Submitted to the Annals of Statistics

LARGE SAMPLE PROPERTIES OF PARTITIONING-BASED SERIES ESTIMATORS

SUPPLEMENTAL APPENDIX

By Matias D. Cattaneo*, Max H. Farrell and Yingjie Feng

Princeton University, University of Chicago, and Princeton University

This supplement gives omitted theoretical proofs of the results discussed in the main paper, additional technical results and methodological discussions, which may be of independent interest, and further simulation evidence. Section SA-1 repeats the setup and assumptions. Section SA-2 states a few important lemmas. Then, Section SA-3 provides an integrated mean squared error expansion for the case of general tensor-product partitions, building on the discussion in the main text. Section SA-4 proves pointwise and uniform stochastic linearizations. Section SA-5 states all uniform inference results, including two feasible inference methods omitted from the main text. Section SA-6 discusses leading examples of partitioning-based estimators, where the main assumptions used in the paper are verified for each case, including splines, wavelets, and piecewise polynomials. Some interesting conceptual and technical asides are given in Section SA-7. Finally, implementation and other numerical issues are discussed in Section SA-8, and complete results from a simulation study are presented in Section SA-9. See also the companion R package `lspartition` detailed in [8] and available at

http://sites.google.com/site/nppackages/lspartition/.

*Financial support from the National Science Foundation (SES 1459931) is gratefully acknowledged.

MSC 2010 subject classifications: Primary 62H10, 62M99, 57R12; secondary 62M99

Keywords and phrases: nonparametric regression, series methods, sieve methods, robust bias correction, uniform inference, strong approximation, tuning parameter selection

Contents

SA-1. Setup and Assumptions. For a d-tuple $\mathbf{q} = (q_1, \ldots, q_d) \in \mathbb{Z}^d_+$, define $[\mathbf{q}] = \sum_{\ell=1}^d q_\ell$, $\mathbf{x}^\mathbf{q} = x_1^{q_1} x_2^{q_2} \cdots x_d^{q_d}$, $\mathbf{q}! = q_1! \cdots q_d!$ and $\partial^\mathbf{q}\mu(\mathbf{x}) = \partial^{[\mathbf{q}]}\mu(\mathbf{x})/\partial x_1^{q_1} \ldots \partial x_d^{q_d}$. Unless explicitly stated otherwise, whenever $\mathbf{x}$ is a boundary point of some closed set, the partial derivative is understood as the limit with $\mathbf{x}$ ranging within it. Let $\mathbf{0} = (0, \cdots, 0)'$ and $\mathbf{1} = (1, \cdots, 1)'$ be the length-d vectors of zeros and ones respectively. The tensor product or Kronecker product operator is $\otimes$, and the entrywise division operator (Hadamard division) is $\oslash$. The smallest integer greater than or equal to u is $\lceil u \rceil$.

We use several norms. For a column vector $\mathbf{v} = (v_1, \ldots, v_M)' \in \mathbb{R}^M$, we let $\|\mathbf{v}\| = (\sum_{i=1}^M v_i^2)^{1/2}$, $\|\mathbf{v}\|_\infty = \max_{1\leq i\leq M} |v_i|$ and $\dim(\mathbf{v}) = M$. For a matrix $\mathbf{A} \in \mathbb{R}^{M\times N}$, $\|\mathbf{A}\|_1 = \max_{1\leq j\leq N} \sum_{i=1}^M |a_{ij}|$, $\|\mathbf{A}\| = \max_i \sigma_i(\mathbf{A})$ and $\|\mathbf{A}\|_\infty = \max_{1\leq i\leq M} \sum_{j=1}^N |a_{ij}|$ for operator norms induced by L_1, L_2, and L_∞ norms respectively, where $\sigma_i(\mathbf{A})$ is the i-th singular value of $\mathbf{A}$. $\lambda_{\max}(\mathbf{A})$ and $\lambda_{\min}(\mathbf{A})$ denote the maximum and minimum eigenvalues of $\mathbf{A}$ when $M = N$. For a real-valued function $g(\mathbf{x})$, $\|g\|_{L_p(\mathcal{X})} = (\int_\mathcal{X} |g(x)|^p \, d\mathbf{x})^{1/p}$ and $\|g\|_{L_\infty(\mathcal{X})} = \operatorname{ess\,sup}_{\mathbf{x}\in\mathcal{X}} |g(\mathbf{x})|$, and let $\mathcal{C}^s(\mathcal{X})$ denote the space of s-times continuously differentiable functions on $\mathcal{X}$.

We also use standard empirical process notation: $\mathbb{E}_n[g] = \mathbb{E}_n[g(\mathbf{x}_i)] = \frac{1}{n}\sum_{i=1}^n g(\mathbf{x}_i)$, $\mathbb{E}[g] = \mathbb{E}[g(\mathbf{x}_i)]$, and $\mathbb{G}_n[g] = \mathbb{G}_n[g(\mathbf{x}_i)] = \frac{1}{\sqrt{n}}\sum_{i=1}^n (g(\mathbf{x}_i) - \mathbb{E}[g(\mathbf{x}_i)])$. For sequences of numbers or random variables, we use $a_n \lesssim b_n$ to denote $\limsup_n |a_n/b_n|$ is finite, and $a_n \lesssim_\mathbb{P} b_n$ or $a_n = O_\mathbb{P}(b_n)$ to denote $\limsup_{\epsilon\to\infty} \limsup_n \mathbb{P}[|a_n/b_n| \geq \epsilon] = 0$. $a_n = o(b_n)$ implies $a_n/b_n \to 0$, and $a_n = o_\mathbb{P}(b_n)$ implies that $a_n/b_n \to_\mathbb{P} 0$ where $\to_\mathbb{P}$ denotes convergence in probability. $a_n \asymp b_n$ implies that $a_n \lesssim b_n$ and $b_n \lesssim a_n$. $\rightsquigarrow$ denotes convergence in distribution, and for two random variables X and Y, $X =_d Y$ implies that they have the same probability distribution.

We set $\mu(\mathbf{x}) := \partial^\mathbf{0}\mu(\mathbf{x})$ and likewise $\widehat{\mu}_j(\mathbf{x}) := \widehat{\partial^\mathbf{0}\mu_j}(\mathbf{x})$ for $j = 0, 1, 2, 3$. Let $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]'$. Finally, let $C, C_1, C_2, \ldots$ denote universal constants which may be different in different uses.

SA-1.1. *Setup.* As a complete reference, this section repeats the setup, assumptions, and notation used in the main paper.

We study the standard nonparametric regression setup, where $\{(y_i, \mathbf{x}_i'), i = 1, \ldots n\}$ is a random sample from the model

$$y_i = \mu(\mathbf{x}_i) + \varepsilon_i, \qquad \mathbb{E}[\varepsilon_i|\mathbf{x}_i] = 0, \qquad \mathbb{E}[\varepsilon_i^2|\mathbf{x}_i] = \sigma^2(\mathbf{x}_i), \tag{SA-1.1}$$

for a scalar response y_i and a d-vector of continuously distributed covariates $\mathbf{x}_i = (x_{1,i}, \ldots, x_{d,i})'$ with compact support $\mathcal{X}$. The object of interest is the unknown regression function $\mu(\cdot)$ and its derivatives.

We focus on *partitioning-based*, or locally-supported, series (linear sieve) least squares regression estimators. Let $\Delta = \{\delta_l : 1 \leq l \leq \bar{\kappa}\}$ be a collection of open and disjoint subsets of $\mathcal{X}$ such that the closure of their union is $\mathcal{X}$. Every $\delta_l \in \Delta$ is restricted to be polyhedral. An important special case is tensor-product partitions, in which case the support of the regressors is of tensor product form and each dimension of $\mathcal{X}$ is partitioned marginally into intervals.

Based on a particular partition, the dictionary of K basis functions, each of order m (e.g., $m = 4$ for cubic splines) is denoted by

$$\mathbf{x}_i \mapsto \mathbf{p}(\mathbf{x}_i) := \mathbf{p}(\mathbf{x}_i; \Delta, m) = (p_1(\mathbf{x}_i; \Delta, m), \cdots, p_K(\mathbf{x}_i; \Delta, m))'.$$

For a point $\mathbf{x} \in \mathcal{X}$ and $\mathbf{q} = (q_1, \ldots, q_d)' \in \mathbb{Z}_+^d$, the partial derivative $\partial^{\mathbf{q}}\mu(\mathbf{x})$ is estimated by least squares regression

$$\widehat{\partial^{\mathbf{q}}\mu}(\mathbf{x}) = \partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})'\widehat{\boldsymbol{\beta}}, \qquad \widehat{\boldsymbol{\beta}} \in \underset{\mathbf{b}\in\mathbb{R}^K}{\arg\min} \sum_{i=1}^{n}(y_i - \mathbf{p}(\mathbf{x}_i)'\mathbf{b})^2,$$

Our assumptions will guarantee that the sample matrix $\mathbb{E}_n[\mathbf{p}(\mathbf{x}_i)\mathbf{p}(\mathbf{x}_i)']$ is nonsingular with probability approaching one in large samples, and thus we write the estimator as

$$\begin{aligned} &\widehat{\partial^{\mathbf{q}}\mu}_0(\mathbf{x}) := \widehat{\boldsymbol{\gamma}}_{\mathbf{q},0}(\mathbf{x})'\mathbb{E}_n\left[\boldsymbol{\Pi}_0(\mathbf{x}_i)y_i\right], \quad \text{where} \\ &\widehat{\boldsymbol{\gamma}}_{\mathbf{q},0}(\mathbf{x})' := \partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})'\mathbb{E}_n[\mathbf{p}(\mathbf{x}_i)\mathbf{p}(\mathbf{x}_i)']^{-1} \quad \text{and} \quad \boldsymbol{\Pi}_0(\mathbf{x}_i) := \mathbf{p}(\mathbf{x}_i). \end{aligned} \tag{SA-1.2}$$

The order m of the basis is usually fixed in practice, and thus the tuning parameter for this class of nonparametric estimators is Δ. As $n \to \infty$, $\bar{\kappa} \to \infty$, and the volume of each δ_l shrinks proportionally to h^d, where $h = \max\{\mathrm{diam}(\delta) : \delta \in \Delta\}$.

SA-1.2. *Assumptions.* We list our main assumptions for completeness. Detailed discussion can be found in the main paper. The first assumption concerns the data generating process.

ASSUMPTION SA-1 (Data Generating Process).

(a) $\{(y_i, \mathbf{x}_i') : 1 \leq i \leq n\}$ *are i.i.d. satisfying (SA-1.1), where* $\mathbf{x}_i$ *has compact connected support* $\mathcal{X} \subset \mathbb{R}^d$ *and an absolutely continuous distribution function. The density of* $\mathbf{x}_i$, $f(\cdot)$, *and the conditional variance of* y_i *given* $\mathbf{x}_i$, $\sigma^2(\cdot)$, *are bounded away from zero and continuous.*

(b) $\mu(\cdot)$ *is* S*-times continuously differentiable, for* $S \geq [\mathbf{q}]$, *and all* $\partial^{\boldsymbol{\varsigma}}\mu(\cdot)$, $[\boldsymbol{\varsigma}] = S$, *are Hölder continuous with exponent* $\varrho > 0$.

ASSUMPTION SA-2 (Quasi-Uniform Partition). *The ratio of the sizes of inscribed and circumscribed balls of each* $\delta \in \Delta$ *is bounded away from zero uniformly in* $\delta \in \Delta$, *and*

$$\frac{\max\{\mathrm{diam}(\delta) : \delta \in \Delta\}}{\min\{\mathrm{diam}(\delta) : \delta \in \Delta\}} \lesssim 1,$$

where $\mathrm{diam}(\delta)$ *denotes the diameter of* δ. *Further, for* $h = \max\{\mathrm{diam}(\delta) : \delta \in \Delta\}$, *assume* $h = o(1)$.

The next assumption requires that the basis is *locally* supported. We employ the notion of *active basis*: a function $p(\cdot)$ on $\mathcal{X}$ is *active* on $\delta \in \Delta$ if it is not identically zero on δ.

Assumption SA-3 (Local Basis).

(a) *For each basis function p_k, $k = 1, \dots, K$, the union of elements of Δ on which p_k is active is a connected set, denoted by $\mathcal{H}_k$. For all $k = 1, \dots, K$, both the number of elements of $\mathcal{H}_k$ and the number of basis functions which are active on $\mathcal{H}_k$ are bounded by a constant.*

(b) *For any $\mathbf{a} = (a_1, \cdots, a_K)' \in \mathbb{R}^K$,*

$$\mathbf{a}' \int_{\mathcal{H}_k} \mathbf{p}(\mathbf{x}; \Delta, m)\mathbf{p}(\mathbf{x}; \Delta, m)' \, d\mathbf{x} \, \mathbf{a} \gtrsim a_k^2 h^d, \qquad k = 1, \dots, K.$$

(c) *Let $[\mathbf{q}] < m$. For an integer $\varsigma \in [[\mathbf{q}], m)$, for all $\boldsymbol{\varsigma}, [\boldsymbol{\varsigma}] \leq \varsigma$,*

$$h^{-[\boldsymbol{\varsigma}]} \lesssim \inf_{\delta \in \Delta} \inf_{\mathbf{x} \in \mathrm{clo}(\delta)} \|\partial^{\boldsymbol{\varsigma}} \mathbf{p}(\mathbf{x}; \Delta, m)\| \leq \sup_{\delta \in \Delta} \sup_{\mathbf{x} \in \mathrm{clo}(\delta)} \|\partial^{\boldsymbol{\varsigma}} \mathbf{p}(\mathbf{x}; \Delta, m)\| \lesssim h^{-[\boldsymbol{\varsigma}]}$$

where $\mathrm{clo}(\delta)$ is the closure of δ, and for $[\boldsymbol{\varsigma}] = \varsigma + 1$,

$$\sup_{\delta \in \Delta} \sup_{\mathbf{x} \in \mathrm{clo}(\delta)} \|\partial^{\boldsymbol{\varsigma}} \mathbf{p}(\mathbf{x}; \Delta, m)\| \lesssim h^{-\varsigma - 1}.$$

We remind readers that Assumption SA-2 and SA-3 implicitly relate the number of approximating series terms, the number of cells in Δ, and the maximum mesh size: $K \asymp \bar{\kappa} \asymp h^{-d}$.

