the $\{l_i+1, l_j+1\}$ element of $\boldsymbol{\Gamma}^{-1}$, we define

$$
\begin{aligned}
\breve{\boldsymbol{V}}_5 = \sum_{l_i=0}^{p}\sum_{l_j=0}^{p}\left[\boldsymbol{\Gamma}^{-1}\right]_{l_i,l_j}\mathbb{E}\Big[(K^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})(X_{h,i})^{l_i}\left(\mu(X_i)-\boldsymbol{r}_p(X_i-\mathsf{x})'\boldsymbol{\beta}_p\right)\Big]&\\
\times\frac{1}{nh}\sum_{j=1}^{n}\Big\{K(X_{h,j})(X_{h,j})^{l_j}\left(Y_j-\boldsymbol{r}_p(X_j-\mathsf{x})'\boldsymbol{\beta}_p\right)\Big\},&\\
\breve{\boldsymbol{V}}_6 = \sum_{l_{i_1}=0}^{p}\sum_{l_{i_2}=0}^{p}\sum_{l_{j_1}=0}^{p}\sum_{l_{j_2}=0}^{p}\left[\boldsymbol{\Gamma}^{-1}\right]_{l_{i_1},l_{j_1}}\left[\boldsymbol{\Gamma}^{-1}\right]_{l_{i_2},l_{j_2}}\mathbb{E}\Big[h^{-1}(K^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})(X_{h,i})^{l_{i_1}+l_{i_2}}\Big]&\\
\times\frac{1}{(nh)^2}\sum_{j=1}^{n}\sum_{k=1}^{n}K(X_{h,j})(X_{h,j})^{l_{j_1}}\left(Y_j-\boldsymbol{r}_p(X_j-\mathsf{x})'\boldsymbol{\beta}_p\right)K(X_{h,k})(X_{h,k})^{l_{j_2}}\left(Y_k-\boldsymbol{r}_p(X_k-\mathsf{x})'\boldsymbol{\beta}_p\right).&
\end{aligned}
$$

Next, using Equation (S.12) to rewrite $\mu^{(\nu)}$, canceling h^ν, and adding and subtracting $\breve{\sigma}_p^{-1}$, write T_p as

$$
\begin{aligned}
T_p &= \hat{\sigma}_p^{-1}\sqrt{nh^{1+2\nu}}(\hat{\theta}_p - \mu^{(\nu)})\\
&= \hat{\sigma}_p^{-1}\sqrt{nh^{1+2\nu}}\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{Y}-\boldsymbol{R}\boldsymbol{\beta}_p\right)/(nh^\nu)\\
&= \hat{\sigma}_p^{-1}s_n\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{Y}-\boldsymbol{R}\boldsymbol{\beta}_p\right)/n\\
&= \breve{\sigma}_p^{-1}s_n\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{Y}-\boldsymbol{R}\boldsymbol{\beta}_p\right)/n + \left(\hat{\sigma}_p^{-1}-\breve{\sigma}_p^{-1}\right)s_n\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{Y}-\boldsymbol{R}\boldsymbol{\beta}_p\right)/n\\
&=: \breve{T} + U_n.
\end{aligned}
$$

Then, referring back to Equation (S.22), we have

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

with

$$
U_n = \left(\hat{\sigma}_p^{-1}-\breve{\sigma}_p^{-1}\right)s_n\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{Y}-\boldsymbol{R}\boldsymbol{\beta}_p\right)/n \tag{S.24}
$$

and

$$\check{T} = \check{\sigma}_p^{-1} s_n \nu! \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1} \boldsymbol{\Omega} \left(\boldsymbol{Y} - \boldsymbol{R}\boldsymbol{\beta}_p \right) / n. \tag{S.25}$$

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 $\boldsymbol{Z}_i$, which is given by

$$\begin{aligned}
\boldsymbol{Z}_i = \Bigg(& \Big\{ (K\boldsymbol{r}_p)(X_{h,i})(Y_i - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p) \Big\}', \\
& \text{vech}\Big\{ (K\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i}) \Big\}', \\
& \text{vech}\Big\{ (K^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})\varepsilon_i^2 \Big\}', \\
& \text{vech}\Big\{ (K^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})(X_{h,i})^0\varepsilon_i \Big\}', \text{vech}\Big\{ (K^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})(X_{h,i})^1\varepsilon_i \Big\}', \\
& \quad \text{vech}\Big\{ (K^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})(X_{h,i})^2\varepsilon_i \Big\}', \ldots, \text{vech}\Big\{ (K^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})(X_{h,i})^p\varepsilon_i \Big\}', \\
& \text{vech}\Big\{ (K^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})\big\{\varepsilon_i\big[\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p\big]\big\} \Big\}' \Bigg)'.
\end{aligned} \tag{S.26}$$

In order of their listing above, these pieces come from (i) the “score” portion of the numerator, (ii) the “Gram” matrix $\boldsymbol{\Gamma}$, (iii) $\boldsymbol{V}_1$, (iv) $\boldsymbol{V}_2$, and (v) $\boldsymbol{V}_4$. Notice that $\check{\boldsymbol{V}}_5$ and $\check{\boldsymbol{V}}_6$ do not add any additional elements to $\boldsymbol{Z}_i$.

Equation (S.22) now follows from Lemma S.1(a), which completes **Step (I)**, if we can show that

$$r_{T_p}^{-1} \mathbb{P}[|U_n| > r_n] = o(1), \tag{S.27}$$

where $r_{T_p} = \max\{s_n^{-2}, \Psi_{T_p,F}^2, s_n^{-1}\Psi_{T_p,F}\}$ and $r_n = o(r_{T_p})$.

We now establish that Equation (S.27) holds. First

$$\frac{1}{\hat{\sigma}_p} = \frac{1}{\check{\sigma}_p}\left(\frac{\hat{\sigma}_p^2}{\check{\sigma}_p^2}\right)^{-1/2} = \frac{1}{\check{\sigma}_p}\left(1 + \frac{\hat{\sigma}_p^2 - \check{\sigma}_p^2}{\check{\sigma}_p^2}\right)^{-1/2},$$

and hence a Taylor expansion gives [1]

$$\frac{1}{\hat{\sigma}_p} = \frac{1}{\breve{\sigma}_p}\left[1 - \frac{1}{2}\frac{\hat{\sigma}_p^2 - \breve{\sigma}_p^2}{\breve{\sigma}_p^2} + \frac{1}{2!}\frac{3}{4}\left(\frac{\hat{\sigma}_p^2 - \breve{\sigma}_p^2}{\breve{\sigma}_p^2}\right)^2 \frac{\breve{\sigma}_p^5}{\bar{\sigma}^5}\right],$$

for a point $\bar{\sigma}^2 \in [\breve{\sigma}_p^2, \hat{\sigma}_p^2]$, and so

$$\hat{\sigma}_p^{-1} - \breve{\sigma}_p^{-1} = -\frac{1}{2}\frac{\hat{\sigma}_p^2 - \breve{\sigma}_p^2}{\breve{\sigma}_p^3} + \frac{3}{8}\frac{\left(\hat{\sigma}_p^2 - \breve{\sigma}_p^2\right)^2}{\bar{\sigma}^5}. \tag{S.28}$$

Plugging this into the definition of U_n gives

$$U_n = \left(-\frac{1}{2\breve{\sigma}_p^3} + \frac{3}{8}\frac{\hat{\sigma}_p^2 - \breve{\sigma}_p^2}{\bar{\sigma}^5}\right)\left(\hat{\sigma}_p^2 - \breve{\sigma}_p^2\right) s_n \nu! \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{Y} - \boldsymbol{R}\boldsymbol{\beta}_p\right)/n.$$

Therefore, if $\left|\hat{\sigma}_p^2 - \breve{\sigma}_p^2\right| = o_{\mathbb{P}}(1)$, the result in (S.27) will hold, and **Step (I)** will be complete, once we have shown that

$$\begin{aligned}
&r_{T_p}^{-1}\mathbb{P}\left[\left|\left(\hat{\sigma}_p^2 - \breve{\sigma}_p^2\right) s_n \nu! \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{Y} - \boldsymbol{R}\boldsymbol{\beta}_p\right)/n\right| > r_n\right] \\
&\quad = r_{T_p}^{-1}\mathbb{P}\left[\left|\left(\nu!^2 \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1}\left(\boldsymbol{V}_3 - 2[\boldsymbol{V}_5 - \breve{\boldsymbol{V}}_5] + [\boldsymbol{V}_6 - \breve{\boldsymbol{V}}_6]\right)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu\right) s_n \nu! \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{Y} - \boldsymbol{R}\boldsymbol{\beta}_p\right)/n\right| > r_n\right] \\
&\quad = o(1).
\end{aligned} \tag{S.29}$$

Recall that $r_{T_p} = \max\{s_n^{-2}, \Psi_{T_p,F}^2, s_n^{-1}\Psi_{T_p,F}\}$ and $r_n = o(r_{T_p})$. This is what we now verify one term at a time.

First, for the $\boldsymbol{V}_3$ term, we claim that

$$\begin{aligned}
&r_{T_p}^{-1}\mathbb{P}\left[\left|\nu!^2 \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1}\boldsymbol{V}_3\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu s_n \nu! \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{Y} - \boldsymbol{R}\boldsymbol{\beta}_p\right)/n\right| > r_n\right] \\
&\quad \leq r_{T_p}^{-1}\mathbb{P}\left[\left|\nu!^2 \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1}\left(\boldsymbol{V}_3 - \mathbb{E}[\boldsymbol{V}_3]\right)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu s_n \nu! \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{Y} - \boldsymbol{M}\right)/n\right| > r_n\right] \\
&\qquad + r_{T_p}^{-1}\mathbb{P}\left[\left|\nu!^2 \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1}\mathbb{E}[\boldsymbol{V}_3]\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu s_n \nu! \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{Y} - \boldsymbol{M}\right)/n\right| > r_n\right] \\
&\qquad + r_{T_p}^{-1}\mathbb{P}\left[\left|\nu!^2 \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1}\left(\boldsymbol{V}_3 - \mathbb{E}[\boldsymbol{V}_3]\right)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu s_n \nu! \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{M} - \boldsymbol{R}\boldsymbol{\beta}_p\right)/n\right| > r_n\right] \\
&\qquad + r_{T_p}^{-1}\mathbb{P}\left[\left|\nu!^2 \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1}\mathbb{E}[\boldsymbol{V}_3]\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu s_n \nu! \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{M} - \boldsymbol{R}\boldsymbol{\beta}_p\right)/n\right| > r_n\right] \\
&\quad = o(1).
\end{aligned} \tag{S.30}$$

[1] It is not necessary to retain higher order terms in the Taylor series, for example via

$$\frac{1}{\hat{\sigma}_p} = \frac{1}{\breve{\sigma}_p}\left[1 - \frac{1}{2}\frac{\hat{\sigma}_p^2 - \breve{\sigma}_p^2}{\breve{\sigma}_p^2} + \frac{1}{2!}\frac{3}{4}\left(\frac{\hat{\sigma}_p^2 - \breve{\sigma}_p^2}{\breve{\sigma}_p^2}\right)^2 - \frac{1}{3!}\frac{15}{8}\left(\frac{\hat{\sigma}_p^2 - \breve{\sigma}_p^2}{\breve{\sigma}_p^2}\right)^3 \frac{\breve{\sigma}_p^7}{\bar{\sigma}^7}\right],$$

because $\breve{\sigma}_p^2$ is constructed exactly to retain all the important terms from $\hat{\sigma}_p^2$. Put differently, because $(\hat{\sigma}_p^2 - \breve{\sigma}_p^2)s_n\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{Y} - \boldsymbol{R}\boldsymbol{\beta}_p\right)/n$ will be shown to be ignorable in the process of verifying Equation (S.27), it is immediate that terms from $(\hat{\sigma}_p^2 - \breve{\sigma}_p^2)^2$ can also be ignored, as they are higher order. A longer Taylor expansion can be useful when computing the terms of the Edgeworth expansion.

For the first term, using the elementary bounds (note that $|e_q| = 1$),

$$
\begin{aligned}
r_{T_p}^{-1}\mathbb{P}&\left[\left|\nu!^2 \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1}\left(\boldsymbol{V}_3 - \mathbb{E}[\boldsymbol{V}_3]\right)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu s_n \nu! \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{Y} - \boldsymbol{M}\right)/n\right| > r_n\right] \\
&\leq r_{T_p}^{-1} 3\mathbb{P}\left[\left|\boldsymbol{\Gamma}^{-1}\right| > C_\Gamma\right] \\
&\quad + r_{T_p}^{-1}\mathbb{P}\left[s_n\left|\boldsymbol{\Omega}\left(\boldsymbol{Y} - \boldsymbol{M}\right)/n\right| > \delta\log(s_n)^{1/2}\right] \\
&\quad + r_{T_p}^{-1}\mathbb{P}\Bigg[\Bigg|\frac{1}{nh}\sum_{i=1}^{n}\Big\{(K^2 \boldsymbol{r}_p \boldsymbol{r}_p')(X_{h,i})\left[\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p\right]^2 \\
&\qquad\qquad - \mathbb{E}\left[(K^2 \boldsymbol{r}_p \boldsymbol{r}_p')(X_{h,i})\left[\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p\right]^2\right]\Big\}\Bigg| > r_n \frac{1}{(|e_q| q! C_\Gamma)^3 \delta \log(s_n)^{1/2}}\Bigg] \\
&= o(1),
\end{aligned}
$$

by Lemmas S.2, S.4, and S.6. In applying the last, take the constant to be $(|e_q| q! C_\Gamma)^{-3}\delta^{-1}$ and note that $r_n = o(r_{T_p})$ may be chosen such that $r_n \log(s_n)^{-1/2}$ vanishes slower than (i.e. is larger than) $\Psi_{T_p,F}^2 s_n^{-2}\log(s_n)^\gamma$, making the probability in the penultimate line bounded by the one in the Lemma. For example, take $r_n = \Psi_{T_p,F} s_n^{-1}\log(s_n)^{-1/2-\gamma}$ and note that

$$
\frac{r_n}{\log(s_n)^{1/2}} = \left(\frac{\Psi_{T_p,F}}{s_n}\right)^2 \log(s_n)^\gamma \left[\left(\frac{s_n}{\Psi_{T_p,F}}\right)^2 \frac{r_n}{\log(s_n)^{1/2+\gamma}}\right] = \left(\frac{\Psi_{T_p,F}}{s_n}\right)^2 \log(s_n)^\gamma \left[\frac{s_n}{\Psi_{T_p,F}}\right],
$$

where factor in square brackets diverges by assumption.

The second term required for result (S.30) obeys

$$
\begin{aligned}
r_{T_p}^{-1}\mathbb{P}&\left[\left|\nu!^2 \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1}\mathbb{E}[\boldsymbol{V}_3]\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu s_n \nu! \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{Y} - \boldsymbol{M}\right)/n\right| > r_n\right] \\
&\leq r_{T_p}^{-1} 3\mathbb{P}\left[\left|\boldsymbol{\Gamma}^{-1}\right| > C_\Gamma\right] \\
&\quad + r_{T_p}^{-1}\mathbb{P}\left[s_n\left|\boldsymbol{\Omega}\left(\boldsymbol{Y} - \boldsymbol{M}\right)/n\right| > \log(s_n)^{1/2}\left\{\frac{s_n^2}{\Psi_{T_p,F}^2} r_n \frac{1}{(|e_q| q! C_\Gamma)^3 \log(s_n)^{1/2}}\right\}\right] \\
&= o(1),
\end{aligned}
$$

using Lemmas S.2 and S.4, as the term in braces diverges (e.g. for $r_n = \Psi_{T_p,F}^2 \log(s_n)^{-1/2}$) and $\mathbb{E}[\boldsymbol{V}_3] = O(\Psi_{T_p,F}^2 s_n^{-2})$ as follows:

$$
\begin{aligned}
\mathbb{E}[\boldsymbol{V}_3] &= \frac{1}{nh}\sum_{i=1}^{n}\mathbb{E}\left[(K^2 \boldsymbol{r}_p \boldsymbol{r}_p')(X_{h,i})\left[\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p\right]^2\right] \\
&= \mathbb{E}\left[h^{-1}(K^2 \boldsymbol{r}_p \boldsymbol{r}_p')(X_{h,i})\left[\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p\right]^2\right] \\
&= \frac{\Psi_{T_p,F}^2}{s_n^2}\mathbb{E}\left[h^{-1}(K^2 \boldsymbol{r}_p \boldsymbol{r}_p')(X_{h,i})\left[\frac{s_n}{\Psi_{T_p,F}}\left(\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p\right)\right]^2\right] \\
&= O\left(\frac{\Psi_{T_p,F}^2}{s_n^2}\right).
\end{aligned}
$$

The third term required for result (S.30) obeys

$$
\begin{aligned}
r_{T_p}^{-1}\mathbb{P}&\left[\left|\nu!^2\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\left(\boldsymbol{V}_3-\mathbb{E}[\boldsymbol{V}_3]\right)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu s_n\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{M}-\boldsymbol{R}\boldsymbol{\beta}_p\right)/n\right|>r_n\right]\\
&\leq r_{T_p}^{-1}3\mathbb{P}\left[\left|\boldsymbol{\Gamma}^{-1}\right|>C_\Gamma\right]\\
&\quad+r_{T_p}^{-1}\mathbb{P}\left[\left|\boldsymbol{\Omega}\left(\boldsymbol{M}-\boldsymbol{R}\boldsymbol{\beta}_p\right)/n\right|>\log(s_n)^{1/2}\right]\\
&\quad+r_{T_p}^{-1}\mathbb{P}\Bigg[\Bigg|\frac{1}{nh}\sum_{i=1}^n\Big\{(K^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})\left[\mu(X_i)-\boldsymbol{r}_p(X_i-\mathsf{x})'\boldsymbol{\beta}_p\right]^2\\
&\qquad\qquad-\mathbb{E}\left[(K^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})\left[\mu(X_i)-\boldsymbol{r}_p(X_i-\mathsf{x})'\boldsymbol{\beta}_p\right]^2\right]\Big\}\Bigg|>r_n\frac{1}{s_n(|e_q|q!C_\Gamma)^3\log(s_n)^{1/2}}\Bigg]\\
&=o(1),
\end{aligned}
$$

by Lemmas S.2, S.5, and S.6. In applying the last, take $\delta=(|e_q|q!C_\Gamma)^{-3}$ and note that $r_n=o(r_{T_p})$ may be chosen such that $r_n\log(s_n)^{-1/2}$ vanishes slower than (i.e. is larger than) $\Psi_{T_p,F}^2s_n^{-2}\log(s_n)^\gamma$, making the probability in the penultimate line bounded by the one in the Lemma. For example, take $r_n=\Psi_{T_p,F}s_n^{-1}\log(s_n)^{-\gamma}$ and note that

$$
\frac{r_n}{s_n\log(s_n)^{1/2}}=\left(\frac{\Psi_{T_p,F}}{s_n}\right)^2\log(s_n)^\gamma\left[\left(\frac{s_n}{\Psi_{T_p,F}}\right)^2\frac{r_n}{s_n\log(s_n)^{1/2+\gamma}}\right]=\left(\frac{\Psi_{T_p,F}}{s_n}\right)^2\log(s_n)^\gamma\left[\frac{1}{\Psi_{T_p,F}\log(s_n)^{1/2+2\gamma}}\right],
$$

where factor in square brackets diverges by assumption.

The fourth term follows the same pattern as the second, using Lemma S.5 in place of Lemma S.4, the same way the third term followed the pattern of the first. This completes the proof of result (S.30).

Turning to the $\boldsymbol{V}_5$ terms, first observe that, when all its components are considered, $\boldsymbol{V}_5$ is a $(p+1)\times(p+1)$ matrix (from $(\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})$) multiplied by a scalar. We write out

$$
\begin{aligned}
\boldsymbol{r}_p'(X_{h,i})\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left[\boldsymbol{Y}-\boldsymbol{R}\boldsymbol{\beta}_p\right]/n&=\frac{1}{nh}\sum_{j=1}^n\left\{\boldsymbol{r}_p'(X_{h,i})\boldsymbol{\Gamma}^{-1}\boldsymbol{r}_p'(X_{h,j})\right\}K(X_{h,j})\left(Y_j-\boldsymbol{r}_p(X_j-\mathsf{x})'\boldsymbol{\beta}_p\right)\\
&=\frac{1}{nh}\sum_{j=1}^n\left\{\sum_{l_i=0}^p\sum_{l_j=0}^p\left[\boldsymbol{\Gamma}^{-1}\right]_{l_i,l_j}(X_{h,i})^{l_i}(X_{h,j})^{l_j}\right\}K(X_{h,j})\left(Y_j-\boldsymbol{r}_p(X_j-\mathsf{x})'\boldsymbol{\beta}_p\right).
\end{aligned}
$$

where $\left[\boldsymbol{\Gamma}^{-1}\right]_{l_i,l_j}$ is the $\{l_i+1,l_j+1\}$ element of $\boldsymbol{\Gamma}^{-1}$, which is well-behaved by Lemma S.2. We make use of this in order to write

$$
\begin{aligned}
\nu!^2\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\Big[\boldsymbol{V}_5\Big]\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu&=\nu!^2\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\frac{1}{nh}\sum_{i=1}^n(K^2\boldsymbol{r}_p\boldsymbol{r}_p'\boldsymbol{r}_p')(X_{h,i})\left[\mu(X_i)-\boldsymbol{r}_p(X_i-\mathsf{x})'\boldsymbol{\beta}_p\right]\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left[\boldsymbol{Y}-\boldsymbol{R}\boldsymbol{\beta}_p\right]/n\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu\\
&=\sum_{l_i=0}^p\sum_{l_j=0}^p\nu!^2\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\left[\boldsymbol{\Gamma}^{-1}\right]_{l_i,l_j}\frac{1}{(nh)^2}\sum_{i=1}^n\sum_{j=1}^n\Big\{(K^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})K(X_{h,j})(X_{h,i})^{l_i}(X_{h,j})^{l_j}
\end{aligned}
$$

$$\times \left[\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p\right]\left(Y_j - \boldsymbol{r}_p(X_j - \mathsf{x})'\boldsymbol{\beta}_p\right)\Big\}\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu$$

$$=: \sum_{l_i=0}^{p}\sum_{l_j=0}^{p} \nu!^2 \boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\Big\{V_{5,1}(l_i,l_j) + V_{5,2}(l_i,l_j)\Big\}\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu, \tag{S.31}$$

where $V_{5,1}(l_i,l_j)$ and $V_{5,2}(l_i,l_j)$ are the "own" and "cross" summands

$$V_{5,1}(l_i,l_j) := \left[\boldsymbol{\Gamma}^{-1}\right]_{l_i,l_j} \frac{1}{(nh)^2}\sum_{i=1}^{n}\Big\{(K^3\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})(X_{h,i})^{l_i+l_j}$$
$$\times \left[\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p\right]\left(Y_i - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p\right)\Big\}$$
$$V_{5,2}(l_i,l_j) := \left[\boldsymbol{\Gamma}^{-1}\right]_{l_i,l_j} \frac{1}{(nh)^2}\sum_{i=1}^{n}\sum_{j\neq i}\Big\{(K^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})K(X_{h,j})(X_{h,i})^{l_i}(X_{h,j})^{l_j}$$
$$\times \left[\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p\right]\left(Y_j - \boldsymbol{r}_p(X_j - \mathsf{x})'\boldsymbol{\beta}_p\right)\Big\}.$$

Recall that the goal is result (S.29). We will study one term of the double sum (S.31), i.e. $V_{5,1}(l_i,l_j)$ and $V_{5,2}(l_i,l_j)$ for a fixed pair $\{l_i,l_j\}$, as all terms are identically handled. If each term is ignorable in the expansion, then it follows that

$$r_{T_p}^{-1}\mathbb{P}\left[\Big|\Big(\nu!^2\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\Big(-2[\boldsymbol{V}_5 - \breve{\boldsymbol{V}}_5\Big)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu\Big) s_n\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{Y} - \boldsymbol{R}\boldsymbol{\beta}_p\right)/n\Big| > r_n\right]$$
$$\leq C \max_{0\leq l_i,l_j\leq p} r_{T_p}^{-1}\mathbb{P}\Bigg[\Bigg|\Big(\nu!^2\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\Big(V_{5,1}(l_i,l_j) + V_{5,2}(l_i,l_j) - \breve{V}_{5,2}(l_i,l_j)\Big)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu\Big)$$
$$\times\; s_n\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{Y} - \boldsymbol{R}\boldsymbol{\beta}_p\right)/n\Bigg| > r_n\Bigg]$$
$$= o(1), \tag{S.32}$$

by Boole's inequality and p fixed.

As hinted at in this display, $\breve{\boldsymbol{V}}_5$ will be constructed from the pieces of $V_{5,2}(l_i,l_j)$ which contribute to the expansion. We first show that the $V_{5,1}(l_i,l_j)$ terms may be ignored. Begin by splitting $(Y_i - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p) = \varepsilon_i + (\mu(X_i) - \boldsymbol{r}_p(X_{-}\mathsf{x})'\boldsymbol{\beta}_p)$ everywhere, as the "variance" and "bias" type pieces have different rates, which must be accounted for:

$$r_{T_p}^{-1}\mathbb{P}\left[\left|\left(\nu!^2\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\left(V_{5,1}(l_i,l_j)\right)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu\right)\; s_n\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{Y} - \boldsymbol{R}\boldsymbol{\beta}_p\right)/n\right| > r_n\right]$$
$$\leq r_{T_p}^{-1}\mathbb{P}\left[\left|\left(\nu!^2\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\left(V_{5,1}(l_i,l_j)\right)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu\right)\; s_n\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{Y} - \boldsymbol{M}\right)/n\right| > r_n\right]$$
$$+ r_{T_p}^{-1}\mathbb{P}\left[\left|\left(\nu!^2\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\left(V_{5,1}(l_i,l_j)\right)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu\right)\; s_n\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{M} - \boldsymbol{R}\boldsymbol{\beta}_p\right)/n\right| > r_n\right]$$
$$\leq r_{T_p}^{-1}\mathbb{P}\Bigg[\Bigg|\Bigg(\nu!^2\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\Bigg(\left[\boldsymbol{\Gamma}^{-1}\right]_{l_i,l_j}\frac{1}{(nh)^2}\sum_{i=1}^{n}\Big\{(K^3\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})(X_{h,i})^{l_i+l_j}$$
$$\times\; \left[\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p\right]^2\Big\}\Bigg)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu\Bigg)\; s_n\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{M} - \boldsymbol{R}\boldsymbol{\beta}_p\right)/n\Bigg| > r_n\Bigg]$$

$$
\begin{aligned}
&+ r_{T_p}^{-1}\mathbb{P}\Bigg[\Bigg|\Bigg(\nu!^2 \boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\Bigg(\left[\boldsymbol{\Gamma}^{-1}\right]_{l_i,l_j}\frac{1}{(nh)^2}\sum_{i=1}^n\Big\{(K^3\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})(X_{h,i})^{l_i+l_j}\\
&\qquad\qquad \times \left[\mu(X_i)-\boldsymbol{r}_p(X_i-\mathsf{x})'\boldsymbol{\beta}_p\right]^2\Big\}\Bigg)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu\Bigg)\, s_n\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{Y}-\boldsymbol{M}\right)/n\Bigg|>r_n\Bigg]\\
&+ r_{T_p}^{-1}\mathbb{P}\Bigg[\Bigg|\Bigg(\nu!^2 \boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\Bigg(\left[\boldsymbol{\Gamma}^{-1}\right]_{l_i,l_j}\frac{1}{(nh)^2}\sum_{i=1}^n\Big\{(K^3\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})(X_{h,i})^{l_i+l_j}\\
&\qquad\qquad \times \left[\mu(X_i)-\boldsymbol{r}_p(X_i-\mathsf{x})'\boldsymbol{\beta}_p\right]\varepsilon_i\Big\}\Bigg)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu\Bigg)\, s_n\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{M}-\boldsymbol{R}\boldsymbol{\beta}_p\right)/n\Bigg|>r_n\Bigg]\\
&+ r_{T_p}^{-1}\mathbb{P}\Bigg[\Bigg|\Bigg(\nu!^2 \boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\Bigg(\left[\boldsymbol{\Gamma}^{-1}\right]_{l_i,l_j}\frac{1}{(nh)^2}\sum_{i=1}^n\Big\{(K^3\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})(X_{h,i})^{l_i+l_j}\\
&\qquad\qquad \times \left[\mu(X_i)-\boldsymbol{r}_p(X_i-\mathsf{x})'\boldsymbol{\beta}_p\right]\varepsilon_i\Big\}\Bigg)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu\Bigg)\, s_n\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{Y}-\boldsymbol{M}\right)/n\Bigg|>r_n\Bigg].
\end{aligned}
$$

For the first (i.e. the first term on the right hand side of the last inequality)

$$
\begin{aligned}
&r_{T_p}^{-1}\mathbb{P}\Bigg[\Bigg|\Bigg(\nu!^2 \boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\Bigg(\left[\boldsymbol{\Gamma}^{-1}\right]_{l_i,l_j}\frac{1}{(nh)^2}\sum_{i=1}^n\Big\{(K^3\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})(X_{h,i})^{l_i+l_j}\\
&\qquad\qquad \times \left[\mu(X_i)-\boldsymbol{r}_p(X_i-\mathsf{x})'\boldsymbol{\beta}_p\right]^2\Big\}\Bigg)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu\Bigg)\, s_n\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{M}-\boldsymbol{R}\boldsymbol{\beta}_p\right)/n\Bigg|>r_n\Bigg]\\
&\quad\leq r_{T_p}^{-1}4\mathbb{P}\left[\left|\boldsymbol{\Gamma}^{-1}\right|>C_\Gamma\right]\\
&\qquad + r_{T_p}^{-1}\mathbb{P}\left[\left|\boldsymbol{\Omega}\left(\boldsymbol{M}-\boldsymbol{R}\boldsymbol{\beta}_p\right)/n\right|>\log(s_n)^{1/2}\right]\\
&\qquad + r_{T_p}^{-1}\mathbb{P}\Bigg[\Bigg|\frac{1}{nh}\sum_{i=1}^n\Big\{(K^3\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})(X_{h,i})^{l_i+l_j}\\
&\qquad\qquad\qquad \times \left[\mu(X_i)-\boldsymbol{r}_p(X_i-\mathsf{x})'\boldsymbol{\beta}_p\right]^2\Big\}\Bigg|>r_n\frac{nh}{s_n(|e_q|q!)^3C_\Gamma^4\log(s_n)^{1/2}}\Bigg]\\
&\quad= o(1),
\end{aligned}
$$

by Lemmas S.2 and S.5, the latter applied twice, and the fact that, for $r_n=\Psi_{T_p,F}s_n\log(s_n)^{-\gamma}$, with any $\gamma>0$

$$
r_n\frac{nh}{s_n(|e_q|q!)^3C_\Gamma^4\log(s_n)^\gamma}\asymp\frac{\Psi_{T_p,F}}{s_n}\log(s_n)^{1/2}\left[\frac{s_n}{\log(s_n)^{1/2+2\gamma}}\right],
$$

and the factor in square brackets diverges. The rest of the $V_{5,1}(l_i,l_j)$ terms are handled by exactly the same steps, but using Lemmas S.4, S.5, and S.7 as needed for the final convergence. This establishes the $V_{5,1}(l_i,l_j)$ part of Equation (S.32).

Turning to the $V_{5,2}(l_i,l_j)$ part of Equation (S.32), we again begin by splitting $(Y_i-\boldsymbol{r}_p(X_i-$

$\mathsf{x})'\boldsymbol{\beta}_p) = \varepsilon_i + (\mu(X_i) - \boldsymbol{r}_p(X\_\mathsf{x})'\boldsymbol{\beta}_p)$ everywhere, just like above,

$$
\begin{aligned}
&r_{T_p}^{-1}\mathbb{P}\left[\left|\left(\nu!^2 \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1}\left(V_{5,2}(l_i,l_j)\right)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu\right)\ s_n \nu! \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{Y} - \boldsymbol{R}\boldsymbol{\beta}_p\right)/n\right| > r_n\right] \\
&\leq r_{T_p}^{-1}\mathbb{P}\Bigg[\Bigg|\Bigg(\nu!^2 \boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\Bigg(\left[\boldsymbol{\Gamma}^{-1}\right]_{l_i,l_j}\frac{1}{(nh)^2}\sum_{i=1}^n\sum_{j\neq i}\Big\{(K^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})K(X_{h,j})(X_{h,i})^{l_i}(X_{h,j})^{l_j} \\
&\qquad\qquad \times\left[\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p\right](\varepsilon_j)\Big\}\Bigg)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu\Bigg)\ s_n\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{Y} - \boldsymbol{M}\right)/n\Bigg| > r_n\Bigg] \\
&\quad + r_{T_p}^{-1}\mathbb{P}\Bigg[\Bigg|\Bigg(\nu!^2 \boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\Bigg(\left[\boldsymbol{\Gamma}^{-1}\right]_{l_i,l_j}\frac{1}{(nh)^2}\sum_{i=1}^n\sum_{j\neq i}\Big\{(K^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})K(X_{h,j})(X_{h,i})^{l_i}(X_{h,j})^{l_j} \\
&\qquad\qquad \times\left[\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p\right]\left(\mu(X_j) - \boldsymbol{r}_p(X_j - \mathsf{x})'\boldsymbol{\beta}_p\right)\Big\}\Bigg)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu\Bigg)\ s_n\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{Y} - \boldsymbol{M}\right)/n\Bigg| > r_n\Bigg] \\
&\quad + r_{T_p}^{-1}\mathbb{P}\Bigg[\Bigg|\Bigg(\nu!^2 \boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\Bigg(\left[\boldsymbol{\Gamma}^{-1}\right]_{l_i,l_j}\frac{1}{(nh)^2}\sum_{i=1}^n\sum_{j\neq i}\Big\{(K^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})K(X_{h,j})(X_{h,i})^{l_i}(X_{h,j})^{l_j} \\
&\qquad\qquad \times\left[\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p\right](\varepsilon_j)\Big\}\Bigg)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu\Bigg)\ s_n\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{M} - \boldsymbol{R}\boldsymbol{\beta}_p\right)/n\Bigg| > r_n\Bigg] \\
&\quad + r_{T_p}^{-1}\mathbb{P}\Bigg[\Bigg|\Bigg(\nu!^2 \boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\Bigg(\left[\boldsymbol{\Gamma}^{-1}\right]_{l_i,l_j}\frac{1}{(nh)^2}\sum_{i=1}^n\sum_{j\neq i}\Big\{(K^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})K(X_{h,j})(X_{h,i})^{l_i}(X_{h,j})^{l_j} \\
&\qquad\qquad \times\left[\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p\right]\left(\mu(X_j) - \boldsymbol{r}_p(X_j - \mathsf{x})'\boldsymbol{\beta}_p\right)\Big\}\Bigg)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu\Bigg)\ s_n\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{M} - \boldsymbol{R}\boldsymbol{\beta}_p\right)/n\Bigg| > r_n\Bigg]
\end{aligned}
$$

For the first term, which has two "variance" terms and one bias-type term:

$$
\begin{aligned}
&r_{T_p}^{-1}\mathbb{P}\Bigg[\Bigg|\Bigg(\nu!^2 \boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\Bigg(\left[\boldsymbol{\Gamma}^{-1}\right]_{l_i,l_j}\frac{1}{(nh)^2}\sum_{i=1}^n\sum_{j\neq i}\Big\{(K^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})K(X_{h,j})(X_{h,i})^{l_i}(X_{h,j})^{l_j} \\
&\qquad\qquad \times\left[\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p\right](\varepsilon_j)\Big\}\Bigg)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu\Bigg)\ s_n\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{Y} - \boldsymbol{M}\right)/n\Bigg| > r_n\Bigg] \\
&\leq r_{T_p}^{-1}4\mathbb{P}\left[\left|\boldsymbol{\Gamma}^{-1}\right| > C_\Gamma\right] \\
&\quad + r_{T_p}^{-1}\mathbb{P}\left[\left|\boldsymbol{\Omega}\left(\boldsymbol{Y} - \boldsymbol{M}\right)/n\right| > C_1 s_n^{-1}\log(s_n)^{1/2}\right] \\
&\quad + r_{T_p}^{-1}\mathbb{P}\left[\left|\frac{1}{nh}\sum_{j=1}^n\Big\{K(X_{h,j})(X_{h,i})^{l_j}\varepsilon_j\Big\}\right| > C_2 s_n^{-1}\log(s_n)^{1/2}\right] \\
&\quad + r_{T_p}^{-1}\mathbb{P}\left[\left|\frac{1}{nh}\sum_{i=1}^n\Big\{(K^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})(X_{h,i})^{l_i}\left[\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p\right]\Big\}\right| > r_n\frac{s_n^2}{s_n(|e_q|q!)^3 C_\Gamma^4 C_1 C_2\log(s_n)}\right] \\
&= o(1),
\end{aligned}
$$

by Lemmas S.2, S.4 applied twice, and S.5. For the last, note that for $r_n = \Psi_{T_p,F} s_n \log(s_n)^{-\gamma}$, with

$\gamma > 0$,

$$r_n \frac{s_n^2}{s_n(|e_q|q!)^3 C_\Gamma^4 C_1 C_2 \log(s_n)} \asymp \frac{\Psi_{T_p,F}}{s_n} \log(s_n)^\gamma \left[\frac{s_n}{\log(s_n)^{1+2\gamma}}\right],$$

and the term in square brackets diverges by assumption.

Turning to the second $V_{5,2}$ term (the third and fourth will be similar), which has one "variance" terms and two bias-type terms:, observe that

$$r_{T_p}^{-1}\mathbb{P}\Bigg[\Bigg|\Bigg(\nu!^2 \boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\Bigg(\left[\boldsymbol{\Gamma}^{-1}\right]_{l_i,l_j}\frac{1}{(nh)^2}\sum_{i=1}^n\sum_{j\neq i}\Bigg\{(K^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})K(X_{h,j})(X_{h,i})^{l_i}(X_{h,j})^{l_j}$$

$$\times\left[\mu(X_i)-\boldsymbol{r}_p(X_i-\mathsf{x})'\boldsymbol{\beta}_p\right]\left(\mu(X_j)-\boldsymbol{r}_p(X_j-\mathsf{x})'\boldsymbol{\beta}_p\right)\Bigg\}\Bigg)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu\Bigg)\; s_n\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{Y}-\boldsymbol{M}\right)/n\Bigg| > r_n\Bigg] \neq o(1),$$

because, compared to the above, Lemma S.4 is applied only once, while Lemma S.5 is needed twice, instead of vice versa. The slower rate in the latter implies that this term can not be ignored. Thus pieces of this will contribute to $\breve{\boldsymbol{V}}_5$. To see which, we will first center some bias terms. Just for notational ease, define the shorthand

$$V_{5,2,i} = (K^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})(X_{h,i})^{l_i}\left[\mu(X_i)-\boldsymbol{r}_p(X_i-\mathsf{x})'\boldsymbol{\beta}_p\right]$$

and

$$V_{5,2,j} = K(X_{h,j})(X_{h,j})^{l_j}\left[\mu(X_j)-\boldsymbol{r}_p(X_j-\mathsf{x})'\boldsymbol{\beta}_p\right].$$

The term in question is then

$$\begin{aligned}
&\Bigg(\nu!^2\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\Bigg(\left[\boldsymbol{\Gamma}^{-1}\right]_{l_i,l_j}\frac{1}{(nh)^2}\sum_{i=1}^n\sum_{j\neq i}V_{5,2,i}V_{5,2,j}\Bigg)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu\Bigg)\; s_n\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{Y}-\boldsymbol{M}\right)/n \\
&= \Bigg(\nu!^2\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\Bigg(\left[\boldsymbol{\Gamma}^{-1}\right]_{l_i,l_j}\mathbb{E}[h^{-1}V_{5,2,i}]\frac{1}{nh}\sum_{j=1}^n V_{5,2,j}\Bigg)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu\Bigg)\; s_n\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{Y}-\boldsymbol{M}\right)/n \\
&\quad + \Bigg(\nu!^2\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\Bigg(\left[\boldsymbol{\Gamma}^{-1}\right]_{l_i,l_j}\frac{1}{nh}\sum_{i=1}^n\left(V_{5,2,i}-\mathbb{E}[V_{5,2,i}]\right)\mathbb{E}[h^{-1}V_{5,2,j}]\Bigg)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu\Bigg)\; s_n\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{Y}-\boldsymbol{M}\right)/n \\
&\quad + \Bigg(\nu!^2\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\Bigg(\left[\boldsymbol{\Gamma}^{-1}\right]_{l_i,l_j}\frac{1}{(nh)^2}\sum_{i=1}^n\sum_{j\neq i}\left(V_{5,2,i}-\mathbb{E}[V_{5,2,i}]\right)\left(V_{5,2,j}-\mathbb{E}[V_{5,2,j}]\right)\Bigg)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu\Bigg)\; s_n\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{Y}-\boldsymbol{M}\right)/n
\end{aligned}$$

The first term here will be incorporated into $\breve{\boldsymbol{V}}_5$, and thus into $\breve{T}$. Note that it is a smooth function of the $\boldsymbol{Z}_i$ from Equation (S.26), which is why we choose the centering the way we do, that is, keeping the term with $\mathbb{E}[h^{-1}V_{5,2,i}]$ instead of $\mathbb{E}[h^{-1}V_{5,2,j}]$. Doing the reverse would force further variables into the vector $\boldsymbol{Z}_i$, and require a stronger Cramér's condition, which we seek to avoid.[2]

[2] Calonico et al. (2018a,b) use such an approach, requiring not only a strengthening of Cramér's condition, but also in the process, ruling out the uniform kernel.

The next term obeys

$$
\begin{aligned}
&r_{T_p}^{-1}\mathbb{P}\left[\left|\left(\nu!^2\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\left(\left[\boldsymbol{\Gamma}^{-1}\right]_{l_i,l_j}\frac{1}{nh}\sum_{i=1}^{n}\left(V_{5,2,i}-\mathbb{E}[V_{5,2,i}]\right)\mathbb{E}[h^{-1}V_{5,2,j}]\right)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu\right)\ s_n\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{Y}-\boldsymbol{M}\right)/n\right|>r_n\right]\\
&\leq r_{T_p}^{-1}4\mathbb{P}\left[\left|\boldsymbol{\Gamma}^{-1}\right|>C_\Gamma\right]\\
&\quad+r_{T_p}^{-1}\mathbb{P}\left[\left|\boldsymbol{\Omega}\left(\boldsymbol{Y}-\boldsymbol{M}\right)/n\right|>C_1s_n^{-1}\log(s_n)^{1/2}\right]\\
&\quad+r_{T_p}^{-1}\mathbb{P}\left[\left|\frac{1}{nh}\sum_{i=1}^{n}\left(V_{5,2,i}-\mathbb{E}[V_{5,2,i}]\right)\right|>r_n\frac{s_n}{C\Psi_{T_p,F}s_n\log(s_n)^{1/2}}\right]\\
&=o(1),
\end{aligned}
$$

by Lemmas S.2, S.4, and S.6, the fact that $\mathbb{E}[h^{-1}V_{5,2,j}]\asymp s_n^{-1}\Psi_{T_p,F}$ (see Section S.2 or the computation for $\mathbb{E}[\boldsymbol{V}_3]$ above), and that for $r_n=\Psi_{T_p,F}s_n^{-1}\log(s_n)^{-\gamma}$, with any $\gamma>0$,

$$
r_n\frac{s_n}{C\Psi_{T_p,F}s_n\log(s_n)^{1/2}}\asymp\frac{\Psi_{T_p,F}}{s_n}\log(s_n)^{\gamma}\left[\frac{1}{\Psi_{T_p,F}\log(s_n)^{1/2+2\gamma}}\right]
$$

the factor in square brackets diverges by assumption.

The final piece of the second $V_{5,2}$ term similarly obeys

$$
\begin{aligned}
&r_{T_p}^{-1}\mathbb{P}\Bigg[\Bigg|\left(\nu!^2\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\left(\left[\boldsymbol{\Gamma}^{-1}\right]_{l_i,l_j}\frac{1}{(nh)^2}\sum_{i=1}^{n}\sum_{j\neq i}\left(V_{5,2,i}-\mathbb{E}[V_{5,2,i}]\right)\left(V_{5,2,j}-\mathbb{E}[V_{5,2,j}]\right)\right)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu\right)\\
&\qquad\qquad\qquad\qquad\times\ s_n\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{Y}-\boldsymbol{M}\right)/n\Bigg|>r_n\Bigg]\\
&\leq r_{T_p}^{-1}4\mathbb{P}\left[\left|\boldsymbol{\Gamma}^{-1}\right|>C_\Gamma\right]\\
&\quad+r_{T_p}^{-1}\mathbb{P}\left[\left|\boldsymbol{\Omega}\left(\boldsymbol{Y}-\boldsymbol{M}\right)/n\right|>C_1s_n^{-1}\log(s_n)^{1/2}\right]\\
&\quad+r_{T_p}^{-1}\mathbb{P}\left[\left|\frac{1}{nh}\sum_{j=1}^{n}\left(V_{5,2,j}-\mathbb{E}[V_{5,2,j}]\right)\right|>\frac{\Psi_{T_p,F}}{s_n}\log(s_n)^{\gamma}\right]+o(1)\\
&\quad+r_{T_p}^{-1}\mathbb{P}\left[\left|\frac{1}{nh}\sum_{i=1}^{n}\left(V_{5,2,i}-\mathbb{E}[V_{5,2,i}]\right)\right|>r_n\frac{s_n}{C\Psi_{T_p,F}s_n\log(s_n)^{1/2+\gamma}}\right]\\
&=o(1),
\end{aligned}
$$

by Lemmas S.2, S.4, and S.6 applied twice, and that for $r_n=\Psi_{T_p,F}s_n^{-1}\log(s_n)^{-\gamma}$, with any $\gamma>0$,

$$
r_n\frac{s_n}{C\Psi_{T_p,F}s_n\log(s_n)^{1/2}}\asymp\frac{\Psi_{T_p,F}}{s_n}\log(s_n)^{\gamma}\left[\frac{1}{\Psi_{T_p,F}\log(s_n)^{1/2+3\gamma}}\right]
$$

the factor in square brackets diverges by assumption. The $o(1)$ factor in the third to last line accounts for the missing term in the sum over the "j" index.

Comparing the first and second $V_{5,2}$ terms, we see the the first was ignorable because it had two "variance" type terms, while the second had only one. This generalizes to the third and fourth $V_{5,2}$ terms, the third being just like the second and the fourth having three bias-type terms. For these, the same centering must be done as was done here. The bounding is then nearly identical. Putting these pieces together, recall the definition of $V_{5,2}(l_i, l_j)$:

$$V_{5,2}(l_i, l_j) := \left[\boldsymbol{\Gamma}^{-1}\right]_{l_i, l_j} \frac{1}{(nh)^2} \sum_{i=1}^{n} \sum_{j \neq i} \Big\{ (K^2 \boldsymbol{r}_p \boldsymbol{r}_p')(X_{h,i}) K(X_{h,j}) (X_{h,i})^{l_i} (X_{h,j})^{l_j} \\ \times \left[\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p\right] \left(Y_j - \boldsymbol{r}_p(X_j - \mathsf{x})'\boldsymbol{\beta}_p\right) \Big\}.$$

Following the logic above, always centering the "i" term first, we define

$$\check{V}_{5,2}(l_i, l_j) := \left[\boldsymbol{\Gamma}^{-1}\right]_{l_i, l_j} \mathbb{E}\left[(K^2 \boldsymbol{r}_p \boldsymbol{r}_p')(X_{h,i}) (X_{h,i})^{l_i} \left(\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p\right)\right] \\ \times \frac{1}{nh} \sum_{j=1}^{n} \Big\{ K(X_{h,j}) (X_{h,j})^{l_j} \left(Y_j - \boldsymbol{r}_p(X_j - \mathsf{x})'\boldsymbol{\beta}_p\right) \Big\}$$

Returning to Equations (S.31), $\check{\boldsymbol{V}}_5$ is defined via

$$\nu!^2 \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1} \left[\check{\boldsymbol{V}}_5\right] \boldsymbol{\Gamma}^{-1} \boldsymbol{e}_\nu := \sum_{l_i=0}^{p} \sum_{l_j=0}^{p} \nu!^2 \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1} \left[\boldsymbol{\Gamma}^{-1}\right]_{l_i, l_j} \mathbb{E}\left[(K^2 \boldsymbol{r}_p \boldsymbol{r}_p')(X_{h,i}) (X_{h,i})^{l_i} \left(\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p\right)\right] \\ \times \frac{1}{nh} \sum_{j=1}^{n} \Big\{ K(X_{h,j}) (X_{h,j})^{l_j} \left(Y_j - \boldsymbol{r}_p(X_j - \mathsf{x})'\boldsymbol{\beta}_p\right) \Big\} \boldsymbol{\Gamma}^{-1} \boldsymbol{e}_\nu.$$

This completes the proof of Equation (S.32).

Lastly, we consider the $\boldsymbol{V}_6 - \check{\boldsymbol{V}}_6$ term of (S.29). Proving this is ignorable will complete **Step (I)**. Begin by expanding the inner product, just as was done for $\boldsymbol{V}_5$:

$$\begin{aligned}
\boldsymbol{V}_6 &= \frac{1}{nh} \sum_{i=1}^{n} (K^2 \boldsymbol{r}_p \boldsymbol{r}_p')(X_{h,i}) \big\{ \boldsymbol{r}_p(X_{h,i})' \boldsymbol{\Gamma}^{-1} \boldsymbol{\Omega} \left[\boldsymbol{Y} - \boldsymbol{R}\boldsymbol{\beta}_p\right]/n \big\}^2 \\
&= \frac{1}{nh} \sum_{i=1}^{n} (K^2 \boldsymbol{r}_p \boldsymbol{r}_p')(X_{h,i}) \left\{ \frac{1}{nh} \sum_{j=1}^{n} \boldsymbol{r}_p(X_{h,i})' \boldsymbol{\Gamma}^{-1} \boldsymbol{r}_p(X_{h,j}) K(X_{h,j}) \left(Y_j - \boldsymbol{r}_p(X_j - \mathsf{x})'\boldsymbol{\beta}_p\right) \right\}^2 \\
&= \frac{1}{nh} \sum_{i=1}^{n} (K^2 \boldsymbol{r}_p \boldsymbol{r}_p')(X_{h,i}) \left\{ \frac{1}{nh} \sum_{j=1}^{n} \sum_{l_i=0}^{p} \sum_{l_j=0}^{p} (X_{h,i})^{l_i} \left[\boldsymbol{\Gamma}^{-1}\right]_{l_i, l_j} (X_{h,j})^{l_j} K(X_{h,j}) \left(Y_j - \boldsymbol{r}_p(X_j - \mathsf{x})'\boldsymbol{\beta}_p\right) \right\}^2 \\
&= \sum_{l_{i_1}=0}^{p} \sum_{l_{i_2}=0}^{p} \sum_{l_{j_1}=0}^{p} \sum_{l_{j_2}=0}^{p} \left[\boldsymbol{\Gamma}^{-1}\right]_{l_{i_1}, l_{j_1}} \left[\boldsymbol{\Gamma}^{-1}\right]_{l_{i_2}, l_{j_2}} \frac{1}{nh} \sum_{i=1}^{n} (K^2 \boldsymbol{r}_p \boldsymbol{r}_p')(X_{h,i}) (X_{h,i})^{l_{i_1} + l_{i_2}} \\
&\qquad \times \frac{1}{(nh)^2} \sum_{j=1}^{n} \sum_{k=1}^{n} K(X_{h,j}) (X_{h,j})^{l_{j_1}} \left(Y_j - \boldsymbol{r}_p(X_j - \mathsf{x})'\boldsymbol{\beta}_p\right) K(X_{h,k}) (X_{h,k})^{l_{j_2}} \left(Y_k - \boldsymbol{r}_p(X_k - \mathsf{x})'\boldsymbol{\beta}_p\right).
\end{aligned}$$

Define

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

Completely analogous steps to those above will show that

$$r_{T_p}^{-1}\mathbb{P}\left[\left|\left(\nu!^2\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\left(\boldsymbol{V}_6 - \check{\boldsymbol{V}}_6\right)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu\right) s_n\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{Y} - \boldsymbol{R}\boldsymbol{\beta}_p\right)/n\right| > r_n\right] = o(1). \tag{S.33}$$

The starting point will again be splitting $(Y_i - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p) = \varepsilon_i + (\mu(X_i) - \boldsymbol{r}_p(X_{-}\mathsf{x})'\boldsymbol{\beta}_p)$ everywhere, which now occurs in three places, giving eight total terms. The most difficult of these will be when all three are bias terms. The rest of the terms will have at least one "variance" type term, and the faster rates of Lemma S.4 can be brought to bear. Thus, we shall only demonstrate the former. For a fixed set of the indexes $l_{i_1}, l_{i_2}, l_{j_1}, l_{j_2}$, let

$$\begin{aligned} V_{6,i} &= (K^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})(X_{h,i})^{l_{i_1}+l_{i_2}}, \\ V_{6,j} &= K(X_{h,j})(X_{h,j})^{l_{j_1}}\left(\mu(X_J) - \boldsymbol{r}_p(X_j - \mathsf{x})'\boldsymbol{\beta}_p\right), \qquad \text{and} \\ V_{6,k} &= K(X_{h,k})(X_{h,k})^{l_{j_2}}\left(\mu(X_k) - \boldsymbol{r}_p(X_k - \mathsf{x})'\boldsymbol{\beta}_p\right). \end{aligned}$$

The term in question, with three "bias" type terms", is:

$$\begin{aligned}
&\nu!^2\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\left(\boldsymbol{V}_6 - \check{\boldsymbol{V}}_6\right)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu s_n\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{M} - \boldsymbol{R}\boldsymbol{\beta}_p\right)/n \\
&= \nu!^2\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\left(\left[\boldsymbol{\Gamma}^{-1}\right]_{l_{i_1},l_{j_1}}\left[\boldsymbol{\Gamma}^{-1}\right]_{l_{i_2},l_{j_2}}\frac{1}{nh}\sum_{i=1}^{n}\left(V_{6,i} - \mathbb{E}[V_{6,i}]\right)\frac{1}{(nh)^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\mathbb{E}[V_{6,j}]\mathbb{E}[V_{6,k}]\right) \\
&\qquad \times\ \boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu s_n\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{M} - \boldsymbol{R}\boldsymbol{\beta}_p\right)/n \\
&\quad + \nu!^2\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\left(\left[\boldsymbol{\Gamma}^{-1}\right]_{l_{i_1},l_{j_1}}\left[\boldsymbol{\Gamma}^{-1}\right]_{l_{i_2},l_{j_2}}\frac{1}{nh}\sum_{i=1}^{n}\left(V_{6,i} - \mathbb{E}[V_{6,i}]\right)\frac{1}{(nh)^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\mathbb{E}[V_{6,j}]\left(V_{6,k} - \mathbb{E}[V_{6,k}]\right)\right) \\
&\qquad \times\ \boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu s_n\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{M} - \boldsymbol{R}\boldsymbol{\beta}_p\right)/n \\
&\quad + \nu!^2\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\left(\left[\boldsymbol{\Gamma}^{-1}\right]_{l_{i_1},l_{j_1}}\left[\boldsymbol{\Gamma}^{-1}\right]_{l_{i_2},l_{j_2}}\frac{1}{nh}\sum_{i=1}^{n}\left(V_{6,i} - \mathbb{E}[V_{6,i}]\right)\frac{1}{(nh)^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\left(V_{6,j} - \mathbb{E}[V_{6,j}]\right)\mathbb{E}[V_{6,k}]\right) \\
&\qquad \times\ \boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu s_n\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{M} - \boldsymbol{R}\boldsymbol{\beta}_p\right)/n \\
&\quad + \nu!^2\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\left(\left[\boldsymbol{\Gamma}^{-1}\right]_{l_{i_1},l_{j_1}}\left[\boldsymbol{\Gamma}^{-1}\right]_{l_{i_2},l_{j_2}}\frac{1}{nh}\sum_{i=1}^{n}\left(V_{6,i} - \mathbb{E}[V_{6,i}]\right)\frac{1}{(nh)^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\left(V_{6,j} - \mathbb{E}[V_{6,j}]\right)\left(V_{6,k} - \mathbb{E}[V_{6,k}]\right)\right) \\
&\qquad \times\ \boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu s_n\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{M} - \boldsymbol{R}\boldsymbol{\beta}_p\right)/n
\end{aligned}$$

The first term is bounded as

$$
\begin{aligned}
&\leq r_{T_p}^{-1} 5\mathbb{P}\left[\left|\boldsymbol{\Gamma}^{-1}\right| > C_\Gamma\right] \\
&\quad + r_{T_p}^{-1}\mathbb{P}\left[\left|\boldsymbol{\Omega}\left(\boldsymbol{M} - \boldsymbol{R}\boldsymbol{\beta}_p\right)/n\right| > C_1 \log(s_n)^\gamma\right] \\
&\quad + r_{T_p}^{-1}\mathbb{P}\left[\left|\frac{1}{nh}\sum_{i=1}^n (V_{6,i} - \mathbb{E}[V_{6,i}])\right| > r_n \frac{1}{C_1 C_\Gamma^5 \nu!^3 |\boldsymbol{e}_\nu|^3 \mathbb{E}[h^{-1}V_{6,j}]\mathbb{E}[h^{-1}V_{6,k}] s_n \log(s_n)^\gamma}\right] \\
&= o(1),
\end{aligned}
$$

by Lemmas S.2, S.5, and S.3. In applying the last, we have used that $\mathbb{E}[h^{-1}V_{6,j}] \asymp \mathbb{E}[h^{-1}V_{6,k}] \asymp s_n^{-1}\Psi_{T_p,F}$ (see Section S.2 or the computation for $\mathbb{E}[\boldsymbol{V}_3]$ above) and $r_n = s_n^{-1}\Psi_{T_p,F}\log(s_n)^{-\gamma}$ for $\gamma > 0$, leaving

$$
r_n \frac{1}{\mathbb{E}[h^{-1}V_{6,j}]\mathbb{E}[h^{-1}V_{6,k}] s_n \log(s_n)^\gamma} \asymp s_n^{-1}\log(s_n)^{1/2}\left[\frac{1}{s_n^{-1}\Psi_{T_p,F}\log(s_n)^{1/2+2\gamma}}\right].
$$

The factor in square brackets diverges by assumption. The second term is

$$
\begin{aligned}
&\nu!^2 \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1}\left(\left[\boldsymbol{\Gamma}^{-1}\right]_{l_{i_1}, l_{j_1}}\left[\boldsymbol{\Gamma}^{-1}\right]_{l_{i_2}, l_{j_2}} \frac{1}{nh}\sum_{i=1}^n (V_{6,i} - \mathbb{E}[V_{6,i}]) \frac{1}{(nh)^2}\sum_{j=1}^n\sum_{k=1}^n \mathbb{E}[V_{6,j}]\left(V_{6,k} - \mathbb{E}[V_{6,k}]\right)\right) \\
&\qquad\qquad \times\ \boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu s_n \nu! \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{M} - \boldsymbol{R}\boldsymbol{\beta}_p\right)/n \\
&\leq r_{T_p}^{-1} 5\mathbb{P}\left[\left|\boldsymbol{\Gamma}^{-1}\right| > C_\Gamma\right] \\
&\quad + r_{T_p}^{-1}\mathbb{P}\left[\left|\boldsymbol{\Omega}\left(\boldsymbol{M} - \boldsymbol{R}\boldsymbol{\beta}_p\right)/n\right| > C_1 \log(s_n)^\gamma\right] \\
&\quad + r_{T_p}^{-1}\mathbb{P}\left[\left|\frac{1}{nh}\sum_{k=1}^n (V_{6,k} - \mathbb{E}[V_{6,k}])\right| > C_2 \frac{\Psi_{T_p,F}}{s_n}\log(s_n)^\gamma\right] \\
&\quad + r_{T_p}^{-1}\mathbb{P}\left[\left|\frac{1}{nh}\sum_{i=1}^n (V_{6,i} - \mathbb{E}[V_{6,i}])\right| > r_n \frac{}{C_1 C_2 C_\Gamma^5 \nu!^3 |\boldsymbol{e}_\nu|^3 \mathbb{E}[h^{-1}V_{6,j}]\Psi_{T_p,F} s_n s_n^{-1}\log(s_n)^{2\gamma}}\right] \\
&= o(1),
\end{aligned}
$$

by nearly identical reasoning, additionally using Lemma S.6. The third term is the identical to this one, and the fourth term is similar, requiring Lemma S.6 twice.

Referring back to the discussion following Equation (S.33), this completes the proof of that result for the case where the bias portion of $(Y_i - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p) = \varepsilon_i + (\mu(X_i) - \boldsymbol{r}_p(X_- \mathsf{x})'\boldsymbol{\beta}_p)$ is retained everywhere, which is the most difficult. All other pieces will follow by similar logic, applying Lemma S.4 when needed. Because this Lemma delivers a faster rate, these other terms will not require strong assumptions. Altogether, this establishes the convergence required by Equation (S.33).

Combining Equations (S.30), (S.32), and (S.33) establishes that $\left|\hat{\sigma}_p^2 - \breve{\sigma}_p^2\right| = o_{\mathbb{P}}(1)$ and (S.29) holds, proving (S.27) and thus completing **Step (I)**.

subsectionStep (II)

We now prove that

$$\boldsymbol{S}_n := \sum_{i=1}^{n} \mathbb{V}[\boldsymbol{Z}_i]^{-1/2}(\boldsymbol{Z}_i - \mathbb{E}[\boldsymbol{Z}_i])/\sqrt{n}$$

obeys an Edgeworth expansion by verifying the conditions of Theorem 3.4 of Skovgaard (1981). Repeating the definition of $\boldsymbol{Z}_i$ from Equation (S.26):

$$\begin{aligned}\boldsymbol{Z}_i = \Bigg(&\Big\{(K\boldsymbol{r}_p)(X_{h,i})(Y_i - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p)\Big\}', \\ &\operatorname{vech}\Big\{(K\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})\Big\}', \\ &\operatorname{vech}\Big\{(K^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})\varepsilon_i^2\Big\}', \\ &\operatorname{vech}\Big\{(K^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})(X_{h,i})^0\varepsilon_i\Big\}', \operatorname{vech}\Big\{(K^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})(X_{h,i})^1\varepsilon_i\Big\}', \\ &\quad \operatorname{vech}\Big\{(K^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})(X_{h,i})^2\varepsilon_i\Big\}', \dots, \operatorname{vech}\Big\{(K^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})(X_{h,i})^p\varepsilon_i\Big\}', \\ &\operatorname{vech}\Big\{(K^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})\big\{\varepsilon_i\big[\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p\big]\big\}\Big\}'\Bigg)'.\end{aligned}$$

First, define

$$\boldsymbol{B} := h\mathbb{V}[\boldsymbol{Z}_i],$$

which may be readily computed, but the constants are not needed here. All that matters at present is that, under our assumptions, $\boldsymbol{B}$ is bounded and bounded away from zero. Write

$$\boldsymbol{S}_n = \sum_{i=1}^{n} \boldsymbol{B}^{-1/2}(\boldsymbol{Z}_i - \mathbb{E}[\boldsymbol{Z}_i])/s_n.$$

By construction, the mean of $\boldsymbol{S}_n$ is zero and the variance is the identity matrix. That is, for any $\boldsymbol{t} \in \mathbb{R}^{\dim(\boldsymbol{Z}_i)}$, $\mathbb{E}[\boldsymbol{t}'\boldsymbol{S}_n] = 0$ and $\mathbb{V}[\boldsymbol{t}'\boldsymbol{S}_n] = |\boldsymbol{t}|^2$.

To verify conditions (I) and (II) of Skovgaard (1981, Theorem 3.4) we first compute the third and fourth moments of $\boldsymbol{Z}_i$, and use these to compute the required directional cumulants of $\boldsymbol{S}_n$. For a nonnegative integer l and $k \in \{3, 4\}$, by a change of variables we find that

$$\mathbb{E}\left[\Big(K(X_{h,i})(X_{h,i})^l\Big)^k\right] = h\int K(u)^k u^{lk} f(\mathsf{x} - uh)du = O(h),$$

under the conditions on the kernel function and the marginal density of X_i, $f(\cdot)$. In exactly the same way, for the remaining pieces of $\boldsymbol{Z}_i$, we find that:

$$\mathbb{E}\left[\Big(K(X_{h,i})(X_{h,i})^l(Y_i - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p)\Big)^k\right] = O(h),$$

$$\mathbb{E}\left[\Big(K(X_{h,i})(X_{h,i})^l\varepsilon_i^2\Big)^k\right] = O(h), \qquad \text{and} \qquad \mathbb{E}\left[\Big(K(X_{h,i})(X_{h,i})^l\varepsilon_i\Big)^k\right] = O(h),$$

$$\mathbb{E}\left[\left(K(X_{h,i})(X_{h,i})^l\varepsilon_i(\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p)\right)^k\right] = O(h),$$

using the assumed moment conditions on ε_i. Therefore, for a $\boldsymbol{t} \in \mathbb{R}^{\dim(\boldsymbol{Z}_i)}$ with $|\boldsymbol{t}| = 1$

$$\mathbb{E}\left[\left(\boldsymbol{t}'\boldsymbol{B}^{-1/2}(\boldsymbol{Z}_i - \mathbb{E}[\boldsymbol{Z}_i])\right)^3\right] = O(h).$$

and

$$\mathbb{E}\left[\left(\boldsymbol{t}'\boldsymbol{B}^{-1/2}(\boldsymbol{Z}_i - \mathbb{E}[\boldsymbol{Z}_i])\right)^4\right] = O(h).$$

Using these, and the fact that the $\boldsymbol{Z}_i$ are i.i.d. and the summands of $\boldsymbol{S}_n$ are mean zero, we have, again for a $\boldsymbol{t} \in \mathbb{R}^{\dim(\boldsymbol{Z}_i)}$ with $|\boldsymbol{t}| = 1$,

$$\mathbb{E}\left[(\boldsymbol{t}'\boldsymbol{S}_n)^3\right] = s_n^{-3}\sum_{i=1}^n \mathbb{E}\left[\left(\boldsymbol{t}'\boldsymbol{B}^{-1/2}(\boldsymbol{Z}_i - \mathbb{E}[\boldsymbol{Z}_i])\right)^3\right] = O(s_n^{-3}nh) = O(s_n^{-1}).$$

The third moment agrees with the third cumulant of $\boldsymbol{S}_n$. The fourth cumulant is

$$\mathbb{E}\left[(\boldsymbol{t}'\boldsymbol{S}_n)^4\right] - 3\mathbb{E}\left[(\boldsymbol{t}'\boldsymbol{S}_n)^2\right]^2.$$

The first term of these two is

$$\begin{aligned}
\mathbb{E}\left[(\boldsymbol{t}'\boldsymbol{S}_n)^4\right] &= s_n^{-4}\binom{4}{2}\sum_{i=1}^n\sum_{j\neq i}\mathbb{E}\left[\left(\boldsymbol{t}'\boldsymbol{B}^{-1/2}(\boldsymbol{Z}_i - \mathbb{E}[\boldsymbol{Z}_i])\right)^2\right]\mathbb{E}\left[\left(\boldsymbol{t}'\boldsymbol{B}^{-1/2}(\boldsymbol{Z}_j - \mathbb{E}[\boldsymbol{Z}_j])\right)^2\right] \\
&\quad + s_n^{-4}\sum_{i=1}^n \mathbb{E}\left[\left(\boldsymbol{t}'\boldsymbol{B}^{-1/2}(\boldsymbol{Z}_i - \mathbb{E}[\boldsymbol{Z}_i])\right)^4\right] \\
&= 3h^{-2}[1 + o(1/n)]\mathbb{E}\left[\left(\boldsymbol{t}'\boldsymbol{B}^{-1/2}(\boldsymbol{Z}_i - \mathbb{E}[\boldsymbol{Z}_i])\right)^2\right]^2 + O(s_n^{-2}).
\end{aligned}$$

By direct computation, the second piece of the fourth cumulant is

$$\mathbb{E}\left[(\boldsymbol{t}'\boldsymbol{S}_n)^2\right]^2 = \left(s_n^{-2}n\mathbb{E}\left[\left(\boldsymbol{t}'\boldsymbol{B}^{-1/2}(\boldsymbol{Z}_i - \mathbb{E}[\boldsymbol{Z}_i])\right)^2\right]\right)^2.$$

This cancels with the corresponding term of $\mathbb{E}\left[(\boldsymbol{t}'\boldsymbol{S}_n)^4\right]$, and thus the fourth cumulant is $O(s_n^{-2})$. Thus, we find that, in the notation of Skovgaard (1981), $\rho_{s,n}(\boldsymbol{t}) \asymp s_n^{-1}$, and so condition (II) of Skovgaard (1981) is satisfied by setting $a_n(\boldsymbol{t}) = Cs_n$ for an appropriate constant C. Recall that $r_n = o(r_{T_p})$, with $r_{T_p} = \max\{s_n^{-2}, \Psi_{T_p,F}^2, s_n^{-1}\Psi_{T_p,F}\}$, i.e. the slowest vanishing of the rates. Thus our r_n is ε_n in the notation of Skovgaard (1981), and condition (I) therein is satisfied because $a_n(\boldsymbol{t})^{-(s-1)} = s_n^{-3} = o(s_n^{-2}) = O(r_n)$.

Next, we verify condition (III''_α) of Skovgaard (1981, Theorem 3.4 and Remark 3.5). Let $\xi_S(\boldsymbol{t})$ be the characteristic function of $\boldsymbol{S}_n$ and $\xi_Z(\boldsymbol{t})$ that of $\boldsymbol{Z}_i$, where $\boldsymbol{t} \in \mathbb{R}^{\dim(\boldsymbol{Z}_i)}$. By the i.i.d. assumption,

$$\begin{aligned}
\xi_S(\boldsymbol{t}) = \mathbb{E}[\exp\{\mathrm{i}\boldsymbol{t}'\boldsymbol{S}_n\}] &= \prod_{i=1}^{n} \mathbb{E}[\exp\{\mathrm{i}\boldsymbol{t}'\boldsymbol{B}^{-1/2}(\boldsymbol{Z}_i - \mathbb{E}[\boldsymbol{Z}_i])/s_n\}] \\
&= \prod_{i=1}^{n} \mathbb{E}\left[\exp\left\{\mathrm{i}\left(\boldsymbol{t}'\boldsymbol{B}^{-1/2}/s_n\right)\boldsymbol{Z}_i\right\}\right]\exp\{-\mathrm{i}\boldsymbol{t}'\boldsymbol{B}^{-1/2}\mathbb{E}[\boldsymbol{Z}_i])/s_n\}.
\end{aligned}$$

The second factor is bounded by one, leaving

$$\xi_S(\boldsymbol{t}) \leq \left[\xi_Z\left(\boldsymbol{t}'\boldsymbol{B}^{-1/2}/s_n\right)\right]^n.$$

Recall that, in the notation of Skovgaard (1981), $a_n(\boldsymbol{t}) = Cs_n^{-1}$, and so condition (III$''_\alpha$) of Theorem 3.4 (and Remark 3.5) is satisfied because

$$\begin{aligned}
\sup_{|\boldsymbol{t}|>\delta C s_n^{-1}} |\xi_S(\boldsymbol{t})| &\leq \sup_{|\boldsymbol{t}|>\delta C s_n^{-1}} \left|\xi_Z\left(\boldsymbol{t}'\boldsymbol{B}^{-1/2}/s_n\right)\right|^n \\
&\leq \sup_{|\boldsymbol{t}_1|>C_1} |\xi_Z(\boldsymbol{t}_1)|^n \\
&= (1 - C_2 h)^n = o(r_n^{-C_3}),
\end{aligned}$$

for any $C_3 > 0$ by the assumption that $nh/\log(nh) \to \infty$. Thus condition (III$''_\alpha$) holds. The penultimate equality holds by Lemma S.9, which verifies that $\boldsymbol{Z}_i$ obeys the n-varying version of Cramér's condition: for h sufficiently small, for all $C_1 > 0$ there is a $C_2 > 0$ such that

$$\sup_{|\boldsymbol{t}|>C_1} |\xi_Z(\boldsymbol{t})| < (1 - C_2 h).$$

Finally, we check condition (IV) of Skovgaard (1981, Theorem 3.4). We aim to prove that

$$\sup_{0<s<1} \frac{\left|\dfrac{d^5}{ds^5}\log\xi_S\left(s\dfrac{\delta a_n(\boldsymbol{t})\boldsymbol{t}}{|\boldsymbol{t}|}\right)\right|}{5!\left|\dfrac{\delta a_n(\boldsymbol{t})\boldsymbol{t}}{|\boldsymbol{t}|}\right|^5} = O(a_n(\boldsymbol{t})^{-3}), \tag{S.34}$$

for some $\delta > 0$, with $a_n(\boldsymbol{t}) = Cs_n$ defined by conditions (I) and (II). For the supremum, as s ranges in $(0,1)$, the quantity $w = s\delta a_n(\boldsymbol{t})$ ranges in $(0, \delta a_n(\boldsymbol{t}))$. Further, by the chain rule

$$\frac{d^5}{ds^5}\log\xi_S\left(s\frac{\delta a_n(\boldsymbol{t})\boldsymbol{t}}{|\boldsymbol{t}|}\right) = \frac{d^5}{dw^5}\log\xi_S\left(\frac{w\boldsymbol{t}}{|\boldsymbol{t}|}\right)(\delta a_n(\boldsymbol{t}))^5.$$

To see why, write $\log\xi_S\left(s\delta a_n(\boldsymbol{t})\boldsymbol{t}/|\boldsymbol{t}|\right)$ as $g(w(s))$, where $w(s) = s\delta a_n(\boldsymbol{t})$ and $g(w) = \log\xi_S\left(w\boldsymbol{t}/|\boldsymbol{t}|\right)$ and then the chain rule gives

$$\frac{d^5}{ds^5}\log\xi_S\left(s\frac{\delta a_n(\boldsymbol{t})\boldsymbol{t}}{|\boldsymbol{t}|}\right) = \frac{d^5 g}{dw^5}\left(\frac{dw}{ds}\right)^5$$

because all the other terms in the chain rule expansion involve higher derivatives of the linear function $w(s) = s\delta a_n(\boldsymbol{t})$ and hence are zero. Therefore

$$\sup_{0<s<1} \frac{\left| \frac{d^5}{ds^5} \log \xi_S \left(s \frac{\delta a_n(\boldsymbol{t})\boldsymbol{t}}{|\boldsymbol{t}|} \right) \right|}{5! \left| \frac{\delta a_n(\boldsymbol{t})\boldsymbol{t}}{|\boldsymbol{t}|} \right|^5} = \sup_{0<w<\delta a_n(\boldsymbol{t})} \frac{\left| \frac{d^5}{dw^5} \log \xi_S \left(\frac{w\boldsymbol{t}}{|\boldsymbol{t}|} \right) (\delta a_n(\boldsymbol{t}))^5 \right|}{5! \left| \frac{\delta a_n(\boldsymbol{t})\boldsymbol{t}}{|\boldsymbol{t}|} \right|^5} = \sup_{0<w<\delta a_n(\boldsymbol{t})} \frac{\left| \frac{d^5}{dw^5} \log \xi_S \left(\frac{w\boldsymbol{t}}{|\boldsymbol{t}|} \right) \right|}{5!},$$

where we have canceled terms and used the fact that $|(\boldsymbol{t}/|\boldsymbol{t}|)| = 1$.

With $a_n(\boldsymbol{t}) = Cs_n$, proving Equation (S.34) is equivalent to showing that

$$\sup_{0<w<\delta a_n(\boldsymbol{t})} \left| \frac{d^5}{dw^5} \log \xi_S \left(\frac{w\boldsymbol{t}}{|\boldsymbol{t}|} \right) \right| = O\left(s_n^{-3}\right).$$

Let $\xi_{\bar{Z}}(\boldsymbol{t})$ be the characteristic function of $(\boldsymbol{Z}_i - \mathbb{E}[\boldsymbol{Z}_i])$. (This is distinct from $\xi_Z(\boldsymbol{t})$, which is the characteristic function of $\boldsymbol{Z}_i$ itself. The two are related via $\xi_{\bar{Z}}(\boldsymbol{t}) = \xi_Z(\boldsymbol{t}) \exp\{-\mathrm{i}\boldsymbol{t}'\mathbb{E}[\boldsymbol{Z}_i]\}$.) By the i.i.d. assumption

$$\log \xi_S \left(\frac{w\boldsymbol{t}}{|\boldsymbol{t}|} \right) = n \log \xi_{\bar{Z}} \left(\frac{w\boldsymbol{B}^{-1/2}\boldsymbol{t}}{|\boldsymbol{t}|s_n} \right).$$

As w varies in $(0, \delta a_n(\boldsymbol{t}))$, the quantity $u = w\boldsymbol{B}^{-1/2}s_n^{-1}$ varies in $(0, C\delta\boldsymbol{B}^{-1/2})$, by the definition of $a_n(\boldsymbol{t})$. Using the same chain rule logic as above,

$$\frac{d^5}{dw^5} \log \xi_{\bar{Z}} \left(\frac{w\boldsymbol{B}^{-1/2}\boldsymbol{t}}{|\boldsymbol{t}|s_n} \right) = \left(\frac{d^5}{du^5} \log \xi_{\bar{Z}} \left(\frac{u\boldsymbol{t}}{|\boldsymbol{t}|} \right) \right) \left(\frac{\boldsymbol{B}^{-1/2}}{s_n} \right)^5.$$

Therefore

$$\begin{aligned}
\sup_{0<w<\delta a_n(\boldsymbol{t})} \left| \frac{d^5}{dw^5} \log \xi_S \left(\frac{w\boldsymbol{t}}{|\boldsymbol{t}|} \right) \right| &= \sup_{0<w<\delta a_n(\boldsymbol{t})} \left| \frac{d^5}{dw^5} n \log \xi_{\bar{Z}} \left(\frac{w\boldsymbol{B}^{-1/2}\boldsymbol{t}}{|\boldsymbol{t}|s_n} \right) \right| \\
&= n \left(\frac{\boldsymbol{B}^{-1/2}}{s_n} \right)^5 \sup_{0<u<C\delta\boldsymbol{B}^{-1/2}} \left| \frac{d^5}{du^5} \log \xi_{\bar{Z}} \left(\frac{u\boldsymbol{t}}{|\boldsymbol{t}|} \right) \right|.
\end{aligned}$$

We aim to show that the final quantity is $O\left(s_n^{-3}\right)$. As $s_n = \sqrt{nh}$ and $\boldsymbol{B}$ is bounded above and below, this will hold if

$$\sup_{0<u<C\delta\boldsymbol{B}^{-1/2}} \left| \frac{d^5}{du^5} \log \xi_{\bar{Z}} \left(\frac{u\boldsymbol{t}}{|\boldsymbol{t}|} \right) \right| = O(h). \tag{S.35}$$

for some $\delta > 0$.

By Corollary 8.2 of Bhattacharya and Rao (1976) for the first inequality and direct calculation for the second,

$$\left| \log \xi_{\bar{Z}} \left(\frac{u\boldsymbol{t}}{|\boldsymbol{t}|} \right) - 1 \right| \leq \frac{1}{2} \left| \frac{u\boldsymbol{t}}{|\boldsymbol{t}|} \right| \mathbb{E}\left[|\boldsymbol{Z}_i - \mathbb{E}[\boldsymbol{Z}_i]|^2 \right] \leq C|u|h. \tag{S.36}$$

Therefore, for h small enough there is a $\delta > 0$ such that $C|u|h < 1/2$ for all u such that $0 < u <$

$C\delta\boldsymbol{B}^{-1/2}$. This allows us to apply Lemma 9.4 of Bhattacharya and Rao (1976), yielding the bound

$$\sup_{0<u<C\delta\boldsymbol{B}^{-1/2}}\left|\frac{d^5}{du^5}\log\xi_{\bar{Z}}\left(\frac{u\boldsymbol{t}}{|\boldsymbol{t}|}\right)\right| \leq C\mathbb{E}\left[|\boldsymbol{Z}_i-\mathbb{E}[\boldsymbol{Z}_i]|^5\right].$$

As the fifth moment of $\boldsymbol{Z}_i$ is $O(h)$, this establishes Equation (S.35) and therefore Equation (S.34), verifying condition (IV) of Skovgaard (1981, Theorem 3.4). All of the conditions of this Theorem are now verified, thus completing **Step (II)**.

Remark S.2. For building intuition it is useful to compare the bound bound in Equation (S.36) and the n-varying version of Cramér's condition established in Lemma S.9. Both reflect the fact that as $h \to 0$, $K(X_{h,i}) \to 0$, and therefore in the limit $\boldsymbol{Z}_i \equiv 0$ is a degenerate random variable. In this case of (S.36), the bound shows that as $h \to 0$, the characteristic function $\log\xi_{\bar{Z}}(u\boldsymbol{t}/|\boldsymbol{t}|) \to 1$. Lemma S.9 shows the same thing, as it is proven therein that

$$\sup_{|\boldsymbol{t}|>C_1}|\xi_Z(\boldsymbol{t})| < (1-C_2h).$$

Notice that in the limit as $h \to 0$, the conventional Cramér's condition fails. Equation (S.36) and Lemma S.9 are in qualitative agreement in this sense. ⌟

S.4.2 Step (III)

We now prove that the expansion for T_p holds and that it holds uniformly over $F \in \mathscr{F}_S$. First, by Equation (S.22) and Lemma S.1(a), T_p will obey the desired expansion (computed formally as in Section S.3.1) if $\breve{T}$ obeys an Edgeworth expansion. Now, $\breve{T}$ is given by

$$\breve{T}\left(s_n^{-1}\sum_{i=1}^{n}\boldsymbol{Z}_i\right) = \breve{T}\left(\mathbb{V}[\boldsymbol{Z}_i]^{1/2}\boldsymbol{S}_n + n\mathbb{E}[\boldsymbol{Z}_i]/s_n\right),$$

which is a smooth function of $\boldsymbol{S}_n := \sum_{i=1}^{n}\mathbb{V}[\boldsymbol{Z}_i]^{-1/2}(\boldsymbol{Z}_i-\mathbb{E}[\boldsymbol{Z}_i])/s_n$. **Step (II)** proved that $\boldsymbol{S}_n$ obeys an Edgeworth expansion, and therefore by Skovgaard (1986) we have that $\breve{T}$ does as well. Equation (S.22) and Lemma S.1(a) deliver the result pointwise for T_p.

To prove that the expansion holds uniformly, first notice that all our results hold pointwise along a sequence $F_n \in \mathscr{F}_S$. That is, the results of Skovgaard (1981) and Skovgaard (1986) hold along this sequence. We thus proceed by arguing as in Romano (2004). Recall that $r_{T_p} = \max\{s_n^{-2}, \Psi^2_{T_p,F}, s_n^{-1}\Psi_{T_p,F}\}$, i.e. the slowest vanishing of the rates. Suppose the result failed. Then we can extract a subsequence $\{F_m \in \mathscr{F}_S\}$ such that

$$r_{T_p}\left|\mathbb{P}_{F_m}[T_p<z]-\Phi(z)-E_{T_p,F_m}(z)\right| \not\to 0.$$

But this contradicts the result above, because T_p obeys the expansion given on $\{F_m \in \mathscr{F}_S\}$.

S.4.3 Lemmas

Our proof of Theorem S.1 relies on the following lemmas. Consistent with the above, we give mainly details for the T_p case, i.e. the proof in Section S.4. The details for $T_{\mathtt{rbc}}$, Section S.5, are entirely analogous. Indeed, though all the results below are stated for a bandwidth sequence h and polynomial degree p, they generalize in the obvious way under the appropriate substitutions and appropriate assumptions.

The first lemma collects high level results regarding the Delta method for Edgeworth expansions, pertaining to **Step (I)**, verifying Equation (S.22).

Lemma S.1.

(a) *Let $U_n := T_p - \breve{T}$. If $r_{T_p}^{-1}\mathbb{P}[|U_n| > r_n] = o(1)$ for a sequence r_n such that $r_n = o(r_{T_p})$, then*

$$\mathbb{P}\left[T_p < z\right] = \mathbb{P}\left[\breve{T} + U_n < z\right] = \mathbb{P}\left[\breve{T} < z\right] + o(r_{T_p}).$$

(b) *If $r_1 = O(r_1')$ and $r_2 = O(r_2')$, for sequences of positive numbers r_1, r_1', r_2, and r_2' and if a sequence of nonnegative random variables obeys $(r_1)^{-1}\mathbb{P}[U_n > r_2] \to 0$ it also holds that $(r_1')^{-1}\mathbb{P}[U_n > r_2'] \to 0$. In particular, $r_1^{-1}\mathbb{P}[|U_n| > r_n] \to 0$ implies $r_{T_p}^{-1}\mathbb{P}[|U_n| > r_n] \to 0$, for r_1 equal in order to any of s_n^{-2}, $\Psi_{T_p,F}^2$, or $s_n^{-1}\Psi_{T_p,F}$, because r_{T_p} is the largest of these, and any $r_n = o(r_{T_p})$. Thus, for different pieces of U_n defined above, we may make different choices for these two sequences, as convenient.*

Proof. Part (a) is the Delta method for Edgeworth expansions, which essentially follows from the fact that the Edgeworth expansion itself is a smooth function. See Hall (1992a, Chapter 2.7) or Maesono (1997, Lemma 2 and Remark following). Part (b) follows from elementary inequalities. □

The next set of results, Lemmas S.2–S.8, give rate bounds on the probability of deviations for various kernel-weighted sample averages. These are used in establishing Equation (S.29) in **Step (I)**. The proofs for all these Lemmas are given in the subsubsection below.

Lemma S.2. *Let the conditions of Theorem S.1 hold. For some $\delta > 0$, a positive integer k, and $C_\Gamma < \infty$, we have*

(a) $r_{T_p}^{-1}\mathbb{P}[|\mathbf{\Gamma} - \tilde{\mathbf{\Gamma}}| > \delta s_n^{-1}\log(s_n)^{1/2}] = o(1)$,

(b) $r_{T_p}^{-1}\mathbb{P}\left[\left|\mathbf{\Gamma}^{-1} - \sum_{j=0}^{k}\left(\mathbf{\Gamma}^{-1}(\tilde{\mathbf{\Gamma}} - \mathbf{\Gamma})\right)^j\tilde{\mathbf{\Gamma}}^{-1}\right| > \delta s_n^{-(k+1)}\log(s_n)^{(k+1)/2}\right] = o(1)$, *and in particular (i.e. $k = 0$)* $r_{T_p}^{-1}\mathbb{P}[|\mathbf{\Gamma}^{-1} - \tilde{\mathbf{\Gamma}}^{-1}| > \delta s_n^{-1}\log(s_n)^{1/2}] = o(1)$, *and*

(c) $r_{T_p}^{-1}\mathbb{P}[\mathbf{\Gamma}^{-1} > C_\Gamma] = o(1)$.

Lemma S.3. *Let the conditions of Theorem S.1 hold. Let $\boldsymbol{A}$ be a fixed-dimension vector or matrix of continuous functions of $X_{h,i}$ that does not depend on n. For some $\delta > 0$,*

$$r_{T_p}^{-1}\mathbb{P}\left[\left|\frac{1}{nh}\sum_{i=1}^{n}\{(K\boldsymbol{A})(X_{h,i}) - \mathbb{E}[(K\boldsymbol{A})(X_{h,i})]\}\right| > \delta s_n^{-1}\log(s_n)^{1/2}\right] \to 0.$$

Further, there is some constant $C_{\boldsymbol{A}} > 0$ such that $r_{T_p}^{-1}\mathbb{P}[\sum_{i=1}^n (K\boldsymbol{A})(X_{h,i})/(nh) > C_{\boldsymbol{A}}] = o(1)$. In particular, $r_{T_p}^{-1}\mathbb{P}[|\boldsymbol{\Lambda}_1 - \tilde{\boldsymbol{\Lambda}}_1| > \delta s_n^{-1}\log(s_n)^{1/2}] = o(1)$. Lemma S.2(a) is also a special case.