The next assumption gives an explicit high-level expression of the leading approximation error which is needed for bias correction and integrated mean squared error (IMSE) expansion. To simplify notation, for each $\mathbf{x} \in \mathcal{X}$ define $\delta_{\mathbf{x}}$ as the element of Δ whose closure contains $\mathbf{x}$ and $h_{\mathbf{x}}$ denotes the diameter of this $\delta_{\mathbf{x}}$.

Assumption SA-4 (Approximation Error). *Let $S \geq m$. For all $\boldsymbol{\varsigma}$ satisfying $[\boldsymbol{\varsigma}] \leq \varsigma$ in Assumption SA-3, there exists $s^* \in \mathcal{S}_{\Delta, m}$, the linear span of $\mathbf{p}(\mathbf{x}; \Delta, m)$, and*

$$\mathscr{B}_{m, \boldsymbol{\varsigma}}(\mathbf{x}) = - \sum_{\boldsymbol{u} \in \Lambda_m} \partial^{\boldsymbol{u}} \mu(\mathbf{x}) h_{\mathbf{x}}^{m - [\boldsymbol{\varsigma}]} B_{\boldsymbol{u}, \boldsymbol{\varsigma}}(\mathbf{x})$$

such that

$$\sup_{\mathbf{x} \in \mathcal{X}} |\partial^{\boldsymbol{\varsigma}} \mu(\mathbf{x}) - \partial^{\boldsymbol{\varsigma}} s^*(\mathbf{x}) + \mathscr{B}_{m, \boldsymbol{\varsigma}}(\mathbf{x})| \lesssim h^{m + \varrho - [\boldsymbol{\varsigma}]} \tag{SA-1.3}$$

and

$$\sup_{\delta \in \Delta} \sup_{\mathbf{x}_1, \mathbf{x}_2 \in \mathrm{clo}(\delta)} \frac{|B_{\boldsymbol{u}, \boldsymbol{\varsigma}}(\mathbf{x}_1) - B_{\boldsymbol{u}, \boldsymbol{\varsigma}}(\mathbf{x}_2)|}{\|\mathbf{x}_1 - \mathbf{x}_2\|} \lesssim h^{-1} \tag{SA-1.4}$$

where $B_{\boldsymbol{u}, \boldsymbol{\varsigma}}(\cdot)$ is a known function that is bounded uniformly over n, and Λ_m is a multi-index set, which depends on the basis, with $[\mathbf{u}] = m$ for $\mathbf{u} \in \Lambda_m$.

Our last assumption concerns the basis used for bias correction. Specifically, for some $\tilde{m} > m$, let $\tilde{\mathbf{p}}(\mathbf{x}) := \tilde{\mathbf{p}}(\mathbf{x}; \tilde{\Delta}, \tilde{m})$ be a basis of order $\tilde{m}$ defined on a partition $\tilde{\Delta}$ with maximum mesh $\tilde{h}$. Objects accented with a tilde always pertain to this secondary basis and partition for bias correction. See the main text for complete discussion.

Assumption SA-5 (Bias Correction). *The partition $\tilde{\Delta}$ satisfies Assumption SA-2, with maximum mesh $\tilde{h}$, and the basis $\tilde{\mathbf{p}}(\mathbf{x}; \tilde{\Delta}, \tilde{m})$, $\tilde{m} > m$, satisfies Assumptions SA-3 and SA-4 with $\tilde{\varsigma} = \tilde{\varsigma}(\tilde{m}) \geq m$ in place of ς. Let $\rho := h/\tilde{h}$, which obeys $\rho \to \rho_0 \in (0, \infty)$. In addition, for $j = 3$, either (i) $\tilde{\mathbf{p}}(\mathbf{x}; \tilde{\Delta}, \tilde{m})$ spans a space containing the span of $\mathbf{p}(\mathbf{x}; \Delta, m)$, and for all $\boldsymbol{u} \in \Lambda_m$, $\partial^{\boldsymbol{u}}\mathbf{p}(\mathbf{x}; \Delta, m) = \mathbf{0}$; or (ii) both $\mathbf{p}(\mathbf{x}; \Delta, m)$ and $\tilde{\mathbf{p}}(\mathbf{x}; \tilde{\Delta}, \widetilde{m})$ reproduce polynomials of degree $[\mathbf{q}]$.*

SA-2. Technical Lemmas. We present a series of technical lemmas that will be used for pointwise and uniform analysis. To begin with, we introduce additional notation to simplify our expressions. The following are frequently used outer-product matrices:

$$\begin{aligned}
\mathbf{Q}_m &= \mathbb{E}[\mathbf{p}(\mathbf{x}_i)\mathbf{p}(\mathbf{x}_i)'], & \widehat{\mathbf{Q}}_m &= \mathbb{E}_n[\mathbf{p}(\mathbf{x}_i)\mathbf{p}(\mathbf{x}_i)'],\\
\mathbf{Q}_{\tilde{m}} &= \mathbb{E}[\tilde{\mathbf{p}}(\mathbf{x}_i)\tilde{\mathbf{p}}(\mathbf{x}_i)'], & \widehat{\mathbf{Q}}_{\tilde{m}} &= \mathbb{E}_n[\tilde{\mathbf{p}}(\mathbf{x}_i)\tilde{\mathbf{p}}(\mathbf{x}_i)'],\\
\mathbf{Q}_{m,\tilde{m}} &= \mathbb{E}[\mathbf{p}(\mathbf{x}_i)\tilde{\mathbf{p}}(\mathbf{x}_i)'], & \widehat{\mathbf{Q}}_{m,\tilde{m}} &= \mathbb{E}_n[\mathbf{p}(\mathbf{x}_i)\tilde{\mathbf{p}}(\mathbf{x}_i)'].
\end{aligned}$$

In the main paper, we define four partitioning-based nonparametric estimators, numbered as $j = 0, 1, 2, 3$, which can be written in exactly the same form:

$$\widehat{\partial^{\mathbf{q}}\mu}_j(\mathbf{x}) := \widehat{\boldsymbol{\gamma}}_{\mathbf{q},j}(\mathbf{x})'\mathbb{E}_n[\boldsymbol{\Pi}_j(\mathbf{x}_i)y_i].$$

We summarize the definitions of these quantities here. For $j = 0$ (classical estimator without bias correction),

$$\widehat{\boldsymbol{\gamma}}_{\mathbf{q},0}(\mathbf{x})' := \partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})'\widehat{\mathbf{Q}}_m^{-1} \quad \text{and} \quad \boldsymbol{\Pi}_0(\mathbf{x}_i) := \mathbf{p}(\mathbf{x}_i). \tag{SA-2.1}$$

For $j = 1$ (high-order-basis bias correction),

$$\widehat{\boldsymbol{\gamma}}_{\mathbf{q},1}(\mathbf{x})' := \partial^{\mathbf{q}}\tilde{\mathbf{p}}(\mathbf{x})'\widehat{\mathbf{Q}}_{\tilde{m}}^{-1}, \quad \text{and} \quad \boldsymbol{\Pi}_1(\mathbf{x}_i) := \tilde{\mathbf{p}}(\mathbf{x}_i). \tag{SA-2.2}$$

For $j = 2$ (least squares bias correction),

$$\begin{aligned}
\widehat{\boldsymbol{\gamma}}_{\mathbf{q},2}(\mathbf{x})' &:= \Big(\widehat{\boldsymbol{\gamma}}_{\mathbf{q},0}(\mathbf{x})', -\widehat{\boldsymbol{\gamma}}_{\mathbf{q},0}(\mathbf{x})'\widehat{\mathbf{Q}}_{m,\tilde{m}}\widehat{\mathbf{Q}}_{\tilde{m}}^{-1} + \widehat{\boldsymbol{\gamma}}_{\mathbf{q},1}(\mathbf{x})'\Big)\\
&\text{and} \quad \boldsymbol{\Pi}_2(\mathbf{x}_i) := \big(\mathbf{p}(\mathbf{x}_i)', \tilde{\mathbf{p}}(\mathbf{x}_i)'\big)'.
\end{aligned} \tag{SA-2.3}$$

For $j = 3$ (plug-in bias correction),

$$\begin{aligned}
\widehat{\boldsymbol{\gamma}}_{\mathbf{q},3}(\mathbf{x})' = \Big(\widehat{\boldsymbol{\gamma}}_{\mathbf{q},0}(\mathbf{x})', \sum_{\boldsymbol{u}\in\Lambda_m} \Big\{&\widehat{\boldsymbol{\gamma}}_{\boldsymbol{u},1}(\mathbf{x})'h_{\mathbf{x}}^{m-[\mathbf{q}]}B_{\boldsymbol{u},\mathbf{q}}(\mathbf{x})\\
&- \widehat{\boldsymbol{\gamma}}_{\mathbf{q},0}'\mathbb{E}_n[\mathbf{p}(\mathbf{x}_i)h_{\mathbf{x}_i}^m B_{\boldsymbol{u},\mathbf{0}}(\mathbf{x}_i)\widehat{\boldsymbol{\gamma}}_{\boldsymbol{u},1}(\mathbf{x}_i)']\Big\}\Big),\\
\text{and} \qquad \boldsymbol{\Pi}_3(\mathbf{x}_i) := \big(\mathbf{p}(\mathbf{x}_i)', \tilde{\mathbf{p}}(\mathbf{x}_i)'\big)'.&
\end{aligned} \tag{SA-2.4}$$

For $j = 0, 1, 2, 3$, $\boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})$ is defined as $\widehat{\boldsymbol{\gamma}}_{\mathbf{q},j}(\mathbf{x})$ in (SA-2.1), (SA-2.2), (SA-2.3) and (SA-2.4) but with sample averages replaced by their population counterparts. Finally, define

$$\begin{aligned}
&\boldsymbol{\Sigma}_j = \mathbb{E}[\boldsymbol{\Pi}_j(\mathbf{x}_i)\boldsymbol{\Pi}_j(\mathbf{x}_i)'\sigma^2(\mathbf{x}_i)], && \bar{\boldsymbol{\Sigma}}_j = \mathbb{E}_n[\boldsymbol{\Pi}_j(\mathbf{x}_i)\boldsymbol{\Pi}_j(\mathbf{x}_i)'\sigma^2(\mathbf{x}_i)] \\
&\widehat{\boldsymbol{\Sigma}}_j = \mathbb{E}_n[\boldsymbol{\Pi}_j(\mathbf{x}_i)\boldsymbol{\Pi}_j(\mathbf{x}_i)'\widehat{\varepsilon}^2_{i,j}], && \widehat{\varepsilon}_{i,j} = y_i - \widehat{\mu}_j(\mathbf{x}_i), \\
&\Omega_j(\mathbf{x}) = \boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})'\boldsymbol{\Sigma}_j\boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x}), \quad \text{and} && \widehat{\Omega}_j(\mathbf{x}) = \widehat{\boldsymbol{\gamma}}_{\mathbf{q},j}(\mathbf{x})'\widehat{\boldsymbol{\Sigma}}_j\widehat{\boldsymbol{\gamma}}_{\mathbf{q},j}(\mathbf{x}).
\end{aligned} \tag{SA-2.5}$$

The next lemma establishes the rate of convergence and boundedness for $\widehat{\mathbf{Q}}_m$ and other relevant matrices, which will be crucial for both pointwise and uniform analysis. Since the orders of bases used to construct bias corrected estimators are fixed and mesh size ratio $\rho \to \rho_0 \in (0, \infty)$ by Assumption SA-5, the same conclusions also apply to $\widehat{\mathbf{Q}}_{\tilde{m}}$, $\mathbf{Q}_{\tilde{m}}$ and other related quantities, though we do not make it explicit in the statement.

Lemma SA-2.1. *Under Assumptions SA-1–SA-3,*

$$h^d \lesssim \lambda_{\min}(\mathbf{Q}_m) \leq \lambda_{\max}(\mathbf{Q}_m) \lesssim h^d, \quad \|\mathbf{Q}_m^{-1}\|_\infty \lesssim h^{-d}.$$

Moreover, when $\frac{\log n}{nh^d} = o(1)$, *we have*

$$\begin{aligned}
&\|\widehat{\mathbf{Q}}_m - \mathbf{Q}_m\|_\infty \lesssim_{\mathbb{P}} h^d\sqrt{\frac{\log n}{nh^d}} = o_{\mathbb{P}}(h^d), \quad \|\widehat{\mathbf{Q}}_m - \mathbf{Q}_m\| \lesssim_{\mathbb{P}} h^d\sqrt{\frac{\log n}{nh^d}} = o_{\mathbb{P}}(h^d), \\
&\|\widehat{\mathbf{Q}}_m\| \lesssim_{\mathbb{P}} h^d, \quad \|\widehat{\mathbf{Q}}_m^{-1}\| \lesssim_{\mathbb{P}} h^{-d}, \quad \|\widehat{\mathbf{Q}}_m^{-1}\|_\infty \lesssim_{\mathbb{P}} h^{-d}, \\
&\|\widehat{\mathbf{Q}}_m^{-1} - \mathbf{Q}_m^{-1}\|_\infty \lesssim_{\mathbb{P}} h^{-d}\sqrt{\frac{\log n}{nh^d}} = o_{\mathbb{P}}(h^{-d}), \\
&\|\widehat{\mathbf{Q}}_m^{-1} - \mathbf{Q}_m^{-1}\| \lesssim_{\mathbb{P}} h^{-d}\sqrt{\frac{\log n}{nh^d}} = o_{\mathbb{P}}(h^{-d}), \quad \textit{and} \\
&\|\bar{\boldsymbol{\Sigma}}_0 - \boldsymbol{\Sigma}_0\| \lesssim_{\mathbb{P}} h^d\sqrt{\frac{\log n}{nh^d}} = o_{\mathbb{P}}(h^d).
\end{aligned}$$

The next lemma shows that the asymptotic error expansion specified in Assumption SA-4 translates into the bias expansion for the classical point estimator $\widehat{\partial^{\mathbf{q}}\mu}_0(\cdot)$, which is (partly) reported in Lemma 3.1 of the main paper. Importantly, the following lemma also gives a *uniform* bound on the conditional bias which is crucial for least squares bias correction ($j = 2$) and uniform inference.