Lemma S.4. *Let the conditions of Theorem S.1 hold. Let $\boldsymbol{A}$ be a fixed-dimension vector or matrix of continuous functions of $X_{h,i}$ that does not depend on n. For some $\delta > 0$,*

$$r_{T_p}^{-1}\mathbb{P}\left[\left|\frac{1}{nh}\sum_{i=1}^n \{(K\boldsymbol{A})(X_{h,i})\varepsilon_i\}\right| > \delta s_n^{-1}\log(s_n)^{1/2}\right] \to 0.$$

In particular, with $\boldsymbol{A} = \boldsymbol{r}_p(X_{h,i})$, $r_{T_p}^{-1}\mathbb{P}\left[|\boldsymbol{\Omega}\left(\boldsymbol{Y} - \boldsymbol{M}\right)/n| > \delta s_n^{-1}\log(s_n)^{1/2}\right]$.

Lemma S.5. *Let the conditions of Theorem S.1 hold. Let $\boldsymbol{A}$ be a fixed-dimension vector or matrix of continuous functions of $X_{h,i}$ that does not depend on n. For any $\delta > 0$, $\gamma > 0$, and positive integer k,*

$$r_{T_p}^{-1}\mathbb{P}\left[\left|\frac{1}{nh}\sum_{i=1}^n \left\{(K\boldsymbol{A})(X_{h,i})\left[\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p\right]^k\right\}\right| > \delta\frac{\Psi_{T_p,F}^{k-1}}{s_n^{k-1}}\log(s_n)^{\gamma}\right] \to 0.$$

In particular, with $k = 1$ and $\boldsymbol{A} = \boldsymbol{r}_p(X_{h,i})$, $r_{T_p}^{-1}\mathbb{P}\left[|\boldsymbol{\Omega}\left(\boldsymbol{M} - \boldsymbol{R}\boldsymbol{\beta}_p\right)/n| > \delta\log(s_n)^{\gamma}\right] \to 0$.

Lemma S.6. *Let the conditions of Theorem S.1 hold. Let $\boldsymbol{A}$ be a fixed-dimension vector or matrix of continuous functions of $X_{h,i}$ that does not depend on n. For any $\delta > 0$, $\gamma > 0$, and positive integer k,*

$$r_{T_p}^{-1}\mathbb{P}\Bigg[\Bigg|\frac{1}{nh}\sum_{i=1}^n \Bigg\{(K\boldsymbol{A})(X_{h,i})\left[\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p\right]^k - \mathbb{E}\left[(K\boldsymbol{A})(X_{h,i})\left[\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p\right]^k\right]\Bigg\}\Bigg| > \delta_2\frac{\Psi_{T_p,F}^{k}}{s_n^{k}}\log(s_n)^{\gamma}\Bigg] \to 0.$$

Lemma S.7. *Let the conditions of Theorem S.1 hold. Let $\boldsymbol{A}$ be a fixed-dimension vector or matrix of continuous functions of $X_{h,i}$ that does not depend on n. For any $\delta > 0$ and $\gamma > 0$,*

$$r_{T_p}^{-1}\mathbb{P}\left[\left|\frac{1}{nh}\sum_{i=1}^n \left\{(K\boldsymbol{A})(X_{h,i})\left[\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p\right]\varepsilon_i\right\}\right| > \delta\frac{\Psi_{T_p,F}}{s_n}\log(s_n)^{\gamma}\right] \to 0.$$

Lemma S.8. *Let the conditions of Theorem S.1 hold. For any $\delta > 0$ and $\gamma > 0$,*

$$r_{T_p}^{-1}\mathbb{P}\Bigg[\Bigg|\frac{1}{nh}\sum_{i=1}^n \Bigg\{(K\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})\Big(K(X_{h,i})\left(\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p\right) - \mathbb{E}\left[K(X_{h,i})\left(\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p\right)\right]\Big)\varepsilon_i\Bigg\}\Bigg| > \delta a_n\log(s_n)^{\gamma}\Bigg] \to 0.$$

where set $a_n = s_n^{-1}\Psi_{T_p,F}$ if $r_{T_p} = s_n^{-2}$; $a_n = s_n^{-2}$ if $r_{T_p} = \Psi_{T_p,F}^2$; or $a_n = s_n^{-3/2}\Psi_{T_p,F}^{1/2}$ if $r_{T_p} = s_n^{-1}\Psi_{T_p,F}$.

Next, we show that the random variable $\boldsymbol{Z}_i$, given in Equation (S.26), obeys the appropriate n-varying version of Cramér's condition. This is used in **Step (II)** to prove that the distribution of the (properly centered and scaled) sample average of $\boldsymbol{Z}_i$ has an Edgeworth expansion. This type of Cramér's condition was first (to our knowledge) used by Hall (1991).

Lemma S.9. *Let the conditions of Theorem S.1 hold. Let $\xi_Z(\boldsymbol{t})$ be the characteristic function of the random variable $\boldsymbol{Z}_i$, given in Equation (S.26). For h sufficiently small, for all $C_1 > 0$ there is a $C_2 > 0$ such that*

$$\sup_{|t|>C_1} |\xi_Z(\boldsymbol{t})| < (1 - C_2 h).$$

Proof of Lemma S.9. Recall the definition of $\boldsymbol{Z}_i$ in Equation (S.26). It is useful to consider $\boldsymbol{Z}_i$ as a function of $(X_{h,i}, Y_i)$ rather than (X_i, Y_i). We compute the characteristic function separately depending on whether X_i is local to x. Note that h is fixed. The characteristic function of $\boldsymbol{Z}_i$ is

$$\xi_Z(\boldsymbol{t}) = \mathbb{E}[\exp\{\mathrm{i}\boldsymbol{t}'\boldsymbol{Z}_i\}] = \mathbb{E}\left[\exp\{\mathrm{i}\boldsymbol{t}'\boldsymbol{Z}_i\}\mathbb{1}\left\{|X_{h,i}| > 1\right\}\right] + \mathbb{E}\left[\exp\{\mathrm{i}\boldsymbol{t}'\boldsymbol{Z}_i\}\mathbb{1}\left\{|X_{h,i}| \le 1\right\}\right]. \tag{S.37}$$

We examine each piece in turn. For the first, begin by noticing that $|X_{h,i}| > 1$ (i.e. $X_i \notin \{\mathsf{x} \pm h\}$), then $K(X_{h,i}) = 0$, in turn implying that $\boldsymbol{Z}_i$ is the zero vector and $\exp\{\mathrm{i}\boldsymbol{t}'\boldsymbol{Z}_i\} = 1$. Therefore

$$\mathbb{E}\left[\exp\{\mathrm{i}\boldsymbol{t}'\boldsymbol{Z}_i\}\mathbb{1}\left\{|X_{h,i}| > 1\right\}\right] = \mathbb{P}[X_i \notin \{\mathsf{x} \pm h\}].$$

By assumption, the density of X is bounded and bounded away from zero in a fixed neighborhood of x. For now consider interior x, we will return to the boundary case at the end. Assume that h is small enough that this neighborhood contains $\{\mathsf{x} \pm h\}$. Then this probability is bounded as

$$\mathbb{P}[X_i \notin \{\mathsf{x} \pm h\}] = 1 - \int_{\mathsf{x}-h}^{\mathsf{x}+h} f(x)dx \le 1 - h2\left(\min_{x\in\{\mathsf{x}\pm h\}} f(x)\right) := 1 - C_3 h. \tag{S.38}$$

Next, consider the event that $|X_{h,i}| \le 1$. Let $f_{xy}(x, y)$ denote the joint density of (X, Y) and explicitly write $\boldsymbol{Z}_i = \boldsymbol{Z}_i(X_{h,i}, Y_i)$. Using the change of variables $U = (X - \mathsf{x})/h$,

$$\begin{aligned}\mathbb{E}\left[\exp\{\mathrm{i}\boldsymbol{t}'\boldsymbol{Z}_i(X_{h,i}, Y_i)\}\mathbb{1}\left\{|X_{h,i}| \le 1\right\}\right] &= \int\int_{\mathsf{x}-h}^{\mathsf{x}+h} \exp\{\mathrm{i}\boldsymbol{t}'\boldsymbol{Z}_i(x, y)\} f_{xy}(x, y)dxdy \\ &= h\int\int_{-1}^{1} \exp\{\mathrm{i}\boldsymbol{t}'\boldsymbol{Z}_i(u, y)\} f_{xy}(\mathsf{x} + uh, y)dudy.\end{aligned}$$

Suppose that K is not the uniform kernel. The assumption that $(1, K\boldsymbol{r}_{3p})(u)'$ is linearly independent implies that $\boldsymbol{Z}_i$ is a set of linearly independent and continuously differentiable functions of (u, y) on $\{[-1, 1]\} \cup \mathbb{R}$. Furthermore, by assumption, the density of (U, Y), as random variables on $\{[-1, 1]\} \cup \mathcal{Y}$, for some $\mathcal{Y} \subset \mathbb{R}$, is strictly positive. Therefore, by (Bhattacharya, 1977, Lemma 1.4), $\boldsymbol{Z}_i = \boldsymbol{Z}_i(U, Y)$ obeys Cramér's condition (as a function of random variables on $\{[-1, 1]\} \cup \mathbb{R}$), and so (Bhattacharya and Rao, 1976, p. 207) there is some $C > 0$ such that

$$\sup_{|t|>C} \left| \int \int_{-1}^{1} \exp\{\mathrm{i}\boldsymbol{t}'\boldsymbol{Z}_i(u,y)\} f_{xy}(\mathsf{x}+uh,y)dudy \right| < 1. \tag{S.39}$$

Collecting Equations (S.37), (S.38), and (S.39) yields the result when the kernel is not uniform.

If K is the uniform kernel, Equation (S.39) will still hold, as follows. Note that one element of $\boldsymbol{Z}_i(U,Y)$ is $K(U)$. For notational ease, let this be the first element, and further write $\boldsymbol{Z}_i(U,Y)$ as $\boldsymbol{Z}_i(U,Y) := 2(K(U), \tilde{\boldsymbol{Z}}_i')'$ and $\boldsymbol{t} \in \mathbb{R}^{\dim(\boldsymbol{Z})}$ as $\boldsymbol{t} = (t_{(1)}, \tilde{\boldsymbol{t}}')'$. Then, because $K(U) \equiv 1/2$ for $U \in [-1,1]$,

$$\begin{aligned}
&\sup_{|t|>C} \left| \int \int_{-1}^{1} \exp\left\{\mathrm{i}\boldsymbol{t}'\boldsymbol{Z}_i(u,y)\right\} f_{xy}(\mathsf{x}+uh,y)dudy \right| \\
&\quad = \sup_{|t|>C} \left| \int \int_{-1}^{1} \exp\left\{\mathrm{i}\boldsymbol{t}' \left[2(K(U), \tilde{\boldsymbol{Z}}_i')'\right]\right\} f_{xy}(\mathsf{x}+uh,y)dudy \right| \\
&\quad = \sup_{|t|>C} \left| \int \int_{-1}^{1} \exp\left\{\mathrm{i}\boldsymbol{t}' \left[(1, \tilde{\boldsymbol{Z}}_i')'\right]\right\} f_{xy}(\mathsf{x}+uh,y)dudy \right| \\
&\quad = \sup_{|t|>C} \left| e^{\mathrm{i}t_1} \int \int_{-1}^{1} \exp\left\{\mathrm{i}\tilde{\boldsymbol{t}}'\tilde{\boldsymbol{Z}}_i\right\} f_{xy}(\mathsf{x}+uh,y)dudy \right|.
\end{aligned}$$

Exactly as above, (Bhattacharya, 1977, Lemma 1.4) applies, but now to $\tilde{\boldsymbol{Z}}_i$, and $|e^{\mathrm{i}t_1}|$ is bounded by one, thus yielding Equation (S.39).

Finally, if x is a boundary point, then all that changes in the above proof are ranges of integration: replace $\mathsf{x} - h$ with zero and remove the factor of 2 in the definition of C_3 in (S.38), and then in the subsequent steps, integrate over $[0,1]$ instead of $[-1,1]$. □

S.4.3.1 Proofs of Lemmas S.2–S.8

Before proving Lemmas S.2–S.7 we first state some generic results that serve as building blocks for the main Lemmas above. Indeed, those results are often are almost immediate consequences of these generic results. The versions of these results for $I_{\mathtt{rbc}}$ are usually omitted, as they are entirely analogous (replacing p and h by $p+1$ and b, as well as other obvious modifications).

Lemma S.10. *Let the conditions of Theorem S.1 hold. Let $g(\cdot)$ and $m(\cdot)$ be generic continuous scalar functions. For some $\delta_1 > 0$, any $\delta_2 > 0$, $\gamma > 0$, and positive integer k, the following hold.*

(a) $s_n^2 \mathbb{P}\left[\left| s_n^{-2} \sum_{i=1}^{n} \{(Km)(X_{h,i})g(X_i) - \mathbb{E}[(Km)(X_{h,i})g(X_i)]\} \right| > \delta_1 s_n^{-1} \log(s_n)^{1/2}\right] \to 0.$

(b) $s_n^2 \mathbb{P}\left[\left| s_n^{-2} \sum_{i=1}^{n} \{(Km)(X_{h,i})g(X_i)\varepsilon_i\} \right| > \delta_1 s_n^{-1} \log(s_n)^{1/2}\right] \to 0.$

(c) $\dfrac{s_n}{\Psi_{T_p,F}} \mathbb{P}\left[\left| s_n^{-2} \sum_{i=1}^{n} (Km)(X_{h,i})g(X_i) \left[\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p\right]^k \right| > \delta_2 \dfrac{\Psi_{T_p,F}^{k-1}}{s_n^{k-1}} \log(s_n)^{\gamma}\right] \to 0.$

(d) $s_n^2 \mathbb{P}\Bigg[\Bigg| s_n^{-2} \sum_{i=1}^n \Big\{ (Km)(X_{h,i}) g(X_i) (\mu(X_i) - \boldsymbol{r}_p(X_i - x)'\boldsymbol{\beta}_p)^k$

$$- \mathbb{E}\left[(Km)(X_{h,i}) g(X_i) (\mu(X_i) - \boldsymbol{r}_p(X_i - x)'\boldsymbol{\beta}_p)^k \right] \Big\} \Bigg| > \delta_2 \left(\frac{\Psi_{T_p,F}}{s_n} \right)^k \log(s_n)^\gamma \Bigg] \to 0.$$

(e) $s_n^2 \mathbb{P}\left[\left| s_n^{-2} \sum_{i=1}^n (Km)(X_{h,i}) g(X_i) \varepsilon_i \left[\mu(X_i) - \boldsymbol{r}_p(X_i - x)'\boldsymbol{\beta}_p \right] \right| > \delta_2 \frac{\Psi_{T_p,F}}{s_n} \log(s_n)^\gamma \right] \to 0.$

(f) $r_{T_p}^{-1} \mathbb{P}\Bigg[\Bigg| \frac{1}{nh} \sum_{i=1}^n \Big\{ (Km)(X_{h,i}) \Big(K(X_{h,i}) \left(\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p \right)$

$$- \mathbb{E}\left[K(X_{h,i}) \left(\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p \right) \right] \Big) \varepsilon_i \Big\} \Bigg| > \delta a_n \log(s_n)^\gamma \Bigg] \to 0,$$

where set $a_n = s_n^{-1}\Psi_{T_p,F}$ *if* $r_{T_p} = s_n^{-2}$; $a_n = s_n^{-2}$ *if* $r_{T_p} = \Psi_{T_p,F}^2$; *or* $a_n = s_n^{-3/2}\Psi_{T_p,F}^{1/2}$ *if* $r_{T_p} = s_n^{-1}\Psi_{T_p,F}$.

Proof of Lemma S.10(a). Because the kernel function has compact support and $g(\cdot)$ and $m(\cdot)$ are continuous, we have

$$|(Km)(X_{h,i})g(X_i) - \mathbb{E}[(Km)(X_{h,i})g(X_i)]| < C_1.$$

Further, by a change of variables and using the assumptions on f, g and m:

$$\begin{aligned}\mathbb{V}[(Km)(X_{h,i})g(X_i)] \leq \mathbb{E}\left[(Km)(X_{h,i})^2 g(X_i)^2 \right] &= \int f(X_i)(Km)(X_{h,i})^2 g(X_i)^2 dX_i \\ &= h \int f(\mathsf{x} + uh) g(\mathsf{x} + uh) (Km)(u)^2 du \leq C_2 h.\end{aligned}$$

Therefore, by Bernstein's inequality

$$\begin{aligned}
&s_n^2 \mathbb{P}\left[\left| \frac{1}{s_n^2} \sum_{i=1}^n \{ (Km)(X_{h,i})g(X_i) - \mathbb{E}[(Km)(X_{h,i})g(X_i)] \} \right| > \delta_1 s_n^{-1} \log(s_n)^{1/2} \right] \\
&\leq 2 s_n^2 \exp\left\{ - \frac{(s_n^4)(\delta_1 s_n^{-1} \log(s_n)^{1/2})^2 / 2}{C_2 s_n^2 + C_1 s_n^2 \delta_1 s_n^{-1} \log(s_n)^{1/2} / 3} \right\} \\
&= 2 \exp\{2 \log(s_n)\} \exp\left\{ - \frac{\delta_1^2 \log(s_n)/2}{C_2 + C_1 \delta_1 s_n^{-1} \log(s_n)^{1/2}/3} \right\} \\
&= 2 \exp\left\{ \log(s_n) \left[2 - \frac{\delta_1^2 / 2}{C_2 + C_1 \delta_1 s_n^{-1} \log(s_n)^{1/2}/3} \right] \right\},
\end{aligned}$$

which vanishes for any δ_1 large enough, as $s_n^{-1} \log(s_n)^{1/2} \to 0$. □

Proof of Lemma S.10(b). For a sequence $a_n \to \infty$ to be given later, define

$$H_i = s_n^{-1} (Km)(X_{h,i}) g(X_i) \left(Y_i \mathbb{1}\{Y_i \leq a_n\} - \mathbb{E}[Y_i \mathbb{1}\{Y_i \leq a_n\} \mid X_i] \right)$$

and

$$T_i = s_n^{-1} (Km)(X_{h,i}) g(X_i) \left(Y_i \mathbb{1}\{Y_i > a_n\} - \mathbb{E}[Y_i \mathbb{1}\{Y_i > a_n\} \mid X_i] \right).$$

By the conditions on $g(\cdot)$ and $t(\cdot)$ and the kernel function,

$$|H_i| < C_1 s_n^{-1} a_n$$

and

$$\begin{aligned}
\mathbb{V}[H_i] = s_n^{-2}\mathbb{V}[(Km)(X_{h,i})g(X_i)Y_i\mathbb{1}\{Y_i \le a_n\}] &\le s_n^{-2}\mathbb{E}\left[(Km)(X_{h,i})^2 g(X_i)^2 Y_i^2 \mathbb{1}\{Y_i \le a_n\}\right] \\
&\le s_n^{-2}\mathbb{E}\left[(Km)(X_{h,i})^2 g(X_i)^2 Y_i^2\right] \\
&= s_n^{-2}\int (Km)(X_{h,i})^2 g(X_i)^2 v(X_i) f(X_i) dX_i \\
&= s_n^{-2} h \int (Km)(u)^2 (gvf)(\mathsf{x} - uh) du \\
&\le C_2/n.
\end{aligned}$$

Therefore, by Bernstein's inequality

$$\begin{aligned}
s_n^2 \mathbb{P}\left[\left|\sum_{i=1}^n H_i\right| > \delta_1 \log(s_n)^{1/2}\right] &\le 2 s_n^2 \exp\left\{-\frac{\delta_1^2 \log(s_n)/2}{C_2 + C_1 s_n^{-1} a_n \delta_1 \log(s_n)^{1/2}/3}\right\} \\
&\le 2\exp\{2\log(s_n)\}\exp\left\{-\frac{\delta_1^2 \log(s_n)/2}{C_2 + C_1 s_n^{-1} a_n \delta_1 \log(s_n)^{1/2}/3}\right\} \\
&\le 2\exp\left\{\log(s_n)\left[2 - \frac{\delta_1^2/2}{C_2 + C_1 s_n^{-1} a_n \delta_1 \log(s_n)^{1/2}/3}\right]\right\},
\end{aligned}$$

which vanishes for δ_1 large enough as long as $s_n^{-1} a_n \log(s_n)^{1/2}$ does not diverge.

Next, let $\pi > 2$ be such that $\mathbb{E}[|Y|^{2+\pi}|X = x]$ is finite in the neighborhood of x, which is possible under Assumption S.1, and then, by Markov's inequality:

$$\begin{aligned}
s_n^2 \mathbb{P}\left[\left|\sum_{i=1}^n T_i\right| > \delta \log(s_n)^{1/2}\right] &\le s_n^2 \frac{1}{\delta^2 \log(s_n)} \mathbb{E}\left[\left|\sum_{i=1}^n T_i\right|^2\right] \\
&\le s_n^2 \frac{1}{\delta_1^2 \log(s_n)} n\mathbb{E}\left[T_i^2\right] \\
&\le s_n^2 \frac{1}{\delta_1^2 \log(s_n)} n\mathbb{V}\left[s_n^{-1}(Km)(X_{h,i})g(X_i)Y_i\mathbb{1}\{Y_i > a_n\}\right] \\
&\le s_n^2 \frac{1}{\delta_1^2 \log(s_n)} n s_n^{-2}\mathbb{E}\left[(Km)(X_{h,i})^2 g(X_i)^2 Y_i^2 \mathbb{1}\{Y_i > a_n\}\right] \\
&\le s_n^2 \frac{1}{\delta_1^2 \log(s_n)} n s_n^{-2}\mathbb{E}\left[(Km)(X_{h,i})^2 g(X_i)^2 |Y_i|^{2+\pi} a_n^{-\pi}\right] \\
&\le s_n^2 \frac{1}{\delta_1^2 \log(s_n)} n s_n^{-2}(C h a_n^{-\pi}) \\
&\le \frac{C}{\delta_1^2}\frac{s_n^2}{\log(s_n) a_n^{\pi}},
\end{aligned}$$

which vanishes if $s_n^2 \log(s_n)^{-1} a_n^{-\pi} \to 0$.

It thus remains to choose a_n such that $s_n^{-1} a_n \log(s_n)^{1/2}$ does not diverge and $s_n^2 \log(s_n)^{-1} a_n^{-\pi} \to 0$. This can be accomplished by setting $a_n = s_n^A$ for any $2/\pi \le A < 1$, which is possible as $\pi > 2$. $\square$

Proof of Lemma S.10(c). By Markov's inequality

$$
\begin{aligned}
&\frac{s_n}{\Psi_{T_p,F}} \mathbb{P}\left[\left| s_n^{-2} \sum_{i=1}^n (Km)(X_{h,i}) g(X_i) \left[\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p\right]^k \right| > \delta_2 (s_n^{-1}\Psi_{T_p,F})^{k-1} \log(s_n)^\gamma\right] \\
&\qquad \le \frac{s_n}{\Psi_{T_p,F}} \left(\frac{s_n}{\Psi_{T_p,F}}\right)^{k-1} \frac{1}{\delta_2 \log(s_n)^\gamma} \mathbb{E}\left[h^{-1}(Km)(X_{h,i}) g(X_i) \left[\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p\right]^k\right] \\
&\qquad \le \frac{1}{\delta_2 \log(s_n)^\gamma} \mathbb{E}\left[h^{-1}(Km)(X_{h,i}) g(X_i) \left[\frac{s_n}{\Psi_{T_p,F}} \left(\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p\right)\right]^k\right] \\
&\qquad = O(\log(s_n)^{-\gamma}) \to 0.
\end{aligned}
$$

This relies on the calculations in Section S.2, and the compact support of the kernel and continuity of $m(\cdot)$ and $g(\cdot)$ to ensure that the expectation is otherwise bounded. $\square$

Proof of Lemma S.10(d). Note that the summand is mean zero and apply Markov's inequality to find

$$
\begin{aligned}
&s_n^2 \mathbb{P}\Bigg[\Bigg| s_n^{-2} \sum_{i=1}^n \Big\{ (Km)(X_{h,i}) g(X_i) (\mu(X_i) - \boldsymbol{r}_p(X_i - x)'\boldsymbol{\beta}_p)^k \\
&\qquad\qquad - \mathbb{E}\left[(Km)(X_{h,i}) g(X_i) (\mu(X_i) - \boldsymbol{r}_p(X_i - x)'\boldsymbol{\beta}_p)^k\right]\Big\}\Bigg| > \delta_2 \left(\frac{\Psi_{T_p,F}}{s_n}\right)^k \log(s_n)^\gamma\Bigg] \\
&\le s_n^2 \left(\frac{s_n}{\Psi_{T_p,F}}\right)^{2k} \frac{1}{\delta_2^2 \log(s_n)^{2\gamma}} s_n^{-2} \mathbb{E}\left[h^{-1}(Km)(X_{h,i}) g(X_i) (\mu(X_i) - \boldsymbol{r}_p(X_i - x)'\boldsymbol{\beta}_p)^{2k}\right] \\
&= \frac{1}{\delta_2^2 \log(s_n)^{2\gamma}} \mathbb{E}\left[h^{-1}(Km)(X_{h,i}) g(X_i) \left[\left(\frac{s_n}{\Psi_{T_p,F}}\right) (\mu(X_i) - \boldsymbol{r}_p(X_i - x)'\boldsymbol{\beta}_p)\right]^{2k}\right] \\
&= o(1).
\end{aligned}
$$

The final line relies on the calculations in Section S.2. $\square$

Proof of Lemma S.10(e). By Markov's inequality, since ε_i is conditionally mean zero, we have

$$
\begin{aligned}
&s_n^2 \mathbb{P}\left[\left| s_n^{-2} \sum_{i=1}^n (Km)(X_{h,i}) g(X_i) \varepsilon_i \left[\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p\right] \right| > \delta_2 (s_n^{-1}\Psi_{T_p,F}) \log(s_n)^\gamma\right] \\
&\qquad \le s_n^2 \frac{1}{\delta_2^2 s_n^{-2} \Psi_{T_p,F}^2 \log(s_n)^{2\gamma}} \frac{1}{s_n^2} \mathbb{E}\left[h^{-1} \left((Km)(X_{h,i}) g(X_i) \varepsilon_i\right)^2 \left[\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p\right]^2\right] \\
&\qquad \le \frac{1}{\delta_2^2 \log(s_n)^{2\gamma}} \mathbb{E}\left[h^{-1} \left((Km)(X_{h,i}) g(X_i) \varepsilon_i\right)^2 \left[\frac{s_n}{\Psi_{T_p,F}} \left(\mu(X_i) - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta}_p\right)\right]^2\right] \\
&\qquad = O(\log(s_n)^{-2\gamma}) \to 0.
\end{aligned}
$$

This relies on the calculations in Section S.2, and the compact support of the kernel and continuity of $m(\cdot)$ and $g(\cdot)$ to ensure that the expectation is otherwise bounded. $\square$

Proof of Lemma S.10(f). By Markov's inequality, since ε_i is conditionally mean zero, we have

$$\begin{aligned}
&r_{T_p}^{-1}\mathbb{P}\Bigg[\Bigg|\frac{1}{nh}\sum_{i=1}^{n}\Big\{(Km)(X_{h,i})\Big(K(X_{h,i})\left(\mu(X_i)-\boldsymbol{r}_p(X_i-\mathsf{x})'\boldsymbol{\beta}_p\right) \\
&\qquad\qquad -\mathbb{E}\left[K(X_{h,i})\left(\mu(X_i)-\boldsymbol{r}_p(X_i-\mathsf{x})'\boldsymbol{\beta}_p\right)\right]\Big)\varepsilon_i\Big\}\Bigg|>\delta a_n\log(s_n)^{\gamma}\Bigg] \\
&\leq \frac{r_{T_p}^{-1}}{a_n^2\log(s_n)^{2\gamma}}\frac{1}{nh}\mathbb{E}\Bigg[h^{-1}(Km)^2(X_{h,i})\Big(K(X_{h,i})\left(\mu(X_i)-\boldsymbol{r}_p(X_i-\mathsf{x})'\boldsymbol{\beta}_p\right) \\
&\qquad\qquad -\mathbb{E}\left[K(X_{h,i})\left(\mu(X_i)-\boldsymbol{r}_p(X_i-\mathsf{x})'\boldsymbol{\beta}_p\right)\right]\Big)^2 v(X_i)\Bigg] \\
&= \frac{r_{T_p}^{-1}}{a_n^2\log(s_n)^{2\gamma}}\frac{1}{nh}\Bigg\{\mathbb{E}\Bigg[h^{-1}(Km)^2(X_{h,i})K(X_{h,i})^2\left(\mu(X_i)-\boldsymbol{r}_p(X_i-\mathsf{x})'\boldsymbol{\beta}_p\right)^2 v(X_i)\Bigg] \\
&\quad -2\mathbb{E}\Bigg[h^{-1}(Km)^2(X_{h,i})K(X_{h,i})\left(\mu(X_i)-\boldsymbol{r}_p(X_i-\mathsf{x})'\boldsymbol{\beta}_p\right)v(X_i)\Bigg]\mathbb{E}\left[K(X_{h,i})\left(\mu(X_i)-\boldsymbol{r}_p(X_i-\mathsf{x})'\boldsymbol{\beta}_p\right)\right] \\
&\quad +\mathbb{E}\Bigg[h^{-1}(Km)^2(X_{h,i})v(X_i)\Bigg]\mathbb{E}\left[K(X_{h,i})\left(\mu(X_i)-\boldsymbol{r}_p(X_i-\mathsf{x})'\boldsymbol{\beta}_p\right)\right]^2\Bigg\} \\
&\asymp \frac{r_{T_p}^{-1}}{a_n^2\log(s_n)^{2\gamma}}\frac{1}{nh}\left(\frac{\Psi_{T_p,F}}{s_n}\right)^2\left\{1+h+h^2\right\} \\
&\asymp \frac{r_{T_p}^{-1}}{a_n^2\log(s_n)^{2\gamma}}\frac{1}{nh}\left(\frac{\Psi_{T_p,F}}{s_n}\right)^2.
\end{aligned}$$

If $r_{T_p}=s_n^{-2}$, this vanishes for $a_n=s_n^{-1}\Psi_{T_p,F}$. If $r_{T_p}=\Psi_{T_p,F}^2$, this vanishes for $a_n=s_n^{-2}$. If $r_{T_p}=s_n^{-1}\Psi_{T_p,F}$, this vanishes for $a_n=s_n^{-3/2}\Psi_{T_p,F}^{1/2}$. This relies on the calculations in Section S.2, and the compact support of the kernel and continuity of $m(\cdot)$ to ensure that the expectation is otherwise bounded. $\square$

Proof of Lemma S.2. A typical element of $\boldsymbol{\Gamma}-\tilde{\boldsymbol{\Gamma}}$ is, for some integer $k\in[0,2p]$,

$$\frac{1}{nh}\sum_{i=1}^{n}\left\{K(X_{h,i})X_{h,i}^k-\mathbb{E}\left[K(X_{h,i})X_{h,i}^k\right]\right\},$$

which has the form treated in Lemma S.10(a). Therefore, by Boole's inequality and p fixed,

$$\begin{aligned}
&r_{T_p}^{-1}\mathbb{P}[|\boldsymbol{\Gamma}-\tilde{\boldsymbol{\Gamma}}|>\delta s_n^{-1}\log(s_n)^{1/2}] \\
&\qquad\leq Cr_{T_p}^{-1}\max_{k\in[0,2p]}\mathbb{P}\left[\left|\frac{1}{nh}\sum_{i=1}^{n}\left\{K(X_{h,i})X_{h,i}^k-\mathbb{E}\left[K(X_{h,i})X_{h,i}^k\right]\right\}\right|>\delta s_n^{-1}\log(s_n)^{1/2}\right]\to 0,
\end{aligned}$$

by Lemma S.1(b). This establishes part (a).

To prove part (b), first note that for any fixed δ_1, part (a) and the sub-multiplicativity of the Frobenius norm imply

$$r_{T_p}^{-1}\mathbb{P}\left[|\boldsymbol{\Gamma}^{-1}(\boldsymbol{\Gamma}-\tilde{\boldsymbol{\Gamma}})|\geq\delta_1\right]\;\leq\;r_{T_p}^{-1}\mathbb{P}\left[|(\boldsymbol{\Gamma}-\tilde{\boldsymbol{\Gamma}})|\geq\delta_1|\boldsymbol{\Gamma}^{-1}|^{-1}\right]\to0,\tag{S.40}$$

because under the maintained assumptions

$$\tilde{\boldsymbol{\Gamma}}=\mathbb{E}\left[h^{-1}(K\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})\right]=h^{-1}\int(K\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})f(X_i)dX_i=\int(K\boldsymbol{r}_p\boldsymbol{r}_p')(u)f(\mathsf{x}+uh)du$$

is bounded away from zero and infinity for n large enough.

Now, on the event $\mathcal{G}_n=\{|\boldsymbol{\Gamma}^{-1}(\boldsymbol{\Gamma}-\tilde{\boldsymbol{\Gamma}})|<1\}$, we use the identity $\boldsymbol{\Gamma}=\tilde{\boldsymbol{\Gamma}}\left(\boldsymbol{I}-\boldsymbol{\Gamma}^{-1}(\tilde{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\right)$ to write $\boldsymbol{\Gamma}^{-1}$ as

$$\boldsymbol{\Gamma}^{-1}=\left(\boldsymbol{I}-\boldsymbol{\Gamma}^{-1}(\tilde{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\right)^{-1}\tilde{\boldsymbol{\Gamma}}^{-1}=\sum_{j=0}^{\infty}\left(\boldsymbol{\Gamma}^{-1}(\tilde{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\right)^j\tilde{\boldsymbol{\Gamma}}^{-1}.$$

Write $a_n=s_n^{-(k+1)}\log(s_n)^{(k+1)/2}$ Using results (S.40) with $\delta_1=1$, we find that $r_{T_p}^{-1}(1-\mathbb{P}[\mathcal{G}_n])=r_{T_p}^{-1}\mathbb{P}[|\boldsymbol{\Gamma}^{-1}(\boldsymbol{\Gamma}-\tilde{\boldsymbol{\Gamma}})|\geq1]\to0$. Therefore

$$\begin{aligned}
&r_{T_p}^{-1}\mathbb{P}\left[\left|\boldsymbol{\Gamma}^{-1}-\sum_{j=0}^{k}\left(\boldsymbol{\Gamma}^{-1}(\tilde{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\right)^j\tilde{\boldsymbol{\Gamma}}^{-1}\right|>\delta a_n\right]\\
&\quad\leq r_{T_p}^{-1}\mathbb{P}\left[\left\{\left|\boldsymbol{\Gamma}^{-1}-\sum_{j=0}^{k}\left(\boldsymbol{\Gamma}^{-1}(\tilde{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\right)^j\tilde{\boldsymbol{\Gamma}}^{-1}\right|>\delta a_n\right\}\cup\mathcal{G}_n\right]+r_{T_p}^{-1}(1-\mathbb{P}[\mathcal{G}_n])\\
&\quad\leq r_{T_p}^{-1}\mathbb{P}\left[\left|\sum_{j=0}^{\infty}\left(\boldsymbol{\Gamma}^{-1}(\tilde{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\right)^j\tilde{\boldsymbol{\Gamma}}^{-1}-\sum_{j=0}^{k}\left(\boldsymbol{\Gamma}^{-1}(\tilde{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\right)^j\tilde{\boldsymbol{\Gamma}}^{-1}\right|>\delta a_n\right]+o(1)\\
&\quad=r_{T_p}^{-1}\mathbb{P}\left[\left|\sum_{j=k+1}^{\infty}\left(\boldsymbol{\Gamma}^{-1}(\tilde{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\right)^j\tilde{\boldsymbol{\Gamma}}^{-1}\right|>\delta a_n\right]+o(1).
\end{aligned}$$

Again using sub-multiplicativity and part (a), $|(\boldsymbol{\Gamma}^{-1}(\tilde{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma}))^j|\leq|\boldsymbol{\Gamma}^{-1}|^j|\tilde{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma}|^j\to0$, and so by dominated convergence and the partial sum formula, the above display is bounded as

$$\begin{aligned}
&\leq r_{T_p}^{-1}\mathbb{P}\left[\sum_{j=k+1}^{\infty}\left|\left(\boldsymbol{\Gamma}^{-1}(\tilde{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\right)^j\right|\left|\tilde{\boldsymbol{\Gamma}}^{-1}\right|>\delta a_n\right]+o(1)\\
&\leq r_{T_p}^{-1}\mathbb{P}\left[\frac{\left|\boldsymbol{\Gamma}^{-1}(\tilde{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\right|^{k+1}}{1-\left|\boldsymbol{\Gamma}^{-1}(\tilde{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\right|}\left|\tilde{\boldsymbol{\Gamma}}^{-1}\right|>\delta a_n\right]+o(1).
\end{aligned}$$

Finally, using result (S.40) with some fixed $\delta_1 < 1$, this last display is bounded by

$$r_{T_p}^{-1}\mathbb{P}\left[\left|\tilde{\mathbf{\Gamma}} - \mathbf{\Gamma}\right|^{k+1} > \left|\tilde{\mathbf{\Gamma}}^{-1}\right|^{-k-2}(1-\delta_1)\delta a_n\right] + r_{T_p}^{-1}\mathbb{P}\left[\left|\mathbf{\Gamma}^{-1}(\tilde{\mathbf{\Gamma}} - \mathbf{\Gamma})\right| \geq \delta_1\right] + o(1) = o(1),$$

where the final convergence follows by part (a).

For part (c), let $C_\Gamma < \infty$ be such that $|\tilde{\mathbf{\Gamma}}^{-1}| < C_\Gamma/2$. Then

$$\begin{aligned} r_{T_p}^{-1}\mathbb{P}[\mathbf{\Gamma}^{-1} > C_\Gamma] &= r_{T_p}^{-1}\mathbb{P}[\left(\mathbf{\Gamma}^{-1} - \tilde{\mathbf{\Gamma}}^{-1}\right) + \tilde{\mathbf{\Gamma}}^{-1} > C_\Gamma] \\ &\leq r_{T_p}^{-1}\mathbb{P}\left[\left|\mathbf{\Gamma}^{-1} - \tilde{\mathbf{\Gamma}}^{-1}\right| > \delta s_n^{-1}\log(s_n)^{1/2}\right] + r_{T_p}^{-1}\mathbb{P}\left[\left|\tilde{\mathbf{\Gamma}}^{-1}\right| > C_\Gamma - \delta s_n^{-1}\log(s_n)^{1/2}\right], \end{aligned}$$

which vanishes because the second term is zero for n large enough such that $\delta s_n^{-1}\log(s_n)^{1/2} < C_\Gamma/2$ and the first is $o(1)$ by part (a). □

Proof of Lemma S.3. The result follows from identical steps to proving Lemma S.2(a), because Lemma S.10(a) also applies. The second conclusion follows from the first exactly the same way Lemma S.2(c) follows from Lemma S.2(a). □

Proof of Lemma S.4. Let $[\boldsymbol{A}]_{j,k}$ be the $\{j,k\}$ entry of $\boldsymbol{A}$. By Boole's inequality, since the dimension of $\boldsymbol{A}$ is fixed, and Lemma S.1(b),

$$\begin{aligned} & r_{T_p}^{-1}\mathbb{P}\left[\left|\frac{1}{nh}\sum_{i=1}^{n}\{(K\boldsymbol{A})(X_{h,i})\varepsilon_i\}\right| > \delta s_n^{-1}\log(s_n)^{1/2}\right] \\ &\quad \leq C r_{T_p}^{-1}\max_{j,k}\mathbb{P}\left[\left|s_n^{-2}\sum_{i=1}^{n}\left\{\left(K\,[\boldsymbol{A}]_{j,k}\right)(X_{h,i})\varepsilon_i\right\}\right| > \delta s_n^{-1}\log(s_n)^{\gamma}\right] \\ &\quad \leq C s_n^2\max_{j,k}\mathbb{P}\left[\left|s_n^{-2}\sum_{i=1}^{n}\left\{\left(K\,[\boldsymbol{A}]_{j,k}\right)(X_{h,i})\varepsilon_i\right\}\right| > \delta s_n^{-1}\log(s_n)^{\gamma}\right], \end{aligned}$$

which vanishes by Lemma S.10(b). □

Proof of Lemma S.5. Exactly as above, but using Lemma S.10(c). □

Proof of Lemma S.6. Exactly as above, but using Lemma S.10(d). □

Proof of Lemma S.7. Exactly as above, but using Lemma S.10(e). □

Proof of Lemma S.8. Exactly as above, but using Lemma S.10(f). □

S.5 Proof of Theorem S.1 with Bias Correction

Proving Theorem S.1 for $T_{\mathtt{rbc}}$ follows the exact same steps as for T_p. The reason being that both are based such similar estimation procedures. To illustrate this point, recall that when $\rho = 1$, $T_{\mathtt{rbc}}$ is the same as T_p but based on a higher degree polynomial. In this special case, there is nothing left to prove: simply apply Theorem S.1 with p replaced with $p+1$. Or, alternatively, re-walk the entire proof replacing p with $p+1$ everywhere.

The more general case, that is, with generic ρ, is not conceptually more difficult, just more cumbersome. There are two chief changes. First, the bias rate changes due to the bias correction, but this is automatically accounted for by the terms of the expansion and the conditions of the theorem. For example, note that the rate $r_{I_{\mathtt{rbc}}}$ automatically includes the new bias rate, as it is defined in general in terms of $\Psi_{T,F}$ Second, there are additional kernel-weighted averages that enter into $T_{\mathtt{rbc}}$ and these will enter into the construction of $\boldsymbol{Z}_i$ and the bounding of remainder terms.

Recall the definitions of the point estimators, standard errors, and t-statistics from Section S.1, specifically Equations (S.8), (S.10), and (S.11):

$$\hat{\mu}_p^{(\nu)} = \frac{1}{nh^\nu}\nu! \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\boldsymbol{Y}, \qquad \hat{\sigma}_p^2 = \nu!^2 \boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}(h\boldsymbol{\Omega}\hat{\boldsymbol{\Sigma}}_p\boldsymbol{\Omega}'/n)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu, \qquad T_p = \frac{\sqrt{nh^{1+2\nu}}(\hat{\mu}_p^{(\nu)} - \mu^{(\nu)})}{\hat{\sigma}_p}$$

$$\hat{\theta}_{\mathtt{rbc}} = \frac{1}{nh^\nu}\nu! \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}_{\mathtt{rbc}}\boldsymbol{Y}, \quad \hat{\sigma}_{\mathtt{rbc}}^2 = \nu!^2 \boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}(h\boldsymbol{\Omega}_{\mathtt{rbc}}\hat{\boldsymbol{\Sigma}}_{\mathtt{rbc}}\boldsymbol{\Omega}_{\mathtt{rbc}}'/n)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu, \quad T_{\mathtt{rbc}} = \frac{\sqrt{nh^{1+2\nu}}(\hat{\theta}_{\mathtt{rbc}} - \mu^{(\nu)})}{\hat{\sigma}_{\mathtt{rbc}}}.$$

Comparing these, we see that the only differences in the change from $\hat{\boldsymbol{\Sigma}}_p$ and $\boldsymbol{\Omega}$ to $\hat{\boldsymbol{\Sigma}}_{\mathtt{rbc}}$ and $\boldsymbol{\Omega}_{\mathtt{rbc}}$, where (to repeat):

- $\hat{\boldsymbol{\Sigma}}_{\mathtt{rbc}} = \mathrm{diag}(\hat{v}(X_i) : i = 1, \ldots, n)$, with $\hat{v}(X_i) = (Y_i - \boldsymbol{r}_{p+1}(X_i - \mathsf{x})'\hat{\boldsymbol{\beta}}_{p+1})^2$,
- $\boldsymbol{\Omega}_{\mathtt{rbc}} = \boldsymbol{\Omega} - \rho^{p+1}\boldsymbol{\Lambda}_1\boldsymbol{e}_{p+1}'\bar{\boldsymbol{\Gamma}}^{-1}\bar{\boldsymbol{\Omega}}$,
- $\rho = h/b$,
- $\boldsymbol{\Lambda}_k = \boldsymbol{\Omega}\left[X_{h,1}^{p+k}, \ldots, X_{h,n}^{p+k}\right]'/n$,
- $X_{b,i} = (X_i - \mathsf{x})/b$,
- $\bar{\boldsymbol{\Gamma}} = \frac{1}{nb}\sum_{i=1}^n (K\boldsymbol{r}_{p+1}\boldsymbol{r}_{p+1}')(X_{b,i})$, and
- $\bar{\boldsymbol{\Omega}} = [(K\boldsymbol{r}_{p+1})(X_{b,1}), (K\boldsymbol{r}_{p+1})(X_{b,2}), \ldots, (K\boldsymbol{r}_{p+1})(X_{b,n})]$.