Lemma SA-2.2 (Conditional Bias). *Let Assumptions SA-1–SA-4 hold. If* $\frac{\log n}{nh^d} = o(1)$, *then,*

$$\mathbb{E}[\widehat{\partial^{\mathbf{q}}\mu}_0(\mathbf{x})|\mathbf{X}] - \partial^{\mathbf{q}}\mu(\mathbf{x}) = \mathscr{B}_{m,\mathbf{q}}(\mathbf{x}) - \widehat{\boldsymbol{\gamma}}_{\mathbf{q},0}(\mathbf{x})'\mathbb{E}_n[\boldsymbol{\Pi}_0(\mathbf{x}_i)\mathscr{B}_{m,\mathbf{0}}(\mathbf{x}_i)] + O_{\mathbb{P}}(h^{m+\varrho-[\mathbf{q}]}),$$

and

$$\sup_{\mathbf{x}\in\mathcal{X}}\left|\mathbb{E}[\widehat{\partial^{\mathbf{q}}\mu}_0(\mathbf{x})|\mathbf{X}]-\partial^{\mathbf{q}}\mu(\mathbf{x})\right|\lesssim_{\mathbb{P}} h^{m-[\mathbf{q}]}.$$

Moreover, if the following condition is also satisfied,

$$\max_{1\leq k\leq K}\int_{\mathcal{H}_k} p_k(\mathbf{x};\Delta,m)\mathscr{B}_{m,\mathbf{0}}(\mathbf{x})\,d\mathbf{x}=o(h^{m+d}), \tag{SA-2.6}$$

then $\|\widehat{\boldsymbol{\gamma}}_{\mathbf{q},0}(\mathbf{x})'\mathbb{E}_n[\mathbf{\Pi}_0(\mathbf{x}_i)\mathscr{B}_{m,\mathbf{0}}(\mathbf{x}_i)]\|_{L_\infty(\mathcal{X})}=o_{\mathbb{P}}(h^{m-[\mathbf{q}]})$.

Equation (SA-2.6) implies that the leading approximation error is approximately orthogonal to $\mathbf{p}(\cdot)$, in which case the L_∞ approximation error coincides with the leading smoothing bias of $\widehat{\partial^{\mathbf{q}}\mu}_0(\cdot)$.

The next lemma shows that $\widehat{\boldsymbol{\gamma}}_{\mathbf{q},j}(\mathbf{x})$ in (SA-2.1), (SA-2.2), (SA-2.3), and (SA-2.4) converge to their population counterpart $\boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})$ in a proper sense, and all the $\boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})$ are sufficiently controlled in terms of both L_∞- and L_2-operator norms.

LEMMA SA-2.3. *Let Assumptions SA-1, SA-2, SA-3, and SA-5 hold. If* $\frac{\log n}{nh^d}=o(1)$*, then,*

$$\sup_{\mathbf{x}\in\mathcal{X}}\|\boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})'\|_\infty\lesssim h^{-d-[\mathbf{q}]},\qquad \sup_{\mathbf{x}\in\mathcal{X}}\|\widehat{\boldsymbol{\gamma}}_{\mathbf{q},j}(\mathbf{x})'-\boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})'\|_\infty\lesssim_{\mathbb{P}} h^{-d-[\mathbf{q}]}\sqrt{\frac{\log n}{nh^d}},$$
$$\sup_{\mathbf{x}\in\mathcal{X}}\|\boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})'\|\lesssim h^{-d-[\mathbf{q}]},\qquad \inf_{\mathbf{x}\in\mathcal{X}}\|\boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})'\|\gtrsim h^{-d-[\mathbf{q}]},\qquad \textit{and}$$
$$\sup_{\mathbf{x}\in\mathcal{X}}\|\widehat{\boldsymbol{\gamma}}_{\mathbf{q},j}(\mathbf{x})'-\boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})'\|\lesssim_{\mathbb{P}} h^{-d-[\mathbf{q}]}\sqrt{\frac{\log n}{nh^d}}.$$

The last lemma in this section proves that the asymptotic variance of classical and bias-corrected estimators is properly bounded. Importantly, we need not only an upper bound, but also a lower bound since the variance term appears in the denominator of t-statistics.

LEMMA SA-2.4. *Let Assumptions SA-1, SA-2, SA-3, and SA-5 hold. If* $\frac{\log n}{nh^d}=o(1)$*,*

$$\sup_{\mathbf{x}\in\mathcal{X}}\Omega_j(\mathbf{x})\lesssim h^{-d-2[\mathbf{q}]}\quad\textit{and}\quad\inf_{\mathbf{x}\in\mathcal{X}}\Omega_j(\mathbf{x})\gtrsim h^{-d-2[\mathbf{q}]}.$$

This lemma also implies that bias correction does not change the asymptotic order of the variance.

SA-3. Integrated Mean Squared Error. We now state an integrated mean squared error (IMSE) result that specializes Theorem 4.1 in the main paper to the case of general tensor-product partitions but is less restricted than Theorem 4.2. In particular, we allow for any $\mathbf{q}$ and do not rely on Assumption 6(c). This result gives the key ingredient for characterizing the limit $\mathscr{V}_{\Delta,\mathbf{q}}$ and $\mathscr{B}_{\Delta,\mathbf{q}}$ in some generality.

To be specific, throughout this section, we assume the support of the regressors is of tensor-product form, that each dimension of $\mathcal{X}$ is partitioned marginally into intervals, and Δ is the tensor product of these intervals. Let $\mathcal{X}_\ell = [\underline{x}_\ell, \overline{x}_\ell]$ be the support of covariate $\ell = 1, 2, \ldots, d$, and partition this into κ_ℓ disjoint subintervals defined by $\{\underline{x}_\ell = t_{\ell,0} < t_{\ell,1} < \cdots < t_{\ell,\kappa_\ell - 1} < t_{\ell,\kappa_\ell} = \overline{x}_\ell\}$. If this partition of $\mathcal{X}_\ell$ is Δ_ℓ, then a complete partition of $\mathcal{X}$ can be formed by tensor products of the one-dimensional partitions: $\Delta = \otimes_{\ell=1}^d \Delta_\ell$, with $\boldsymbol{\kappa} = (\kappa_1, \kappa_2, \ldots, \kappa_d)'$ collecting the number of subintervals in each dimension of $\mathbf{x}_i$, and $\bar{\kappa} = \kappa_1 \kappa_2 \cdots \kappa_d$. A generic cell of this partition is the rectangle

$$\delta_{l_1 \ldots l_d} = \{\mathbf{x} : t_{\ell,l_\ell} < x_\ell < t_{\ell,l_\ell+1}, \quad 0 \le l_\ell \le \kappa_\ell - 1 \quad 1 \le \ell \le d\}. \tag{SA-3.1}$$

Assumption SA-2 can be verified by choosing the knot positions appropriately, often dividing $\mathcal{X}_\ell$ uniformly or by quantiles.

As before, $\delta_{\mathbf{x}}$ is the subrectangle in Δ containing $\mathbf{x}$, and we write $\mathbf{b}_{\mathbf{x}}$ for the vector collecting the interval lengths of $\delta_{\mathbf{x}}$ (as defined in Assumption 6). $\mathbf{t}_{\mathbf{x}}^L$ denotes the start point of $\delta_{\mathbf{x}}$. For $\ell = 1, \ldots, d$, for a generic cell $\delta_{l_1 \ldots l_d}$ as in (SA-3.1), we write $b_{\ell,l_\ell} = t_{\ell,l_\ell+1} - t_{\ell,l_\ell}$, $l_\ell = 0, \ldots, \kappa_\ell - 1$, $b_\ell = \max_{0 \le l_\ell \le \kappa_\ell - 1} b_{\ell,l_\ell}$ and $b = \max_{1 \le \ell \le d} b_\ell$ ($b \asymp h$ by Assumption SA-2).

Theorem SA-3.1 (IMSE for Tensor-Product Partitions). *Suppose that the conditions in Theorem 4.1 and Assumptions 6(a)–(b) hold. Then, for any arbitrary sequence of points $\{\boldsymbol{\tau}_k\}_{k=1}^K$ such that $\boldsymbol{\tau}_k \in \operatorname{supp}(p_k(\cdot))$ for each $k = 1, \ldots, K$,*

$$\mathscr{V}_{\Delta,\mathbf{q}} = \mathscr{V}_{\boldsymbol{\kappa},\mathbf{q}} + o(h^{-d-2[\mathbf{q}]}) \qquad \textit{and} \qquad \mathscr{B}_{\Delta,\mathbf{q}} = \mathscr{B}_{\boldsymbol{\kappa},\mathbf{q}} + o(h^{2m-2[\mathbf{q}]}), \quad \textit{with}$$

$$\mathscr{V}_{\boldsymbol{\kappa},\mathbf{q}} = \boldsymbol{\kappa}^{\mathbf{1}+2\mathbf{q}} \sum_{k=1}^K \Big[\frac{\sigma^2(\boldsymbol{\tau}_k) w(\boldsymbol{\tau}_k)}{f(\boldsymbol{\tau}_k)} \prod_{\ell=1}^d g_\ell(\boldsymbol{\tau}_k)\Big] \mathbf{e}_k' \mathbf{H}_{\mathbf{0}}^{-1} \mathbf{H}_{\mathbf{q}} \mathbf{e}_k \operatorname{vol}(\delta_{\boldsymbol{\tau}_k}) \asymp h^{-d-2[\mathbf{q}]},$$

$$\mathscr{B}_{\boldsymbol{\kappa},\mathbf{q}} = \sum_{\boldsymbol{u}_1, \boldsymbol{u}_2 \in \Lambda_m} \boldsymbol{\kappa}^{-(\boldsymbol{u}_1 + \boldsymbol{u}_2 - 2\mathbf{q})} \Big\{\mathscr{B}_{\boldsymbol{u}_1,\boldsymbol{u}_2,\mathbf{q}} + \mathbf{v}_{\boldsymbol{u}_1,\mathbf{0}}' \mathbf{H}_{\mathbf{0}}^{-1} \mathbf{H}_{\mathbf{q}} \mathbf{H}_{\mathbf{0}}^{-1} \mathbf{v}_{\boldsymbol{u}_2,\mathbf{0}} - 2 \mathbf{v}_{\boldsymbol{u}_1,\mathbf{q}}' \mathbf{H}_{\mathbf{0}}^{-1} \mathbf{v}_{\boldsymbol{u}_2,\mathbf{0}}\Big\} \lesssim h^{2m-2[\mathbf{q}]},$$

where the K-dimensional vector $\mathbf{e}_k$ is the k-th unit vector (i.e., $\mathbf{e}_k$ has a 1 in the k-th position and 0 elsewhere), $\mathbf{g}(\mathbf{x}) = (g_1(\mathbf{x}), \ldots, g_d(\mathbf{x}))'$ is discussed below,

$$\mathbf{H}_{\mathbf{q}} = \boldsymbol{\kappa}^{-2\mathbf{q}} \int_{[0,1]^d} \partial^{\mathbf{q}} \mathbf{p}(\mathbf{x}) \partial^{\mathbf{q}} \mathbf{p}(\mathbf{x})' d\mathbf{x},$$

$$\mathscr{B}_{\boldsymbol{u}_1,\boldsymbol{u}_2,\mathbf{q}} = \eta_{\boldsymbol{u}_1,\boldsymbol{u}_2,\mathbf{q}} \int_{[0,1]^d} \frac{\partial^{\boldsymbol{u}_1} \mu(\mathbf{x}) \partial^{\boldsymbol{u}_2} \mu(\mathbf{x})}{\mathbf{g}(\mathbf{x})^{\boldsymbol{u}_1 + \boldsymbol{u}_2 - 2\mathbf{q}}} w(\mathbf{x}) d\mathbf{x},$$

and the K-dimensional vector $\mathbf{v}_{\mathbf{u},\mathbf{q}}$ has k-th typical element

$$\frac{\sqrt{w(\boldsymbol{\tau}_k)} \partial^{\boldsymbol{u}} \mu(\boldsymbol{\tau}_k)}{\boldsymbol{\kappa}^{\mathbf{q}} \mathbf{g}(\boldsymbol{\tau}_k)^{\boldsymbol{u}-\mathbf{q}}} \int_{\mathcal{X}} \frac{h_{\mathbf{x}}^{m-[\mathbf{q}]}}{\mathbf{b}_{\mathbf{x}}^{\boldsymbol{u}-\mathbf{q}}} \partial^{\mathbf{q}} p_k(\mathbf{x}) B_{\boldsymbol{u},\mathbf{q}}(\mathbf{x}) d\mathbf{x}.$$

This theorem takes advantage of the assumed tensor-product structure of the partition Δ to express the leading bias and variance as proportional to the number of subintervals used for each regressor. The accompanying constants are expressed as sums over the local contributions of the basis function $\mathbf{p}(\cdot)$ used, and are easy to be shown bounded in general (they may still not converge to any well-defined limit at this level of generality).

Assumption 6(a) used in Theorem SA-3.1 (and in Theorem 4.2) slightly strengthens the quasi-uniform condition imposed in Assumption SA-2. It can be verified by choosing knot positions appropriately. Specifically, let Δ be a tensor-product partition (see Equation (SA-3.1)) with (marginal) knots chosen as

$$t_{\ell,l} = G_\ell^{-1}\left(\frac{l}{\kappa_\ell}\right), \qquad l = 0, 1, \dots, \kappa_\ell, \qquad \ell = 1, 2, \dots, d,$$

where $G_\ell(\cdot)$ is a univariate continuously differentiable distribution function and $G_\ell^{-1}(v) = \inf\{x \in \mathbb{R} : G_\ell(x) \geq v\}$. In this case the function $g_\ell(\cdot)$ in Assumption 6(a) is simply the density of $G_\ell(\cdot)$. Two examples commonly used in practice are: (i) evenly-spaced partitions, denoted by $\Delta_{\texttt{ES}}$, where $G_\ell(x) = x$ and $g_\ell(x) = 1$, and (ii) quantile-spaced partitions, denoted by $\Delta_{\texttt{QS}}$, where $G_\ell(x) = \widehat{F}_\ell(x)$ with $\widehat{F}_\ell(x)$ the empirical distribution function for the ℓ-th covariate, $\ell = 1, 2, \dots, d$. For the case of quantile-spaced partitioning, if $\widehat{F}_\ell(x)$ convergences to $F_\ell(x) = \mathbb{P}[x_\ell \leq x]$ in a suitable sense, $g_\ell(x) = dF_\ell(x)/dx$, i.e., the marginal density of x_ℓ. See [1] and [2] for a slightly more high-level condition in the context of univariate B-splines.

Assumption 6(b) in Theorem SA-3.1 (and in Theorem 4.2) involves a scaling factor $\frac{h_{\mathbf{x}}^{2m-2[\mathbf{q}]}}{\mathbf{b}_{\mathbf{x}}^{\boldsymbol{u}_1+\boldsymbol{u}_2-2[\mathbf{q}]}}$. In Section SA-1 we do not add specific restrictions on the shape of cells in Δ and thus the diameter of a cell is used to conveniently express the order of approximation error (denoted by h), but for tensor-product partitions the approximation error is usually characterized by the lengths of intervals on each axis (denoted by $\mathbf{b}_{\mathbf{x}}$), as illustrated in Section SA-6. Therefore, the scaling factor makes the above theorem immediately apply to the three leading examples in Section SA-6 based on tensor-product partitioning schemes (splines, wavelets, and piecewise polynomials).

Using Theorem SA-3.1, the IMSE-optimal tuning parameter selector for partitioning-based series estimators becomes

$$\boldsymbol{\kappa}_{\texttt{IMSE},\mathbf{q}} = \operatorname*{arg\,min}_{\boldsymbol{\kappa}\in\mathbb{Z}_{++}^d} \left\{\frac{1}{n}\mathscr{V}_{\boldsymbol{\kappa},\mathbf{q}} + \mathscr{B}_{\boldsymbol{\kappa},\mathbf{q}}\right\},$$

which is still given in implicit form because the leading constants are not known to converge at this level of generality. It follows that $\boldsymbol{\kappa}_{\texttt{IMSE},\mathbf{q}} = (\kappa_{\texttt{IMSE},\mathbf{q},1}, \dots, \kappa_{\texttt{IMSE},\mathbf{q},d})'$ with $\kappa_{\texttt{IMSE},\mathbf{q},\ell} \asymp n^{\frac{1}{2m+d}}$, $\ell = 1, 2, \cdots, d$.

SA-4. Rates of Convergence. In this section, we discuss rates of convergence for the classical estimator and three bias-corrected estimators. Some results, such as the L_2 and uniform convergence rates for $\widehat{\partial^{\mathbf{q}}\mu}_0(\mathbf{x})$, are also reported in Theorem 4.3 of the main paper.

To begin with, we first report pointwise linearization which is an important intermediate step towards pointwise asymptotic normality.

LEMMA SA-4.1 (Pointwise Linearization). *Let Assumptions SA-1–SA-5 hold. If* $\frac{\log n}{nh^d} = o(1)$*, then for each* $\mathbf{x} \in \mathcal{X}$ *and* $j = 0, 1, 2, 3$*,*

$$\widehat{\partial^{\mathbf{q}}\mu}_j(\mathbf{x}) - \partial^{\mathbf{q}}\mu(\mathbf{x}) = \boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})'\mathbb{E}_n[\boldsymbol{\Pi}_j(\mathbf{x}_i)\varepsilon_i] + R_{1n,\mathbf{q}}(\mathbf{x}) + R_{2n,\mathbf{q}}(\mathbf{x}), \quad \textit{where}$$

$$R_{1n,\mathbf{q}}(\mathbf{x}) := \left(\widehat{\boldsymbol{\gamma}}_{\mathbf{q},j}(\mathbf{x}) - \boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})'\right)\mathbb{E}_n[\boldsymbol{\Pi}_j(\mathbf{x}_i)\varepsilon_i] \lesssim_{\mathbb{P}} \frac{\sqrt{\log n}}{nh^{d+[\mathbf{q}]}},$$

$$R_{2n,\mathbf{q}}(\mathbf{x}) := \mathbb{E}[\widehat{\partial^{\mathbf{q}}\mu}_j(\mathbf{x})|\mathbf{X}] - \partial^{\mathbf{q}}\mu(\mathbf{x}) \lesssim_{\mathbb{P}} h^{m-[\mathbf{q}]}.$$

Furthermore, for $j = 1, 2, 3$*,* $R_{2n,\mathbf{q}}(\mathbf{x}) \lesssim_{\mathbb{P}} h^{m+\varrho-[\mathbf{q}]}$*.*

We can use Lemma SA-4.1 to establish the limiting distribution of the standardized t-statistics as given in Theorem 5.1 of the main paper. The proof is available in Section SA-10.9.

REMARK SA-4.1. The term denoted by $R_{2n,\mathbf{q}}(\mathbf{x})$ captures the conditional bias of the point estimator. We establish a sharp bound on it by applying Bernstein's maximal inequality to control the *largest* element of $\mathbb{E}_n[\boldsymbol{\Pi}_0(\mathbf{x}_i)\mathscr{B}_{m,\mathbf{0}}(\mathbf{x}_i)]$ and employing the bound on the *uniform norm* of $\widehat{\mathbf{Q}}_m^{-1}$ derived in Lemma SA-2.1. This result improves on previous results in the literature and, in particular, confirms a conjecture posed by [4, Comment 4.2(ii), p. 352], for partitioning-based series estimation in general. To be precise, in their setup, $\ell_k c_k$ can be understood as a uniform bound on the L_2 approximation error (orthogonal to the approximating basis). Thus, using our results, we obtain

$$\mathbf{p}(\mathbf{x})'(\widehat{\mathbf{Q}}^{-1} - \mathbf{Q}^{-1})\mathbb{G}_n[\mathbf{p}(\mathbf{x}_i)\ell_k c_k] \lesssim_{\mathbb{P}} \sqrt{\frac{\log n}{nh^d}} \cdot \sqrt{\frac{\log n}{h^d}}\ell_k c_k,$$

which coincides with Equation (4.15) of [4], up to a normalization, thereby improving on the approximation established in their Equation (4.12). See our proof of Lemmas SA-2.2 and SA-4.1 for more details. ♦

For uniform convergence, we need to upgrade the above lemma to a uniform version at the cost of some stronger conditions:

LEMMA SA-4.2 (Uniform Linearization). *Let Assumptions SA-1–SA-5 hold. For each* $j = 0, 1, 2, 3$*, define* $R_{1n,\mathbf{q}}(\mathbf{x})$ *and* $R_{2n,\mathbf{q}}(\mathbf{x})$ *as in Lemma SA-4.1. Assume one of the following holds:*

(i) $\mathbb{E}[|\varepsilon_i|^{2+\nu}] < \infty$ *for some* $\nu > 0$*, and* $\frac{n^{\frac{2}{2+\nu}}(\log n)^{\frac{2\nu}{4+2\nu}}}{nh^d} \lesssim 1$*; or*

(ii) $\mathbb{E}[|\varepsilon_i|^3 \exp(|\varepsilon_i|)] < \infty$*, and* $\frac{(\log n)^3}{nh^d} \lesssim 1$*.*

Then, $\sup_{\mathbf{x}\in\mathcal{X}} |R_{1n,\mathbf{q}}(\mathbf{x})| \lesssim_{\mathbb{P}} \frac{\log n}{nh^{d+[\mathbf{q}]}} =: \bar{R}_{1n,\mathbf{q}}$*. For* $j = 0$*,* $\sup_{\mathbf{x}\in\mathcal{X}} |R_{2n,\mathbf{q}}(\mathbf{x})| \lesssim_{\mathbb{P}} h^{m-[\mathbf{q}]} =: \bar{R}_{2n,\mathbf{q}}$*, and for* $j = 1, 2, 3$*,* $\sup_{\mathbf{x}\in\mathcal{X}} |R_{2n,\mathbf{q}}(\mathbf{x})| \lesssim_{\mathbb{P}} h^{m+\varrho-[\mathbf{q}]} =: \bar{R}_{2n,\mathbf{q}}$*.*

Given Lemma SA-4.1 and SA-4.2, we are ready to derive the desired rates of convergence.

THEOREM SA-4.1 (Convergence Rates). *Let Assumption SA-1, SA-2, and SA-3 hold. Also assume that* $\sup_{\mathbf{x}\in\mathcal{X}} |\partial^{\mathbf{q}}\mu(\mathbf{x}) - \partial^{\mathbf{q}}s^*(\mathbf{x})| \lesssim h^{m-[\mathbf{q}]}$ *with* s^* *defined in Assumption SA-4. Then, if* $\frac{\log n}{nh^d} = o(1)$*,*

$$\int_{\mathcal{X}} \left(\widehat{\partial^{\mathbf{q}}\mu}_0(\mathbf{x}) - \partial^{\mathbf{q}}\mu(\mathbf{x})\right)^2 w(\mathbf{x})d\mathbf{x} \lesssim_{\mathbb{P}} \frac{1}{nh^{d+2[\mathbf{q}]}} + h^{2(m-[\mathbf{q}])}.$$

In addition, under the conditions of Lemma SA-4.2, for each $j = 0, 1, 2, 3$*,*

$$\sup_{\mathbf{x}\in\mathcal{X}} \left|\widehat{\partial^{\mathbf{q}}\mu}_j(\mathbf{x}) - \partial^{\mathbf{q}}\mu(\mathbf{x})\right| \lesssim_{\mathbb{P}} h^{-(d/2+[\mathbf{q}])}\sqrt{\frac{\log n}{n}} + \bar{R}_{1n,\mathbf{q}} + \bar{R}_{2n,\mathbf{q}} =: R^{\mathtt{uc}}_{\mathbf{q},j}$$

where $\bar{R}_{1n,\mathbf{q}}$ *and* $\bar{R}_{2n,\mathbf{q}}$ *are defined in Lemma SA-4.2.*

Relying on the uniform convergence results, we can further show the consistency of variance estimates $\widehat{\Omega}_j(\mathbf{x})$ defined in Section SA-2. The results in the following theorem are partly reported in Theorem 5.2 in the main paper.

THEOREM SA-4.2 (Variance Estimate). *Let Assumptions SA-1–SA-5 hold.*

(i) Assume that $\mathbb{E}[|\varepsilon_i|^{2+\nu}] < \infty$ *for some* $\nu > 0$*, and* $\frac{n^{\frac{2}{2+\nu}}(\log n)^{\frac{2\nu}{4+2\nu}}}{nh^d} = o(1)$*. Then for each* $j = 0, 1, 2, 3$*,*

$$\|\widehat{\mathbf{\Sigma}}_j - \mathbf{\Sigma}_j\| \lesssim_{\mathbb{P}} h^d \left(R^{\mathtt{uc}}_{\mathbf{0},j} + \frac{n^{\frac{1}{2+\nu}}(\log n)^{\frac{\nu}{4+2\nu}}}{\sqrt{nh^d}}\right) = o_{\mathbb{P}}(h^d), \quad \text{and}$$

$$\sup_{\mathbf{x}\in\mathbf{X}} \left|\widehat{\Omega}_j(\mathbf{x}) - \Omega_j(\mathbf{x})\right| \lesssim_{\mathbb{P}} h^{-d-2[\mathbf{q}]} \left(R^{\mathtt{uc}}_{\mathbf{0},j} + \frac{n^{\frac{1}{2+\nu}}(\log n)^{\frac{\nu}{4+2\nu}}}{\sqrt{nh^d}}\right) = o_{\mathbb{P}}(h^{-d-2[\mathbf{q}]})$$

where $R^{\mathtt{uc}}_{\mathbf{0},j}$ *is the uniform convergence rate given in Theorem SA-4.1 with* $\mathbf{q} = \mathbf{0}$*.*

(ii) *Assume that* $\mathbb{E}[|\varepsilon_i|^3\exp(|\varepsilon_i|)] < \infty$, *and* $\frac{(\log n)^3}{nh^d} = o(1)$. *Then for each* $j = 0,1,2,3$,

$$\|\widehat{\mathbf{\Sigma}}_j - \mathbf{\Sigma}_j\| \lesssim_{\mathbb{P}} h^d \left(R^{\mathtt{uc}}_{\mathbf{0},j} + \frac{(\log n)^{3/2}}{\sqrt{nh^d}} \right) = o_{\mathbb{P}}(h^d), \quad \textit{and}$$

$$\sup_{\mathbf{x}\in\mathcal{X}} \left| \widehat{\Omega}_j(\mathbf{x}) - \Omega_j(\mathbf{x}) \right| \lesssim_{\mathbb{P}} h^{-d-2[\mathbf{q}]} \left(R^{\mathtt{uc}}_{\mathbf{0},j} + \frac{(\log n)^{3/2}}{\sqrt{nh^d}} \right) = o_{\mathbb{P}}(h^{-d-2[\mathbf{q}]}).$$

SA-5. Uniform Inference.

SA-5.1. *Strong Approximation.* Now we move on to uniform inference. The following t-statistic processes are of interest:

$$\widehat{T}_j(\mathbf{x}) = \frac{\widehat{\partial^{\mathbf{q}}\mu}_j(\mathbf{x}) - \mu(\mathbf{x})}{\sqrt{\widehat{\Omega}_j(\mathbf{x})/n}}, \quad \mathbf{x}\in\mathcal{X}, \quad j=0,1,2,3.$$

Let r_n be a positive non-vanishing sequence, which will be used to denote the approximation error rate in the following analysis. As the first step, we employ Lemma SA-2.3 and Theorem SA-4.2 to show that the sampling and estimation uncertainty of $\widehat{\boldsymbol{\gamma}}_{\mathbf{q},j}(\mathbf{x})$ and $\widehat{\Omega}_j(\mathbf{x})$ are negligible uniformly over $\mathbf{x}\in\mathcal{X}$. Specifically, define

$$t_j(\mathbf{x}) = \frac{\boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})'}{\sqrt{\Omega_j(\mathbf{x})}} \mathbb{G}_n[\mathbf{\Pi}_j(\mathbf{x}_i)\varepsilon_i], \quad \mathbf{x}\in\mathcal{X}, \quad j=0,1,2,3.$$

The following lemma, which appears as Lemma 6.1 in the main paper, shows that $\widehat{T}_j(\cdot)$ can be approximated by $t_j(\cdot)$ uniformly.