Notice that these are the same as their counterparts for T_p, but with $b = h\rho^{-1}$ in place of h and $p+1$ in place of p. With these comparisons in mind, we briefly discuss the three steps of Section S.4, highlighting key pieces.

For **Step (I)**, first observe that the "numerator", or $\hat{\theta}_{\mathtt{rbc}}$, portion of the t-statistic is once again already a smooth function of well-behaved random variables, albeit different ones that for T_p. Terms will be added to $\boldsymbol{Z}_i$ to reflect this. In particular, $\boldsymbol{\Lambda}_1$, $\bar{\boldsymbol{\Gamma}}$, and $\bar{\boldsymbol{\Omega}}$ are present. Importantly, Lemma S.2 applies to $\bar{\boldsymbol{\Gamma}}$ with $b = h\rho^{-1}$ in place of h and $p+1$ in place of p.

Turning to the Studentization, Equation (S.23) expands the quantity $(h\boldsymbol{\Omega}\hat{\boldsymbol{\Sigma}}_p\boldsymbol{\Omega}'/n)$ and this needs to be adapted to account instead for $(h\boldsymbol{\Omega}_{\mathtt{rbc}}\hat{\boldsymbol{\Sigma}}_{\mathtt{rbc}}\boldsymbol{\Omega}_{\mathtt{rbc}}'/n)$, which requires two changes. The fundamental issue remains the estimated residuals and thus the terms represented by $\boldsymbol{V}_1 - \boldsymbol{V}_6$ will remain conceptually the same. The first change, which is automatically accounted for by the rate assumptions of the Theorem and the terms of the expansion, are that the bias is now lower because the residuals are estimated with a $p+1$ degree fit. This matches the numerator bias, and thus the calculations are as above. Second, whereas the summands of each term of $\boldsymbol{V}_1 - \boldsymbol{V}_6$ include $(K^2 \boldsymbol{r}_p \boldsymbol{r}_p')(X_{h,i})$ stemming from the pre- and post-multiplying by $\boldsymbol{\Omega}$, now we multiply by $\boldsymbol{\Omega}_{\texttt{rbc}}$, which means the new versions of $\boldsymbol{V}_1 - \boldsymbol{V}_6$ have

$$\Big((K\boldsymbol{r}_p)(X_{h,i}) - \rho^{p+1}\boldsymbol{\Lambda}_1 \boldsymbol{e}_{p+1}' \bar{\boldsymbol{\Gamma}}^{-1}(K\boldsymbol{r}_{p+1})(X_{b,i})\Big)\Big((K\boldsymbol{r}_p)(X_{h,i}) - \rho^{p+1}\boldsymbol{\Lambda}_1 \boldsymbol{e}_{p+1}' \bar{\boldsymbol{\Gamma}}^{-1}(K\boldsymbol{r}_{p+1})(X_{b,i})\Big)'.$$

This is mostly a change in notation and increased complexity of all terms, which now will include many more factors that much be accounted for. This does not affect the rates or the identity of the important terms: in other words the expansion is not fundamentally changed. Notice that in estimating the residuals $\hat{v}(X_i) = (Y_i - \boldsymbol{r}_{p+1}(X_i - \mathsf{x})'\hat{\boldsymbol{\beta}}_{p+1})^2$ is used, and not, as might also be plausible, any further bias correction (such as $\hat{v}(X_i) = (Y_i - \boldsymbol{r}_{p+1}(X_i - \mathsf{x})'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}_{\texttt{rbc}}\boldsymbol{Y}/(nh))^2$. This means no other terms appear.

We illustrate with one example. Consider the first term bounded in Equation (S.30). For $\boldsymbol{V}_3$ defined following Equation (S.23) it was shown following Equation (S.30) that

$$r_{I_{\texttt{rbc}}}^{-1}\mathbb{P}\left[\left|\nu!^2 \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1}\left(\boldsymbol{V}_3 - \mathbb{E}[\boldsymbol{V}_3]\right)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu s_n \nu! \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\left(\boldsymbol{Y} - \boldsymbol{M}\right)/n\right| > r_n\right] \to 0.$$

The corresponding bound required here is

$$r_{I_{\texttt{rbc}}}^{-1}\mathbb{P}\left[\left|\nu!^2 \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1}\left(\boldsymbol{V}_{3,\texttt{rbc}} - \mathbb{E}[\boldsymbol{V}_{3,\texttt{rbc}}]\right)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu s_n \nu! \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}_{\texttt{rbc}}\left(\boldsymbol{Y} - \boldsymbol{M}\right)/n\right| > r_n\right] \to 0. \tag{S.41}$$

The analogue of $\boldsymbol{V}_3$ is given by applying the two changes above: the bias term and replacing $(K^2 \boldsymbol{r}_p \boldsymbol{r}_p')(X_{h,i})$ with the expression above, yielding what we will call $\boldsymbol{V}_{3,\texttt{rbc}}$:

$$\begin{aligned}
\boldsymbol{V}_{3,\texttt{rbc}} = {} & \frac{1}{nh}\sum_{i=1}^n (K^2 \boldsymbol{r}_p \boldsymbol{r}_p')(X_{h,i})\left[\mu(X_i) - \boldsymbol{r}_{p+1}(X_i - \mathsf{x})'\boldsymbol{\beta}_{p+1}\right]^2 \\
& + \rho^{2p+2}\boldsymbol{\Lambda}_1 \boldsymbol{e}_{p+1}' \bar{\boldsymbol{\Gamma}}^{-1}\left\{\frac{1}{nh}\sum_{i=1}^n (K^2 \boldsymbol{r}_{p+1}\boldsymbol{r}_{p+1}')(X_{b,i})\left[\mu(X_i) - \boldsymbol{r}_{p+1}(X_i - \mathsf{x})'\boldsymbol{\beta}_{p+1}\right]^2\right\}\bar{\boldsymbol{\Gamma}}^{-1}\boldsymbol{e}_{p+1}\boldsymbol{\Lambda}_1' \\
& + \rho^{p+1}\boldsymbol{\Lambda}_1 \boldsymbol{e}_{p+1}' \bar{\boldsymbol{\Gamma}}^{-1}\left\{\frac{1}{nh}\sum_{i=1}^n (K\boldsymbol{r}_{p+1})(X_{b,i})(K\boldsymbol{r}_p)(X_{h,i})\left[\mu(X_i) - \boldsymbol{r}_{p+1}(X_i - \mathsf{x})'\boldsymbol{\beta}_{p+1}\right]^2\right\} \\
& + \rho^{p+1}\left\{\frac{1}{nh}\sum_{i=1}^n (K\boldsymbol{r}_p)(X_{h,i})(K\boldsymbol{r}_{p+1}')(X_{b,i})\left[\mu(X_i) - \boldsymbol{r}_{p+1}(X_i - \mathsf{x})'\boldsymbol{\beta}_{p+1}\right]^2\right\}\bar{\boldsymbol{\Gamma}}^{-1}\boldsymbol{e}_{p+1}\boldsymbol{\Lambda}_1'.
\end{aligned}$$

Verifying Equation (S.41) now amounts to repeating the original logic (for the first term of Equation (S.30)) four times, once for each line here.

First, observe that all the conclusions of Lemma S.2 hold in exactly the same way for $\bar{\boldsymbol{\Gamma}}$ (substituting b and $p+1$ for h and p respectively, as needed), and thus the same type of bounds can be applied whenever necessary. Second, Lemma S.3 implies that we can bound and remove the $\boldsymbol{\Lambda}_1$ everywhere as well, just as was originally done with $\boldsymbol{\Gamma}^{-1}$. These two together imply that Lemma

S.4 holds for $\boldsymbol{\Omega}_{\texttt{rbc}}$ in place of $\boldsymbol{\Omega}$ (again with b and $p+1$ where necessary).

For the first term listed of $\boldsymbol{V}_{3,\texttt{rbc}}$ the original logic now goes through almost as written, simply with additional bounds for $\boldsymbol{\Lambda}_1$ and $\bar{\boldsymbol{\Gamma}}$. Lemma S.6 applies just the same, only p is replaced by $p+1$ but this is accounted for automatically by the generic rates.

For the remaining three terms listed of $\boldsymbol{V}_{3,\texttt{rbc}}$, the argument is much the same. The only additional complexity is the bandwidth b (or ρ). However, because b does not vanish faster than h, this will not cause a problem. Firstly, pre-multiplication by ρ to a positive power can only reduce the asymptotic order because $\rho \not\to \infty$. Secondly, for the factors enclosed in braces in each of the three terms, Lemma S.6 will still hold. Checking the proof of Lemma S.10(d), which gives Lemma S.6, we can see that we simply must substitute the appropriate bias calculations of Section S.2.

For the second term listed of $\boldsymbol{V}_{3,\texttt{rbc}}$ this is immediate, since the form is identical and we only need to substitute b and $p+1$ for h and p respectively, after re-writing so the averaging is done according to nb instead of nh.

$$\rho^{2p+1}\boldsymbol{\Lambda}_1\boldsymbol{e}_{p+1}'\bar{\boldsymbol{\Gamma}}^{-1}\left\{\frac{1}{nb}\sum_{i=1}^{n}(K^2\boldsymbol{r}_{p+1}\boldsymbol{r}_{p+1}')(X_{b,i})\left[\mu(X_i)-\boldsymbol{r}_{p+1}(X_i-\mathsf{x})'\boldsymbol{\beta}_{p+1}\right]^2\right\}\bar{\boldsymbol{\Gamma}}^{-1}\boldsymbol{e}_{p+1}\boldsymbol{\Lambda}_1'.$$

For the third and fourth terms listed of $\boldsymbol{V}_{3,\texttt{rbc}}$, the only potential further complication is that the summand includes both $X_{h,i}$ and $X_{b,i}$. However, because $X_{b,i}=\rho X_{h,i}$, all applications of changing variables can proceed as usual, as typified by, for smooth functions m_1 and m_2 (c.f. Lemma S.10)

$$h^{-1}\mathbb{E}[(Km_1)(X_{h,i})(Km_2)(X_{b,i})]=\int_{-1}^{1}(Km_1)(u)(Km_2)(\rho u)f(\mathsf{x}+uh)du,$$

which is just as well behaved as usual.

Collecting all of these results establishes the convergence of Equation (S.41). This illustrates that although the notational complexity is increased and there are more terms to keep track of, there is nothing fundamentally different in **Step (I)** for $T_{\texttt{rbc}}$. We omit the rest of the details.

Moving to **Step (II)**, the proof proceeds in almost exactly the same way as in Section S.4.1, but now the quantity $\boldsymbol{Z}_i$ is different. Collecting all the changes described above (the inclusion of $\bar{\boldsymbol{\Gamma}}$, Lp_1, and $\bar{\boldsymbol{\Omega}}$, the change in estimated residuals to $\hat{\boldsymbol{\Sigma}}_{\texttt{rbc}}$, and the premultiplication by $\boldsymbol{\Omega}_{\texttt{rbc}}$), the new $\boldsymbol{Z}_i$ is now the collection (deleting duplicate entries)

$$\begin{aligned}\boldsymbol{Z}_{i,\texttt{rbc}}=\Bigg(&\boldsymbol{Z}_{i,\texttt{rbc}}^{\text{numer}},\ \boldsymbol{Z}_{i,\texttt{rbc}}^{\text{denom}}\Big[(K^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})\Big],\ \boldsymbol{Z}_{i,\texttt{rbc}}^{\text{denom}}\Big[(K^2\boldsymbol{r}_{p+1}\boldsymbol{r}_{p+1}')(X_{b,i})\Big],\\ &\boldsymbol{Z}_{i,\texttt{rbc}}^{\text{denom}}\Big[(K\boldsymbol{r}_p)(X_{h,i})(K\boldsymbol{r}_{p+1}')(X_{b,i})\Big]\Bigg)',\end{aligned}\tag{S.42}$$

where

$$\boldsymbol{Z}_{i,\texttt{rbc}}^{\text{numer}}=\Bigg(\Big\{(K\boldsymbol{r}_p)(X_{h,i})(Y_i-\boldsymbol{r}_{p+1}(X_i-\mathsf{x})'\boldsymbol{\beta}_{p+1})\Big\}',$$

$$\left\{(K\boldsymbol{r}_{p+1})(X_{b,i})(Y_i - \boldsymbol{r}_{p+1}(X_i - \mathsf{x})'\boldsymbol{\beta}_{p+1})\right\}',$$
$$\operatorname{vech}\left\{(K\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})\right\}',$$
$$\operatorname{vech}\left\{(K\boldsymbol{r}_{p+1}\boldsymbol{r}_{p+1}')(X_{b,i})\right\}',$$
$$\left.\operatorname{vech}\left\{(K\boldsymbol{r}_p)(X_{h,i})(X_{h,i})^{p+1}\right\}',\right)$$

and for a matrix depending on $(X_{h,i}, X_{b,i})$, the function $\boldsymbol{Z}_{i,\mathtt{rbc}}^{\text{denom}}\Big[\boldsymbol{\kappa}(X_{h,i}, X_{b,i})\Big]$ is

$$\boldsymbol{Z}_{i,\mathtt{rbc}}^{\text{denom}}\Big[\boldsymbol{\kappa}(X_{h,i}, X_{b,i})\Big] = \Bigg(\operatorname{vech}\left\{\boldsymbol{\kappa}(X_{h,i}, X_{b,i})\varepsilon_i^2\right\}',$$
$$\operatorname{vech}\left\{\boldsymbol{\kappa}(X_{h,i}, X_{b,i})(X_{b,i})^0\varepsilon_i\right\}', \operatorname{vech}\left\{(K^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})(X_{b,i})^1\varepsilon_i\right\}',$$
$$\operatorname{vech}\left\{\boldsymbol{\kappa}(X_{h,i}, X_{b,i})(X_{b,i})^2\varepsilon_i\right\}', \ldots, \operatorname{vech}\left\{(K^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})(X_{b,i})^{p+1}\varepsilon_i\right\}',$$
$$\operatorname{vech}\left\{\boldsymbol{\kappa}(X_{h,i}, X_{b,i})\left\{\varepsilon_i\left[\mu(X_i) - \boldsymbol{r}_{p+1}(X_i - \mathsf{x})'\boldsymbol{\beta}_{p+1}\right]\right\}\right\}'\Bigg).$$

$\boldsymbol{Z}_{i,\mathtt{rbc}}$ is notationally intimidating, but comparing this to the original $\boldsymbol{Z}_i$ of Equation (S.26), we see that nothing fundamentally different has been added: the additions are mostly just repetition to account for the higher degree local polynomial. Notice that if $\rho = 1$, i.e. $h = b$, then many of the elements are duplicated (or contained in others) and can be removed: examples include the first, third, and fifth lines of $\boldsymbol{Z}_{i,\mathtt{rbc}}^{\text{numer}}$ and all of $\boldsymbol{Z}_{i,\mathtt{rbc}}^{\text{denom}}\Big[(K^2\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i})\Big]$. (Note also that in estimating the residuals $\hat{v}(X_i) = (Y_i - \boldsymbol{r}_{p+1}(X_i - \mathsf{x})'\hat{\boldsymbol{\beta}}_{p+1})^2$ is used, and not, as might also be plausible, any further bias correction (such as $\hat{v}(X_i) = (Y_i - \boldsymbol{r}_{p+1}(X_i - \mathsf{x})'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}_{\mathtt{rbc}}\boldsymbol{Y}/(nh))^2$. This means no other terms appear.)

Because, by assumption, $\rho \not\to \infty$, the asymptotic orders do not change. Therefore, verifying conditions (I), (II), and (IV) of Theorem 3.4 of Skovgaard (1981) are nearly identical for this new $\boldsymbol{Z}_{i,\mathtt{rbc}}$. For condition ($\text{III}_\alpha''$) of Skovgaard (1981, Theorem 3.4 and Remark 3.5) the crucial ingredient is Lemma S.9, which continues to hold in exactly the same way.

Finally, **Step (III)** carries over essentially without change, completing the proof of Theorem S.1 with bias correction.

S.6 Computing the Terms of the Expansion

Computing the terms of the Edgeworth expansion of Theorem S.1, listed in Section S.3.1, is straightforward but tedious. We give a short summary here, following the essential steps of (Hall, 1992a, Chapter 2). In what follows, will always discard higher order terms (those that will not appear in the Theorem) and write $A \overset{o}{=} B$ to denote $A = B + o((nh)^{-1} + (nh)^{-1/2}\Psi_{T,F} + \Psi_{T,F}^2)$. Let $\tilde{\boldsymbol{G}}$ stand in for $\tilde{\boldsymbol{\Gamma}}$ or $\tilde{\bar{\boldsymbol{\Gamma}}}$, $\tilde{p}$ stand in for p or $p+1$, and d_n stand in for h or b, all depending on if $T = T_p$ or $T_{\mathtt{rbc}}$.

Note however, that h is still used in many places, in particular for stabilizing fixed-n expectations, for $T_{\texttt{rbc}}$. We will also need the notation defined in Section S.3.1.

The steps to compute the expansion are as follows. First, we compute a Taylor expansion of T around nonrandom denominators. Then we compute the first four moments of this expansion. These are then combined into cumulants, which determine the terms of the expansion.

The Taylor expansion is

$$T \overset{o}{=} \left\{1 - \frac{1}{2\tilde{\sigma}_T^2}\left(W_{T,1} + W_{T,2} + W_{T,3}\right) + \frac{3}{8\tilde{\sigma}_T^4}\left(W_{T,1} + W_{T,2} + W_{T,3}\right)^2\right\} \\ \times \ \tilde{\sigma}_T^{-1}\left\{N_{T,1} + N_{T,2} + N_{T,3} + B_{T,1}\right\},$$

where

$$W_{T,1} = \frac{1}{nh}\sum_{i=1}^{n}\left\{\ell_T^0(X_i)^2\left(\varepsilon_i^2 - v(X_i)\right)\right\} - 2\frac{1}{n^2h^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\{\ell_T^0(X_i)^2\boldsymbol{r}_{\tilde{p}}(X_{d_n,i})'\tilde{\boldsymbol{G}}^{-1}(K\boldsymbol{r}_{\tilde{p}})(X_{d_n,i})\varepsilon_i\varepsilon_j\right\} \\ + \frac{1}{n^3h^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\left\{\ell_T^0(X_i)^2\boldsymbol{r}_{\tilde{p}}(X_{d_n,i})'\tilde{\boldsymbol{G}}^{-1}(K\boldsymbol{r}_{\tilde{p}})(X_{d_n,i})\varepsilon_j\varepsilon_k\right\},$$

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

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

$$B_{T,1} = s_n\frac{1}{nh}\sum_{i=1}^{n}\ell_T^0(X_i)[\mu(X_i) - \boldsymbol{r}_{\tilde{p}}(X_i - x)'\boldsymbol{\beta}_{\tilde{p}}],$$

$$N_{T,1} = s_n\frac{1}{nh}\sum_{i=1}^{n}\ell_T^0(X_i)\varepsilon_i,$$

$$N_{T,2} = s_n\frac{1}{(nh)^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\ell_T^1(X_i,X_j)\varepsilon_i,$$

$$N_{T,3} = s_n\frac{1}{(nh)^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\ell_T^2(X_i,X_j,X_k)\varepsilon_i,$$

with the final line defining $\ell_T^2(X_i,X_j,X_k)$ in the obvious way following ℓ_T^1, i.e. taking account of the next set of remainders. Terms involving $\ell_T^2(X_i,X_j,X_k)$ are higher-order, which is why it is not needed in Section S.3.1. To concretize the notation, note that $\Psi_{T,F} = \mathbb{E}[B_{T,1}]$, and, for example for T_p we are defining,

$$N_{T_p,1} = s_n\nu!\boldsymbol{e}_\nu'\tilde{\boldsymbol{\Gamma}}^{-1}\boldsymbol{\Omega}(\boldsymbol{Y}\boldsymbol{M})/n,$$
$$N_{T_p,2} = s_n\nu!\boldsymbol{e}_\nu'\tilde{\boldsymbol{\Gamma}}^{-1}(\tilde{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma})\tilde{\boldsymbol{\Gamma}}^{-1}\boldsymbol{\Omega}(\boldsymbol{Y}\boldsymbol{M})/n,$$
$$N_{T_p,3} = s_n\nu!\boldsymbol{e}_\nu'\tilde{\boldsymbol{\Gamma}}^{-1}(\tilde{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma})\tilde{\boldsymbol{\Gamma}}^{-1}(\tilde{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma})\tilde{\boldsymbol{\Gamma}}^{-1}\boldsymbol{\Omega}(\boldsymbol{Y}\boldsymbol{M})/n.$$

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

$$\mathbb{E}[T] \overset{o}{=} \tilde{\sigma}_T^{-1}\mathbb{E}\left[B_{T,1}\right] - \frac{1}{2\tilde{\sigma}_T^2}\mathbb{E}\left[W_{T,1}N_{T,1}\right],$$

$$\begin{aligned}\mathbb{E}[T^2] \overset{o}{=} & \frac{1}{\tilde{\sigma}_T^2}\mathbb{E}\left[N_{T,1}^2 + N_{T,2}^2 + 2N_{T,1}N_{T,2} + 2N_{T,1}N_{T,3}\right] \\ & - \frac{1}{\tilde{\sigma}_T^4}\mathbb{E}\left[W_{T,1}N_{T,1}^2 + W_{T,2}N_{T,1}^2 + W_{T,3}N_{T,1}^2 + 2W_{T,2}N_{T,1}N_{T,2}\right] \\ & + \frac{1}{\tilde{\sigma}_T^6}\mathbb{E}\left[W_{T,1}^2N_{T,1}^2 + W_{T,2}^2N_{T,1}^2\right] + \frac{1}{\tilde{\sigma}_T^2}\mathbb{E}\left[B_{T,1}^2\right] - \frac{1}{\tilde{\sigma}_T^4}\mathbb{E}\left[W_{T,1}N_{T,1}B_{T,1}\right],\end{aligned}$$

$$\mathbb{E}[T^3] \overset{o}{=} \frac{1}{\tilde{\sigma}_T^3}\mathbb{E}\left[N_{T,1}^3\right] - \frac{3}{2\tilde{\sigma}_T^5}\mathbb{E}\left[W_{T,1}N_{T,1}^3\right] + \frac{3}{\tilde{\sigma}_T^3}\mathbb{E}\left[N_{T,1}^2B_{T,1}\right],$$

and

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

Computing each factor, we get the following results. For these terms below, indexes i, j, and k are always distinct (i.e. $X_{h,i} \neq X_{h,j} \neq X_{h,k}$).

$$\begin{aligned}\mathbb{E}\left[B_{T,1}\right] &= \Psi_{T,F}, \\ \mathbb{E}\left[W_{T,1}N_{T,1}\right] &\overset{o}{=} s_n^{-1}\mathbb{E}\left[h^{-1}\ell_T^0(X_i)^3\varepsilon_i^3\right], \\ \mathbb{E}\left[N_{T,1}^2\right] &\overset{o}{=} \tilde{\sigma}_T^2, \\ \mathbb{E}\left[N_{T,1}N_{T,2}\right] &\overset{o}{=} s_n^{-2}\mathbb{E}\left[h^{-1}\ell_T^1(X_i,X_i)\ell_T^0(X_i)\varepsilon_i^2\right], \\ \mathbb{E}\left[N_{T,2}^2\right] &\overset{o}{=} s_n^{-1}\mathbb{E}\left[h^{-2}\ell_T^1(X_i,X_j)^2\varepsilon_i^2\right], \\ \mathbb{E}\left[N_{T,2}N_{T,3}\right] &\overset{o}{=} s_n^{-2}\mathbb{E}\left[h^{-2}\ell_v^2(X_i,X_j,X_j)\ell_T^0(X_i)\varepsilon_i^2\right], \\ \mathbb{E}\left[W_{T,1}N_{T,1}^2\right] &\overset{o}{=} s_n^{-2}\Bigg\{\mathbb{E}\left[h^{-1}\ell_T^0(X_i)^4\left(\varepsilon_i^4 - v(X_i)^2\right)\right] \\ &\quad - 2\tilde{\sigma}_T^2\mathbb{E}\left[h^{-1}\ell_T^0(X_i)^2\boldsymbol{r}_{\tilde{p}}(X_{d_n,i})'\tilde{\boldsymbol{G}}^{-1}(K\boldsymbol{r}_{\tilde{p}})(X_{d_n,i})\varepsilon_i^2\right] \\ &\quad - 4\mathbb{E}\left[h^{-1}\ell_T^0(X_i)^4\boldsymbol{r}_{\tilde{p}}(X_{d_n,i})'\tilde{\boldsymbol{G}}^{-1}\varepsilon_i^2\right]\mathbb{E}\left[h^{-1}(K\boldsymbol{r}_{\tilde{p}})(X_{d_n,i})\ell_T^0(X_i)\varepsilon_i^2\right]\end{aligned}$$

$$
\begin{aligned}
&+ \tilde{\sigma}_T^2 \mathbb{E}\left[h^{-2}\ell_T^0(X_i)^2\left(\boldsymbol{r}_{\tilde{p}}(X_{d_n,i})'\tilde{\boldsymbol{G}}^{-1}(K\boldsymbol{r}_{\tilde{p}})(X_{d_n,j})\right)^2\varepsilon_j^2\right] \\
&+ 2\mathbb{E}\left[h^{-1}\ell_T^0(X_j)^2\left(\mathbb{E}\left[h^{-1}\boldsymbol{r}_{\tilde{p}}(X_{d_n,j})'\tilde{\boldsymbol{G}}^{-1}(K\boldsymbol{r}_{\tilde{p}})(X_{d_n,i})\ell_T^0(X_i)\varepsilon_i^2|X_j\right]\right)^2\right]\Bigg\},
\end{aligned}
$$

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

The so-called approximate cumulants of T, denoted here by $\kappa_{T,k}$ for the k^{th} cumulant, can now be directly calculated from these approximate moments using standard formulas (Hall, 1992a, Equation (2.6)). It is useful to list these and collect their asymptotic orders. For the first two, we split them into two subterms each, by their different asymptotic order.

$$\begin{aligned}
\kappa_{T,1} &= \mathbb{E}[T] := \kappa_{T,1,1} + \kappa_{T,1,2} \overset{o}{=} s_n^{-1} + \Psi_{T,F}, \\
\kappa_{T,2} &= \mathbb{E}[T^2] - \mathbb{E}[T]^2 := 1 + \kappa_{T,2,1} + \kappa_{T,2,2} \overset{o}{=} 1 + s_n^{-2} + s_n^{-1}\Psi_{T,F}, \\
\kappa_{T,3} &= \mathbb{E}[T^3] - 3\mathbb{E}[T^2]\mathbb{E}[T] + 2\mathbb{E}[T]^3 \overset{o}{=} s_n^{-1}, \\
\kappa_{T,4} &= \mathbb{E}[T^4] - 4\mathbb{E}[T^3]\mathbb{E}[T] - 3\mathbb{E}[T^2]^2 + 12\mathbb{E}[T^2]\mathbb{E}[T]^2 - 6\mathbb{E}[T]^4 \overset{o}{=} s_n^{-2}.
\end{aligned}$$

Next, our equivalent of (Hall, 1992a, Equation (2.22)) would be the exponential of

$$\begin{aligned}
&\kappa_{T,1}(it) + \frac{1}{2}(it)^2(\kappa_{T,2} - 1) + \frac{1}{3!}(it)^3\kappa_{T,3} + \frac{1}{4!}(it)^4\kappa_{T,4} \\
&\quad + \frac{1}{2}(it)^2\left(\kappa_{T,1,1}^2 + 2\kappa_{T,1,1}\kappa_{T,1,2}\kappa_{T,1,2}^2\right) + \frac{1}{2}\frac{1}{3!^2}(it)^6\kappa_{T,3}^2 \\
&\quad + \frac{1}{2}2\frac{1}{3!}(it)(it)^3\left(\kappa_{T,1,1}\kappa_{T,3} + \kappa_{T,1,2}\kappa_{T,3}\right).
\end{aligned}$$

Then, the final computation is done by following (Hall, 1992a, p. 44f, Equations (2.17)). We find that the Edgeworth expansion, with asymptotic order listed in parentheses at right, is given by

$$\begin{aligned}
\Phi(z) - \phi(z)\Bigg\{ &\left[\kappa_{T,1,1} + \frac{1}{3!}(z^2 - 1)\kappa_{T,3}\right] && (s_n^{-1}) \\
&\left[\kappa_{T,1,2}\right] && (\Psi_{T,F}) \\
&\Bigg[\frac{1}{2}z\kappa_{T,1,1}^2 + \frac{1}{2}\frac{1}{3!^2}z(z^4 - 10z^2 + 15)\kappa_{T,3}^2 \\
&\qquad + \frac{1}{2}2\frac{1}{3!}z(z^2 - 3)\kappa_{T,1,1}\kappa_{T,3} + \frac{1}{2}z\kappa_{T,2,1} + \frac{1}{4!}z(z^2 - 3)\kappa_{T,4}\Bigg] && (s_n^{-2}) \\
&\left[\frac{1}{2}z\kappa_{T,1,2}^2\right] && (\Psi_{T,F}^2) \\
&\left[\frac{1}{2}z2\kappa_{T,1,1}\kappa_{T,1,2} + \frac{1}{2}2\frac{1}{3!}z(z^2 - 3)\kappa_{T,1,2}\kappa_{T,3} + \frac{1}{2}z\kappa_{T,2,2}\right]\Bigg\}. && (s_n^{-1}\Psi_{T,F})
\end{aligned}$$

This is exactly the result of Theorem S.1 and these terms, in the order displayed, are exactly the $\omega_k(T,z), k = 1, 2, 3, 4, 5$ of Section S.3.1.

S.7 Notes on Alternative Standard Errors

The proofs above are based on specific standard errors. In particular, we use the fixed-n form of the variance from Equation (S.9), namely

$$\sigma_p^2 = \nu!^2 \boldsymbol{e}_\nu' \boldsymbol{\Gamma}^{-1}(h\boldsymbol{\Omega}\boldsymbol{\Sigma}\boldsymbol{\Omega}'/n)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu,$$

and estimate $\boldsymbol{\Sigma}$ using regression residuals, $\hat{\boldsymbol{\Sigma}}_p = \mathrm{diag}(\hat{v}(X_i) : i = 1, \ldots, n)$, with $\hat{v}(X_i) = (Y_i - \boldsymbol{r}_p(X_i - \mathsf{x})'\hat{\boldsymbol{\beta}}_p)^2$ for $\hat{\boldsymbol{\beta}}_p$ defined in Equation (S.4). This is the HC0 variance estimator. We discuss two types of alternatives here: (i) different estimators of essentially the same fixed-n object and (ii) different population standardizations altogether. If other standard errors are used, the results may change. The type and severity of the change will depend on the choice of standard error. In particular, the coverage error rate can be slower, but not faster. This is because the Studentization and standardization do not affect the rate of any term besides the $\lambda_{I,F}\omega_{3,I,F}$ term, and thus $\lambda_{I,F} \equiv 0$ is the most that can be accomplished through variance estimation.

Within the fixed-n form, we consider two alternative estimators of (essentially) the conditional variances of Equation (S.9): the HCk class estimators and nearest-neighbor based estimators.

First, motivated by the fact that the least-squares residuals are on average too small, we could implement one of the HCk class of heteroskedasticity-consistent standard errors (MacKinnon, 2013) beyond HC0. In particular, HC0, HC1, HC2, and HC3 are allowed in the `nprobust` package (Calonico et al., 2019). These are defined as follows. First, $\hat{\sigma}^2_p$ (and $\hat{\sigma}^2_{\mathtt{rbc}}$) defined above and treated in the proofs is the HC0 estimator, employing the estimated residuals unweighted: $\hat{\varepsilon}_i^2 = \hat{v}(X_i) = (Y_i - \boldsymbol{r}_p(X_i - \mathsf{x})'\hat{\boldsymbol{\beta}}_p)^2$. Then, for $k = 1, 2, 3$, the $\hat{\sigma}^2_p$-HCk estimator is obtained by dividing $\hat{\varepsilon}_i^2$ by, respectively, $(n - 2\,\mathrm{trace}(\boldsymbol{Q}_p) + \mathrm{trace}(\boldsymbol{Q}_p'\boldsymbol{Q}_p))/n$, $(1 - \boldsymbol{Q}_{p,ii})$, and $(1 - \boldsymbol{Q}_{p,ii})^2$, where $\boldsymbol{Q}_{p,ii}$ is the i-th diagonal element of the projection matrix $\boldsymbol{Q}_p := \check{\boldsymbol{R}}'(\check{\boldsymbol{R}}'\boldsymbol{W}\check{\boldsymbol{R}})^{-1}\check{\boldsymbol{R}}'\boldsymbol{W} = \check{\boldsymbol{R}}'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}/n$. The corresponding estimators $\hat{\sigma}^2_{\mathtt{rbc}}$-HC$k$ are the same way, substituting the appropriate pieces.

These estimators may perform better in small samples, a conjecture backed by simulation studies elsewhere. Adapting the proofs to allow for HC1, HC2, and HC3 would be notationally extremely cumbersome, but is conceptually straightforward. The building block of each is the matrix $\boldsymbol{Q}_p$, which is almost already a function of $\boldsymbol{Z}_i$ from (S.26); it is not difficult to see that Cramér's condition is plausible for this object. It is important to note that the rates in the expansion would not change, only the constants (through the terms of (S.23)).

A second option, still using the fixed-n form and also designed to improve upon the least squares residuals, is to use a nearest-neighbor-based estimator with a fixed number of neighbors (Muller and Stadtmuller, 1987). This is also allowed in our software (Calonico et al., 2019). For a fixed, positive integer J, let $X_{j(i)}$ denote the j-th closest observation to X_i, $j = 1, \ldots, J$. Set $\hat{v}(X_i) = \frac{J}{J+1}(Y_i - \sum_{j=1}^{J} Y_{j(i)}/J)^2$. This estimate is unbiased for $v(X_i)$, and although $\hat{v}(\cdot)$ is inconsistent, the resulting $\hat{\sigma}^2_p = \nu!^2\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}(h\boldsymbol{\Omega}\hat{\boldsymbol{\Sigma}}_{NN}\boldsymbol{\Omega}'/n)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu$ provides valid Studentization (as would the analogous $\hat{\sigma}^2_{\mathtt{rbc}}$). This approach, however, falls outside our proofs. Lemma S.9 would not verify Cramér's condition for this estimator. A modified approach to verifying condition (III$''_\alpha$) of Skovgaard (1981) would be required and Assumption S.2 would not be sufficient.

Finally, as discussed above, on may use a different form of standardization altogether. As argued in the main text and above, using variance forms other than (S.9) can be detrimental to coverage by injecting terms with $\lambda_{I,F} \neq 0$. Examples were given in Section S.3.1. The most common option would be to employ the asymptotic approximation to the conditional variance:

$$\sigma^2 \to_{\mathbb{P}} \frac{v(\mathsf{x})}{f(\mathsf{x})}\mathcal{V},$$

where $f(\cdot)$ is the marginal density of X and $\mathcal{V}$ is a known constant depending only on the equivalent kernel (and thus $\mathcal{V}_p$ and $\mathcal{V}_{\texttt{rbc}}$ would be different); see Fan and Gijbels (1996, Theorem 3.1). Estimating this quantity requires estimating the conditional variance function and the (inverse of the) density at a single point, the point of interest x. If both of these are based on kernel methods using the same kernel and bandwidth h, then Theorem S.1 allows for this choice. It is clear that the expansion of the Studentization, Equation (S.23), will change dramatically, as will the elements of $\boldsymbol{Z}_i$. However, the latter change will be relatively innocuous as far as the proof is concerned, because Lemma S.9 covers the objects already. But the change to Equation (S.23) will result in additional terms, with potentially slower rates, appearing the Edgeworth expansion. See the discussion in Section S.3.1.

There are certainly many other options for (first-order) valid Studentization. Other population choices include (i) using $\hat{v}(X_i) = (Y_i - \hat{m}(\mathsf{x}))^2$; (ii) using local or assuming global heteroskedasticity; (iii) using other nonparametric estimators for $v(X_i)$, relying on new tuning parameters. None of these can be recommended based on our results. As above, some can be accommodated into our proof more or less directly, depending on the implementation details.

S.8 Check Function Loss

In the main text, it was pointed out that coverage error can be measured by the check function loss:

$$\sup_{F \in \mathscr{F}_S} \mathcal{L}\Big(\mathbb{P}_F[\mu^{(\nu)} \in I] - (1-\alpha)\Big), \qquad \mathcal{L}(e) = \mathcal{L}_\tau(e) = e\left(\tau - \mathbb{1}\{e < 0\}\right)$$

Using the check function loss allows the researcher, through their choice of τ, to evaluate inference procedures according to their preferences against over- and under-coverage. Setting $\tau = 1/2$ recovers the above, symmetric measure of coverage error. Guarding more against undercoverage (a preference for conservative intervals) requires choosing a $\tau < 1/2$. For example, setting $\tau = 1/3$ encodes the belief that undercoverage is twice as bad as the same amount of overcoverage.

Using this loss will affect the constants of the optimal bandwidths and kernels (dependent on how these are optimized, such as for length, coverage error, or trading these off) but the rates will not be impacted. This is due to standard properties of the check function, which, for completeness, we spell out in the following result.

Lemma S.11. $\mathcal{L}(e) = e\left(\tau - \mathbb{1}\{e < 0\}\right)$ *obeys:*

(a) $\mathcal{L}(ae) = a\mathcal{L}(e)$ *for* $a > 0$*,*

(b) $\mathcal{L}(e) \leq (\tau + 1)|e|$*, and*

(c) $\mathcal{L}(e_1 + e_2) \leq \mathcal{L}(e_1) + \mathcal{L}(e_2)$ *for* $a > 0$.

Proof. The first property follows because $\mathcal{L}(ae) = (ae)\,(\tau - \mathbb{1}\{(ae) < 0\})$ and, as $a > 0$, $\mathbb{1}\{(ae) < 0\} = \mathbb{1}\{e < 0\}$. The second uses the obvious bounds. The third, the triangle inequality, holds as follows.

$$\begin{aligned}
\mathcal{L}(e_1 + e_2) &= (e_1 + e_2)\,(\tau - \mathbb{1}\{(e_1 + e_2) < 0\}) \\
&= e_1\,(\tau - \mathbb{1}\{e_1 < 0\}) + e_2\,(\tau - \mathbb{1}\{e_2 < 0\}) \\
&\quad + e_1 \mathbb{1}\{e_1 < 0\} + e_2 \mathbb{1}\{e_2 < 0\} - (e_1 + e_2)\mathbb{1}\{(e_1 + e_2) < 0\}.
\end{aligned}$$

In the second equality, the first line is exactly $\mathcal{L}(e_1)+\mathcal{L}(e_2)$. The second line is nonpositive. To this, consider four cases. (1) If $e_1 \geq 0$ and $e_2 \geq 0$, then all the indicators are zero and the second line is zero. (2) If $e_1 < 0$ and $e_2 < 0$, then all the indicators are one and the second line is $e_1 + e_2 - (e_1 + e_2)$ and is again zero. (3) If $e_1 \geq 0$, $e_2 < 0$, and $e_1 \geq |e_2|$, then $\mathbb{1}\{e_1 < 0\} = \mathbb{1}\{(e_1 + e_2) < 0\} = 0$, and the second line is $e_2 < 0$. (4) If $e_1 \geq 0$, $e_2 < 0$, and $e_1 < |e_2|$, then $\mathbb{1}\{e_2 < 0\} = \mathbb{1}\{(e_1 + e_2) < 0\} = 1$, and the second line is $e_2 - (e_1 + e_2) = -e_1 < 0$. □

S.9 Simulation Results and Numerical Details

S.9.1 Simulation Study

In this section we present the complete results from our simulation study addressing the finite-sample performance of the methods described in the main paper. All results are qualitatively consistent with the main theoretical results of our paper.

We study model (S.1) with X_i uniformly distributed on $[-1, 1]$, ε distributed independently standard normal, and

$$\mu(x) = \frac{\sin(3\pi x/2)}{1 + 18x^2(\operatorname{sgn}(x) + 1)},$$

where $\operatorname{sgn}(x) - 1, 0$ or -1 according to $x > 0$, $x = 0$ or $x < 0$, respectively.

We consider $5,000$ simulation replications, where for each replication we generate data as i.i.d. draws of size $n = \{100, 250, 500, 750, 1000, 2000\}$. The point of evaluation is one of six equally spaced evaluation points $\mathsf{x} \in \{-1, -0.6, -0.2, 0.2, 0.6, 1\}$ using the Epanechnikov and Uniform kernel, setting $p = 1$ (for $\nu = 0$) and $p = 2$ (for $\nu = 1$). Finally, we evaluate the performance of the confidence intervals using several bandwidth choices. First, we use $\hat{h}_{\mathtt{rbc}}$, a data-driven version of the inference-optimal bandwidth $h_{\mathtt{rbc}}$. We also consider the analogous version for undersmoothing confidence intervals, $\hat{h}_{\mathtt{us}}$, and the standard choice in practice, $\hat{h}_{\mathtt{mse}}$. In all cases, robust bias correction is implemented using $\rho = \rho^*$.

We report empirical coverage probabilities and average interval length of nominal 95% confidence interval for $\mu(\mathsf{x})$ and $\mu^{(1)}(\mathsf{x})$ based on robust bias correction and undersmoothing.

First, in Figures S.1, S.3, and S.5 we present empirical coverage probabilities for $\nu = 0$ using the Epanechnikov kernel for each evaluation point and choice of bandwidth selector, as a function on the different sample sizes considered. Overall, we can see that robust bias correction yields close to accurate coverage, improving over undersmoothing in almost every case. Performance is highly superior at points where the functions present high curvature and also at the boundary. Performance is never worse even when the function is quite linear. We obtain similar findings when looking at the results for $\nu = 1$ in Figure S.2, S.4, and S.6, where robust bias correction outperforms undersmoothing even more.