Lemma SA-5.1 (Hats Off). *Let Assumptions SA-1–SA-5 hold. Assume one of the following holds:*

(i) $\mathbb{E}[|\varepsilon_i|^{2+\nu}] < \infty$ *for some* $\nu > 0$, *and* $\frac{n^{\frac{2}{2+\nu}}(\log n)^{\frac{2+2\nu}{2+\nu}}}{nh^d} = o(r_n^{-2})$; *or*

(ii) $\mathbb{E}[|\varepsilon_i|^3\exp(|\varepsilon_i|)] < \infty$, *and* $\frac{(\log n)^4}{nh^d} = o(r_n^{-2})$.

Furthermore, for $j=0$, *assume* $nh^{d+2m} = o(r_n^{-2})$; *and for* $j=1,2,3$, *assume* $nh^{d+2m+2\varrho} = o(r_n^{-2})$. *Then for each* $j=0,1,2,3$, $\sup_{\mathbf{x}\in\mathcal{X}}|\widehat{T}_j(\mathbf{x}) - t_j(\mathbf{x})| = o_{\mathbb{P}}(r_n^{-1})$.

The proof can be found in Section 8.2 of the main paper. Now our task reduces to construct valid distributional approximation to $t_j(\cdot)$ for each $j=0,1,2,3$, in a proper sense. Specifically, we want to show that on a sufficiently rich enough probability space there exists a copy $t_j'(\cdot)$ of $t_j(\cdot)$ and a Gaussian process $Z_j(\cdot)$ such that $\sup_{\mathbf{x}\in\mathcal{X}}|t_j'(\mathbf{x}) - Z_j(\mathbf{x})| = o_{\mathbb{P}}(r_n^{-1})$. When such a construction is possible, we can use $Z_j(\cdot)$ to approximate the distribution of $t_j(\cdot)$, as well as $\widehat{T}_j(\cdot)$ in view of Lemma SA-5.1. To save our notation, we denote this strong approximation by

$\widehat{T}_j(\cdot) =_d Z_j(\cdot) + o_{\mathbb{P}}(r_n^{-1})$ in $\mathcal{L}^\infty(\mathcal{X})$ where $\mathcal{L}^\infty(\mathcal{X})$ refers to the set of all uniformly bounded real functions on $\mathcal{X}$ equipped with uniform norm.

We will show in the following that there exists an unconditional Gaussian approximating process

$$Z_j(\mathbf{x}) = \frac{\boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})'\boldsymbol{\Sigma}_j^{1/2}}{\sqrt{\Omega_j(\mathbf{x})}}\mathbf{N}_{K_j}, \quad \mathbf{x} \in \mathcal{X}, \quad \text{for} \quad j = 0, 1, 2, 3,$$

where $\mathbf{N}_{K_j}$ is a K_j-dimensional standard Normal vector with $K_j = \dim(\boldsymbol{\Pi}_j(\cdot))$. As discussed in the main paper, we may have two strategies to construct strong approximations. The next theorem employs KMT coupling techniques.

THEOREM SA-5.1 (Strong Approximation: KMT Coupling). *Let the conditions of Lemma SA-5.1 hold, and for $j = 2, 3$, also assume that $\frac{(\log n)^{3/2}}{\sqrt{nh^d}} = o_{\mathbb{P}}(r_n^{-2})$. In addition, assume one of the following holds:*

(i) $\sup_{x\in\mathcal{X}} \mathbb{E}[|\varepsilon_i|^{2+\bar{\nu}}|\mathbf{x}_i = \mathbf{x}] < \infty$ *for some* $\bar{\nu} > 0$, *and* $\frac{n^{2/(2+\bar{\nu})}}{nh^d} = o(r_n^{-2})$;
(ii) $\sup_{\mathbf{x}\in\mathcal{X}} \mathbb{E}[|\varepsilon_i|^3 \exp(|\varepsilon_i|)|\mathbf{x}_i = \mathbf{x}] < \infty$.

Then for $d = 1$ and $j = 0, 1, 2, 3$, $\widehat{T}_j(\cdot) =_d Z_j(\cdot) + o_{\mathbb{P}}(r_n^{-1})$ in $\mathcal{L}^\infty(\mathcal{X})$. When $d > 1$ and $\mathbf{p}(\cdot)$ is a Haar basis, the same result still holds for $j = 0$.

Unfortunately, extending this to general multi-dimensional cases needs some non-trivial results which are not available in current strong approximation literature, to the best of our knowledge. Thus, when $d \geq 2$, we employ an improved version of Yurinskii's inequality due to [3], which allows us to replace the Euclidean norm in the original version of Yurinskii's inequality by a sup-norm. Since basis functions considered in this paper are locally supported, a metric of distance between random vectors in terms of sup-norm leads to a weaker rate restriction than that previously used in literature, though still stronger than what we obtained in Theorem SA-5.1. The result is reported in Theorem 6.2 of the main paper. We restate it in the following for completeness. The proof can be found in Section 8.4 of the main paper.

THEOREM SA-5.2 (Strong Approximation: Yurinskii's coupling). *Let the conditions of Lemma SA-5.1 hold. In addition, assume that $\sup_{\mathbf{x}\in\mathcal{X}} \mathbb{E}[|\varepsilon_i|^3|\mathbf{x}_i = \mathbf{x}] < \infty$ and $\frac{(\log n)^4}{nh^{3d}} = o(r_n^{-6})$. Then for $j = 0, 1, 2, 3$, $\widehat{T}_j(\cdot) =_d Z_j(\cdot) + o_{\mathbb{P}}(r_n^{-1})$ in $\mathcal{L}^\infty(\mathcal{X})$.*

SA-5.2. *Implementation.* $\{Z_j(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\}$ is still an infeasible approximation of $\{\widehat{T}_j(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\}$. Our next objective is to construct practicably feasible distributional approximations. To be precise, we construct Gaussian processes $\{\widehat{Z}_j(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\}$, with distributions known conditional on the data $(\mathbf{y}, \mathbf{X})$, and there exists a copy $\widehat{Z}'_j(\cdot)$ of $\widehat{Z}_j(\cdot)$ in a sufficiently rich probability space such that (i) $\widehat{Z}'_j(\cdot) =_d \widehat{Z}_j(\cdot)$

conditional on the data $(\mathbf{y}, \mathbf{X})$ and (ii) for some positive non-vanishing sequence r_n and for all $\eta > 0$,

$$\mathbb{P}^*\left[\sup_{\mathbf{x}\in\mathcal{X}} |\widehat{Z}_j'(\mathbf{x}) - Z_j(\mathbf{x})| \geq \eta r_n^{-1}\right] = o_{\mathbb{P}}(1),$$

where $\mathbb{P}^*[\cdot] = \mathbb{P}[\cdot|\mathbf{y}, \mathbf{X}]$ denotes the probability operator conditional on the data. When such a feasible process exists, we write $\widehat{Z}_j(\cdot) =_{d^*} Z_j(\cdot) + o_{\mathbb{P}^*}(r_n^{-1})$ in $\mathcal{L}^\infty(\mathcal{X})$. From a practical perspective, sampling from $\widehat{Z}_j(\cdot)$, conditional on the data, is possible and provides a valid distributional approximation.

Our first construction is a direct plug-in approach using the conclusion of Theorem SA-5.1 and SA-5.2. All unknown objects are replaced by consistent estimators already used in the feasible t-statistics:

$$\widehat{Z}_j(\mathbf{x}) = \frac{\widehat{\boldsymbol{\gamma}}_{\mathbf{q},j}(\mathbf{x})'\widehat{\boldsymbol{\Sigma}}_j^{1/2}}{\sqrt{\widehat{\Omega}_j(\mathbf{x})}}\mathbf{N}_{K_j}, \qquad \mathbf{x} \in \mathcal{X}, \qquad j = 0, 1, 2, 3$$

The validity of this method is shown in Theorem 6.3 of the main paper. We restate it in the following whose proof is available in Section 8.5 of the main paper.

THEOREM SA-5.3 (Plug-in Approximation). *Let the conditions in Lemma SA-5.1 hold. Furthermore, for $j = 2, 3$,*

(i) when $\mathbb{E}[|\varepsilon_i|^{2+\nu}] < \infty$ for some $\nu > 0$, assume that $\frac{n^{\frac{1}{2+\nu}}(\log n)^{\frac{4+3\nu}{4+2\nu}}}{\sqrt{nh^d}} = o(r_n^{-2})$;

(ii) when $\mathbb{E}[|\varepsilon_i|^3 \exp(|\varepsilon_i|)] < \infty$, assume that $\frac{(\log n)^{5/2}}{\sqrt{nh^d}} = o(r_n^{-2})$.

Then for each $j = 0, 1, 2, 3$, $\widehat{Z}_j(\cdot) =_{d^*} Z_j(\cdot) + o_{\mathbb{P}^*}(r_n^{-1})$ in $\mathcal{L}^\infty(\mathcal{X})$.

Our second construction, which is not reported in the main paper, is inspired by the intermediate approximation in Lemma 8.2 of the main paper. Again, $\boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})$ and $\Omega_j(\mathbf{x})$ can be simply replaced by their sample analogues, and ζ_i's can be generated by sampling from an n-dimensional standard Gaussian vector independent of the data. The key unknown quantity is the conditional variance function $\sigma^2(\mathbf{x}) = \mathbb{E}[\varepsilon_i^2|\mathbf{x}_i = \mathbf{x}]$. If the errors are homoskedastic, one can simply let $\sigma^2(\cdot) = 1$. If not, we need an estimator $\widehat{\sigma}(\cdot)$ of the conditional variance satisfying a mild condition on uniform convergence, and then we can construct

$$\widehat{z}_j(\mathbf{x}) = \frac{\widehat{\boldsymbol{\gamma}}_{\mathbf{q},j}(\mathbf{x})'}{\sqrt{\widehat{\Omega}_j(\mathbf{x})}}\mathbb{G}_n[\boldsymbol{\Pi}_j(\mathbf{x}_i)\zeta_i\widehat{\sigma}(\mathbf{x}_i)], \quad \mathbf{x} \in \mathcal{X}, \quad j = 0, 1, 2, 3.$$

In practice one may simply construct a nonparametric estimator of $\sigma^2(\cdot)$ by using, for example, regression splines or smoothing splines. See [14, Chapter 22.4] for more details. We do not elaborate on this issue, but the next theorem shows the validity of this approach.

THEOREM SA-5.4. *Let the conditions in Lemma SA-5.1 hold, and* $\hat{\sigma}^2(\cdot)$ *is an estimator of* $\sigma^2(\cdot)$ *such that* $\max_{1\leq i\leq n}|\hat{\sigma}^2(\mathbf{x}_i)-\sigma^2(\mathbf{x}_i)| = o_{\mathbb{P}}(r_n^{-1}(\log n)^{-1/2})$. *For* $j=2,3$, *further assume* $\frac{(\log n)^{3/2}}{\sqrt{nh^d}} = o(r_n^{-2})$. *Then for each* $j=0,1,2,3$, $\widehat{z}_j(\cdot) =_{d^*} Z_j(\cdot) + o_{\mathbb{P}^*}(r_n^{-1})$ *in* $\mathcal{L}^\infty(\mathcal{X})$.

Our third approach to approximating the infeasible $Z_j(\cdot)$ employs an easy-to-implement wild bootstrap procedure. Specifically, for i.i.d. bounded random variables $\{\omega_i : 1 \leq i \leq n\}$ with $\mathbb{E}[\omega_i]=0$ and $\mathbb{E}[\omega_i^2]=1$ independent of the data, we construct bootstrapped t-statistics:

$$\widehat{z}_j^*(\mathbf{x}) = \frac{\widehat{\boldsymbol{\gamma}}_{\mathbf{q},j}(\mathbf{x})'}{\sqrt{\widehat{\Omega}_j^*(\mathbf{x})}}\mathbb{G}_n[\boldsymbol{\Pi}_j(\mathbf{x}_i)\omega_i\widehat{\varepsilon}_{i,j}], \qquad \mathbf{x}\in\mathcal{X}, \qquad j=0,1,2,3,$$

where $\widehat{\varepsilon}_{i,j}$ is defined in (SA-2.5), and the bootstrap Studentization $\widehat{\Omega}_j^*(\mathbf{x})$ is constructed using $\widehat{\boldsymbol{\Sigma}}_j^* = \mathbb{E}_n[\boldsymbol{\Pi}_j(\mathbf{x}_i)\boldsymbol{\Pi}_j(\mathbf{x}_i)'\omega_i^2\widehat{\varepsilon}_{i,j}^2]$.

In comparison with previous plug-in approximations, the bootstrap method requires an additional Gaussian approximation step as in either Theorem SA-5.1 or Theorem SA-5.2 (conditional on the data). These results are stated in the following two theorems. $\mathbb{E}^*[\cdot]$ denotes the expectation conditional on the data.

THEOREM SA-5.5 (Wild Bootstrap: KMT). *Let the conditions of Lemma SA-5.1 hold. In addition,*

(i) when $\mathbb{E}[|\varepsilon_i|^{2+\nu}]<\infty$ *for some* $\nu>0$, *assume that for* $j=2,3$, $\frac{n^{\frac{1}{2+\nu}}(\log n)^{\frac{4+3\nu}{4+2\nu}}}{\sqrt{nh^d}} = o(r_n^{-2})$;

(ii) when $\mathbb{E}[|\varepsilon_i|^3\exp(|\varepsilon_i|)]<\infty$, *assume that there exists some constant* $C>0$ *such that for* $1\leq i\leq n$,

$$C\mathbb{E}^*[|\varepsilon_i\omega_i|^3\exp(|\varepsilon_i\omega_i|)] \leq \mathbb{E}^*[|\varepsilon_i\omega_i|^2] \quad a.s.,$$

and for $j=2,3$, $\frac{(\log n)^{5/2}}{\sqrt{nh^d}} = o(r_n^{-2})$.

Then for $d=1$ *and* $j=0,1,2,3$, $\widehat{z}_j^*(\cdot) =_{d^*} Z_j(\cdot) + o_{\mathbb{P}^*}(r_n^{-1})$ *in* $\mathcal{L}^\infty(\mathcal{X})$. *When* $d>1$ *and* $\mathbf{p}(\cdot)$ *is a Haar basis, the same result still holds for* $j=0$.

The additional condition in (ii) of the above theorem is generally difficult to verify in practice unless some restrictive conditions, e.g. bounded support of ε_i, are imposed.

THEOREM SA-5.6 (Wild Bootstrap: Yurinskii). *Let the conditions in Theorem SA-5.2 hold. In addition, for* $j=2,3$,

(i) when $\mathbb{E}[|\varepsilon_i|^{2+\nu}]<\infty$ *for some* $\nu>0$, *assume that* $\frac{n^{\frac{1}{2+\nu}}(\log n)^{\frac{4+3\nu}{4+2\nu}}}{\sqrt{nh^d}} = o(r_n^{-2})$;

(ii) when $\mathbb{E}[|\varepsilon_i|^3 \exp(|\varepsilon_i|)] < \infty$*, assume that* $\frac{(\log n)^{5/2}}{\sqrt{nh^d}} = o(r_n^{-2})$.

Then for each $j = 0, 1, 2, 3$, $\widehat{z}_j^*(\cdot) =_{d^*} Z_j(\cdot) + o_{\mathbb{P}^*}(r_n^{-1})$ *in* $\mathcal{L}^\infty(\mathcal{X})$.

SA-5.3. *Application: Confidence Bands.* An immediate application of Theorems SA-5.1–SA-5.6 is to construct confidence bands for the regression function or its derivatives. Specifically, for $j = 0, 1, 2, 3$, and $\alpha \in (0, 1)$, we seek a quantile $q_j(\alpha)$ such that

$$\mathbb{P}\left[\sup_{\mathbf{x}\in\mathcal{X}} |\widehat{T}_j(\mathbf{x})| \leq q_j(\alpha)\right] = 1 - \alpha + o(1),$$

which then can be used to construct uniform $100(1-\alpha)$-percent confidence bands for $\partial^{\mathbf{q}}\mu(\mathbf{x})$ of the form

$$\left[\widehat{\partial^{\mathbf{q}}\mu}_j(\mathbf{x}) \pm q_j(\alpha)\sqrt{\widehat{\Omega}_j(\mathbf{x})/n} \;:\; \mathbf{x} \in \mathcal{X}\right].$$

The following theorem establishes a valid distributional approximation for the suprema of the t-statistic processes $\{\widehat{T}_j(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\}$ using [9, Lemma 2.4] to convert our strong approximation results into convergence of distribution functions in terms of Kolmogorov distance.

THEOREM SA-5.7 (Confidence Band). *Let the conditions of Theorem SA-5.1 or Theorem SA-5.2 hold with* $r_n = \sqrt{\log n}$. *If the corresponding conditions of Theorem SA-5.3 for each* $j = 0, 1, 2, 3$ *hold, then*

$$\sup_{u\in\mathbb{R}}\left|\mathbb{P}\left[\sup_{\mathbf{x}\in\mathcal{X}} |\widehat{T}_j(\mathbf{x})| \leq u\right] - \mathbb{P}^*\left[\sup_{\mathbf{x}\in\mathcal{X}} |\widehat{Z}_j(\mathbf{x})| \leq u\right]\right| = o_{\mathbb{P}}(1).$$

If the corresponding conditions in Theorem SA-5.4 hold for each $j = 0, 1, 2, 3$*, then*

$$\sup_{u\in\mathbb{R}}\left|\mathbb{P}\left[\sup_{\mathbf{x}\in\mathcal{X}} |\widehat{T}_j(\mathbf{x})| \leq u\right] - \mathbb{P}^*\left[\sup_{\mathbf{x}\in\mathcal{X}} |\widehat{z}_j(\mathbf{x})| \leq u\right]\right| = o_{\mathbb{P}}(1).$$

If the corresponding conditions in Theorem SA-5.5 or SA-5.6 hold for each $j = 0, 1, 2, 3$*, then*

$$\sup_{u\in\mathbb{R}}\left|\mathbb{P}\left[\sup_{\mathbf{x}\in\mathcal{X}} |\widehat{T}_j(\mathbf{x})| \leq u\right] - \mathbb{P}^*\left[\sup_{\mathbf{x}\in\mathcal{X}} |\widehat{z}_j^*(\mathbf{x})| \leq u\right]\right| = o_{\mathbb{P}}(1).$$

SA-6. Examples. We now illustrate how Assumptions SA-3, SA-4, and SA-5 are verified for popular basis choices, including B-splines, wavelets, and piecewise polynomials, and show when (SA-2.6) holds. Note well that even in the cases where (SA-2.6) holds, we will not make use of this property neither herein nor in our software implementation [8], since both bias terms of Lemma SA-2.2 may be important in finite samples.

The first three subsections illustrate each of these assuming rectangular support, $\mathcal{X} = \otimes_{\ell=1}^d \mathcal{X}_\ell \subset \mathbb{R}^d$, and correspondingly use tensor-product partitions. We note that these three in fact overlap in the special cases of $m = 1$ on a tensor-product partition (see Equation SA-3.1), leading to so-called zero-degree splines, Haar wavelets, and regressograms (piecewise constant). The final subsection considers a general partition. Recall that whenever Δ covers only strict subset of $\mathcal{X}$, all our results hold on that subset; the first three subsections may also be interpreted in this light.

SA-6.1. *B-Splines on Tensor-Product Partitions.* A univariate spline is a piecewise polynomial satisfying certain smoothness constraints. For some integer $m_\ell \geq 1$, let $\mathcal{S}_{\Delta_\ell, m_\ell}$ be the set of splines of order m_ℓ with univariate partition Δ_ℓ, i.e.,

$$\mathcal{S}_{\Delta_\ell, m_\ell} = \Big\{ s \in \mathcal{C}^{m_\ell - 2}(\mathcal{X}_\ell) \colon s(x) \text{ is a polynomial of degree } (m_\ell - 1) \text{ on each } [t_{\ell, l_\ell}, t_{\ell, l_\ell + 1}] \Big\}.$$

If $m_\ell = 1$, it is interpreted as the set of piecewise constant functions that are discontinuous on Δ_ℓ. $\mathcal{S}_{\Delta_\ell, m_\ell}$ is a vector space and can be spanned by many equivalent representing bases. B-splines as a local basis are well studied in literature and enjoy many nice properties [20, 23].

Define an extended knot sequence $\Delta_{\ell,e}$ such that

$$t_{\ell,-m_\ell+1} = t_{\ell,-m_\ell+2} = \cdots = t_{\ell,0} < t_{\ell,1} < \cdots < t_{\ell,\kappa_\ell - 1} < t_{\ell,\kappa_\ell} = \cdots = t_{\ell,\kappa_\ell + m_\ell - 1},$$

Then, the m_ℓ-th order B-spline with knot sequence $\Delta_{\ell,e}$ is

$$p_{l_\ell, m_\ell}(x_\ell) = (t_{\ell, l_\ell} - t_{\ell, l_\ell - m_\ell})[t_{\ell, l_\ell - m_\ell}, \cdots, t_{\ell, l_\ell}]_t (t - x_\ell)_+^{m_\ell - 1}, \; l_\ell = 1, \ldots, K_\ell = \kappa_\ell + m_\ell - 1,$$

where $(a)_+ = \max\{a, 0\}$ and for a function $g(t,x)$, $[t_1, t_2, \cdots, t_l]_t \, g(t,x)$ denotes the divided difference of $g(t,x)$ with respect to t, given a sequence of knots $t_1 \leq t_2 \leq \cdots \leq t_l$.

When there is no chance of confusion, we shall write $p_{l_\ell}(x_\ell)$ instead of $p_{l_\ell, m_\ell}(x_\ell)$. Accordingly, the space of tensor-product polynomial splines of order $\mathbf{m} = (m_1, \ldots, m_d)$ with partition Δ is spanned by tensor products of univariate B-splines

$$\mathcal{S}_{\Delta, \mathbf{m}} = \otimes_{\ell=1}^d \mathcal{S}_{\Delta_\ell, m_\ell} = \operatorname{span}\{p_{l_1}(x_1) p_{l_2}(x_2) \cdots p_{l_d}(x_d)\}_{l_1 = 1, \ldots, l_d = 1}^{K_1, \ldots, K_d}.$$

We have a total of $K = \prod_{\ell=1}^d K_\ell$ basis functions. The order of univariate basis could vary across dimensions, but for simplicity we assume that $m_1 = \cdots = m_d = m$ and hence write $\mathcal{S}_{\Delta, m} = \mathcal{S}_{\Delta, \mathbf{m}}$. Also, let $p_{l_1 \ldots l_d}(\mathbf{x}) = p_{l_1}(x_1) \cdots p_{l_d}(x_d)$ to further simplify notation.

Arrange this set of basis functions by first increasing l_d from 1 to K_d with other l_ℓ's fixed at 1 and then increasing l_ℓ sequentially. As a result, we construct a one-to-one correspondence φ mapping from $\{(l_1, \ldots, l_d) : 1 \leq l_\ell \leq K_\ell, \ell = 1, \ldots, d\}$ to $\{1, \ldots, K\}$. Then, we can write $p_k(\mathbf{x}) = p_{\varphi^{-1}(k)}(\mathbf{x})$, $k = 1, \ldots, K$, which is consistent with our notation in Section SA-1.

The following lemma shows that Assumptions SA-3 and SA-4 hold for B-splines.

LEMMA SA-6.1 (B-Splines Estimators). *Let $\mathbf{p}(\mathbf{x})$ be a tensor-product B-Spline basis of order m, and suppose Assumptions SA-1 and SA-2 hold with $m \le S$. Then:*

- **(a)** $\mathbf{p}(\mathbf{x})$ *satisfies Assumption SA-3.*
- **(b)** *If, in addition,*

$$\max_{0\le l_\ell \le \kappa_\ell - 2} |b_{\ell,l_\ell+1} - b_{\ell,l_\ell}| = O(b^{1+\varrho}), \qquad \ell = 1, \dots, d, \tag{SA-6.1}$$

then Assumption SA-4 holds with $\varsigma = m-1$ and

$$\mathscr{B}_{m,\boldsymbol{\varsigma}}(\mathbf{x}) = -\sum_{\boldsymbol{u}\in\Lambda_m} \frac{\partial^{\boldsymbol{u}}\mu(\mathbf{x})h_{\mathbf{x}}^{m-[\boldsymbol{\varsigma}]}}{(\boldsymbol{u}-\boldsymbol{\varsigma})!} \frac{\mathbf{b}_{\mathbf{x}}^{\boldsymbol{u}-\boldsymbol{\varsigma}}}{h_{\mathbf{x}}^{m-[\boldsymbol{\varsigma}]}} B^{\mathsf{S}}_{\boldsymbol{u}-\boldsymbol{\varsigma}}\Big((\mathbf{x}-\mathbf{t}^L_{\mathbf{x}}) \oslash \mathbf{b}_{\mathbf{x}}\Big)$$

where $\Lambda_m = \{\boldsymbol{u} \in \mathbb{Z}^d_+ : [\boldsymbol{u}] = m, \text{ and } u_\ell = m \text{ for some } 1 \le \ell \le d\}$ and $B^{\mathsf{S}}_{\boldsymbol{u}}(\mathbf{x})$ is the product of univariate Bernoulli polynomials; that is, $B^{\mathsf{S}}_{\boldsymbol{u}}(\mathbf{x}) := \prod_{\ell=1}^d B_{u_\ell}(x_\ell)$ with $B_{u_\ell}(\cdot)$ being the u_ℓ-th Bernoulli polynomial and $B^{\mathsf{S}}_{\boldsymbol{u}}(\cdot) = 0$ if $\boldsymbol{u}$ contains negative elements. Furthermore, Equation (SA-2.6) holds.
- **(c)** *Let $\tilde{\mathbf{p}}(\mathbf{x})$ be a tensor-product B-Spline basis of order $\tilde{m} > m$ on the same partition Δ, and assume $\tilde{m} \le S$. Then Assumption SA-5 is satisfied.*

Equation (SA-6.1) gives a precise definition of the strong quasi-uniform condition on the partition scheme. Assumption SA-2 requires that the volume of all cells vanishes at the same rate but allows for any constant proportionality between neighboring cells. Presently, cells are further restricted to be of the same volume asymptotically, and further, a specific rate is required that is related to the smoothness of $\mu(\cdot)$. Note that, for example, equally spaced knots satisfy this condition trivially. For other schemes, additional work may be needed. Under (SA-6.1), [2] obtained an expression of the leading asymptotic error of univariate splines, which was later used by [27] and [28], among others. Lemma SA-6.1 extends previous results to the multi-dimensional case, in addition to showing that the high-level conditions in Assumptions SA-3 and SA-4 hold for B-Splines.

When (SA-6.1) fails, it may be possible to obtain results, but additional cumbersome notation would be needed and the results may be less useful. However, it is straightforward to verify that without (SA-6.1), there still exists $s^* \in \mathcal{S}_{\Delta,m}$ such that for all $[\boldsymbol{\varsigma}] \le \varsigma$,

$$\sup_{\mathbf{x}\in\mathcal{X}} |\partial^{\boldsymbol{\varsigma}}\mu - \partial^{\boldsymbol{\varsigma}} s^*| \lesssim h^{m-[\boldsymbol{\varsigma}]}, \tag{SA-6.2}$$

where recall that $\mathcal{S}_{\Delta,m}$ denotes the linear span of $\mathbf{p}(\mathbf{x};\Delta,m)$. However, this result does not allow for bias correction or IMSE expansion.