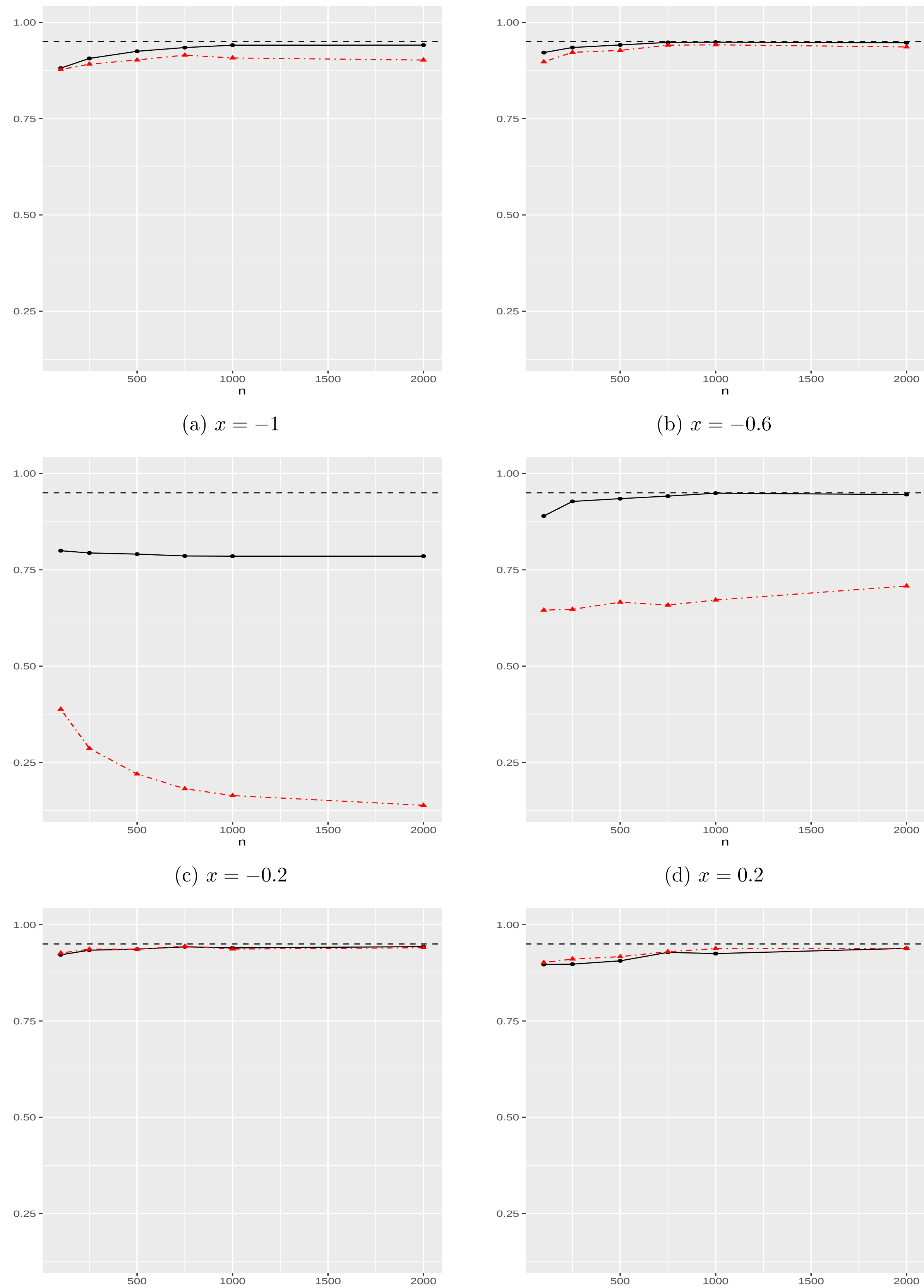
We compare confidence interval performance in terms of length, taking coverage into account by looking at RBC and US confidence intervals implemented with their corresponding coverage error optimal bandwidth choices ($\hat{h}_{\mathtt{rbc}}$ and $\hat{h}_{\mathtt{us}}$, respectively), which is when they perform best in terms of coverage. Figures S.13 and S.14 present the results for $\nu = 0$ and $\nu = 1$, respectively, using the Epanechnikov kernel. We find that, in most cases, RBC confidence intervals are, on average, not larger than US, and sometimes even shorter.

Finally, we report the average (over simulations) of the estimated bandwidths in Figures S.17 and S.18.

All the information used to generate the plots can be found in Tables S.3 and S.4 (for coverage probabilities), and S.5 and S.6 (for average length).

We find similar results for the performance of RBC and US confidence intervals when using the Uniform kernel, as shown in the remaining figures and tables, corresponding exactly to those for the Epanechnikov kernel.

Figure S.1: Empirical Coverage for 95% Confidence Intervals
Epanechnikov Kernel, $\hat{h}_{\texttt{rbc}}$, $\nu = 0$



(a) $x = -1$

(b) $x = -0.6$

(c) $x = -0.2$

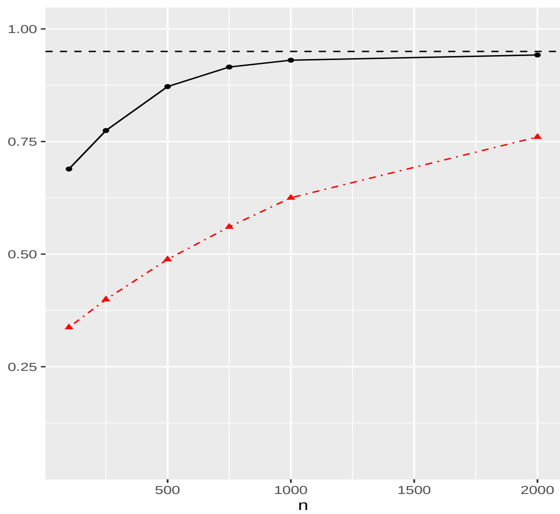
(d) $x = 0.2$

(e) $x = 0.6$

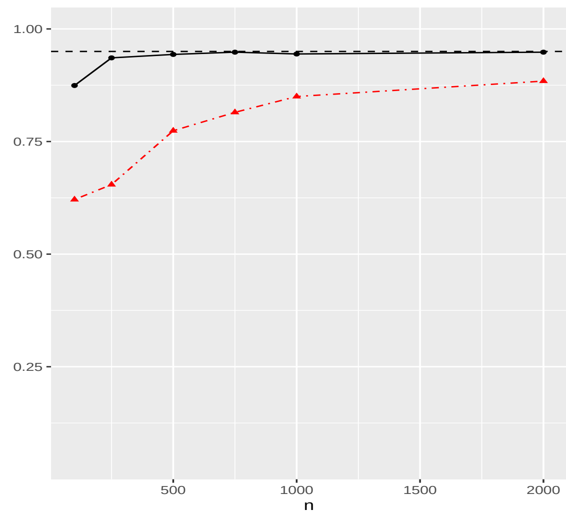
(f) $x = 1$

Notes: Robust Bias Correction, Undersmoothing

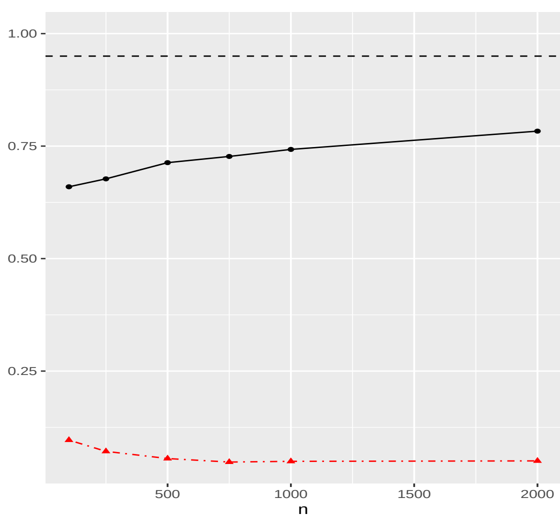
Figure S.2: Empirical Coverage for 95% Confidence Intervals
Epanechnikov Kernel, $\hat{h}_{\texttt{rbc}}$, $\nu = 1$



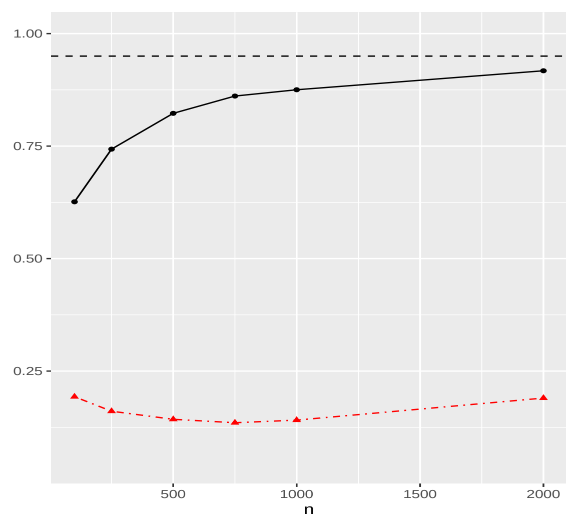
(a) $x = -1$



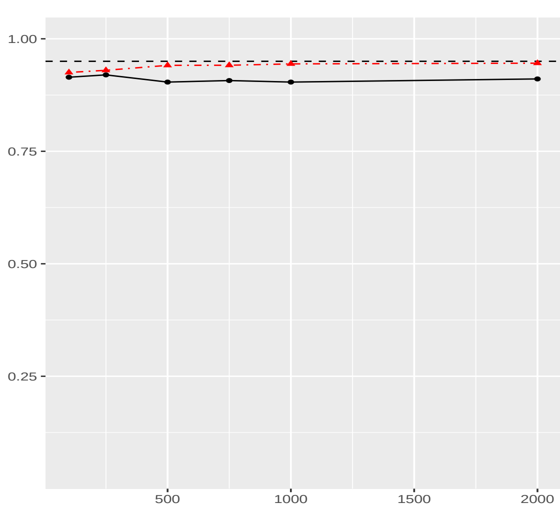
(b) $x = -0.6$



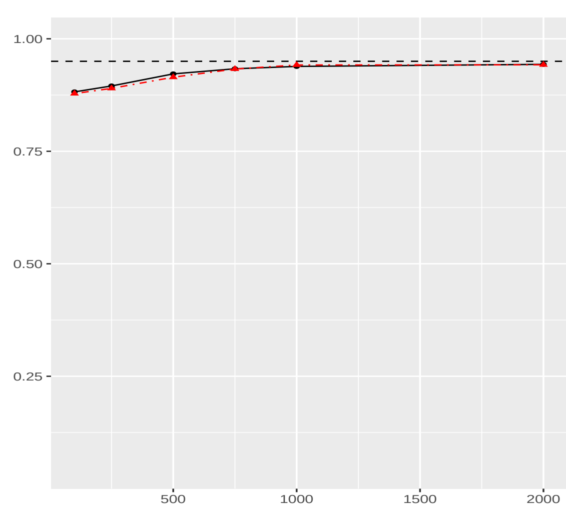
(c) $x = -0.2$



(d) $x = 0.2$



(e) $x = 0.6$

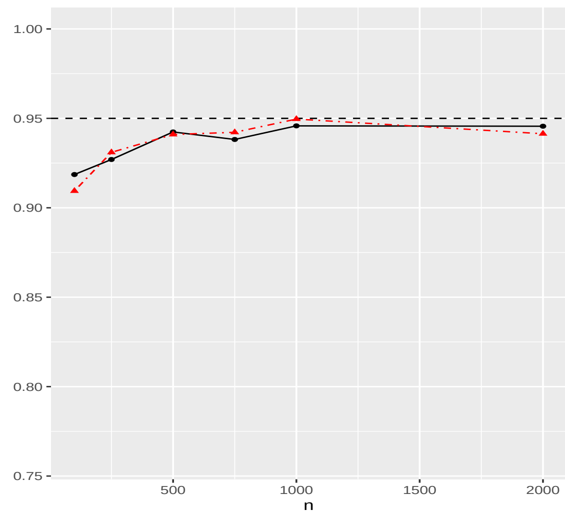


(f) $x = 1$

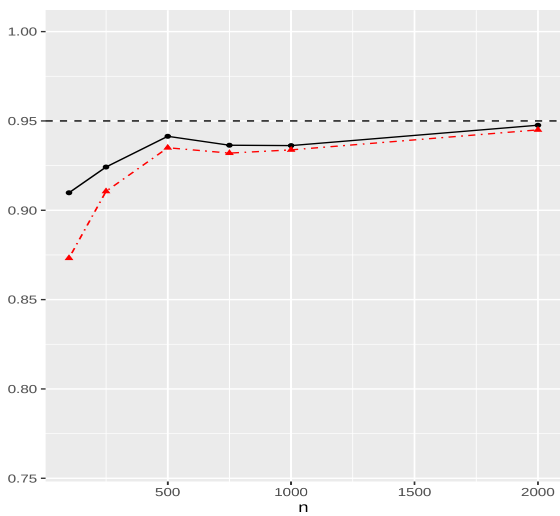
Notes: —Robust Bias Correction, -·- Undersmoothing

Figure S.3: Empirical Coverage for 95% Confidence Intervals
Epanechnikov Kernel, $\hat{h}_{\mathtt{us}}$, $\nu = 0$

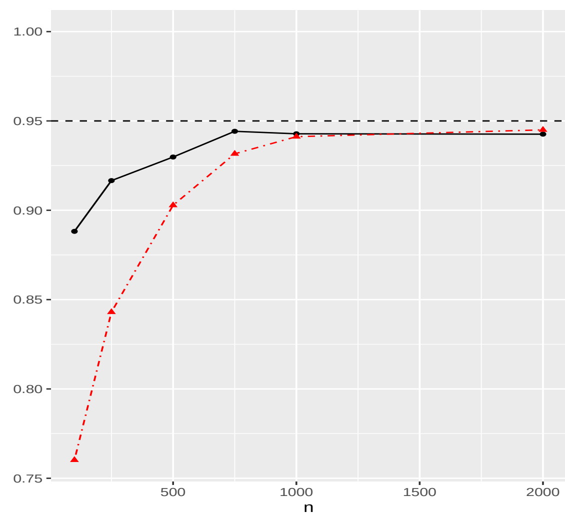
(a) $x = -1$



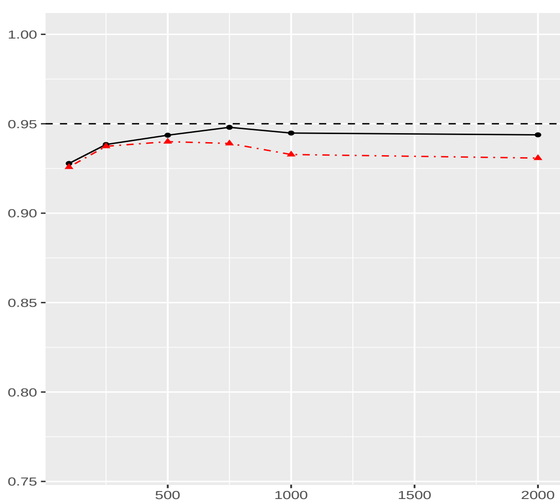
(b) $x = -0.6$



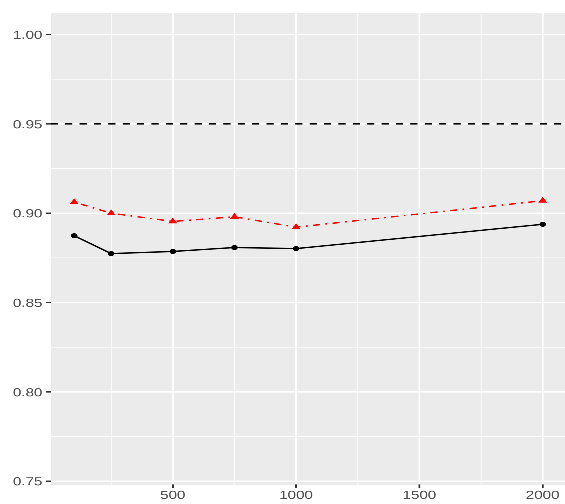
(c) $x = -0.2$



(d) $x = 0.2$



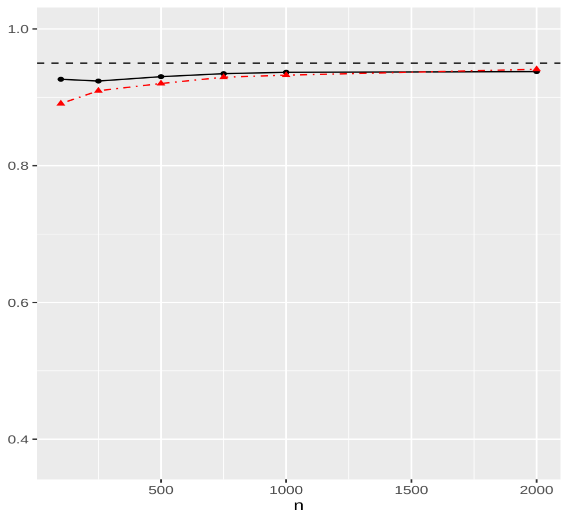
(e) $x = 0.6$



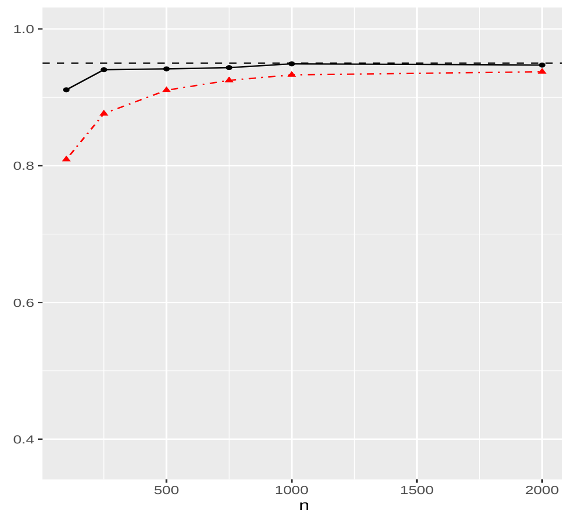
(f) $x = 1$

Notes: —Robust Bias Correction, -·- Undersmoothing

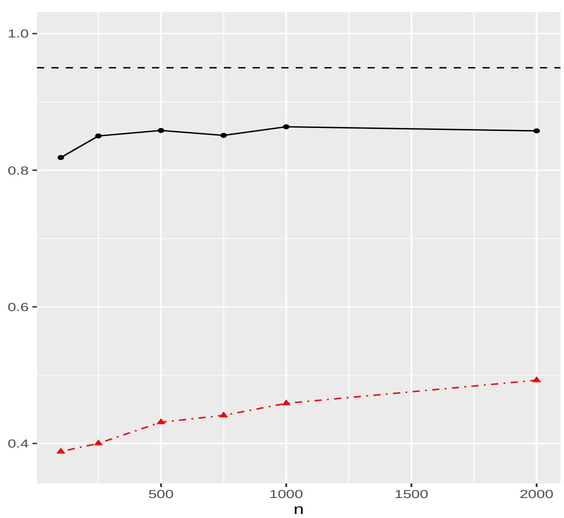
Figure S.4: Empirical Coverage for 95% Confidence Intervals
Epanechnikov Kernel, $\hat{h}_{\mathtt{us}}$, $\nu = 1$



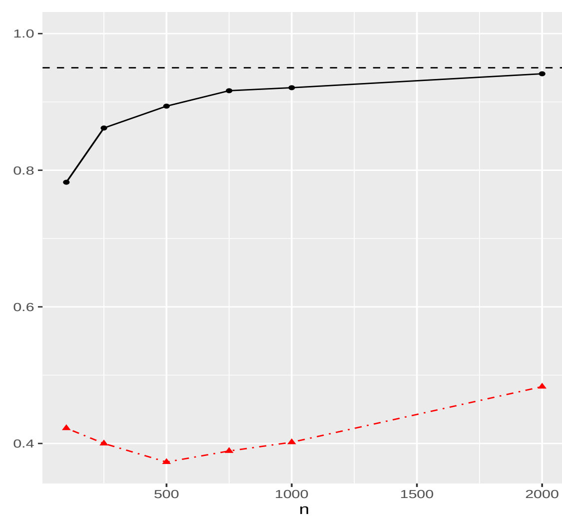
(a) $x = -1$



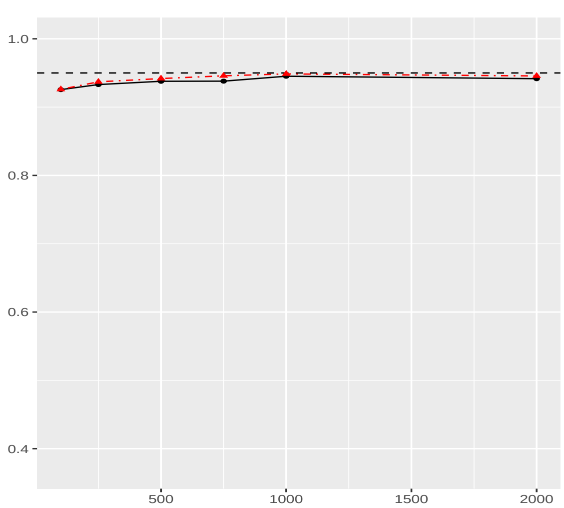
(b) $x = -0.6$



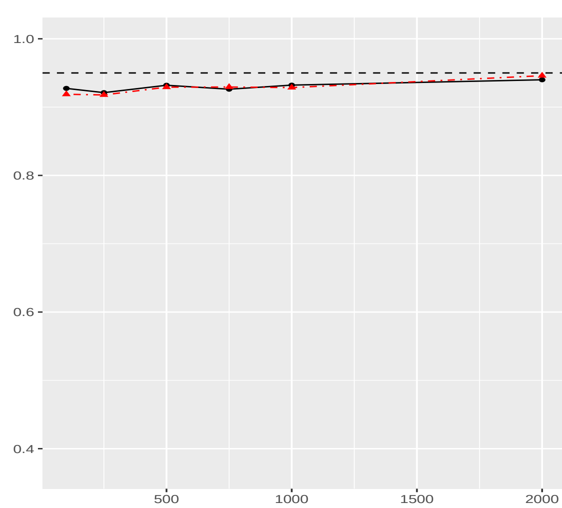
(c) $x = -0.2$



(d) $x = 0.2$



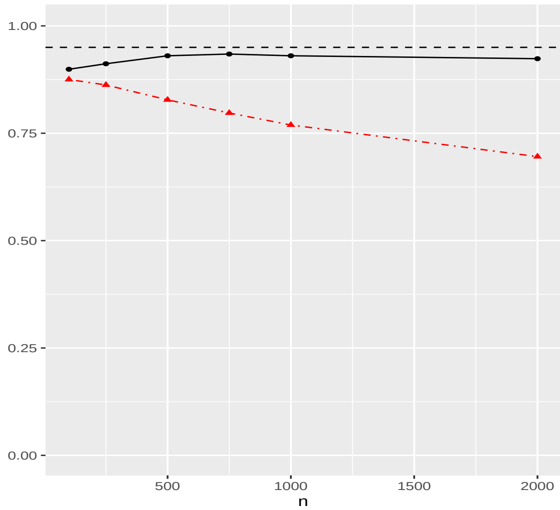
(e) $x = 0.6$



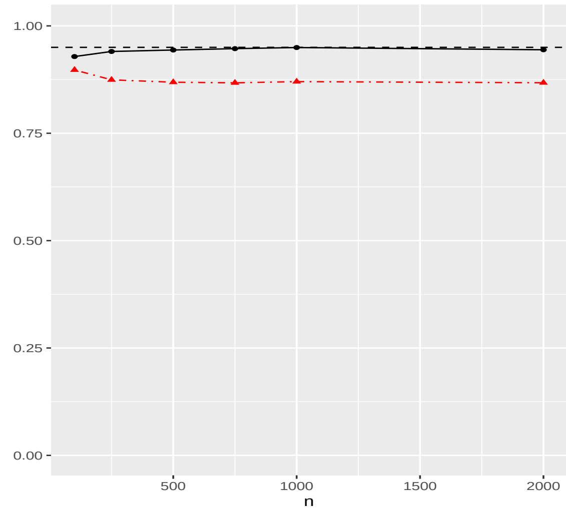
(f) $x = 1$

Notes: —Robust Bias Correction, -·- Undersmoothing

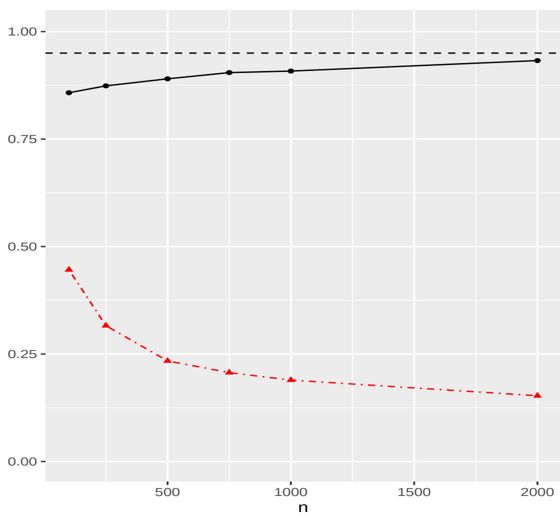
Figure S.5: Empirical Coverage for 95% Confidence Intervals
Epanechnikov Kernel, $\hat{h}_{\texttt{mse}}$, $\nu = 0$



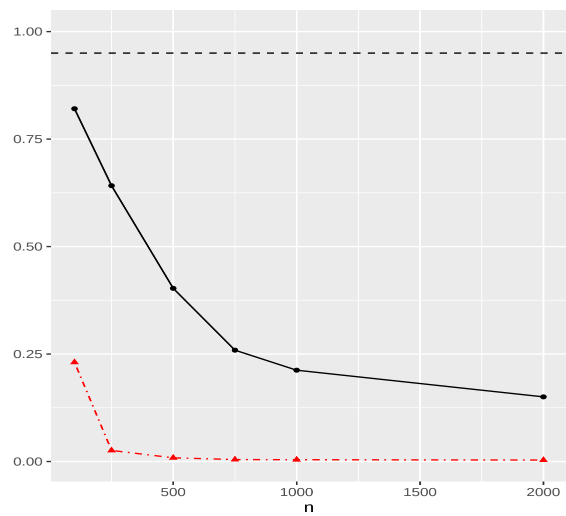
(a) $x = -1$



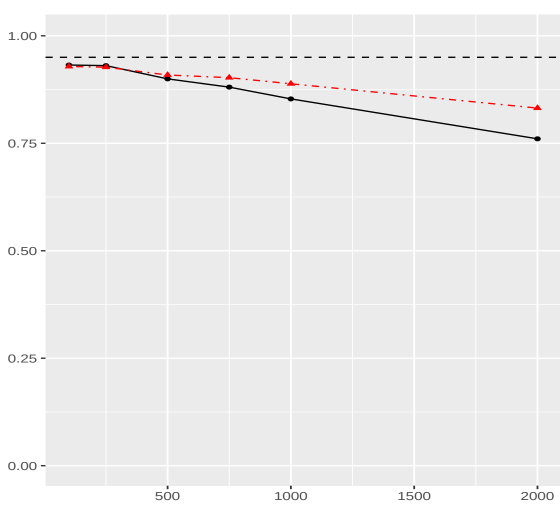
(b) $x = -0.6$



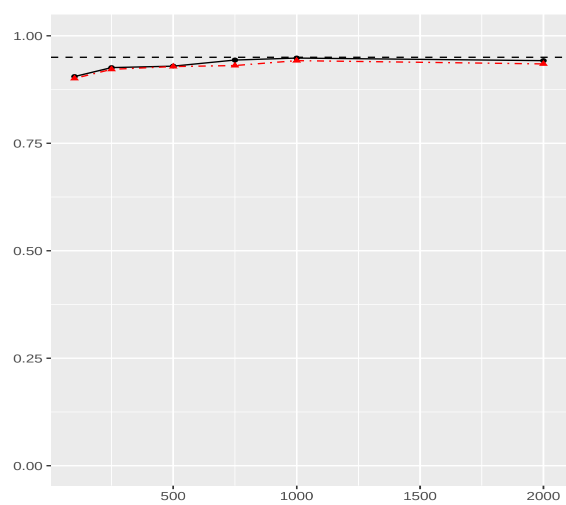
(c) $x = -0.2$



(d) $x = 0.2$



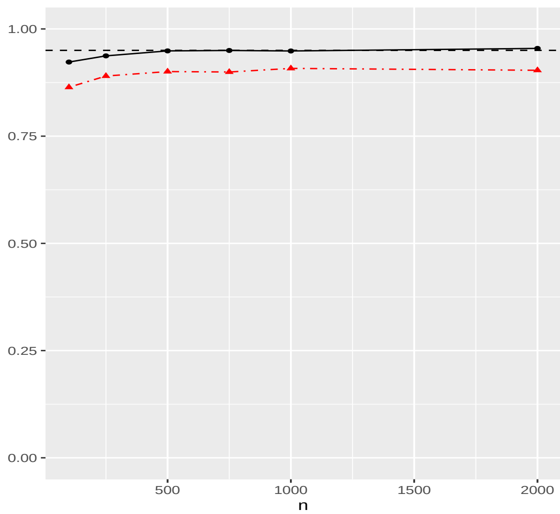
(e) $x = 0.6$



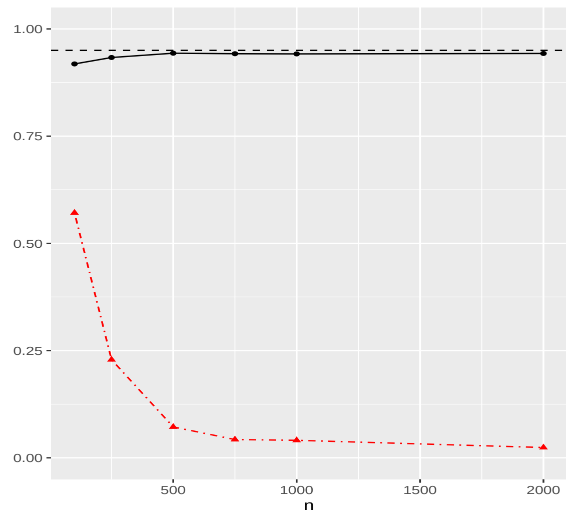
(f) $x = 1$

Notes: —— Robust Bias Correction, -·-·- Undersmoothing

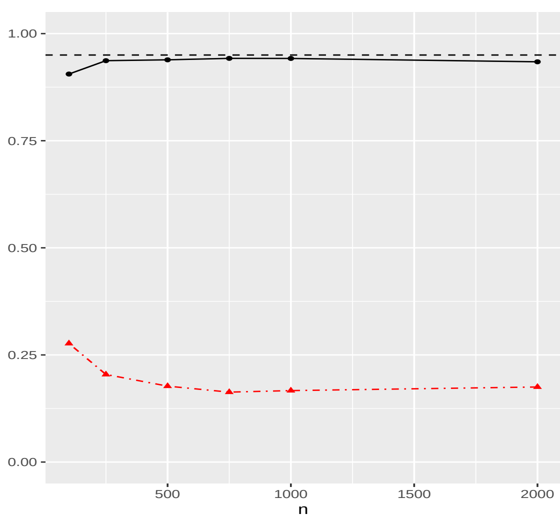
Figure S.6: Empirical Coverage for 95% Confidence Intervals
Epanechnikov Kernel, $\hat{h}_{\mathtt{mse}}$, $\nu = 1$



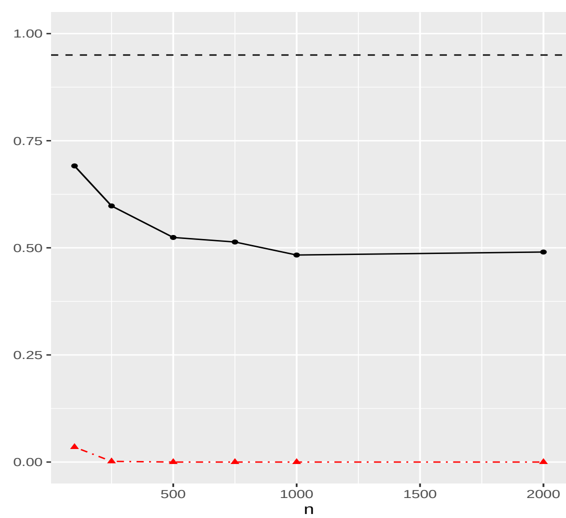
(a) $x = -1$



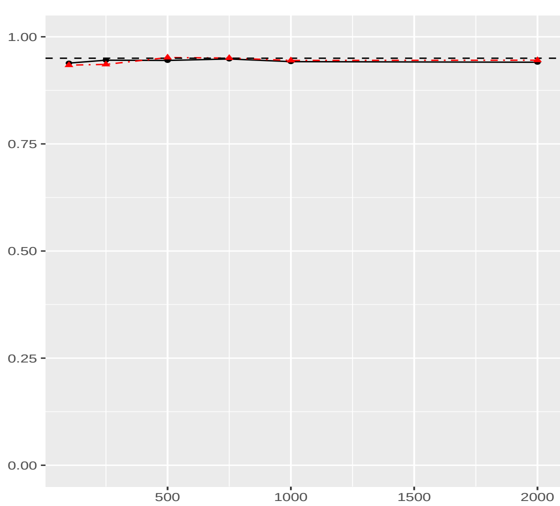
(b) $x = -0.6$



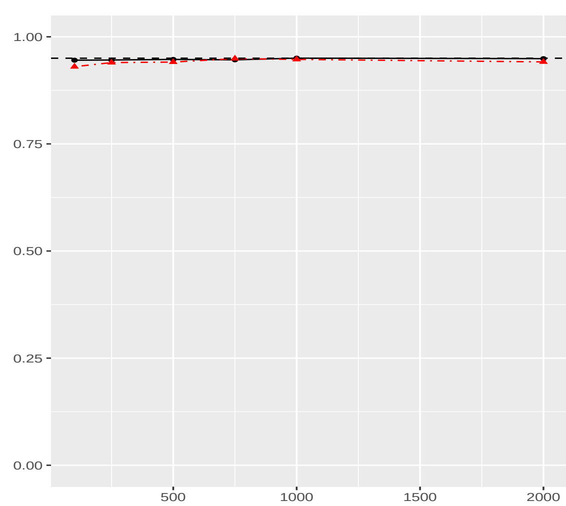
(c) $x = -0.2$



(d) $x = 0.2$



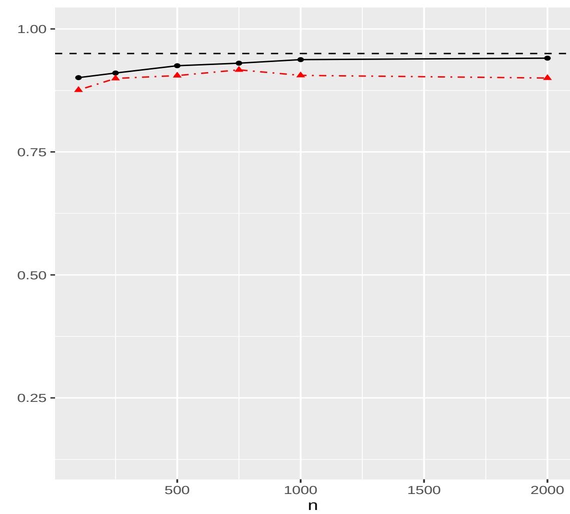
(e) $x = 0.6$



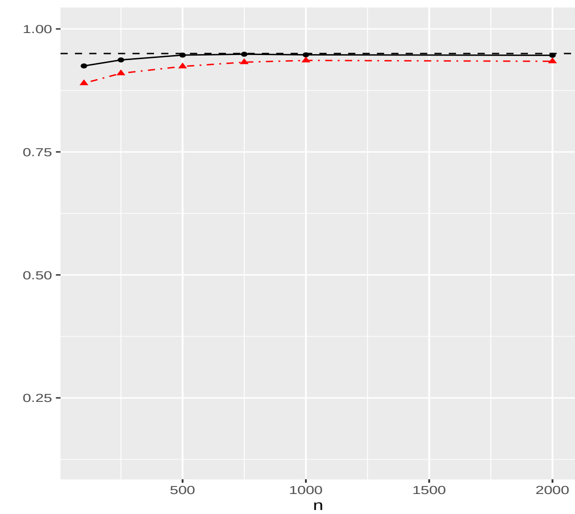
(f) $x = 1$

Notes: —Robust Bias Correction, -·-Undersmoothing

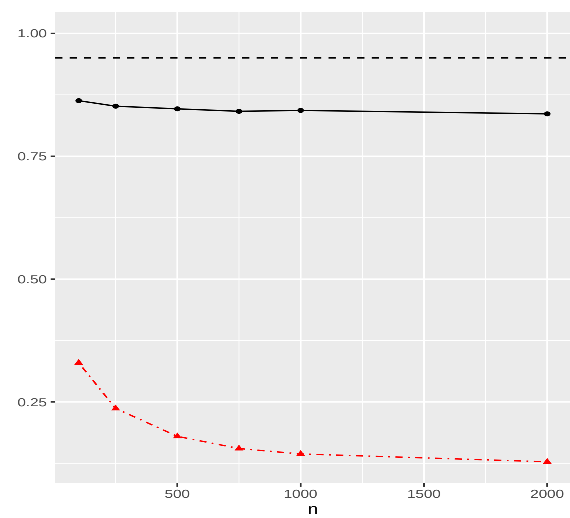
Figure S.7: Empirical Coverage for 95% Confidence Intervals
Uniform Kernel, $\hat{h}_{\texttt{rbc}}$, $\nu = 0$



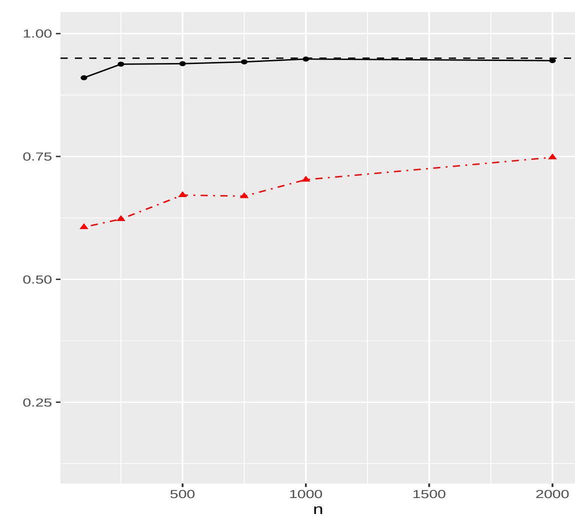
(a) $x = -1$



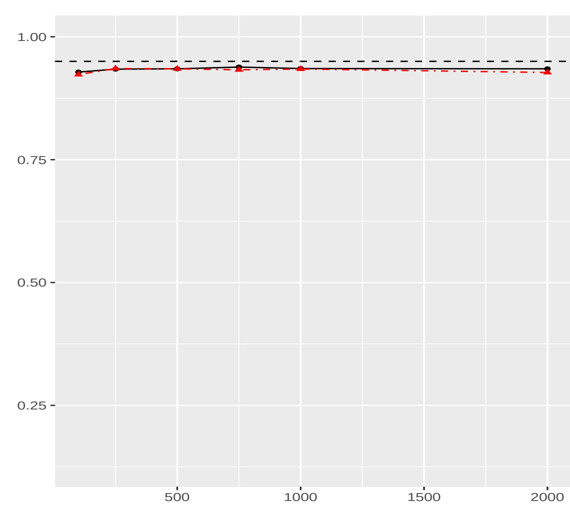
(b) $x = -0.6$



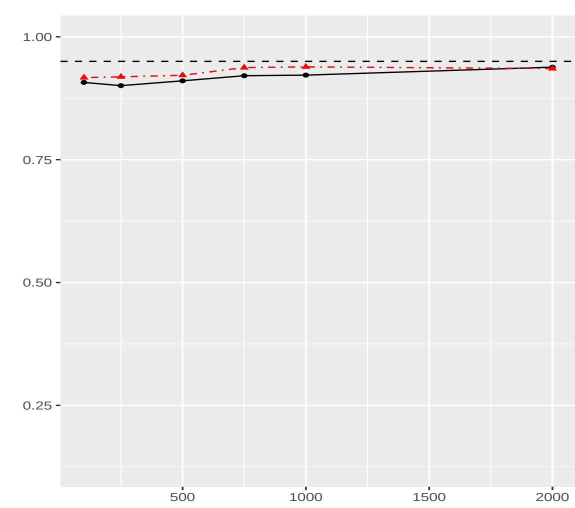
(c) $x = -0.2$



(d) $x = 0.2$



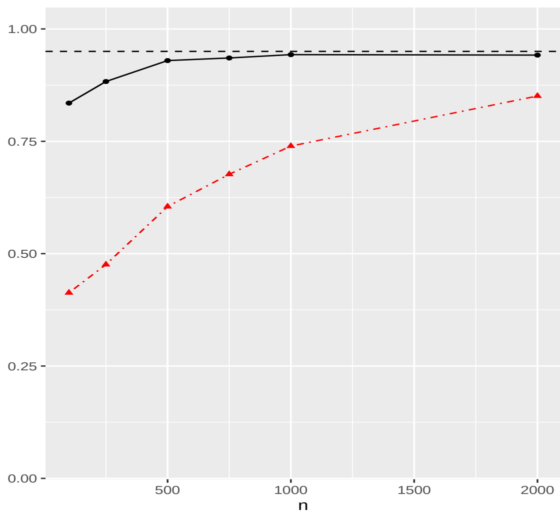
(e) $x = 0.6$



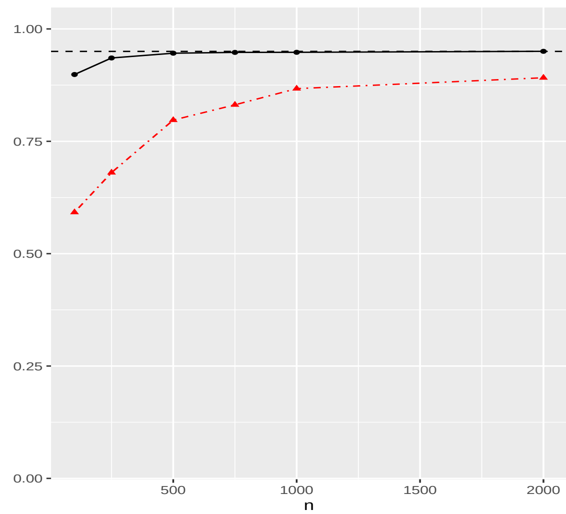
(f) $x = 1$

Notes: —— Robust Bias Correction, -·-· Undersmoothing

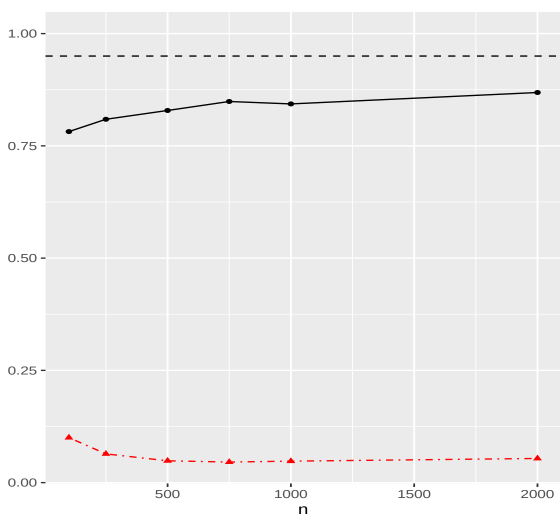
Figure S.8: Empirical Coverage for 95% Confidence Intervals
Uniform Kernel, $\hat{h}_{\texttt{rbc}}$, $\nu = 1$