Finally, Λ_m contains only the multi-indices corresponding to m-th order partial derivatives of $\mu(\mathbf{x})$. This is due to the fact that, as a variant of polynomial approximation, the total order of tensor-product B-splines is not fixed at m, i.e. some higher-order components are retained that approximate terms with m-th order cross partial derivatives. This feature distinguishes tensor-product splines from multivariate splines which do control the total order of bases.

SA-6.2. *Wavelets on Tensor-Product Partitions.* Our results apply to compactly supported wavelets, such as Cohen-Daubechies-Vial wavelets [12]. For more background details see [19, 17, 11], and references therein.

To describe these estimators, we first introduce the general definition of wavelets and then show that a large class of orthogonal wavelets satisfy our high-level assumptions. For some integer $m \geq 1$, we call ϕ a (univariate) scaling function or "father wavelet" of degree $m-1$ if (i) $\int_{\mathbb{R}} \phi(x_\ell)\, dx_\ell = 1$, (ii) ϕ and all its derivatives up to order $m-1$ decrease rapidly as $|x_\ell| \to \infty$, and (iii) $\{\phi(x_\ell - l) : l \in \mathbb{Z}\}$ forms a Riesz basis for a closed subspace of $L_2(\mathbb{R})$. A real-valued function ψ is called a (univariate) "mother wavelet" of degree $m-1$ if (i) $\int_{\mathbb{R}} x_\ell^v \psi(x_\ell) dx_\ell = 0$ for $0 \leq v \leq m-1$, (ii) ψ and all its derivatives decrease rapidly as $|x_\ell| \to \infty$, and (iii) $\{2^{s/2}\psi(2^s x_\ell - l) : s, l \in \mathbb{Z}\}$ forms a Riesz basis of $L_2(\mathbb{R})$.

We restrict our attention to orthogonal wavelets with compact support. The father wavelet ϕ and the mother wavelet ψ are both compactly supported, and for any integer $s_0 \geq 0$, any function in $L_2(\mathbb{R})$ admits the following $(m-1)$-regular wavelet multiresolutional expansion:

$$\mu(x_\ell) = \sum_{l=-\infty}^{\infty} a_{s_0 l}\phi_{s_0 l}(x_\ell) + \sum_{s=s_0}^{\infty}\sum_{l=-\infty}^{\infty} b_{sl}\psi_{sl}(x_\ell),\ x_\ell \in \mathbb{R}, \quad \text{where}$$

$$\begin{aligned} a_{sl} &= \int_{\mathbb{R}} \mu(x_\ell)\phi_{sl}(x_\ell)dx_\ell, \qquad \phi_{sl}(x_\ell) = 2^{s/2}\phi(2^s x_\ell - l), \\ b_{sl} &= \int_{\mathbb{R}} \mu(x_\ell)\psi_{sl}(x_\ell)dx_\ell, \qquad \psi_{sl}(x_\ell) = 2^{s/2}\psi(2^s x_\ell - l), \end{aligned}$$

and $\{\phi_{s_0 l}, l \in \mathbb{Z}; \psi_{sl}, s \geq s_0, l \in \mathbb{Z}\}$ is an orthonormal basis of $L_2(\mathbb{R})$. Accordingly, to construct an orthonormal basis of $L_2([0,1])$, one can pick those basis functions supported in the interior of $[0,1]$ and add some boundary correction functions. For details of construction of such a basis see, for example, [12]. With a slight abuse of notation, in what follows we write $\{\phi_{s_0 l}, l \in \mathcal{L}_{s_0}; \psi_{sl}, s \geq s_0, l \in \mathcal{L}_s\}$ for an orthonormal basis of $L_2([0,1])$ rather than $L_2(\mathbb{R})$ where $\mathcal{L}_{s_0}$ and $\mathcal{L}_s$ are some proper index sets depending on s_0 and s respectively. Again, we use tensor product to form a multidimensional wavelet basis, and then it fits into the context of our analysis.

The compactness of the father wavelet is needed to verify Assumption SA-3. To see this, consider a standard multiresolutional analysis setting. A large space, say $L_2([0,1])$, is decomposed into a sequence of nested subspaces $\{0\} \subset \cdots \subset V_{-1} \subset V_0 \subset V_1 \cdots \subset L_2([0,1])$. Generally, $\{\phi_{sl}, l \in \mathcal{L}_s\}$ constitutes a basis for V_s, and $\{\psi_{sl}, l \in \mathcal{L}_s\}$ forms a basis for the orthogonal complement W_s of V_s in V_{s+1}. Since the support of ϕ is compact and $\{\phi_{sl}\}$ is generated simply by dilation and translation of ϕ, $[0,1]$ can be viewed as implicitly partitioned. Specifically, by increasing the resolution level from s to $s+1$, the length of the support of ϕ_{sl} is halved. Hence, it is equivalent to placing additional partitioning knots at the midpoint of each subinterval. In addition, as the sample size n grows, we allow the resolution level s to increase (thus written as s_n when needed), but each basis function $\phi_{s_n l}$ is supported by only a finite number of subintervals since the generating scaling function is compactly supported.

The above discussion associates the generating process of wavelets at different levels with a partitioning scheme. It is easy to see that Assumption SA-2 is automatically satisfied in this case, and given a resolution level s_n, the mesh width $b = 2^{-s_n}$. On the other hand, given a starting level s_0, $\{\phi_{s_0 l}, l \in \mathcal{L}_{s_0}; \psi_{sl}, s_0 \leq s \leq s_n, l \in \mathcal{L}_s\}$ does not satisfy Assumption SA-3, since some functions in such a basis have increasing supports as the resolution level $s_n \to \infty$. Nevertheless, they span the same space as $\{\phi_{s_n,l}, l \in \mathcal{L}_{s_n}\}$ and hence there exists a linear transformation that connects the two equivalent bases. Least squares estimators are invariant to such a transformation. Therefore, we will focus on the following tensor-product (father) wavelet basis

$$\mathbf{p}(\mathbf{x}) := \otimes_{\ell=1}^{d} 2^{-s_n/2} \boldsymbol{\phi}_{s_n}(x_\ell) \tag{SA-6.3}$$

where $\boldsymbol{\phi}_{s_n}$ is a vector containing all functions in $\{\phi_{s_n,l}, l \in \mathcal{L}_{s_n}\}$. By multiplying the basis by $2^{-s_n/2}$, we drop the normalizing constants that originally appear in the construction of orthonormal basis. As the next lemma shows, this large class of orthogonal wavelet bases satisfy our assumptions.

LEMMA SA-6.2 (Wavelets Estimators). *Let ϕ and ψ be a scaling function and a wavelet function of degree $m-1$ with $q+1$ continuous derivatives, $\mathbf{p}(\mathbf{x})$ be the tensor-product orthogonal (father) wavelet basis of degree $m-1$ generated by ϕ, and suppose Assumption SA-1 holds with $m \leq S$.*

(a) *$\mathbf{p}(\mathbf{x})$ satisfies Assumption SA-3.*

(b) *Assumption SA-4 holds with $\varsigma = q$ and*

$$\mathscr{B}_{m,\boldsymbol{\varsigma}}(\mathbf{x}) = -\sum_{\boldsymbol{u} \in \Lambda_m} \frac{\partial^{\boldsymbol{u}} \mu(\mathbf{x}) h^{m-[\boldsymbol{\varsigma}]}}{\boldsymbol{u}!} \frac{b^{m-[\boldsymbol{\varsigma}]}}{h^{m-[\boldsymbol{\varsigma}]}} B^{\mathtt{W}}_{\boldsymbol{u},\boldsymbol{\varsigma}}(\mathbf{x}/b)$$

where $\Lambda_m = \{\boldsymbol{u} \in \mathbb{Z}_+^d : [\boldsymbol{u}] = m,$ and $u_\ell = m$ for some $1 \leq \ell \leq d\}$ and $B^{\mathtt{W}}_{\boldsymbol{u},\boldsymbol{\varsigma}}(\mathbf{x}) = \sum_{s \geq 0} \partial^{\boldsymbol{\varsigma}} \xi_{\boldsymbol{u},s}(\mathbf{x})$ converges uniformly, with $\xi_{\boldsymbol{u},s}(\cdot)$ being a linear combination of products of univariate father wavelet ϕ and the mother wavelet ψ; its exact form is notationally cumbersome and is given in Equation (SA-10.14). Furthermore, Equation (SA-2.6) holds.

(c) *Let $\tilde{\phi}$ be a scaling function of degree $\tilde{m}-1$ with $m+1$ continuous derivatives for some $\tilde{m} > m$, $\tilde{\mathbf{p}}(\mathbf{x})$ be the tensor-product orthogonal wavelet basis generated by $\tilde{\phi}$ that has the same resolution level as $\mathbf{p}(\mathbf{x})$, and assume $\tilde{m} \leq S$. Then Assumption SA-5 is satisfied.*

In addition to verifying that our high-level assumptions hold for wavelets, this result gives a novel asymptotic error expansion for multidimensional compactly supported wavelets. Our derivation employs the ideas in [24] and exploits the tensor-product structure of the wavelet basis. The end result is similar to tensor-product splines, in the sense that the total order of the approximating basis is not fixed at m and thus Λ_m is the same as that for B-splines.

SA-6.3. *Generalized Regressograms on Tensor-Product Partitions.* To construct the generalized regressogram, we first define the piecewise polynomials. What will distinguish these from splines is that (i) each polynomial is supported on exactly one cell, and relatedly (ii) no continuity is assumed between cells. First, for some fixed integer $m \in \mathbb{Z}_+$, let $\mathbf{r}(x_\ell) = (1, x_\ell, \ldots, x_\ell^{m-1})'$ denote a vector of powers up to degree $m-1$. To extend it to a multidimensional basis, we take the tensor product of $\mathbf{r}(x_\ell)$, denoted by a column vector $\mathbf{R}(\mathbf{x})$. The total order of such a basis is not fixed, and its behavior is more similar to tensor-product B-splines. Following [7, and references therein], we exclude all terms with degree greater than $m-1$ in $\mathbf{R}(\mathbf{x})$. Hence the remaining elements in $\mathbf{R}(\mathbf{x})$ are given by $\mathbf{x}^{\boldsymbol{\alpha}} = x_1^{\alpha_1} \cdots x_d^{\alpha_d}$ for a unique d-tuple $\boldsymbol{\alpha}$ such that $[\boldsymbol{\alpha}] \leq m-1$. Then we "localize" this basis by restricting it to a particular subrectangle $\delta_{l_1 \cdots l_d}$. Specifically, we write $\mathbf{p}_{l_1 \ldots l_d}(\mathbf{x}) = \mathbb{1}_{\delta_{l_1 \ldots l_d}}(\mathbf{x})\mathbf{R}(\mathbf{x})$, where $\mathbb{1}_{\delta_{l_1 \ldots l_d}}(\mathbf{x})$ is equal to 1 if $\mathbf{x} \in \delta_{l_1 \ldots l_d}$ and 0 otherwise. Finally, we rotate the basis by centering each basis function at the start point of the corresponding cell and scale it by interval lengths. We can arrange the basis functions according to a particular ordering φ: first order the subrectangles the same way as that for B-splines, and then within each subrectangle $\delta_{l_1 \ldots l_d}$ the basis functions in $\mathbf{R}(\mathbf{x})$ are ordered ascendingly in $\boldsymbol{\alpha}$ and $\ell = 1, \ldots, d$. Using the same notation as in Section SA-1, we still write the basis as $\mathbf{p}(\mathbf{x}) = (p_1(\mathbf{x}), \cdots, p_K(\mathbf{x}))'$.

The following lemma shows that Assumptions SA-3 and SA-4 hold for generalized regressograms.

LEMMA SA-6.3 (Generalized Regressograms). *Let $\mathbf{p}(\mathbf{x})$ be the rotated piecewise polynomial basis of degree $m-1$ based on Legendre polynomials, and suppose that Assumptions SA-1 and SA-2 hold with $m \leq S$. Then,*

(a) $\mathbf{p}(\mathbf{x})$ *satisfies Assumption SA-3.*

(b) *Assumption SA-4 holds with $\varsigma = m-1$ and*

$$\mathscr{B}_{m,\boldsymbol{\varsigma}}(\mathbf{x}) = -\sum_{\boldsymbol{u} \in \Lambda_m} \frac{\partial^{\boldsymbol{u}} \mu(\mathbf{x}) h_{\mathbf{x}}^{m-[\boldsymbol{\varsigma}]}}{(\boldsymbol{u}-\boldsymbol{\varsigma})!} \frac{\mathbf{b}_{\mathbf{x}}^{\boldsymbol{u}-\boldsymbol{\varsigma}}}{h_{\mathbf{x}}^{m-[\boldsymbol{\varsigma}]}} B^{\mathsf{P}}_{\boldsymbol{u}-\boldsymbol{\varsigma}}\Big((\mathbf{x}-\mathbf{t}_{\mathbf{x}}^L) \oslash \mathbf{b}_{\mathbf{x}}\Big),$$

where $\Lambda_m = \{\boldsymbol{u} : [\boldsymbol{u}] = m\}$ and

$$B^{\mathsf{P}}_{\boldsymbol{u}}(\mathbf{x}) := \prod_{\ell=1}^{d} \binom{2u_\ell}{u_\ell}^{-1} P_{u_\ell}(x_\ell),$$

with $P_{u_\ell}(\cdot)$ being the u_ℓ-th shifted Legendre polynomial orthogonal on $[0,1]$, and $B^{\mathsf{P}}_{\boldsymbol{u}}(\cdot) = 0$ if $\boldsymbol{u}$ contains negative elements. Furthermore, Equation (SA-2.6) holds.

(c) *Let $\tilde{\mathbf{p}}(\mathbf{x})$ be a piecewise polynomial basis of degree $\tilde{m}-1$ on the same partition Δ for some $\tilde{m} > m$, and assume $\tilde{m} \le S$. Then Assumption SA-5 is satisfied.*

The leading asymptotic error obtained in Lemma SA-6.3 differs from the one in [7] because it is expressed in terms of orthogonal polynomials. Here we employ Legendre polynomials $\bar{P}_m(x)$ orthogonal on $[-1, 1]$ with respect to the Lebesgue measure, and then shift them to $P_m(x) = \bar{P}_m(2x-1)$. Thus the shifted Legendre polynomials are orthogonal on $[0, 1]$. Further, we allow for more general partitioning schemes.