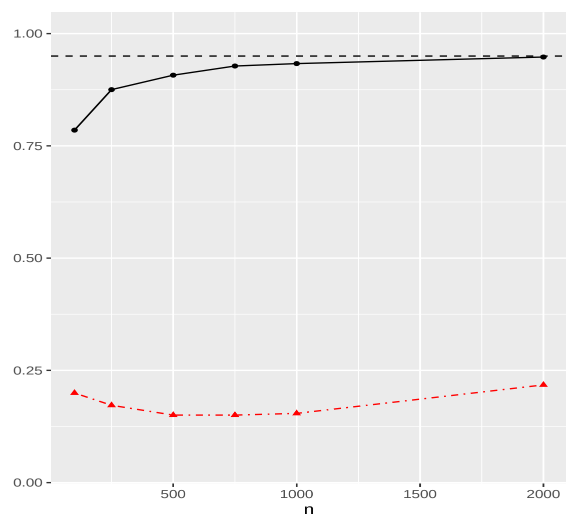
(a) $x = -1$



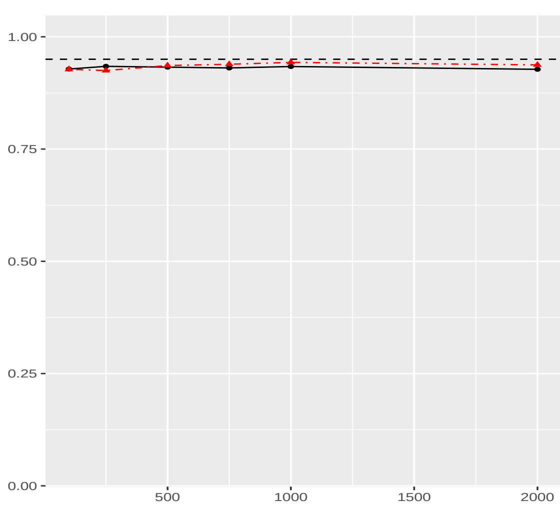
(b) $x = -0.6$



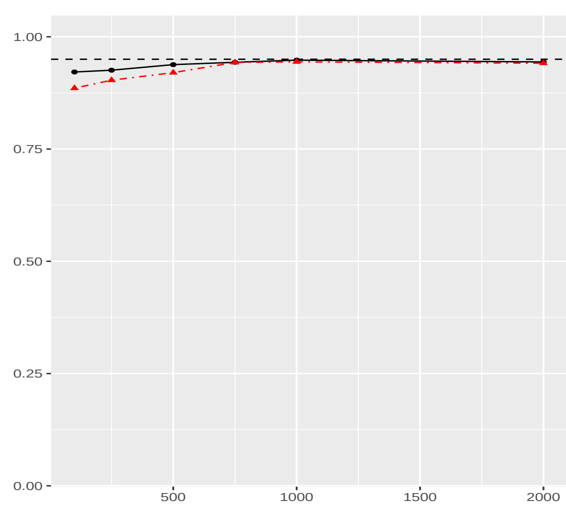
(c) $x = -0.2$



(d) $x = 0.2$



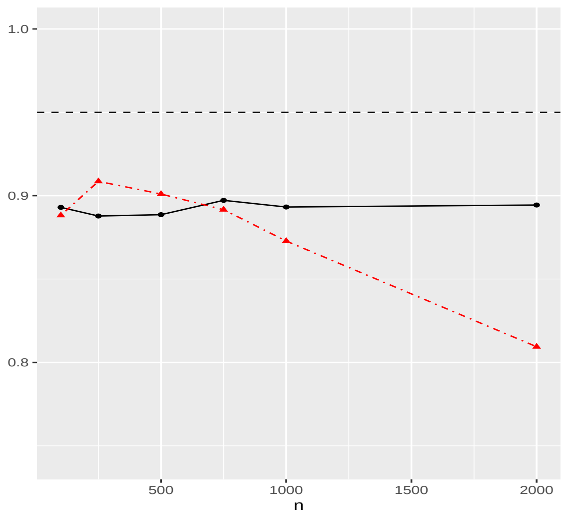
(e) $x = 0.6$



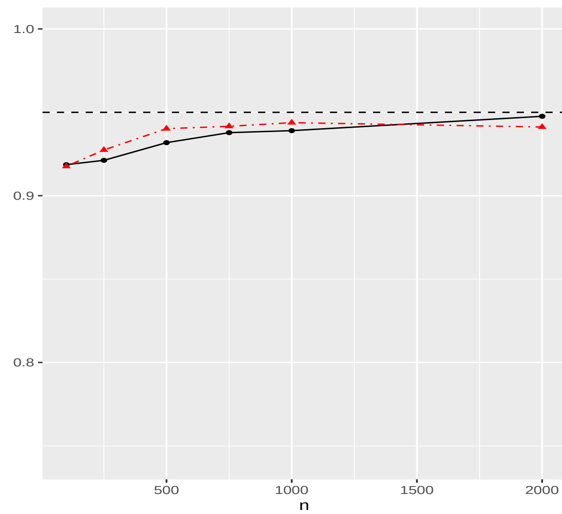
(f) $x = 1$

Notes: —— Robust Bias Correction, -·-· Undersmoothing

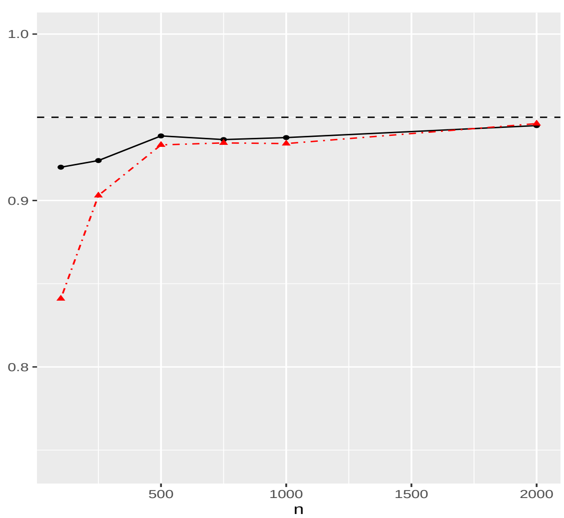
Figure S.9: Empirical Coverage for 95% Confidence Intervals
Uniform Kernel, $\hat{h}_{\mathtt{us}}$, $\nu = 0$



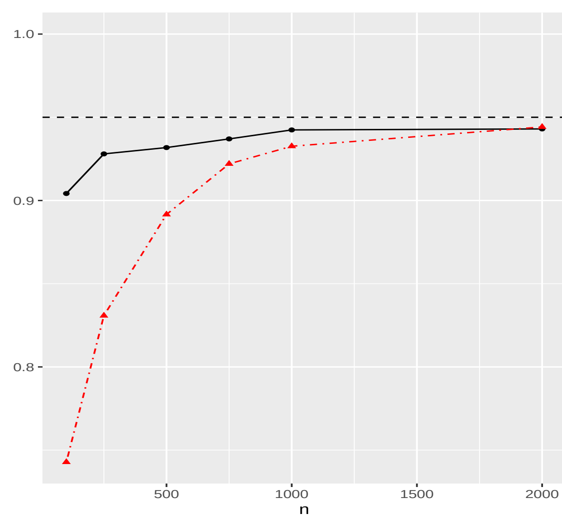
(a) $x = -1$



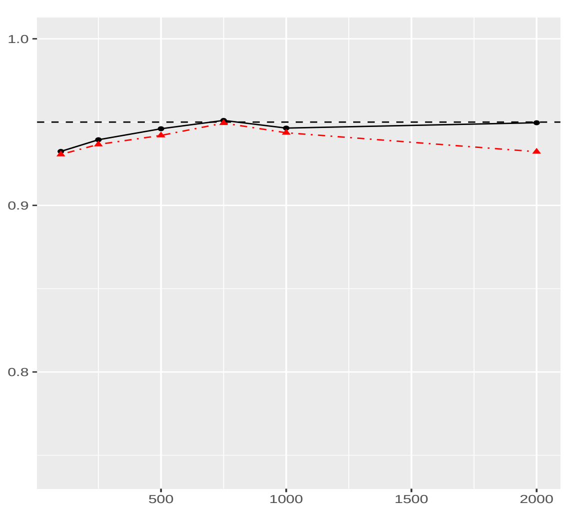
(b) $x = -0.6$



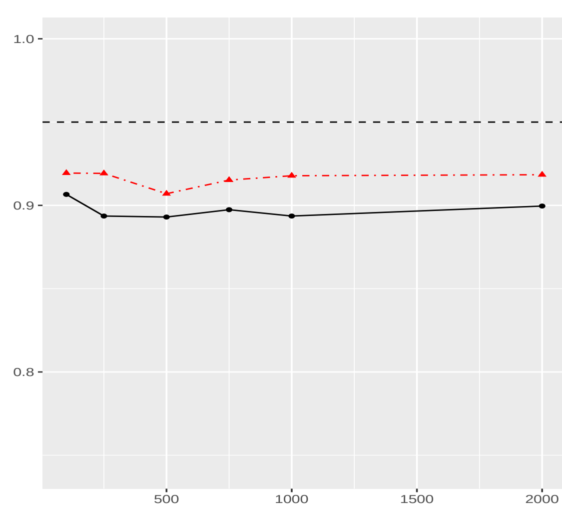
(c) $x = -0.2$



(d) $x = 0.2$



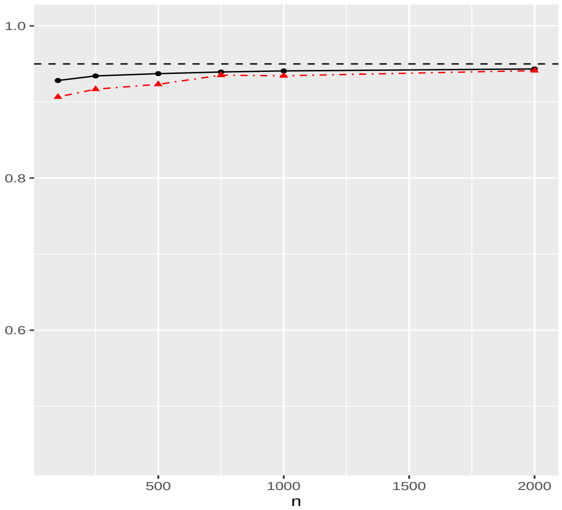
(e) $x = 0.6$



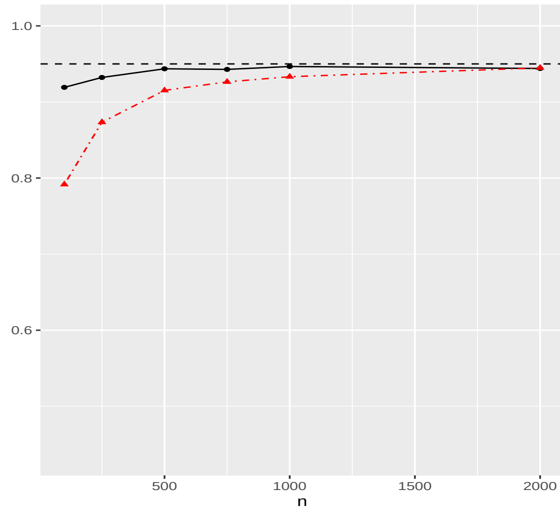
(f) $x = 1$

Notes: —Robust Bias Correction, -·- Undersmoothing

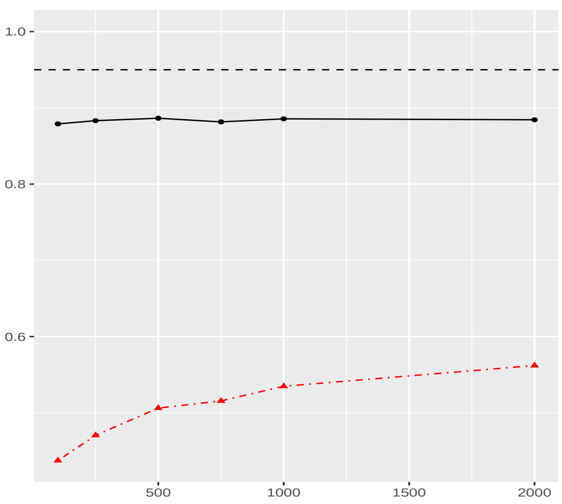
Figure S.10: Empirical Coverage for 95% Confidence Intervals
Uniform Kernel, $\hat{h}_{\mathtt{us}}$, $\nu = 1$



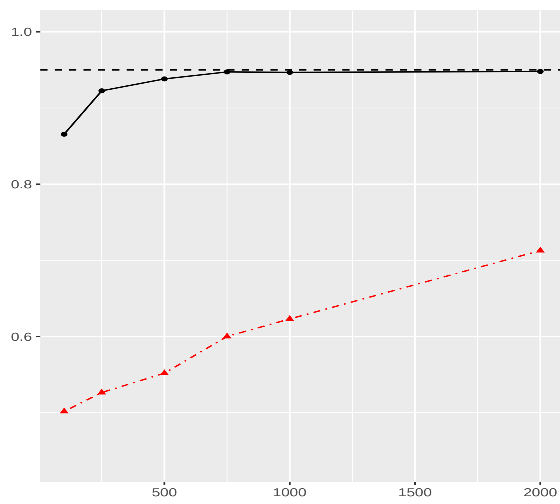
(a) $x = -1$



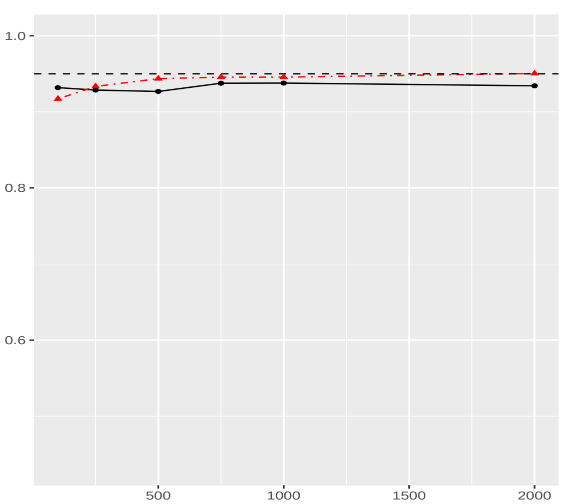
(b) $x = -0.6$



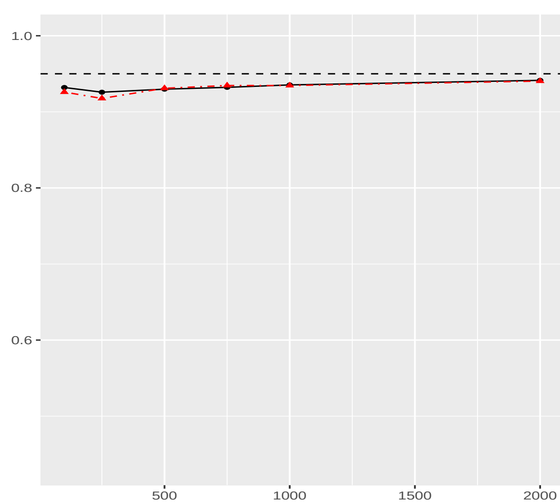
(c) $x = -0.2$



(d) $x = 0.2$



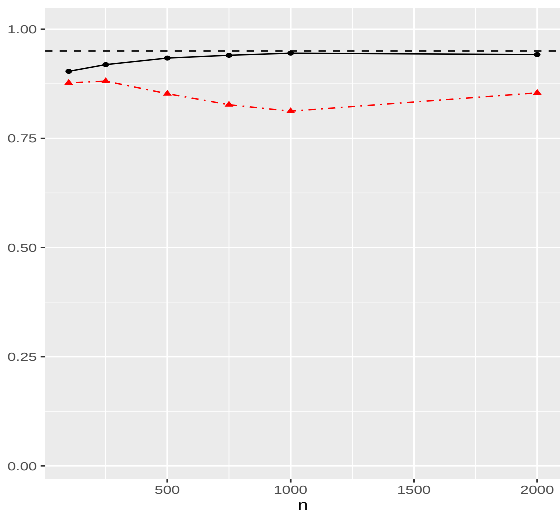
(e) $x = 0.6$



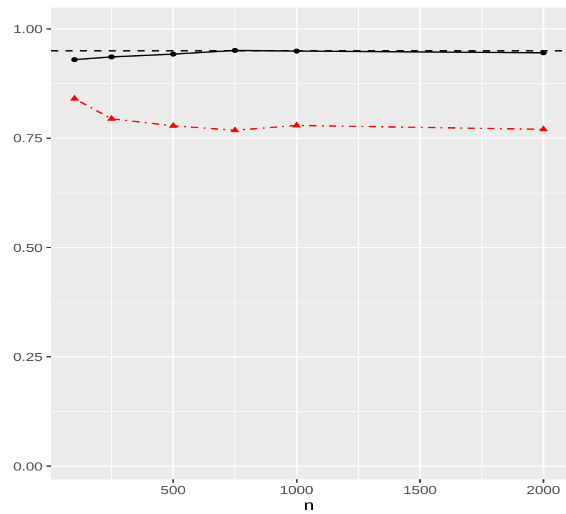
(f) $x = 1$

Notes: —— Robust Bias Correction, -·-· Undersmoothing

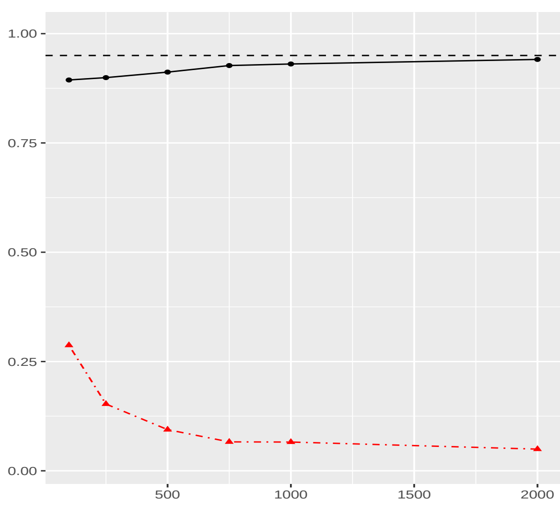
Figure S.11: Empirical Coverage for 95% Confidence Intervals
Uniform Kernel, $\hat{h}_{\texttt{mse}}$, $\nu = 0$



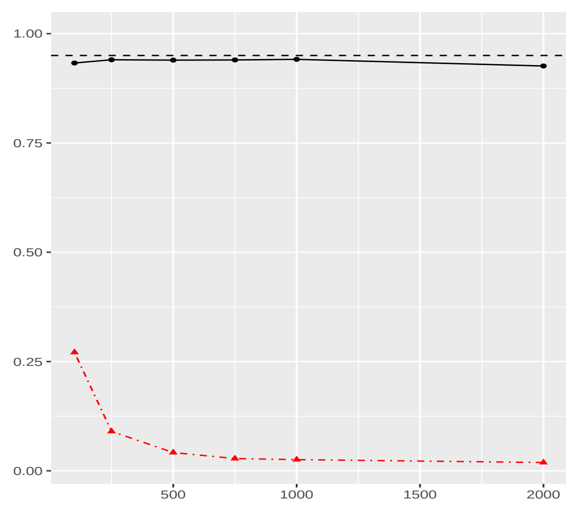
(a) $x = -1$



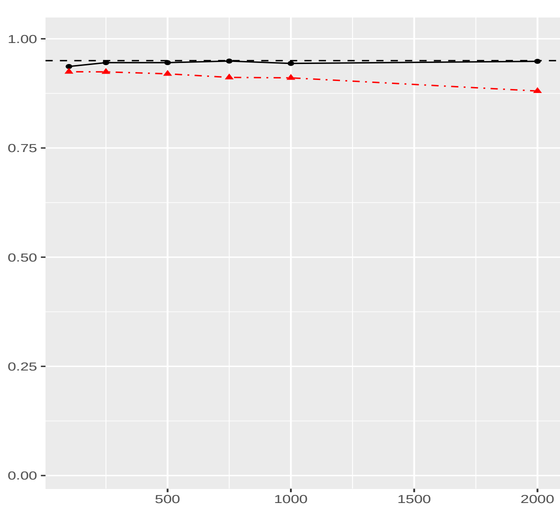
(b) $x = -0.6$



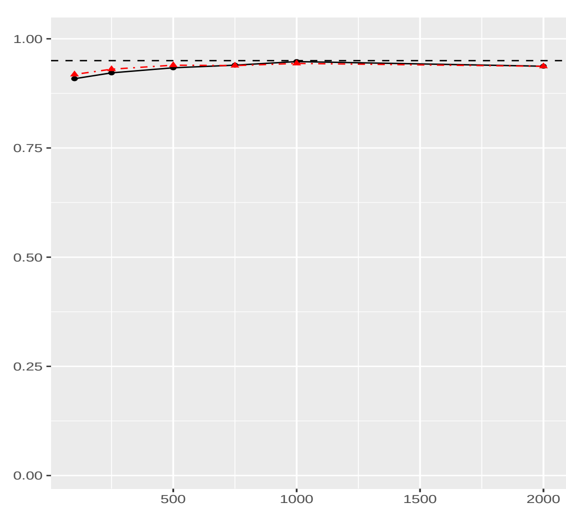
(c) $x = -0.2$



(d) $x = 0.2$



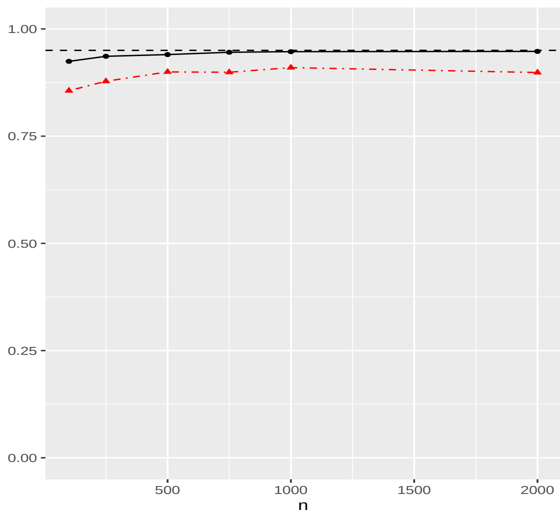
(e) $x = 0.6$



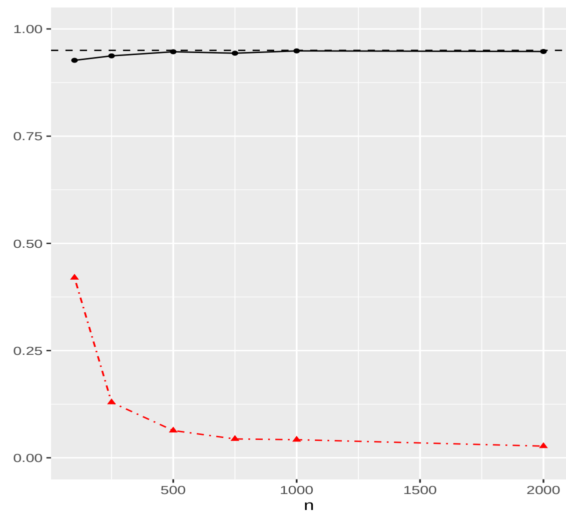
(f) $x = 1$

Notes: —Robust Bias Correction, -·- Undersmoothing

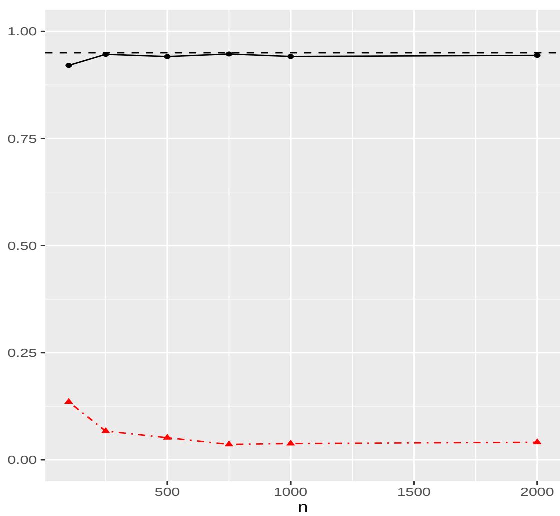
Figure S.12: Empirical Coverage for 95% Confidence Intervals
Uniform Kernel, $\hat{h}_{\texttt{mse}}$, $\nu = 1$



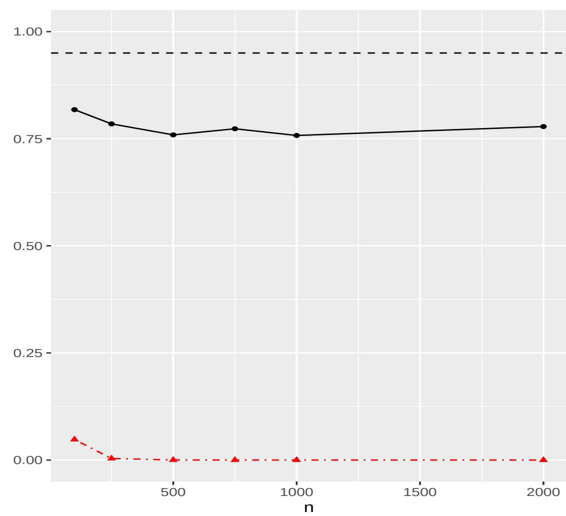
(a) $x = -1$



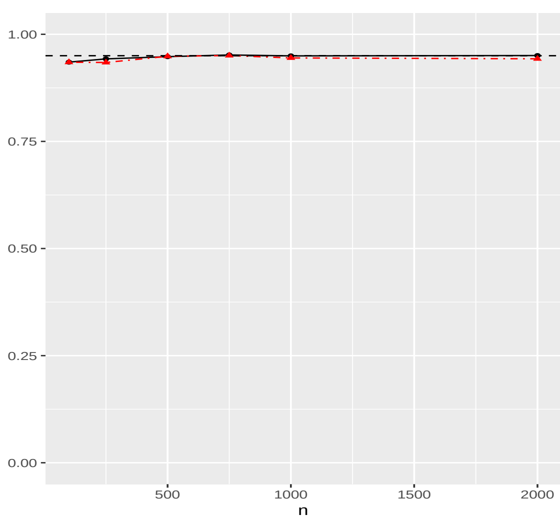
(b) $x = -0.6$



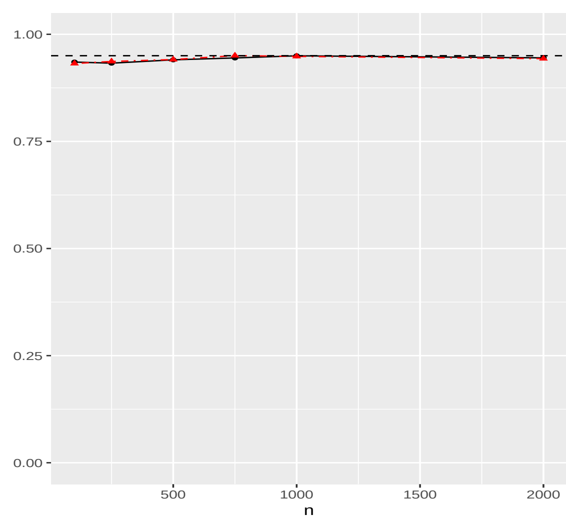
(c) $x = -0.2$



(d) $x = 0.2$



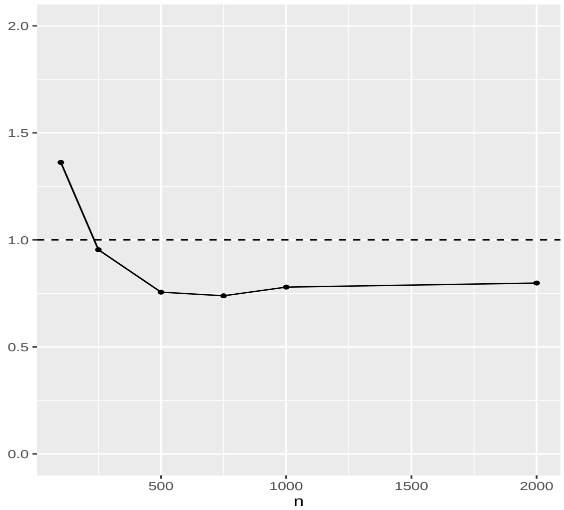
(e) $x = 0.6$



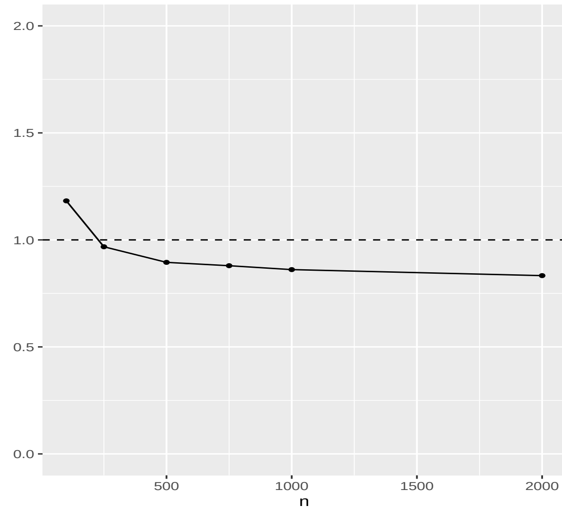
(f) $x = 1$

Notes: —— Robust Bias Correction, -·-· Undersmoothing

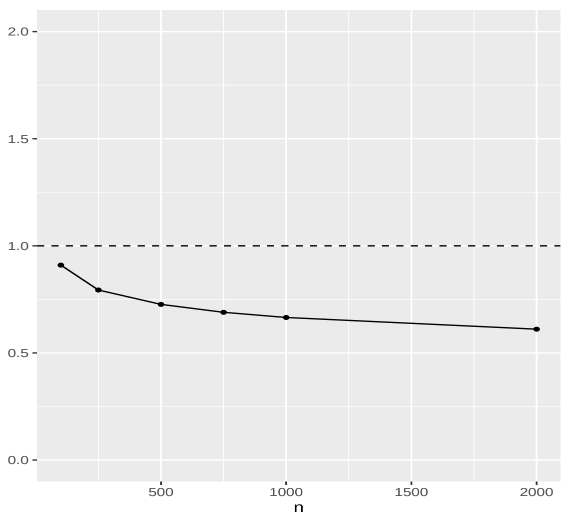
Figure S.13: Average Interval Length for 95% Confidence Intervals
Epanechnikov Kernel, $\nu = 0$



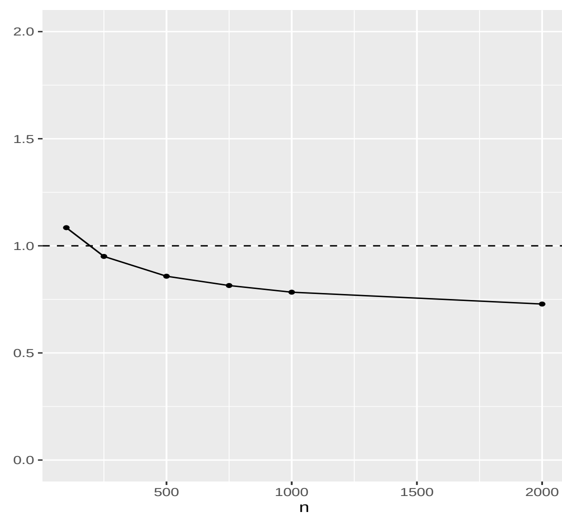
(a) $x = -1$



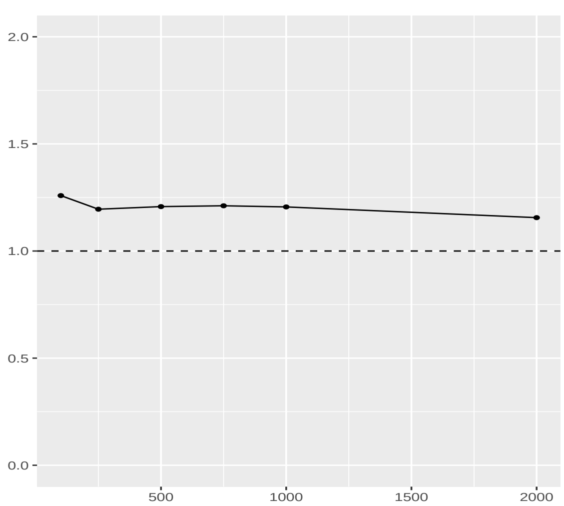
(b) $x = -0.6$



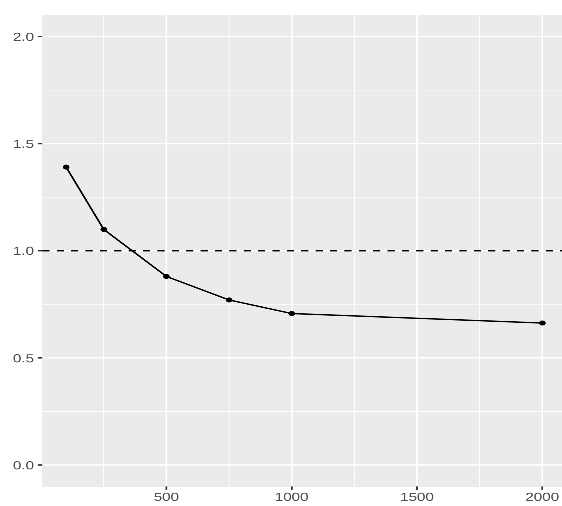
(c) $x = -0.2$



(d) $x = 0.2$



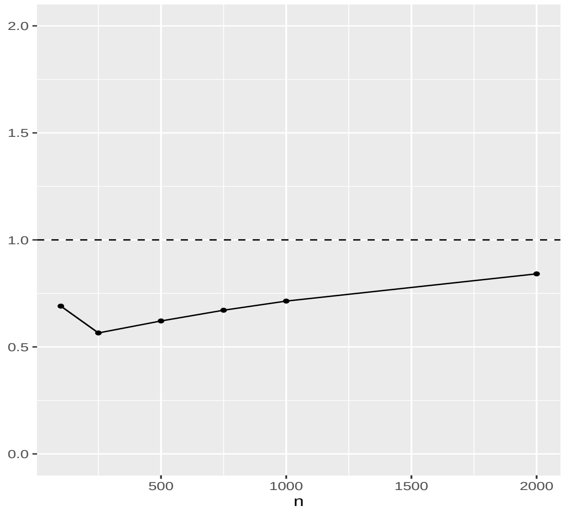
(e) $x = 0.6$



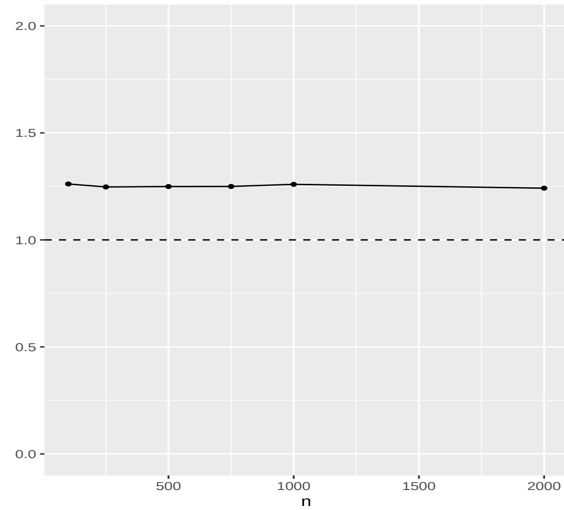
(f) $x = 1$

Notes: ratio of Average Interval Length for $I_{\texttt{rbc}}(\hat{h}_{\texttt{rbc}})$ over $I_{\texttt{us}}(\hat{h}_{\texttt{us}})$

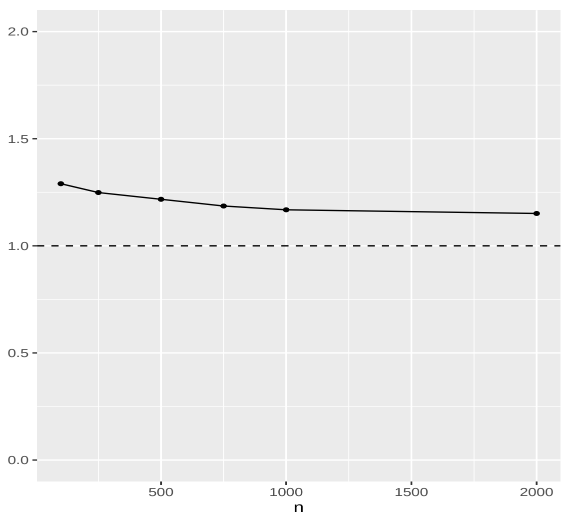
Figure S.14: Average Interval Length for 95% Confidence Intervals
Epanechnikov Kernel, $\nu = 1$



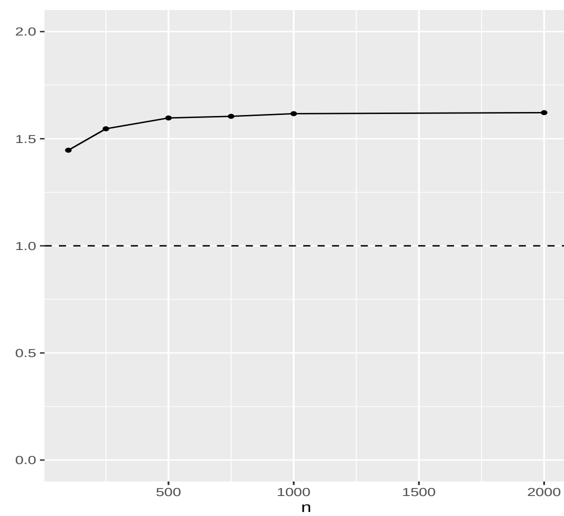
(a) $x = -1$



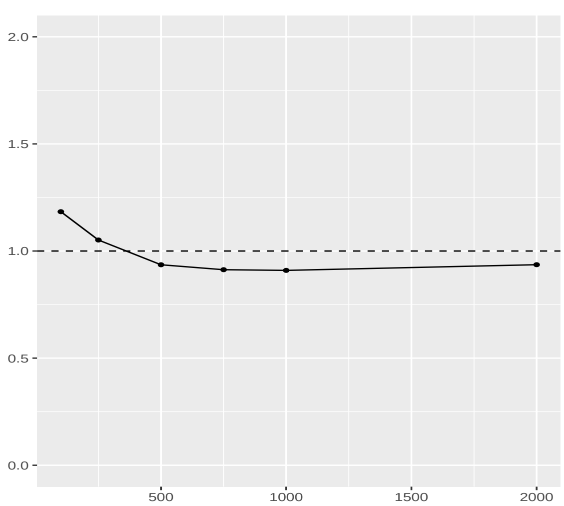
(b) $x = -0.6$



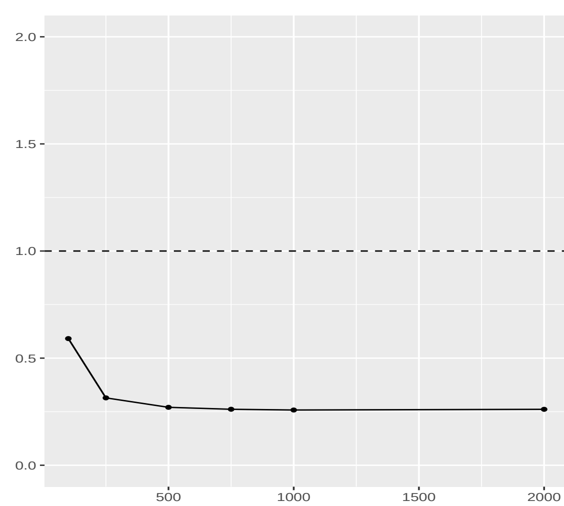
(c) $x = -0.2$



(d) $x = 0.2$



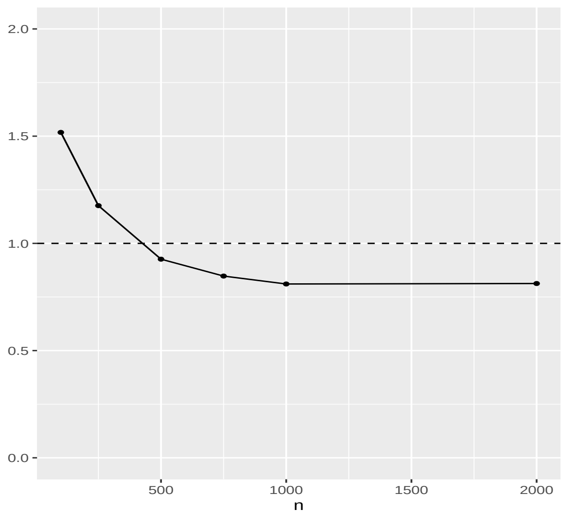
(e) $x = 0.6$



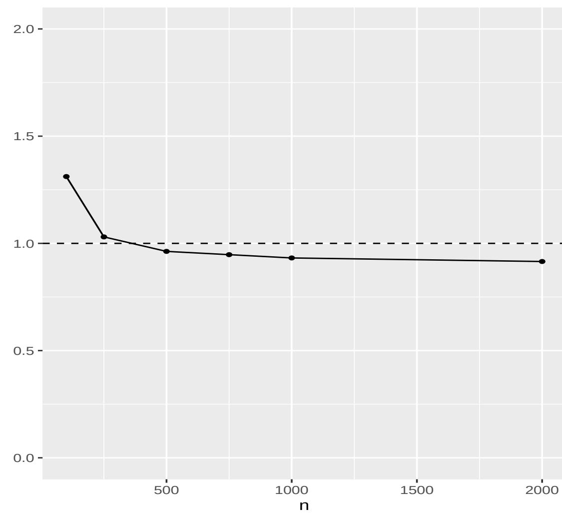
(f) $x = 1$

Notes: ratio of Average Interval Length for $I_{\mathtt{rbc}}(\hat{h}_{\mathtt{rbc}})$ over $I_{\mathtt{us}}(\hat{h}_{\mathtt{us}})$

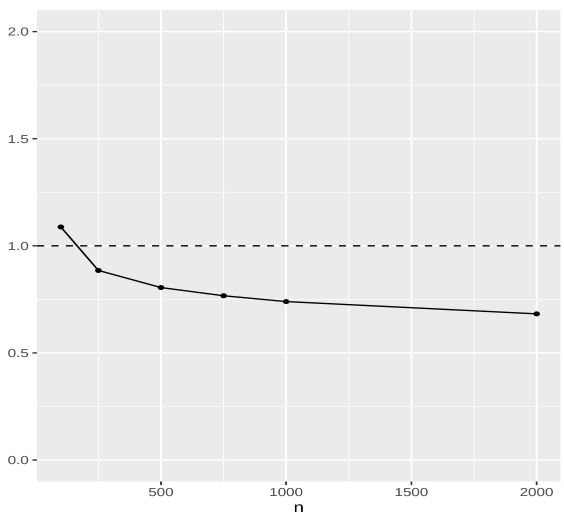
Figure S.15: Average Interval Length for 95% Confidence Intervals
Uniform Kernel, $\nu = 0$



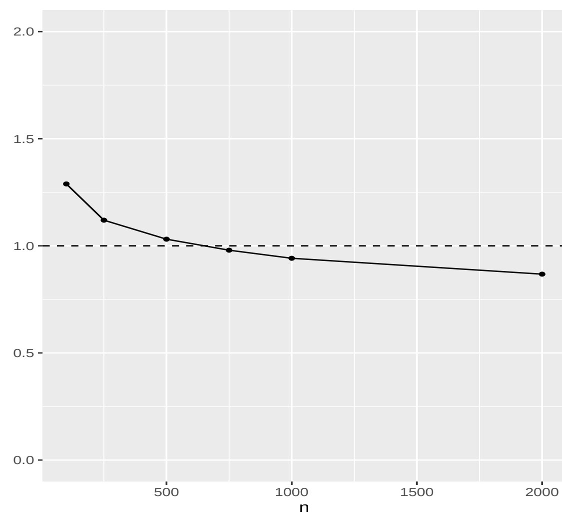
(a) $x = -1$



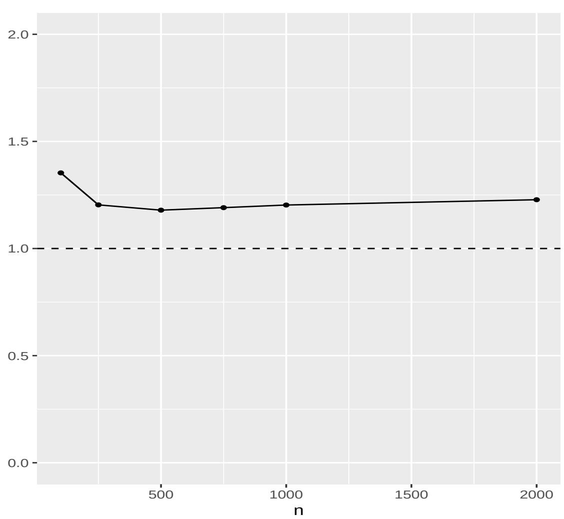
(b) $x = -0.6$



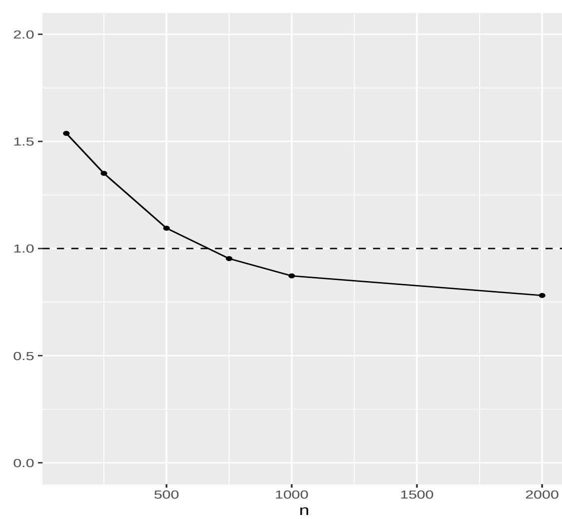
(c) $x = -0.2$



(d) $x = 0.2$



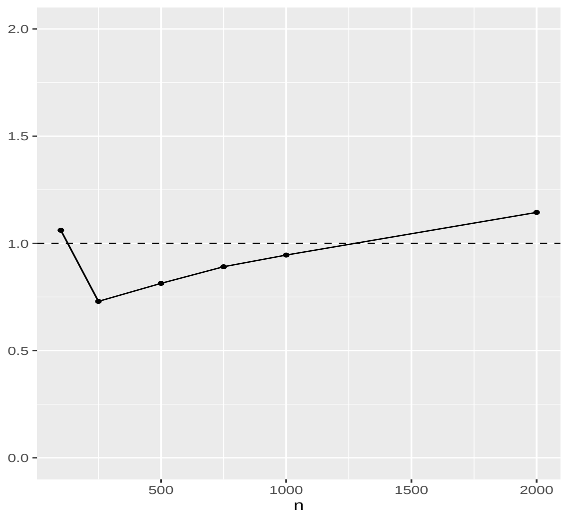
(e) $x = 0.6$



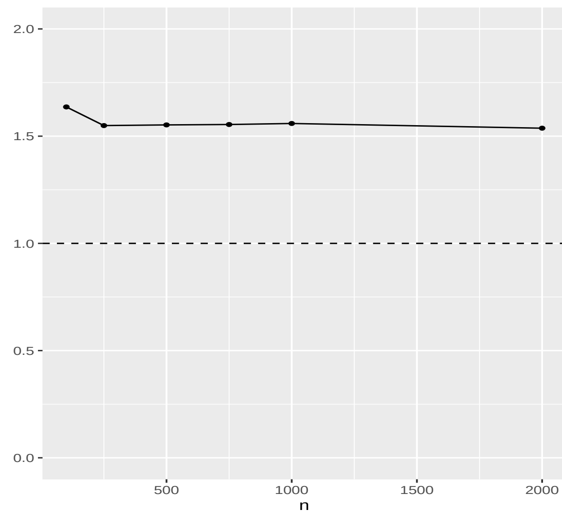
(f) $x = 1$

Notes: ratio of Average Interval Length for $I_{\mathtt{rbc}}(\hat{h}_{\mathtt{rbc}})$ over $I_{\mathtt{us}}(\hat{h}_{\mathtt{us}})$

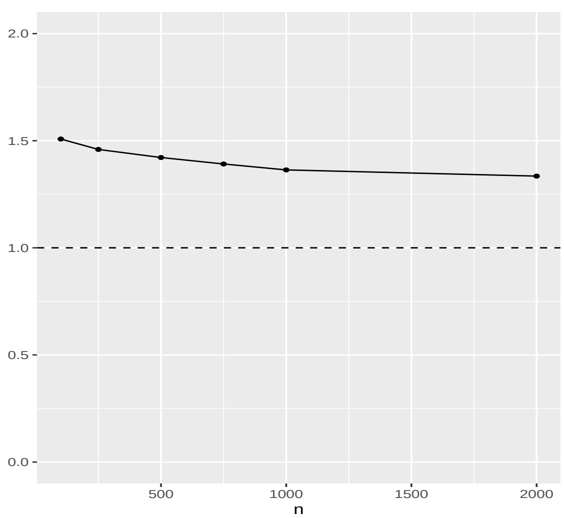
Figure S.16: Average Interval Length for 95% Confidence Intervals
Uniform Kernel, $\nu = 1$



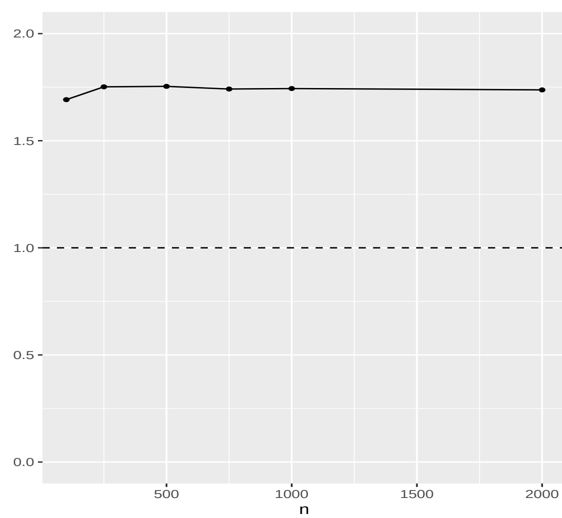
(a) $x = -1$



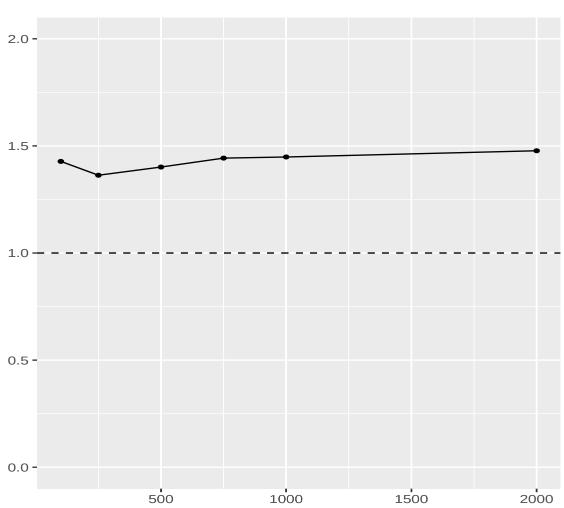
(b) $x = -0.6$



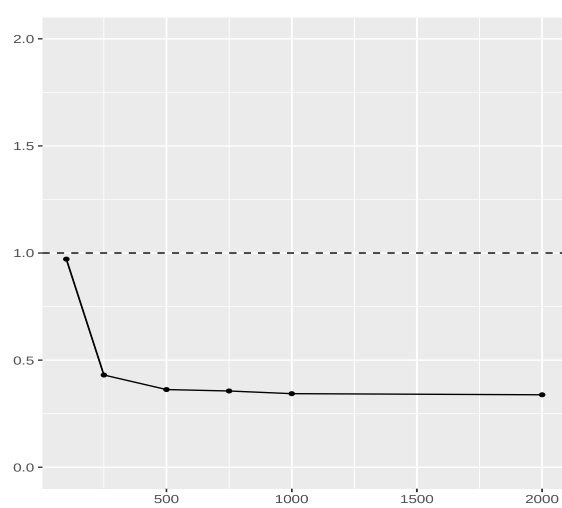
(c) $x = -0.2$



(d) $x = 0.2$



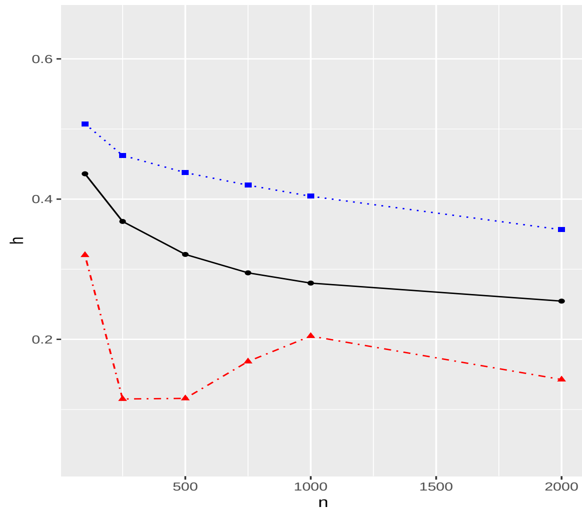
(e) $x = 0.6$



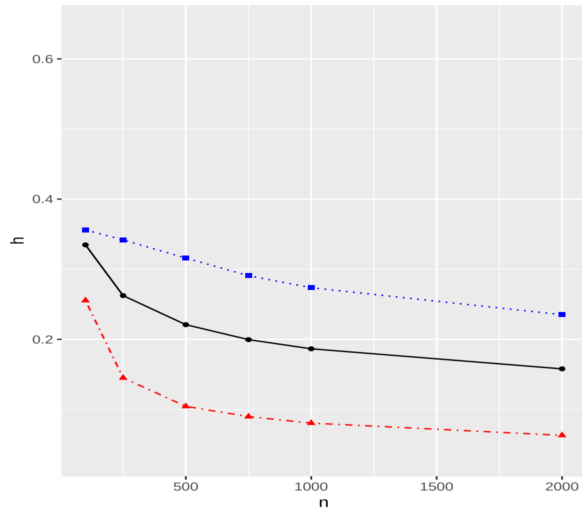
(f) $x = 1$

Notes: ratio of Average Interval Length for $I_{\mathtt{rbc}}(\hat{h}_{\mathtt{rbc}})$ over $I_{\mathtt{us}}(\hat{h}_{\mathtt{us}})$

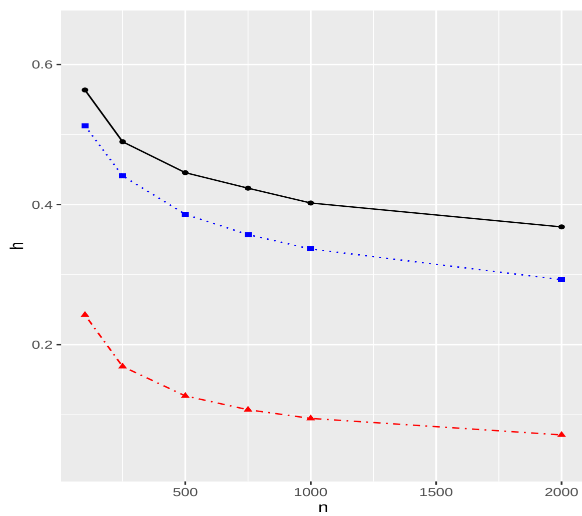
Figure S.17: Average Estimated Bandwidths, Epanechnikov Kernel, $\nu = 0$



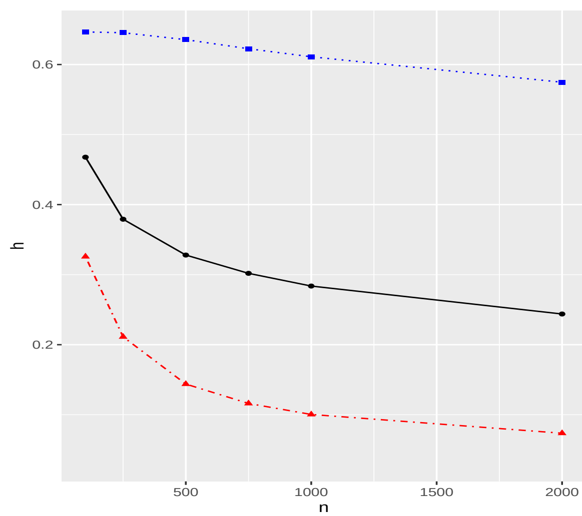
(a) $x = -1$



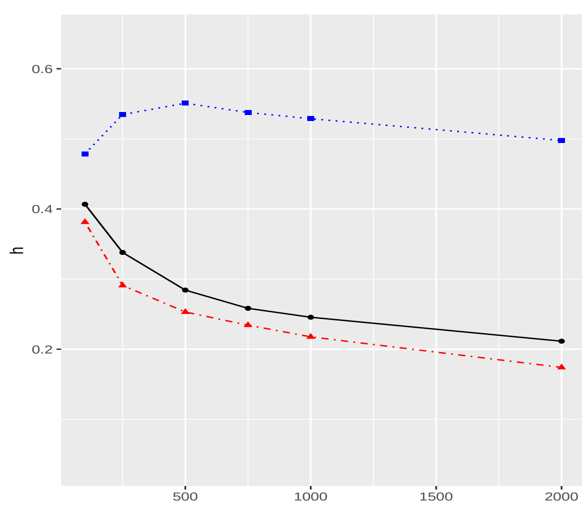
(b) $x = -0.6$



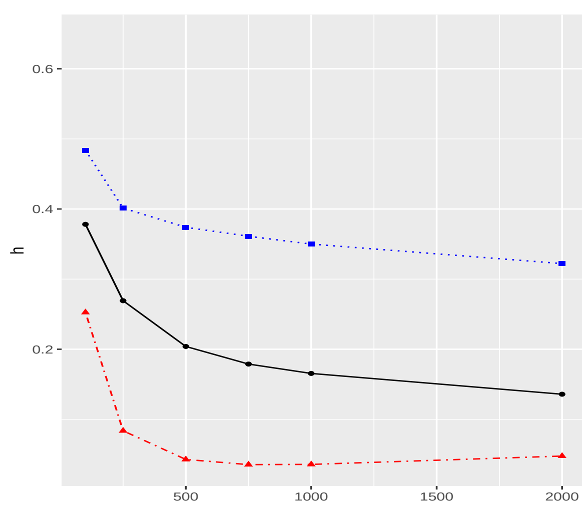
(c) $x = -0.2$



(d) $x = 0.2$



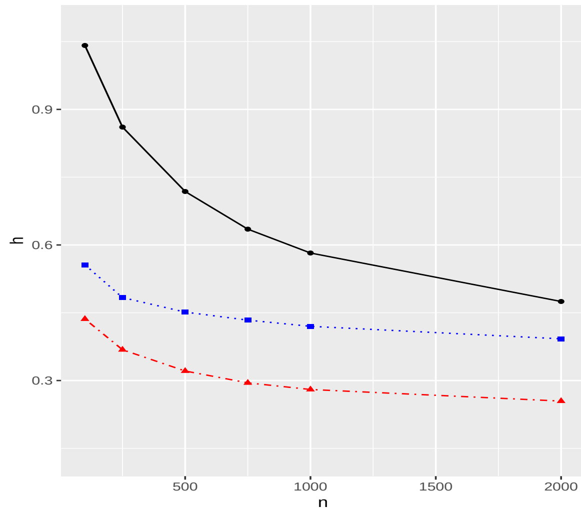
(e) $x = 0.6$



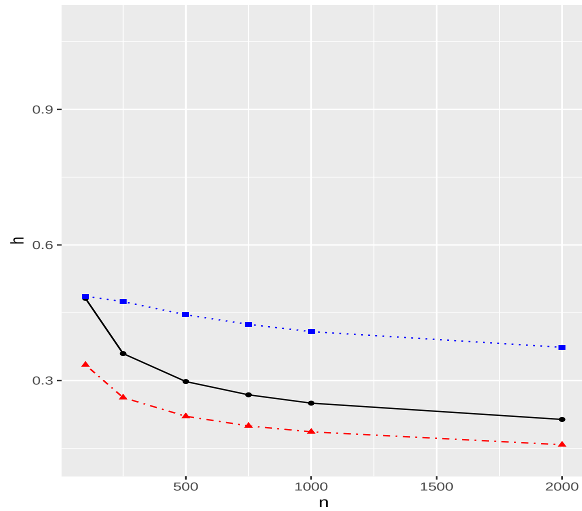
(f) $x = 1$

Notes: $\hat{h}_{\mathtt{rbc}}$, $\hat{h}_{\mathtt{us}}$, $\hat{h}_{\mathtt{mse}}$

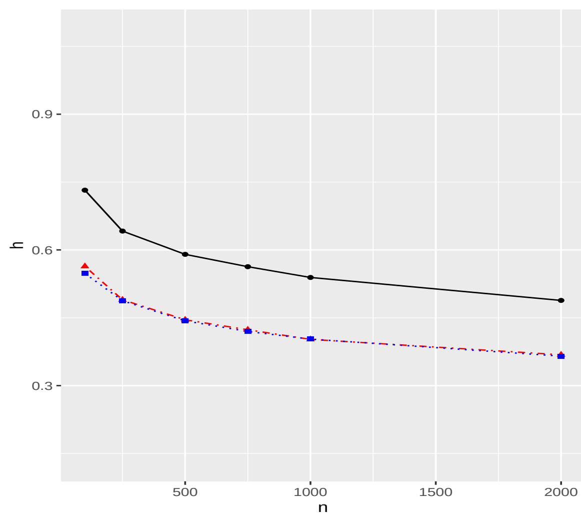
Figure S.18: Average Estimated Bandwidths, Epanechnikov Kernel, $\nu = 1$



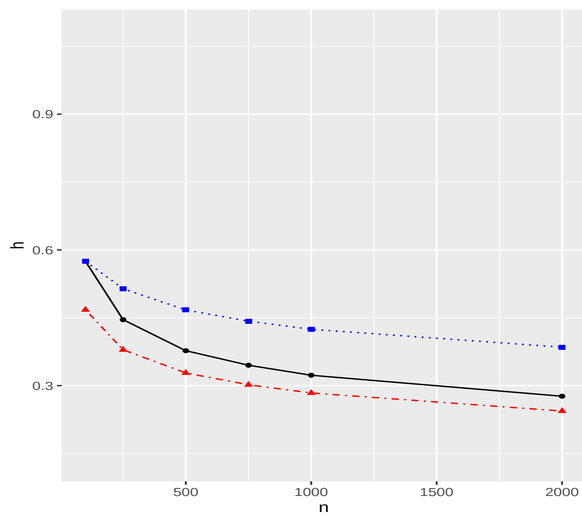
(a) $x = -1$



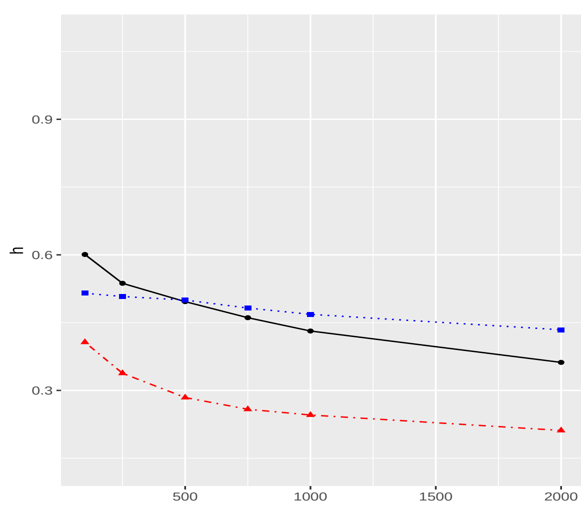
(b) $x = -0.6$



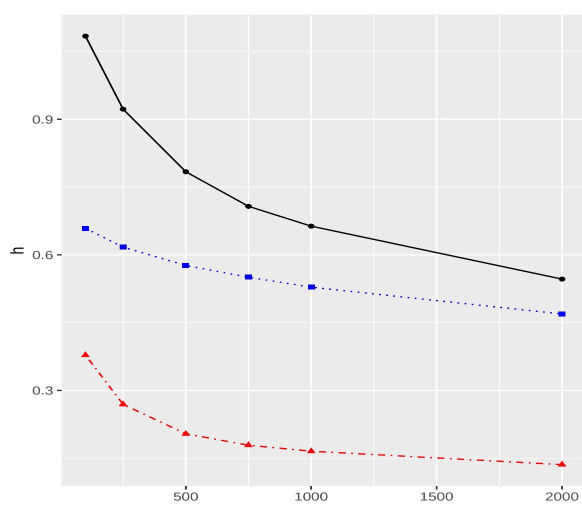
(c) $x = -0.2$



(d) $x = 0.2$



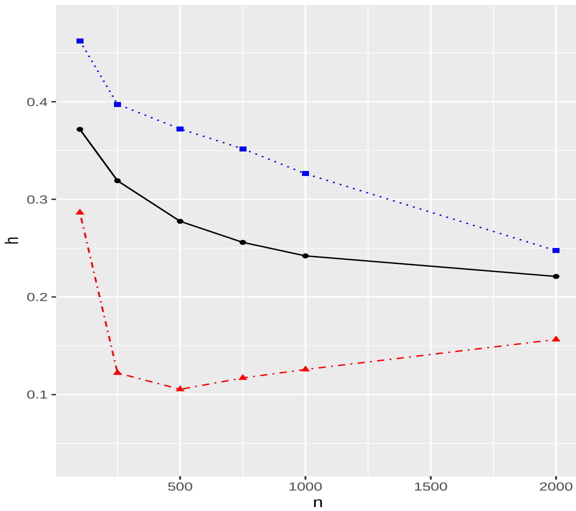
(e) $x = 0.6$



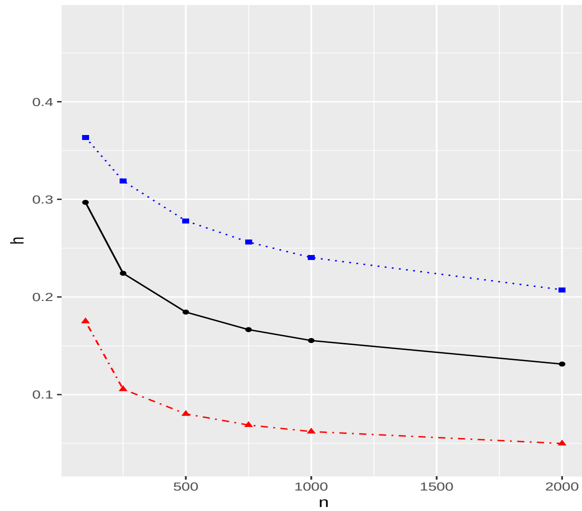
(f) $x = 1$

Notes: $\hat{h}_{\texttt{rbc}}$, $\hat{h}_{\texttt{us}}$, $\hat{h}_{\texttt{mse}}$

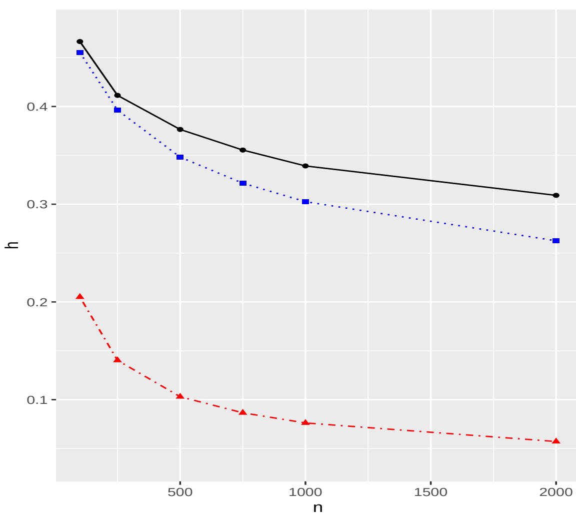
Figure S.19: Average Estimated Bandwidths, Uniform Kernel, $\nu = 0$



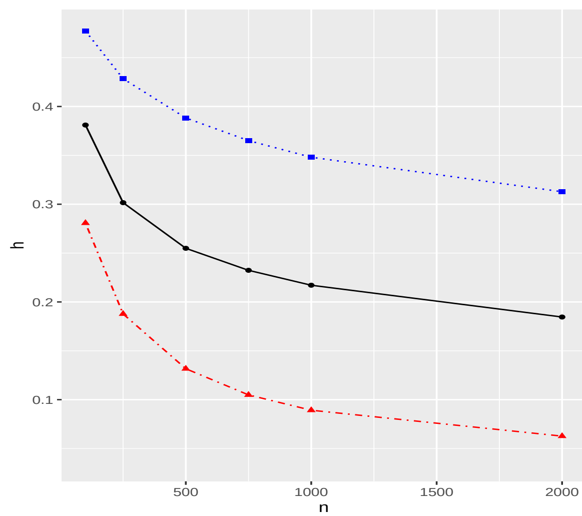
(a) $x = -1$



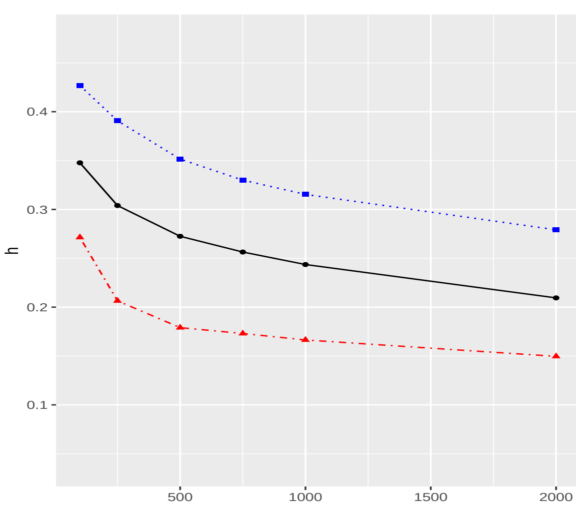
(b) $x = -0.6$



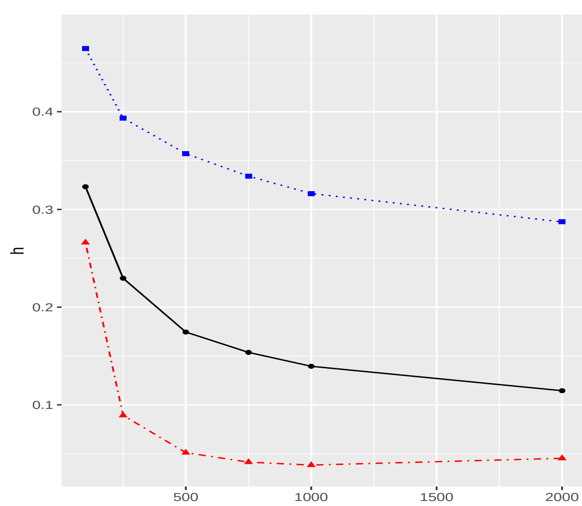
(c) $x = -0.2$



(d) $x = 0.2$



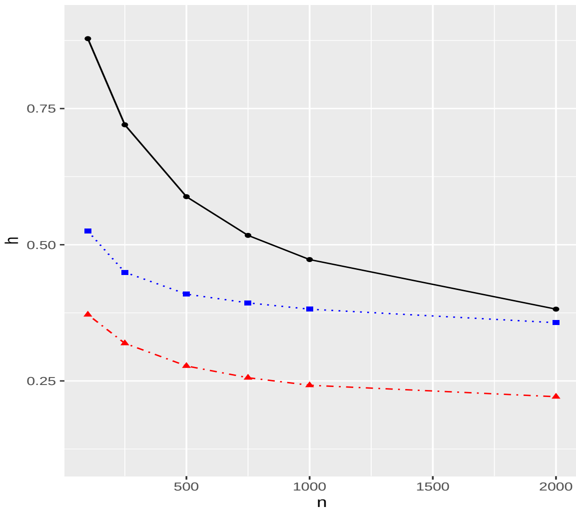
(e) $x = 0.6$



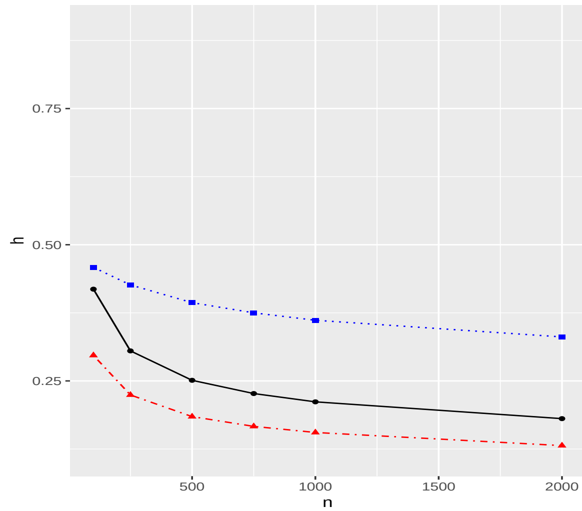
(f) $x = 1$

Notes: —— $\hat{h}_{\texttt{rbc}}$, - · - $\hat{h}_{\texttt{us}}$, ⋯ $\hat{h}_{\texttt{mse}}$

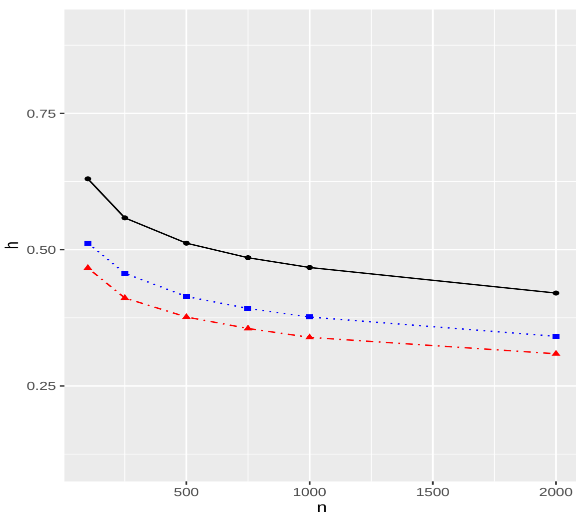
Figure S.20: Average Estimated Bandwidths, Uniform Kernel, $\nu = 1$



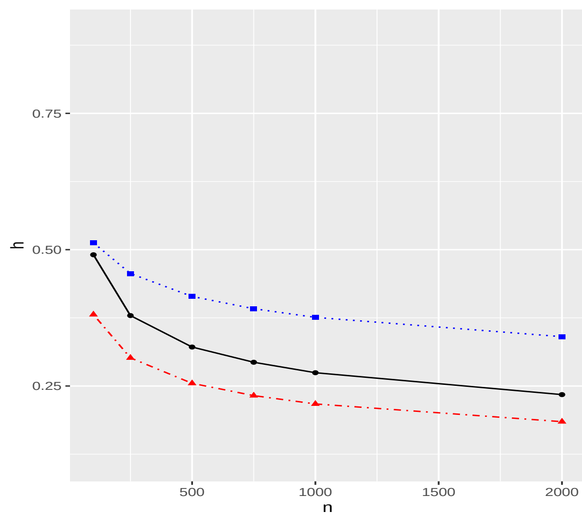
(a) $x = -1$



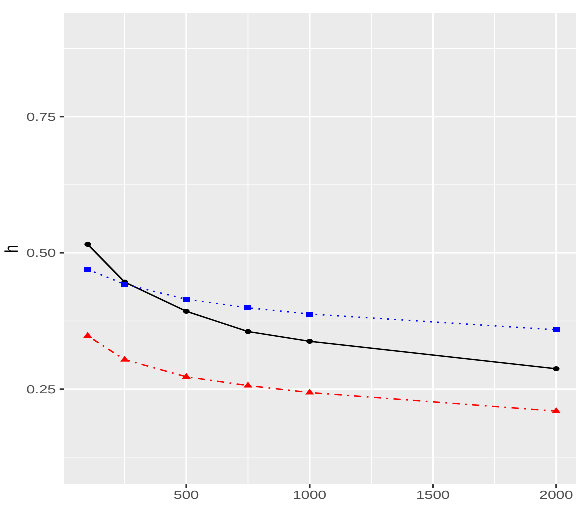
(b) $x = -0.6$



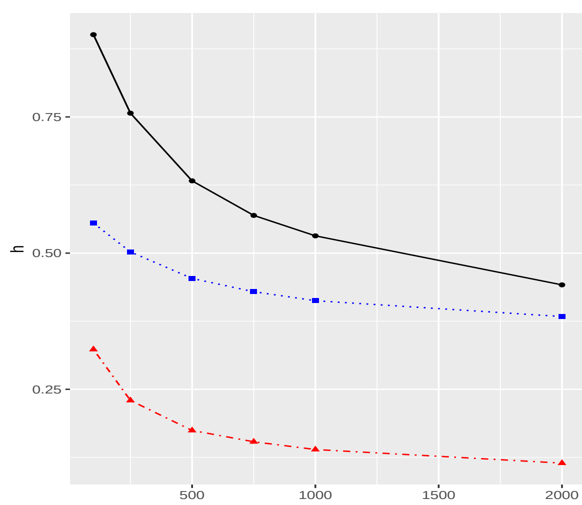
(c) $x = -0.2$



(d) $x = 0.2$



(e) $x = 0.6$



(f) $x = 1$

Notes: $\hat{h}_{\texttt{rbc}}$, $\hat{h}_{\texttt{us}}$, $\hat{h}_{\texttt{mse}}$

Table S.3: Empirical Coverage Probabilities, 95% Confidence Intervals, $\nu = 0$, Epanechnikov Kernel

	$h_{\texttt{RBC}}$			$h_{\texttt{US}}$			$h_{\texttt{MSE}}$		
	h	RBC	US	h	RBC	US	h	RBC	US
x =**-1**									
100	0.436	0.881	0.877	0.320	0.873	0.889	0.507	0.899	0.875
250	0.368	0.906	0.892	0.115	0.879	0.893	0.462	0.912	0.862
500	0.321	0.925	0.902	0.116	0.879	0.893	0.438	0.930	0.828
750	0.295	0.935	0.915	0.168	0.881	0.880	0.420	0.934	0.797
1000	0.280	0.941	0.908	0.205	0.887	0.860	0.404	0.930	0.769
2000	0.255	0.941	0.902	0.143	0.920	0.902	0.356	0.924	0.696
x =**-0.6**									
100	0.335	0.922	0.898	0.255	0.919	0.909	0.356	0.929	0.897
250	0.262	0.935	0.922	0.145	0.927	0.931	0.342	0.940	0.874
500	0.221	0.941	0.927	0.104	0.942	0.941	0.316	0.944	0.869
750	0.200	0.948	0.941	0.090	0.938	0.942	0.291	0.947	0.867
1000	0.186	0.949	0.942	0.081	0.946	0.950	0.274	0.950	0.870
2000	0.158	0.947	0.936	0.063	0.946	0.941	0.235	0.944	0.868
x =**-0.2**									
100	0.564	0.800	0.388	0.242	0.910	0.873	0.512	0.858	0.446
250	0.490	0.794	0.286	0.169	0.924	0.911	0.441	0.874	0.316
500	0.446	0.791	0.220	0.127	0.941	0.935	0.386	0.890	0.234
750	0.423	0.786	0.182	0.107	0.936	0.932	0.357	0.905	0.207
1000	0.402	0.785	0.164	0.095	0.936	0.934	0.337	0.908	0.189
2000	0.368	0.785	0.139	0.071	0.948	0.945	0.293	0.933	0.153
x =**0.2**									
100	0.468	0.890	0.645	0.326	0.888	0.760	0.647	0.821	0.231
250	0.379	0.928	0.647	0.211	0.917	0.843	0.645	0.642	0.026
500	0.328	0.935	0.666	0.144	0.930	0.903	0.635	0.403	0.009
750	0.302	0.941	0.658	0.116	0.944	0.932	0.623	0.259	0.005
1000	0.284	0.949	0.672	0.100	0.943	0.941	0.611	0.212	0.004
2000	0.244	0.945	0.708	0.074	0.943	0.945	0.575	0.150	0.003
x =**0.6**									
100	0.407	0.922	0.926	0.381	0.928	0.926	0.479	0.932	0.929
250	0.338	0.934	0.936	0.291	0.938	0.937	0.535	0.931	0.927
500	0.284	0.937	0.936	0.253	0.944	0.940	0.551	0.900	0.909
750	0.258	0.943	0.944	0.234	0.948	0.939	0.538	0.881	0.903
1000	0.246	0.940	0.937	0.218	0.945	0.933	0.529	0.853	0.888
2000	0.211	0.943	0.940	0.174	0.944	0.931	0.498	0.760	0.832
x =**1**									
100	0.378	0.897	0.902	0.253	0.887	0.906	0.484	0.905	0.901
250	0.269	0.898	0.911	0.084	0.877	0.900	0.401	0.926	0.922
500	0.204	0.906	0.917	0.043	0.879	0.895	0.374	0.929	0.928
750	0.179	0.928	0.930	0.035	0.881	0.898	0.361	0.944	0.931
1000	0.165	0.925	0.938	0.036	0.880	0.892	0.350	0.948	0.942
2000	0.136	0.939	0.939	0.048	0.894	0.907	0.322	0.942	0.935

Table S.4: Empirical Coverage Probabilities, 95% Confidence Intervals, $\nu = 1$, Epanechnikov Kernel

	$h_{\texttt{RBC}}$			$h_{\texttt{US}}$			$h_{\texttt{MSE}}$		
	h	RBC	US	h	RBC	US	h	RBC	US
$x =$**-1**									
100	1.041	0.689	0.337	0.436	0.926	0.891	0.556	0.923	0.864
250	0.861	0.774	0.399	0.368	0.924	0.910	0.484	0.937	0.890
500	0.718	0.872	0.488	0.321	0.930	0.920	0.451	0.949	0.901
750	0.635	0.915	0.561	0.295	0.935	0.929	0.434	0.950	0.900
1000	0.582	0.931	0.625	0.280	0.937	0.932	0.420	0.949	0.908
2000	0.475	0.942	0.760	0.255	0.938	0.941	0.392	0.955	0.904
$x =$**-0.6**									
100	0.482	0.874	0.621	0.335	0.911	0.809	0.486	0.918	0.571
250	0.360	0.936	0.654	0.262	0.940	0.876	0.475	0.933	0.229
500	0.298	0.943	0.774	0.221	0.942	0.910	0.446	0.943	0.072
750	0.268	0.948	0.815	0.200	0.943	0.925	0.424	0.942	0.043
1000	0.250	0.944	0.850	0.186	0.949	0.933	0.408	0.942	0.041
2000	0.214	0.948	0.884	0.158	0.947	0.937	0.373	0.943	0.024
$x =$**-0.2**									
100	0.732	0.660	0.097	0.564	0.819	0.388	0.548	0.906	0.277
250	0.642	0.677	0.072	0.490	0.850	0.400	0.488	0.937	0.204
500	0.590	0.713	0.056	0.446	0.858	0.431	0.443	0.939	0.177
750	0.563	0.727	0.048	0.423	0.851	0.441	0.420	0.942	0.163
1000	0.539	0.743	0.049	0.402	0.864	0.459	0.403	0.942	0.167
2000	0.488	0.783	0.050	0.368	0.858	0.493	0.365	0.934	0.175
$x =$**0.2**									
100	0.575	0.626	0.193	0.468	0.782	0.422	0.575	0.691	0.035
250	0.446	0.743	0.161	0.379	0.862	0.400	0.514	0.598	0.002
500	0.377	0.823	0.143	0.328	0.894	0.373	0.467	0.524	0.000
750	0.345	0.861	0.135	0.302	0.916	0.389	0.442	0.514	0.000
1000	0.323	0.875	0.141	0.284	0.921	0.402	0.425	0.483	0.000
2000	0.277	0.917	0.190	0.244	0.941	0.483	0.385	0.490	0.000
$x =$**0.6**									
100	0.601	0.915	0.925	0.407	0.926	0.926	0.515	0.939	0.934
250	0.537	0.920	0.930	0.338	0.933	0.937	0.508	0.945	0.936
500	0.496	0.904	0.941	0.284	0.938	0.942	0.500	0.945	0.952
750	0.461	0.907	0.941	0.258	0.938	0.946	0.482	0.948	0.950
1000	0.431	0.904	0.944	0.246	0.945	0.949	0.468	0.942	0.945
2000	0.362	0.911	0.946	0.211	0.942	0.946	0.434	0.941	0.945
$x =$**1**									
100	1.084	0.882	0.878	0.378	0.927	0.919	0.659	0.945	0.930
250	0.922	0.895	0.890	0.269	0.921	0.918	0.617	0.946	0.940
500	0.784	0.922	0.915	0.204	0.932	0.929	0.577	0.947	0.941
750	0.707	0.933	0.933	0.179	0.926	0.929	0.551	0.946	0.949
1000	0.663	0.939	0.942	0.165	0.932	0.929	0.528	0.950	0.947
2000	0.546	0.943	0.942	0.136	0.940	0.946	0.469	0.949	0.941

Table S.5: Average Interval Length, 95% Confidence Intervals, $\nu = 0$, Epanechnikov Kernel

	$h_{\texttt{RBC}}$			$h_{\texttt{US}}$			$h_{\texttt{MSE}}$		
	h	RBC	US	h	RBC	US	h	RBC	US
x =**-1**									
100	0.436	2.442	1.674	0.320	2.761	1.793	0.507	2.330	1.656
250	0.368	1.713	1.239	0.115	2.781	1.795	0.462	1.492	1.089
500	0.321	1.282	0.940	0.116	2.585	1.697	0.438	1.084	0.796
750	0.295	1.090	0.801	0.168	2.210	1.474	0.420	0.907	0.667
1000	0.280	0.966	0.710	0.205	1.799	1.240	0.404	0.801	0.589
2000	0.255	0.711	0.524	0.143	1.222	0.891	0.356	0.605	0.446
x =**-0.6**									
100	0.335	1.020	0.762	0.255	1.173	0.863	0.356	0.983	0.734
250	0.262	0.715	0.537	0.145	0.985	0.739	0.342	0.633	0.473
500	0.221	0.547	0.411	0.104	0.812	0.611	0.316	0.461	0.346
750	0.200	0.467	0.352	0.090	0.706	0.531	0.291	0.390	0.293
1000	0.186	0.419	0.315	0.081	0.646	0.487	0.274	0.349	0.262
2000	0.158	0.322	0.242	0.063	0.513	0.386	0.235	0.266	0.200
x =**-0.2**									
100	0.564	0.786	0.592	0.242	1.171	0.864	0.512	0.799	0.601
250	0.490	0.533	0.401	0.169	0.890	0.671	0.441	0.543	0.409
500	0.446	0.396	0.298	0.127	0.724	0.545	0.386	0.411	0.309
750	0.423	0.334	0.251	0.107	0.644	0.485	0.357	0.349	0.263
1000	0.402	0.297	0.223	0.095	0.592	0.446	0.337	0.312	0.234
2000	0.368	0.221	0.166	0.071	0.481	0.362	0.293	0.236	0.178
x =**0.2**									
100	0.468	0.844	0.632	0.326	1.050	0.778	0.647	0.711	0.532
250	0.379	0.589	0.443	0.211	0.822	0.619	0.645	0.451	0.338
500	0.328	0.447	0.336	0.144	0.692	0.520	0.635	0.321	0.241
750	0.302	0.379	0.286	0.116	0.619	0.466	0.623	0.265	0.200
1000	0.284	0.339	0.255	0.100	0.575	0.433	0.611	0.232	0.174
2000	0.244	0.258	0.194	0.074	0.471	0.355	0.575	0.169	0.127
x =**0.6**									
100	0.407	0.942	0.706	0.381	0.994	0.749	0.479	0.864	0.643
250	0.338	0.647	0.489	0.291	0.718	0.542	0.535	0.525	0.399
500	0.284	0.495	0.373	0.253	0.544	0.410	0.551	0.367	0.282
750	0.258	0.422	0.318	0.234	0.463	0.348	0.538	0.304	0.234
1000	0.246	0.375	0.283	0.218	0.414	0.311	0.529	0.266	0.205
2000	0.211	0.285	0.215	0.174	0.328	0.247	0.498	0.196	0.150
x =**1**									
100	0.378	2.547	1.725	0.253	2.847	1.832	0.484	2.397	1.704
250	0.269	2.034	1.440	0.084	2.893	1.850	0.401	1.625	1.182
500	0.204	1.609	1.169	0.043	2.825	1.828	0.374	1.172	0.861
750	0.179	1.391	1.018	0.035	2.760	1.806	0.361	0.971	0.715
1000	0.165	1.250	0.917	0.036	2.670	1.767	0.350	0.856	0.630
2000	0.136	0.966	0.710	0.048	2.069	1.458	0.322	0.627	0.462

Table S.6: Average Interval Length, 95% Confidence Intervals, $\nu = 1$, Epanechnikov Kernel

	$h_{\mathtt{RBC}}$			$h_{\mathtt{US}}$			$h_{\mathtt{MSE}}$		
	h	RBC	US	h	RBC	US	h	RBC	US
$x =$**-1**									
100	1.041	19.701	9.445	0.436	69.744	28.528	0.556	48.016	22.617
250	0.861	16.338	8.028	0.368	61.142	28.904	0.484	34.436	16.763
500	0.718	14.947	7.391	0.321	49.275	24.059	0.451	26.558	13.049
750	0.635	14.553	7.190	0.295	44.292	21.675	0.434	22.961	11.300
1000	0.582	14.242	7.042	0.280	40.630	19.944	0.420	20.849	10.271
2000	0.475	13.336	6.587	0.255	32.206	15.851	0.392	16.238	8.016
$x =$**-0.6**									
100	0.482	6.098	3.158	0.335	9.635	4.835	0.486	5.571	2.846
250	0.360	5.109	2.557	0.262	8.041	4.096	0.475	3.562	1.750
500	0.298	4.512	2.289	0.221	7.088	3.612	0.446	2.666	1.273
750	0.268	4.255	2.165	0.200	6.681	3.405	0.424	2.287	1.096
1000	0.250	4.084	2.079	0.186	6.358	3.243	0.408	2.055	0.997
2000	0.214	3.627	1.846	0.158	5.737	2.921	0.373	1.597	0.804
$x =$**-0.2**									
100	0.732	3.023	1.531	0.564	4.600	2.343	0.548	4.033	2.033
250	0.642	2.216	1.127	0.490	3.476	1.775	0.488	2.994	1.513
500	0.590	1.757	0.896	0.446	2.832	1.443	0.443	2.428	1.231
750	0.563	1.535	0.781	0.423	2.541	1.294	0.420	2.153	1.091
1000	0.539	1.407	0.716	0.402	2.364	1.204	0.403	1.977	1.004
2000	0.488	1.156	0.588	0.368	1.971	1.004	0.365	1.617	0.822
$x =$**0.2**									
100	0.575	3.998	2.031	0.468	5.462	2.765	0.575	3.747	1.885
250	0.446	3.521	1.783	0.379	4.477	2.277	0.514	2.765	1.396
500	0.377	3.133	1.591	0.328	3.853	1.962	0.467	2.244	1.137
750	0.345	2.909	1.478	0.302	3.554	1.813	0.442	1.988	1.009
1000	0.323	2.770	1.408	0.284	3.362	1.713	0.425	1.827	0.929
2000	0.277	2.456	1.250	0.244	2.973	1.515	0.385	1.493	0.760
$x =$**0.6**									
100	0.601	4.652	2.707	0.407	7.736	3.930	0.515	5.184	2.714
250	0.537	3.328	1.792	0.338	6.157	3.165	0.508	3.305	1.709
500	0.496	2.585	1.354	0.284	5.415	2.762	0.500	2.348	1.187
750	0.461	2.313	1.196	0.258	4.969	2.534	0.482	2.000	0.994
1000	0.431	2.179	1.124	0.246	4.692	2.396	0.468	1.791	0.883
2000	0.362	1.953	1.001	0.211	4.095	2.087	0.434	1.377	0.680
$x =$**1**									
100	1.084	18.377	8.838	0.378	77.836	31.082	0.659	36.723	17.371
250	0.922	14.556	7.134	0.269	102.575	46.277	0.617	24.053	11.744
500	0.784	12.866	6.341	0.204	99.887	47.579	0.577	18.347	9.047
750	0.707	12.195	6.022	0.179	96.458	46.670	0.551	15.971	7.885
1000	0.663	11.490	5.680	0.165	91.635	44.564	0.528	14.687	7.261
2000	0.546	10.592	5.232	0.136	82.575	40.567	0.469	12.323	6.085

Table S.7: Empirical Coverage Probabilities, 95% Confidence Intervals, $\nu = 0$, Uniform Kernel

	$h_{\texttt{RBC}}$			$h_{\texttt{US}}$			$h_{\texttt{MSE}}$		
	h	RBC	US	h	RBC	US	h	RBC	US
x =**-1**									
100	0.372	0.901	0.876	0.286	0.893	0.888	0.462	0.903	0.877
250	0.319	0.910	0.899	0.122	0.888	0.909	0.397	0.919	0.881
500	0.277	0.925	0.905	0.105	0.889	0.901	0.372	0.934	0.852
750	0.256	0.930	0.917	0.117	0.897	0.892	0.352	0.940	0.827
1000	0.242	0.938	0.906	0.126	0.893	0.873	0.326	0.945	0.812
2000	0.221	0.941	0.900	0.156	0.894	0.809	0.247	0.942	0.854
x =**-0.6**									
100	0.297	0.925	0.890	0.175	0.919	0.918	0.364	0.930	0.841
250	0.224	0.937	0.910	0.106	0.921	0.927	0.319	0.936	0.794
500	0.184	0.947	0.924	0.080	0.932	0.940	0.278	0.942	0.778
750	0.167	0.949	0.932	0.069	0.938	0.942	0.256	0.951	0.768
1000	0.155	0.947	0.936	0.062	0.939	0.944	0.240	0.949	0.779
2000	0.131	0.946	0.934	0.050	0.948	0.941	0.207	0.945	0.770
x =**-0.2**									
100	0.466	0.863	0.330	0.205	0.920	0.841	0.455	0.894	0.287
250	0.411	0.852	0.237	0.140	0.924	0.903	0.396	0.899	0.153
500	0.377	0.846	0.180	0.103	0.939	0.933	0.348	0.912	0.094
750	0.355	0.841	0.155	0.087	0.937	0.935	0.322	0.927	0.066
1000	0.339	0.843	0.144	0.076	0.938	0.934	0.303	0.931	0.066
2000	0.309	0.836	0.128	0.057	0.945	0.946	0.263	0.941	0.050
x =**0.2**									
100	0.381	0.910	0.606	0.281	0.904	0.743	0.477	0.933	0.271
250	0.302	0.938	0.623	0.188	0.928	0.831	0.428	0.940	0.090
500	0.255	0.939	0.671	0.132	0.932	0.892	0.388	0.939	0.042
750	0.232	0.942	0.669	0.105	0.937	0.922	0.365	0.940	0.028
1000	0.217	0.948	0.703	0.089	0.942	0.933	0.348	0.941	0.026
2000	0.185	0.945	0.748	0.063	0.943	0.944	0.313	0.926	0.019
x =**0.6**									
100	0.348	0.928	0.923	0.271	0.932	0.931	0.427	0.937	0.925
250	0.304	0.934	0.935	0.206	0.939	0.937	0.391	0.945	0.924
500	0.273	0.935	0.935	0.179	0.946	0.942	0.352	0.945	0.920
750	0.256	0.938	0.933	0.173	0.951	0.949	0.330	0.949	0.912
1000	0.244	0.935	0.934	0.166	0.946	0.944	0.315	0.944	0.911
2000	0.209	0.935	0.927	0.150	0.950	0.932	0.279	0.948	0.880
x =**1**									
100	0.323	0.907	0.917	0.266	0.907	0.919	0.464	0.909	0.918
250	0.230	0.901	0.919	0.089	0.894	0.919	0.393	0.922	0.930
500	0.175	0.910	0.922	0.051	0.893	0.907	0.357	0.934	0.940
750	0.154	0.921	0.937	0.041	0.897	0.915	0.334	0.940	0.939
1000	0.139	0.922	0.939	0.038	0.894	0.918	0.316	0.948	0.944
2000	0.115	0.938	0.935	0.045	0.900	0.918	0.287	0.937	0.937

Table S.8: Empirical Coverage Probabilities, 95% Confidence Intervals, $\nu = 1$, Uniform Kernel

	$h_{\texttt{RBC}}$			$h_{\texttt{US}}$			$h_{\texttt{MSE}}$		
	h	RBC	US	h	RBC	US	h	RBC	US
x =**-1**									
100	0.878	0.835	0.413	0.372	0.928	0.907	0.525	0.924	0.856
250	0.720	0.883	0.476	0.319	0.934	0.917	0.449	0.936	0.878
500	0.588	0.930	0.605	0.277	0.937	0.923	0.409	0.940	0.900
750	0.517	0.935	0.677	0.256	0.939	0.935	0.393	0.946	0.899
1000	0.473	0.943	0.739	0.242	0.941	0.934	0.382	0.947	0.910
2000	0.382	0.942	0.851	0.221	0.943	0.941	0.357	0.948	0.898
x =**-0.6**									
100	0.418	0.899	0.592	0.297	0.919	0.792	0.458	0.927	0.420
250	0.305	0.935	0.680	0.224	0.932	0.873	0.426	0.937	0.129
500	0.251	0.946	0.797	0.184	0.944	0.915	0.394	0.947	0.064
750	0.227	0.948	0.831	0.167	0.943	0.926	0.375	0.943	0.044
1000	0.212	0.948	0.867	0.155	0.947	0.933	0.361	0.949	0.042
2000	0.181	0.950	0.891	0.131	0.944	0.945	0.331	0.947	0.027
x =**-0.2**									
100	0.630	0.782	0.100	0.466	0.879	0.437	0.512	0.921	0.135
250	0.558	0.809	0.064	0.411	0.883	0.470	0.457	0.946	0.067
500	0.512	0.829	0.049	0.377	0.886	0.506	0.414	0.941	0.052
750	0.485	0.849	0.046	0.355	0.882	0.516	0.392	0.947	0.036
1000	0.467	0.843	0.048	0.339	0.886	0.535	0.376	0.941	0.038
2000	0.420	0.869	0.054	0.309	0.884	0.562	0.341	0.944	0.041
x =**0.2**									
100	0.491	0.785	0.200	0.381	0.866	0.502	0.513	0.818	0.048
250	0.379	0.875	0.172	0.302	0.923	0.526	0.456	0.785	0.004
500	0.321	0.907	0.150	0.255	0.938	0.552	0.414	0.759	0.000
750	0.294	0.928	0.150	0.232	0.947	0.600	0.392	0.773	0.000
1000	0.274	0.933	0.154	0.217	0.947	0.623	0.376	0.758	0.000
2000	0.234	0.948	0.218	0.185	0.948	0.713	0.340	0.778	0.000
x =**0.6**									
100	0.516	0.928	0.928	0.348	0.932	0.917	0.470	0.935	0.935
250	0.446	0.934	0.925	0.304	0.929	0.933	0.442	0.943	0.934
500	0.393	0.932	0.936	0.273	0.927	0.944	0.415	0.948	0.949
750	0.356	0.931	0.939	0.256	0.938	0.946	0.399	0.952	0.950
1000	0.338	0.934	0.943	0.244	0.938	0.946	0.388	0.950	0.945
2000	0.287	0.928	0.937	0.209	0.934	0.950	0.359	0.950	0.943
x =**1**									
100	0.901	0.922	0.886	0.323	0.932	0.926	0.555	0.935	0.932
250	0.757	0.926	0.903	0.230	0.926	0.918	0.502	0.933	0.937
500	0.633	0.938	0.920	0.175	0.930	0.931	0.454	0.940	0.941
750	0.569	0.943	0.943	0.154	0.932	0.935	0.429	0.945	0.950
1000	0.532	0.948	0.944	0.139	0.935	0.935	0.413	0.950	0.948
2000	0.442	0.944	0.941	0.115	0.941	0.940	0.384	0.945	0.943

Table S.9: Average Interval Length, 95% Confidence Intervals, $\nu = 0$, Uniform Kernel

	$h_{\texttt{RBC}}$			$h_{\texttt{US}}$			$h_{\texttt{MSE}}$		
	h	RBC	US	h	RBC	US	h	RBC	US
x =**-1**									
100	0.372	2.561	1.638	0.286	2.649	1.688	0.462	2.566	1.636
250	0.319	1.948	1.274	0.122	2.593	1.656	0.397	1.694	1.116
500	0.277	1.457	0.965	0.105	2.453	1.573	0.372	1.230	0.816
750	0.256	1.231	0.818	0.117	2.266	1.453	0.352	1.034	0.688
1000	0.242	1.092	0.725	0.126	2.106	1.347	0.326	0.931	0.620
2000	0.221	0.798	0.532	0.156	1.507	0.982	0.247	0.753	0.502
x =**-0.6**									
100	0.297	1.107	0.741	0.175	1.260	0.844	0.364	1.026	0.690
250	0.224	0.802	0.535	0.106	1.169	0.779	0.319	0.682	0.455
500	0.184	0.619	0.413	0.080	0.964	0.643	0.278	0.513	0.342
750	0.167	0.528	0.352	0.069	0.836	0.557	0.256	0.434	0.290
1000	0.155	0.474	0.316	0.062	0.762	0.509	0.240	0.389	0.259
2000	0.131	0.364	0.243	0.050	0.597	0.398	0.207	0.296	0.197
x =**-0.2**									
100	0.466	0.894	0.598	0.205	1.234	0.822	0.455	0.879	0.585
250	0.411	0.604	0.404	0.140	1.020	0.682	0.396	0.594	0.395
500	0.377	0.449	0.300	0.103	0.836	0.558	0.348	0.447	0.298
750	0.355	0.380	0.254	0.087	0.744	0.495	0.322	0.381	0.254
1000	0.339	0.338	0.225	0.076	0.683	0.457	0.303	0.340	0.226
2000	0.309	0.252	0.168	0.057	0.555	0.370	0.263	0.258	0.172
x =**0.2**									
100	0.381	0.972	0.646	0.281	1.132	0.754	0.477	0.863	0.572
250	0.302	0.684	0.457	0.188	0.914	0.611	0.428	0.573	0.381
500	0.255	0.524	0.349	0.132	0.762	0.508	0.388	0.425	0.283
750	0.232	0.446	0.298	0.105	0.684	0.456	0.365	0.358	0.239
1000	0.217	0.401	0.267	0.089	0.637	0.426	0.348	0.318	0.212
2000	0.185	0.306	0.204	0.063	0.529	0.353	0.313	0.238	0.159
x =**0.6**									
100	0.348	1.045	0.702	0.271	1.150	0.772	0.427	0.936	0.635
250	0.304	0.710	0.477	0.206	0.879	0.590	0.391	0.611	0.411
500	0.273	0.527	0.354	0.179	0.668	0.447	0.352	0.454	0.304
750	0.256	0.446	0.299	0.173	0.561	0.374	0.330	0.383	0.256
1000	0.244	0.395	0.265	0.166	0.492	0.329	0.315	0.339	0.227
2000	0.209	0.301	0.201	0.150	0.368	0.245	0.279	0.257	0.171
x =**1**									
100	0.323	2.608	1.664	0.266	2.666	1.696	0.464	2.555	1.631
250	0.230	2.307	1.477	0.089	2.702	1.707	0.393	1.733	1.136
500	0.175	1.838	1.206	0.051	2.632	1.679	0.357	1.263	0.838
750	0.154	1.578	1.044	0.041	2.568	1.656	0.334	1.059	0.705
1000	0.139	1.427	0.946	0.038	2.547	1.636	0.316	0.943	0.628
2000	0.115	1.101	0.733	0.045	2.171	1.410	0.287	0.694	0.463

Table S.10: Average Interval Length, 95% Confidence Intervals, $\nu = 1$, Uniform Kernel

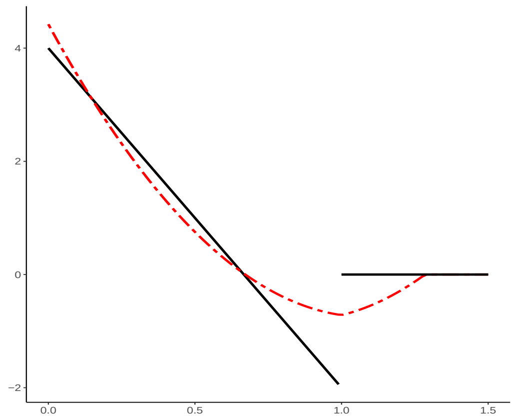
	$h_{\texttt{RBC}}$			$h_{\texttt{US}}$			$h_{\texttt{MSE}}$		
	h	RBC	US	h	RBC	US	h	RBC	US
$x =$**-1**									
100	0.878	28.499	11.007	0.372	74.089	26.860	0.525	58.507	21.792
250	0.720	23.646	9.391	0.319	85.698	32.415	0.449	42.941	16.854
500	0.588	22.170	8.862	0.277	69.500	27.253	0.409	33.847	13.454
750	0.517	21.707	8.686	0.256	61.629	24.361	0.393	29.132	11.609
1000	0.473	21.202	8.489	0.242	56.624	22.424	0.382	26.357	10.515
2000	0.382	20.050	8.033	0.221	44.037	17.519	0.357	20.483	8.194
$x =$**-0.6**									
100	0.418	7.793	3.196	0.297	11.921	4.762	0.458	6.335	2.673
250	0.305	6.840	2.707	0.224	11.095	4.415	0.426	4.159	1.683
500	0.251	6.225	2.472	0.184	10.016	4.010	0.394	3.181	1.265
750	0.227	5.844	2.334	0.167	9.395	3.760	0.375	2.752	1.095
1000	0.212	5.596	2.236	0.155	8.965	3.589	0.361	2.509	0.998
2000	0.181	4.983	1.991	0.131	8.115	3.242	0.331	2.022	0.807
$x =$**-0.2**									
100	0.630	3.974	1.598	0.466	6.568	2.636	0.512	4.846	1.878
250	0.558	2.909	1.166	0.411	4.975	1.994	0.457	3.552	1.399
500	0.512	2.315	0.930	0.377	4.070	1.629	0.414	2.876	1.144
750	0.485	2.045	0.818	0.355	3.667	1.469	0.392	2.543	1.011
1000	0.467	1.873	0.750	0.339	3.431	1.373	0.376	2.336	0.931
2000	0.420	1.551	0.621	0.309	2.908	1.162	0.341	1.910	0.763
$x =$**0.2**									
100	0.491	5.462	2.151	0.381	8.147	3.229	0.513	4.813	1.859
250	0.379	4.802	1.897	0.302	6.861	2.742	0.456	3.550	1.395
500	0.321	4.238	1.686	0.255	6.070	2.417	0.414	2.874	1.141
750	0.294	3.947	1.572	0.232	5.664	2.267	0.392	2.548	1.014
1000	0.274	3.766	1.502	0.217	5.400	2.160	0.376	2.342	0.934
2000	0.234	3.362	1.342	0.185	4.837	1.935	0.340	1.914	0.766
$x =$**0.6**									
100	0.516	5.937	2.723	0.348	10.278	4.159	0.470	6.156	2.648
250	0.446	4.422	1.876	0.304	8.031	3.244	0.442	4.028	1.669
500	0.393	3.684	1.516	0.273	6.540	2.629	0.415	3.008	1.218
750	0.356	3.405	1.382	0.256	5.857	2.359	0.399	2.565	1.034
1000	0.338	3.188	1.292	0.244	5.466	2.202	0.388	2.313	0.933
2000	0.287	2.833	1.137	0.209	4.776	1.918	0.359	1.842	0.740
$x =$**1**									
100	0.901	27.285	10.546	0.323	77.275	28.086	0.555	53.679	20.014
250	0.757	21.694	8.600	0.230	136.702	50.383	0.502	36.186	14.246
500	0.633	19.479	7.765	0.175	139.797	53.718	0.454	28.845	11.503
750	0.569	18.470	7.391	0.154	132.358	51.889	0.429	25.410	10.147
1000	0.532	17.516	7.020	0.139	129.817	50.990	0.413	23.265	9.305
2000	0.442	15.794	6.319	0.115	117.431	46.691	0.384	18.243	7.290

S.9.2 Numerical Computations

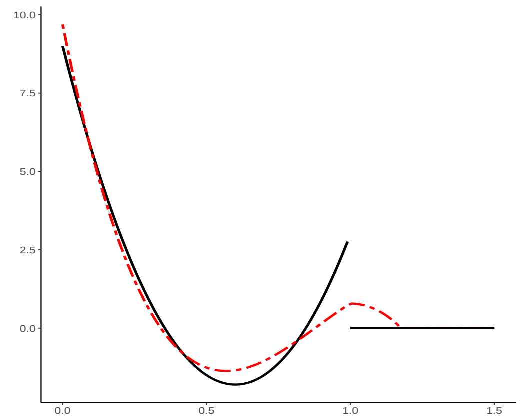
In the main text we discussed the optimization of ρ by minimizing the L_2 distance to the known optimal kernel shape in various contexts. These optimal kernel shapes are shown in the figures below for both the Triangular and Epanechnikov kernels, at interior and boundary points, for levels and derivatives. In each case the black line shows $\mathcal{K}^*_{p+1}(u)$ while the dash-dotted red line is $\mathcal{K}_{\texttt{rbc}}(u; K, \rho^*, \nu)$.

Figure S.21: $\mathcal{K}^*_{p+1}(u)$ vs. $\mathcal{K}_{\texttt{rbc}}(u; K, \rho^*, \nu)$, $\nu = 0$

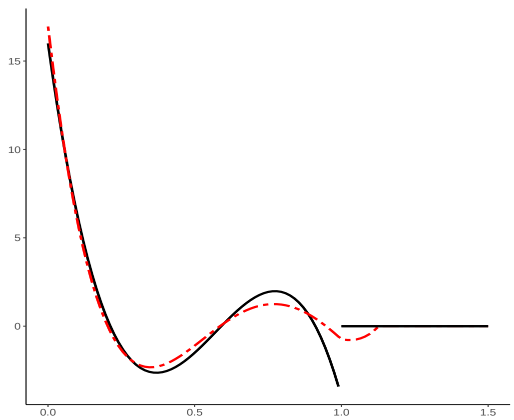
(a) Triangular Kernel, Boundary Point



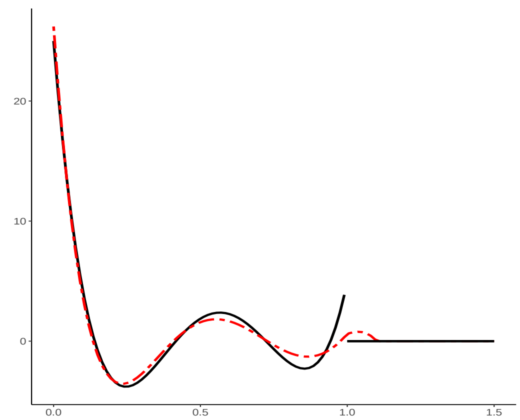
(b) $p = 0$



(c) $p = 1$

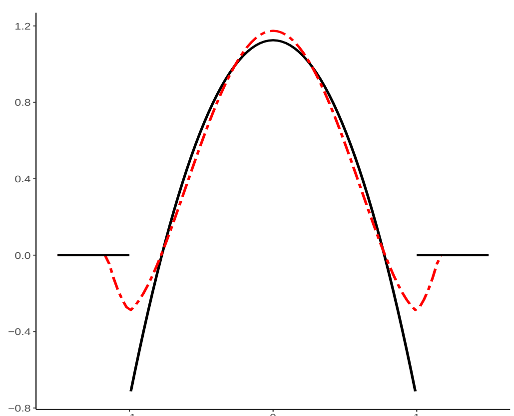


(d) $p = 2$

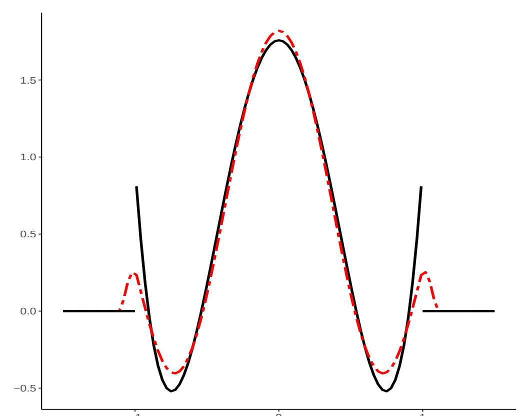


(e) $p = 3$

(f) Epanechnikov Kernel, Interior Point



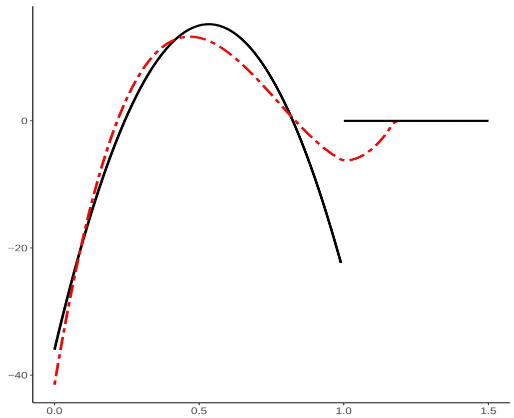
(g) $p = 1$



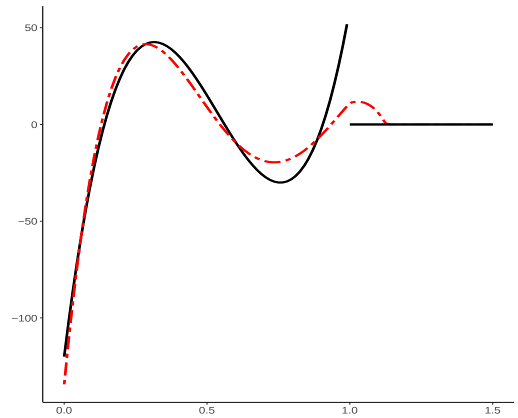
(h) $p = 3$

Figure S.22: $\mathcal{K}^*_{p+1}(u)$ vs. $\mathcal{K}_{\texttt{rbc}}(u; K, \rho^*, \nu)$, $\nu = 1$

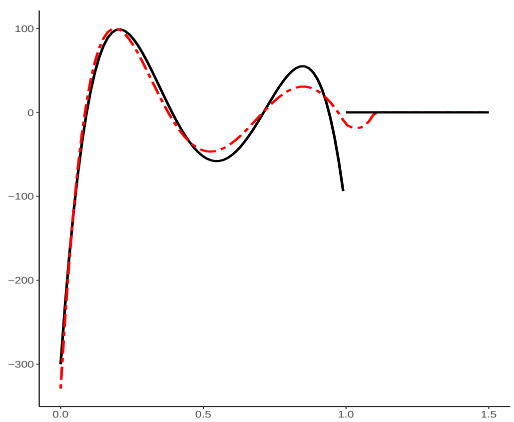
(a) Triangular Kernel, Boundary Point



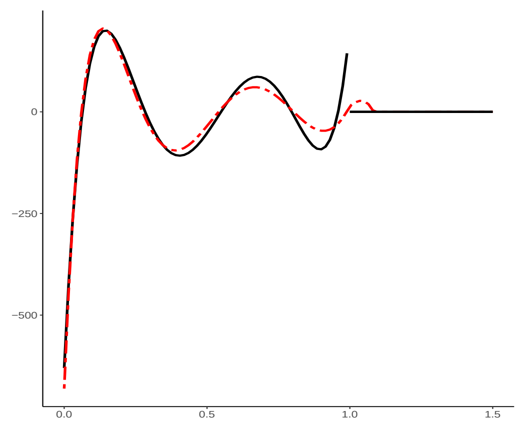
(b) $p = 1$



(c) $p = 2$

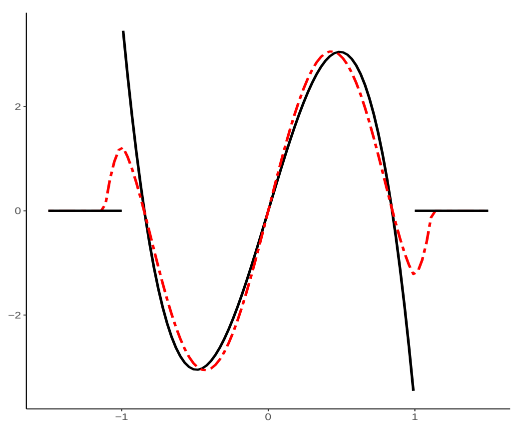


(d) $p = 3$

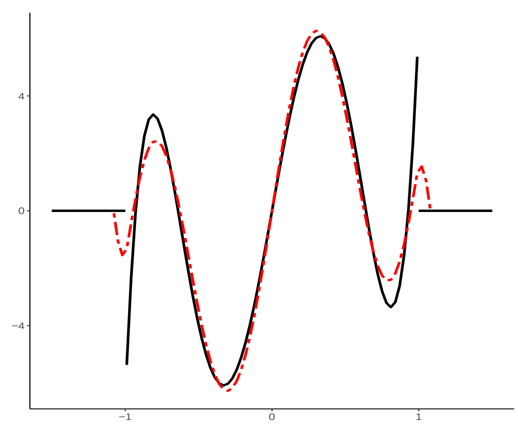


(e) $p = 4$

(f) Epanechnikov Kernel, Interior Point



(g) $p = 2$



(h) $p = 4$

S.10 List of Notation

Below is a (hopefully) complete list of the notation used in this Part, group by Section, roughly in order of introduction. This is intended only as a reference. Each object is redefined below when it is needed.

Asymptotic orders and their in-probability versions hold uniformly in $\mathscr{F}_S$, as required by our framework; e.g., $A_n = o_{\mathbb{P}}(a_n)$ means $\sup_{F\in\mathscr{F}_S} \mathbb{P}_F[|A_n/a_n| > \epsilon] \to 0$ for every $\epsilon > 0$.

Local Polynomial Regression, t-Statistics, and Confidence Intervals

- $\{(Y_1, X_1), \ldots, (Y_n, X_n)\}$ is a random sample distributed according to F, the data-generating process. F is assumed to belong to a class $\mathscr{F}_S$
- $\mu^{(\nu)} = \mu_F^{(\nu)}(\mathsf{x}) := \frac{\partial^\nu}{\partial x^\nu}\mathbb{E}_F\left[Y \mid X = x\right]\big|_{x=\mathsf{x}}$, where $\nu \le S$, where $\mu(\cdot)$ possess at least S derivatives.
- $\mu_F(\mathsf{x}) = \mu_F^{(0)}(\mathsf{x}) = \mathbb{E}_F[Y \mid X = \mathsf{x}]$
- Where it causes no confusion the point of evaluation x will be omitted as an argument, so that for a function $g(\cdot)$ we will write $g := g(\mathsf{x})$
- $\hat{\mu}^{(\nu)} = \nu!\boldsymbol{e}_\nu'\hat{\boldsymbol{\beta}}_p = \frac{1}{nh^\nu}\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\boldsymbol{Y}$
- $\hat{\boldsymbol{\beta}}_p = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^{p+1}} \sum_{i=1}^n (Y_i - \boldsymbol{r}_p(X_i - \mathsf{x})'\boldsymbol{\beta})^2 K\left(X_{h,i}\right)$
- $\hat{\boldsymbol{\beta}}_{p+1} = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^{(p+1)+1}} \sum_{i=1}^n (Y_i - \boldsymbol{r}_{p+1}(X_i - \mathsf{x})'\boldsymbol{\beta})^2 K\left(X_{b,i}\right)$
- $\boldsymbol{e}_k$ 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
- h is a bandwidth sequence that vanishes as n diverges
- p is an integer greater than ν, with $p - \nu$ odd
- $\boldsymbol{r}_p(u) = (1, u, u^2, \ldots, u^p)'$
- $X_{h,i} = (X_i - \mathsf{x})/h$, for a bandwidth h and point of interest x
- to save space, products of functions will often be written together, with only one argument, for example
$$(K\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i}) := K(X_{h,i})r_p(X_{h,i})r_p(X_{h,i})' = K\left(\frac{X_i - \mathsf{x}}{h}\right)\boldsymbol{r}_p\left(\frac{X_i - \mathsf{x}}{h}\right)\boldsymbol{r}_p\left(\frac{X_i - \mathsf{x}}{h}\right)',$$
- $\boldsymbol{\Gamma} = \frac{1}{nh}\sum_{i=1}^n (K\boldsymbol{r}_p\boldsymbol{r}_p')(X_{h,i}) = (\check{\boldsymbol{R}}'\boldsymbol{W}\check{\boldsymbol{R}})/n$
- $\boldsymbol{\Omega} = [(K\boldsymbol{r}_p)(X_{h,1}), (K\boldsymbol{r}_p)(X_{h,2}), \ldots, (K\boldsymbol{r}_p)(X_{h,n})] = \check{\boldsymbol{R}}'\boldsymbol{W}$
- $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$
- $\boldsymbol{R} = [\boldsymbol{r}_p(X_1 - \mathsf{x}), \cdots, \boldsymbol{r}_p(X_n - \mathsf{x})]'$
- $\boldsymbol{W} = \mathrm{diag}\left(h^{-1}K(X_{h,i}) : i = 1, \ldots, n\right)$

- $\boldsymbol{H} = \operatorname{diag}\left(1, h, h^2, \ldots, h^p\right)$
- $\check{\boldsymbol{R}} = \boldsymbol{R}\boldsymbol{H}^{-1} = [\boldsymbol{r}_p(X_{h,1}), \cdots, \boldsymbol{r}_p(X_{h,n})]'$
- $\operatorname{diag}(a_i : i = 1, \ldots, k)$ denote the $k \times k$ diagonal matrix constructed using the elements $a_1, a_2, \cdots, a_k$
- $\boldsymbol{\Lambda}_k = \boldsymbol{\Omega}\left[X_{h,1}^{p+k}, \ldots, X_{h,n}^{p+k}\right]'/n$, where, in particular $\boldsymbol{\Lambda}_1$ was denoted $\boldsymbol{\Lambda}$ in the main text
- b is a bandwidth sequence that vanishes as n diverges
- $X_{b,i} = (X_i - \mathsf{x})/b$, for a bandwidth b and point of interest x, exactly like $X_{h,i}$ but with b in place of h
- $\bar{\boldsymbol{\Omega}} = [(K\boldsymbol{r}_{p+1})(X_{b,1}), (K\boldsymbol{r}_{p+1})(X_{b,2}), \ldots, (K\boldsymbol{r}_{p+1})(X_{b,n})]$, exactly like $\boldsymbol{\Omega}$ but with b in place of h and $p+1$ in place of p
- $\bar{\boldsymbol{\Gamma}} = \frac{1}{nb}\sum_{i=1}^{n}(K\boldsymbol{r}_{p+1}\boldsymbol{r}_{p+1}')(X_{b,i})$, exactly like $\boldsymbol{\Gamma}$ but with b in place of h and $p+1$ in place of p, and
- $\bar{\boldsymbol{\Lambda}}_k = \bar{\boldsymbol{\Omega}}\left[X_{b,1}^{p+1+k}, \ldots, X_{b,n}^{p+1+k}\right]'/n$, exactly like $\boldsymbol{\Lambda}_k$ but with b in place of h and $p+1$ in place of p (implying $\bar{\boldsymbol{\Omega}}$ in place of $\boldsymbol{\Omega}$)
- $\hat{\mu}^{(\nu)} = \frac{1}{nh^\nu}\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\boldsymbol{Y}$

  $\hat{\theta}_{\mathtt{rbc}} = \hat{\mu}^{(\nu)} - h^{p+1-\nu}\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Lambda}_1\frac{\hat{\mu}^{(p+1)}}{(p+1)!} = \frac{1}{nh^\nu}\nu!\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}_{\mathtt{rbc}}\boldsymbol{Y}$
- $\boldsymbol{\Omega}_{\mathtt{rbc}} = \boldsymbol{\Omega} - \rho^{p+1}\boldsymbol{\Lambda}_1\boldsymbol{e}_{p+1}'\bar{\boldsymbol{\Gamma}}^{-1}\bar{\boldsymbol{\Omega}}$
- $\rho = h/b$, the ratio of the two bandwidth sequences
- $\Sigma = \operatorname{diag}(v(X_i) : i = 1, \ldots, n)$, with $v(x) = \mathbb{V}[Y|X = x]$
- $\sigma_p^2 = \nu!^2\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}(h\boldsymbol{\Omega}\boldsymbol{\Sigma}\boldsymbol{\Omega}'/n)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu$

  $\sigma_{\mathtt{rbc}}^2 = \nu!^2\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}(h\boldsymbol{\Omega}_{\mathtt{rbc}}\boldsymbol{\Sigma}\boldsymbol{\Omega}_{\mathtt{rbc}}'/n)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu$
- $\hat{\sigma}_p^2 = \nu!^2\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}(h\boldsymbol{\Omega}\hat{\boldsymbol{\Sigma}}_p\boldsymbol{\Omega}'/n)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu$

  $\hat{\sigma}_{\mathtt{rbc}}^2 = \nu!^2\boldsymbol{e}_\nu'\boldsymbol{\Gamma}^{-1}(h\boldsymbol{\Omega}_{\mathtt{rbc}}\hat{\boldsymbol{\Sigma}}_{\mathtt{rbc}}\boldsymbol{\Omega}_{\mathtt{rbc}}'/n)\boldsymbol{\Gamma}^{-1}\boldsymbol{e}_\nu$
- $\hat{\boldsymbol{\Sigma}}_p = \operatorname{diag}(\hat{v}(X_i) : i = 1, \ldots, n)$, with $\hat{v}(X_i) = (Y_i - \boldsymbol{r}_p(X_i - \mathsf{x})'\hat{\boldsymbol{\beta}}_p)^2$ for $\hat{\boldsymbol{\beta}}_p$ defined in Equation (S.4), and
- $\hat{\boldsymbol{\Sigma}}_{\mathtt{rbc}} = \operatorname{diag}(\hat{v}(X_i) : i = 1, \ldots, n)$, with $\hat{v}(X_i) = (Y_i - \boldsymbol{r}_{p+1}(X_i - \mathsf{x})'\hat{\boldsymbol{\beta}}_{p+1})^2$ for $\hat{\boldsymbol{\beta}}_{p+1}$ defined exactly as in Equation (S.4) but with $p+1$ in place of p and b in place of h.
- $T_p = \dfrac{\sqrt{nh^{1+2\nu}}(\hat{\mu}_p^{(\nu)} - \mu^{(\nu)})}{\hat{\sigma}_p}$

  $T_{\mathtt{rbc}} = \dfrac{(\hat{\theta}_{\mathtt{rbc}} - \mu^{(\nu)})}{\hat{\vartheta}_{\mathtt{rbc}}} = \dfrac{\sqrt{nh^{1+2\nu}}(\hat{\theta}_{\mathtt{rbc}} - \mu^{(\nu)})}{\hat{\sigma}_{\mathtt{rbc}}}$
- $I_p = \left[\hat{\mu}_p^{(\nu)} - z_u\hat{\sigma}_p/\sqrt{nh^{1+2\nu}}\ ,\ \hat{\mu}_p^{(\nu)} - z_l\hat{\sigma}_p/\sqrt{nh^{1+2\nu}}\right]$

  $I_{\mathtt{rbc}} = \left[\hat{\theta}_{\mathtt{rbc}} - z_u\hat{\vartheta}_{\mathtt{rbc}}\ ,\ \hat{\theta}_{\mathtt{rbc}} - z_l\hat{\vartheta}_{\mathtt{rbc}}\right] = \left[\hat{\theta}_{\mathtt{rbc}} - z_u\hat{\sigma}_{\mathtt{rbc}}/\sqrt{nh^{1+2\nu}}\ ,\ \hat{\theta}_{\mathtt{rbc}} - z_l\hat{\sigma}_{\mathtt{rbc}}/\sqrt{nh^{1+2\nu}}\right]$

Bias and the Role of Smoothness

- $\boldsymbol{\beta}_k$ (usually $k = p$ or $k = p+1$) as the $k+1$ vector with $(j+1)$ element equal to $\mu^{(j)}(\mathsf{x})/j!$ for $j = 0, 1, \ldots, k$ as long as $j \leq S$, and zero otherwise
- $\boldsymbol{B}_k$ as the n-vector with i^{th} entry $[\mu(X_i) - \boldsymbol{r}_k(X_i - \mathsf{x})'\boldsymbol{\beta}_k]$
- $\boldsymbol{M} = [\mu(X_1), \ldots, \mu(X_n)]'$
- $\rho = h/b$, the ratio of the two bandwidth sequences
- $\tilde{\boldsymbol{\Gamma}} = \mathbb{E}[\boldsymbol{\Gamma}]$, $\tilde{\bar{\boldsymbol{\Gamma}}} = \mathbb{E}[\bar{\boldsymbol{\Gamma}}]$, $\tilde{\boldsymbol{\Lambda}}_k = \mathbb{E}[\boldsymbol{\Lambda}_k]$, $\tilde{\bar{\boldsymbol{\Lambda}}}_k = \mathbb{E}[\bar{\boldsymbol{\Lambda}}_k]$, and so forth. A tilde always denotes a fixed-n expectation, and all expectations are fixed-n calculations unless explicitly denoted otherwise. The dependence on $\mathscr{F}_S$ is suppressed notationally.
- $\Psi_{T,F} = \Psi_{I,F}$, the fixed-n bias for interval I or t-statistic T. They are identical for all I and F, e.g., $\Psi_{\texttt{rbc},F} = \Psi_{I_{\texttt{rbc}},F} = \Psi_{T_{\texttt{rbc}},F}$. See Equation (S.15)
- $\psi_{T,F} = \psi_{I,F}$, the constant portion of the fixed-n bias for interval I or t-statistic T. They are identical for all I and F, e.g., $\psi_{\texttt{rbc},F} = \psi_{I_{\texttt{rbc}},F} = \psi_{T_{\texttt{rbc}},F}$. See Tables S.1 and S.2

Main Results and Proofs

- See Section S.3.1 for definitions of all terms in the Edgeworth expansion.
- C shall be a generic conformable constant that may take different values in different places. Note that C may be a vector or matrix but will generally not be denoted by a bold symbol. If more than one constant is needed, C_1, C_2, ..., will be used.
- **Norms.** Unless explicitly noted otherwise, $|\cdot|$ will be the Euclidean/Frobenius norm: for a scalar $c \in \mathbb{R}^1$, $|c|$ is the absolute value; for a vector $\boldsymbol{c}$, $|\boldsymbol{c}| = \sqrt{\boldsymbol{c}'\boldsymbol{c}}$; for a matrix $\boldsymbol{C}$, $|\boldsymbol{C}| = \sqrt{\text{trace}(\boldsymbol{C}'\boldsymbol{C})}$.
- $s_n = \sqrt{nh}$.
- $r_T = \max\{s_n^{-2}, \Psi_{T,F}^2, s_n^{-1}\Psi_{T,F}\}$, i.e. the slowest vanishing of the rates, and
- r_n as a generic sequence that obeys $r_n = o(r_T)$.

S.11 Supplement References

Bhattacharya, R. N. (1977), "Refinements of the Multidimensional Central Limit Theorem and Applications," *Annals of Probability*, 5, 1–27.

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

Calonico, S., Cattaneo, M. D., and Farrell, M. H. (2018a), "On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference," *Journal of the American Statistical Association*, 113, 767–779.

——— (2018b), "Supplement to 'On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference'," *Journal of the American Statistical Association*, 113, 767–779.

——— (2019), "`nprobust`: Nonparametric Kernel-Based Estimation and Robust Bias-Corrected Inference," *Journal of Statistical Software*, 91, 1–33.

Chen, S. X., and Qin, Y. S. (2002), "Confidence Intervals Based on Local Linear Smoother," *Scandinavian Journal of Statistics*, 29, 89–99.

Fan, J., and Gijbels, I. (1996), *Local Polynomial Modelling and its Applications*, London: Chapman and Hall.

Hall, P. (1991), "Edgeworth Expansions for Nonparametric Density Estimators, with Applications," *Statistics*, 22, 215–232.

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

——— (1992b), "Effect of Bias Estimation on Coverage Accuracy of Bootstrap Confidence Intervals for a Probability Density," *Annals of Statistics*, 20, 675–694.

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

Maesono, Y. (1997), "Edgeworth expansions of a studentized U-Statistic and a jackknife estimator of variance," *Journal of Statistical Planning and Inference*, 61, 61–84.

Muller, H.-G., and Stadtmuller, U. (1987), "Estimation of Heteroscedasticity in Regression Analysis," *Annals of Statistics*, 15, 610–625.

Romano, J. P. (2004), "On non-parametric testing, the uniform behaviour of the t-test, and related problems," *Scandinavian Journal of Statistics*, 31, 567–584.

Skovgaard, I. M. (1981), "Transformation of an Edgeworth Expansion by a Sequence of Smooth Functions," *Scandinavian Journal of Statistics*, 8, 207–217.

——— (1986), "On Multivariate Edgeworth Expansions," *International Statistical Review*, 54, 169–186.