SA-6.4. *Generalized Regressograms on General Partitions.* Moving away from tensor-product partitions will impede verification of Assumption SA-4 in general. A more typical result, such as Equation (SA-6.2) can be more easily proven in generality. In practice, however, many data-driven partitioning selection procedures, such as those induced by regression trees, do not necessarily lead to a tensor-product partition. There is also a large literature in approximation theory discussing bases constructed on triangulations or other general partition schemes [e.g., 20]. To demonstrate the power of our theory, we show presently that for the generalized regressogram we can obtain concrete results on general partitions.

The basis is as described above, but now allowing for cells of general shape. Let $\mathcal{X}$ be convex and $\mathbf{t}_\delta$ be an arbitrary point in each $\delta \in \Delta$. Hence, $\mathbf{t}_{\delta_\mathbf{x}}$ is simply a point in the cell containing $\mathbf{x}$. Construct the rotated piecewise polynomial basis as in Section SA-6.3, but centering the basis at $\mathbf{t}_\delta$ and rescaling it by the diameter of δ.

LEMMA SA-6.4. *Let $\mathbf{p}(\mathbf{x})$ be the rotated piecewise polynomial basis of degree $m-1$ on a general partition Δ, and suppose that Assumptions SA-1 and SA-2 hold with $m \le S$. Then,*

(a) *$\mathbf{p}(\mathbf{x})$ satisfies Assumption SA-3.*

(b) *Assumption SA-4 holds with $\varsigma = m-1$ and*

$$\mathscr{B}_{m,\boldsymbol{\varsigma}}(\mathbf{x}) = -\sum_{\boldsymbol{u}\in\Lambda_m} \frac{\partial^{\boldsymbol{u}}\mu(\mathbf{x})h_{\mathbf{x}}^{m-[\boldsymbol{\varsigma}]}}{(\boldsymbol{u}-\boldsymbol{\varsigma})!}\Big(\frac{\mathbf{x}-\mathbf{t}_{\delta_\mathbf{x}}}{h_\mathbf{x}}\Big)^{\boldsymbol{u}-\boldsymbol{\varsigma}}\mathbb{1}(\boldsymbol{u}\ge\boldsymbol{\varsigma}),$$

where $\Lambda_m = \{\boldsymbol{u} : [\boldsymbol{u}] = m\}$ and $\boldsymbol{u} \ge \boldsymbol{\varsigma} \Leftrightarrow u_1 \ge \varsigma_1, \cdots, u_d \ge \varsigma_d$.

(c) *Let $\tilde{\mathbf{p}}(\mathbf{x})$ be a piecewise polynomial of degree $\tilde{m}-1$ on the same partition Δ for some $\tilde{m} > m$, and assume $\tilde{m} \le S$. Then Assumption SA-5 is satisfied.*

It is a challenging task to construct orthogonal polynomial bases on non-rectangular domains, which makes Equation (SA-2.6) hard to be satisfied when employing partitioning estimators on general partitions. Thus, our more general characterization (and correction) of the bias is quite useful in this case.

SA-7. Discussion and Extensions.

SA-7.1. *Connecting Splines and Piecewise Polynomials.* There is an important connection between splines and piecewise polynomial bases. Essentially, the former can be viewed as a piecewise polynomial basis with certain continuity restrictions. Therefore, the estimators based on the two bases are linked by utilizing the well-known results about regressions with linear constraints. See [6] for an illustration of cubic splines as a special case of restricted least squares.

Formally, let us start with the (rotated) piecewise polynomials $\mathbf{p}(\cdot)$ discussed in Section SA-6. For expositional simplicity, we only discuss tensor-product partitions here. Clearly, $\mathbf{p}(\cdot)$ spans a vector space $\mathcal{P}$ containing all "piecewise" polynomials with degree no greater than $m-1$ on $\mathcal{X}$:

$$\mathcal{P} := \Big\{ s(\cdot) : \; s(\cdot) = \sum_{k=1}^{K} a_k p_k(\cdot), \; a_k \in \mathbb{R} \Big\}.$$

Functions in this space are continuous within each subrectangle, but might have "jumps" along boundaries of cells. Then by imposing certain continuity restrictions on functions in $\mathcal{P}$, we can construct a subspace $\mathcal{S} \subset \mathcal{P}$

$$\mathcal{S} := \Big\{ s(\cdot) : \; s(\cdot) = \sum_{k=1}^{K} a_k p_k(\cdot), \; a_k \in \mathbb{R}, \text{ and } \partial^{\boldsymbol{\varsigma}} s(\cdot) \in \mathcal{C}(\mathcal{X}), \; \forall [\boldsymbol{\varsigma}] \leq \iota \Big\}.$$

where $\iota \leq m-2$ is a positive integer that controls the smoothness of the basis (and thus the smoothness of the estimated function). Since the derivatives of polynomials are linear in coefficients, the continuity constraints are linear.

Now consider the following restricted least squares:

$$\widehat{\boldsymbol{\beta}}_{\mathtt{r}} = \underset{\mathbf{b} \in \mathbb{R}^K}{\arg\min} \, \|\mathbf{y} - \mathbf{P}\mathbf{b}\|^2 \quad \text{s.t. } \mathbf{R}\mathbf{b} = 0$$

where $\mathbf{P} = (\mathbf{p}(\mathbf{x}_1), \ldots, \mathbf{p}(\mathbf{x}_n))'$ and $\mathbf{R}$ is a $\vartheta_\iota \times K$ restriction matrix. ϑ_ι denotes the number of restrictions depending on the required smoothness ι. If there are no redundant constraints, $\mathbf{R}$ has full row rank. It is well known that given the restriction matrix $\mathbf{R}$, the least squares estimator can be written as

$$\widehat{\boldsymbol{\beta}}_{\mathtt{r}} = [\mathbf{I} - (\mathbf{P}'\mathbf{P})^{-1}\mathbf{R}'(\mathbf{R}(\mathbf{P}'\mathbf{P})^{-1}\mathbf{R}')^{-1}\mathbf{R}](\mathbf{P}'\mathbf{P})^{-1}\mathbf{P}'\mathbf{y}.$$

Since the unrestricted least squares estimator $\widehat{\boldsymbol{\beta}}_{\mathtt{ur}} = (\mathbf{P}'\mathbf{P})^{-1}\mathbf{P}'\mathbf{y}$, the above equation implies that $\widehat{\boldsymbol{\beta}}_{\mathtt{r}} = [\mathbf{I} - (\mathbf{P}'\mathbf{P})^{-1}\mathbf{R}'(\mathbf{R}(\mathbf{P}'\mathbf{P})^{-1}\mathbf{R}')^{-1}\mathbf{R}]\widehat{\boldsymbol{\beta}}_{\mathtt{ur}} =: (\mathbf{I} - \mathbf{U})\widehat{\boldsymbol{\beta}}_{\mathtt{ur}}$. Therefore,

$$\widehat{\mu}_{\mathtt{r}}(\mathbf{x}) = \mathbf{p}(\mathbf{x})'(\mathbf{I} - \mathbf{U})\widehat{\boldsymbol{\beta}}_{\mathtt{ur}} = \widehat{\mu}_{\mathtt{ur}}(\mathbf{x}) - \mathbf{p}(\mathbf{x})'\mathbf{U}\widehat{\boldsymbol{\beta}}_{\mathtt{ur}} \tag{SA-7.1}$$

where $\widehat{\mu}_{\mathtt{ur}}(\mathbf{x}) := \mathbf{p}(\mathbf{x})'\widehat{\boldsymbol{\beta}}_{\mathtt{ur}}$ is the unrestricted estimator. With this relation we can derive expressions of bias and variance for the restricted estimator. Clearly, the conditional variance $\mathbb{V}[\widehat{\mu}_{\mathtt{r}}(\mathbf{x})|\mathbf{X}] = \mathbf{p}(\mathbf{x})'(\mathbf{I} - \mathbf{U})\mathbb{V}[\widehat{\mu}_{\mathtt{ur}}(\mathbf{x})|\mathbf{X}](\mathbf{I} - \mathbf{U})'\mathbf{p}(\mathbf{x})$. On the other hand, as shown in Lemma SA-6.3, there exists $s^*(\mathbf{x}) := \mathbf{p}(\mathbf{x})'\boldsymbol{\beta}^*$ such that $\|\mu - s^* + \mathscr{B}_{m,\mathbf{0}}\|_{L_\infty(\mathcal{X})} = o(h^m)$. Hence

$$\begin{aligned}
&\mathbb{E}[\widehat{\mu}_{\mathbf{r}}(\mathbf{x})|\mathbf{X}] - \mu(\mathbf{x})\\
&= \mathbf{p}(\mathbf{x})'(\mathbf{I}-\mathbf{U})(\mathbf{P}'\mathbf{P})^{-1}\mathbf{P}'[\mathbf{P}\boldsymbol{\beta}^* - \boldsymbol{\mathscr{B}}_{m,\mathbf{0}} + o(h^m)] - \mu(\mathbf{x})\\
&= \mathscr{B}_{m,\mathbf{0}}(\mathbf{x}) - \mathbf{p}(\mathbf{x})'(\mathbf{I}-\mathbf{U})(\mathbf{P}'\mathbf{P})^{-1}\mathbf{P}'[\boldsymbol{\mathscr{B}}_{m,\mathbf{0}} + o(h^m)] - \mathbf{p}(\mathbf{x})'\mathbf{U}\boldsymbol{\beta}^* + o(h^m)
\end{aligned}$$

where $\boldsymbol{\mathscr{B}}_{m,\mathbf{0}} = (\mathscr{B}_{m,\mathbf{0}}(\mathbf{x}_1), \dots, \mathscr{B}_{m,\mathbf{0}}(\mathbf{x}_n))'$. $\mathbf{p}(\mathbf{x})'\mathbf{U}\boldsymbol{\beta}^*$ can be viewed as a measure of to what extent the continuity constraints are satisfied. When $s^*(\mathbf{x})$ is continuous up to order ι, $\mathbf{R}\boldsymbol{\beta}^*$ is exactly 0 and thus this term vanishes.

To repeat the previous analysis for such a restricted estimator, the main challenge is to analyze the asymptotic properties of the outer product of the restriction matrix. Once we know its limiting eigenvalue distributions (e.g., bounds on extreme eigenvalues), $\mathbf{U}$ can be properly bounded, and then the conclusions in previous sections may be established with similar proofs. Unfortunately, for general multidimensional cases, it is difficult and tedious to specify a non-redundant set of continuity constraints and analyze the eigenvalue distributions of $\mathbf{R}\mathbf{R}'$, thus damping the usefulness of such a method, whereas when $d = 1$, the restriction matrix is quite straightforward and well bounded.

Formally, let $\mathbf{R}$ be a restriction matrix corresponding to $\bar{\iota} = \iota + 1$ continuity constraints at each partitioning knot, implying that $s \in \mathcal{S}$ has ι continuous derivatives. We write the ℓth restriction at the kth knot as $\mathbf{r}_{k\ell}$, corresponding to the requirement that the $(\ell-1)$th derivatives of functions in $\mathcal{S}$ are continuous. Explicitly, the entire restriction matrix admits the following structure:

$$\mathbf{R} = \begin{bmatrix} \mathbf{r}'_{11} \\ \vdots \\ \mathbf{r}'_{1\bar{\iota}} \\ \vdots \\ \mathbf{r}'_{\kappa 1} \\ \vdots \\ \mathbf{r}'_{\kappa\bar{\iota}} \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix} a\\ a\\ a\end{matrix}}\right\} \text{Restrictions at the 1st knot} \\ \\ \left.\vphantom{\begin{matrix} a\\ a\\ a\end{matrix}}\right\} \text{Restrictions at the } \kappa\text{th knot} \end{matrix} \tag{SA-7.2}$$

As κ increases, the dimension of $\mathbf{R}$ also grows. Moreover, ι cannot exceed $m-2$ since when m continuity constraints are imposed, $\mathcal{S}$ degenerates to the space of global polynomials of degree no greater than $m-1$.

The asymptotic behavior of the restricted estimator is closely related to the extreme eigenvalues of $\mathbf{R}\mathbf{R}'$. For a general restriction matrix $\mathbf{R}$ with fixed dimensions, if it contains non-redundant constraints only, the minimum eigenvalue of $\mathbf{R}\mathbf{R}'$ is nonzero. When the number of constraints $\vartheta_\iota \to \infty$, however, the limit of the minimum eigenvalue does not have to be nonzero, and its limiting behavior depends on the specific structure of constraints. The next lemma shows that for the particular restricted estimators considered here, the eigenvalues of $\mathbf{R}\mathbf{R}'$ are indeed bounded and bounded away from zero uniformly over the number of knots for $\iota \leq m-2$.

LEMMA SA-7.1 (Restriction Matrix). *Let* $\mathbf{R}$ *be the restriction matrix described as* (SA-7.2) *with* $\iota \leq m-2$. *Then*

$$1 \lesssim \lambda_{\min}(\mathbf{R}\mathbf{R}') \leq \lambda_{\max}(\mathbf{R}\mathbf{R}') \lesssim 1.$$

The proof of this lemma employs the specific structure of the restriction matrix. Generally, the outer product of $\mathbf{R}$ takes the following form:

$$\mathbf{R}\mathbf{R}' = \begin{bmatrix} \mathbf{A} & \mathbf{B} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{B}' & \mathbf{A} & \mathbf{B} & \cdots & \mathbf{0} \\ \vdots & & \ddots & & \vdots \\ \mathbf{0} & \cdots & \mathbf{B}' & \mathbf{A} & \mathbf{B} \\ \mathbf{0} & \cdots & & \mathbf{B}' & \mathbf{A} \end{bmatrix} \tag{SA-7.3}$$

where

$$\mathbf{A} = \begin{bmatrix} \mathbf{r}_{11}'\mathbf{r}_{11} & \mathbf{r}_{11}'\mathbf{r}_{12} & \cdots & \mathbf{r}_{11}'\mathbf{r}_{1\bar{\iota}} \\ \mathbf{r}_{12}'\mathbf{r}_{11} & \mathbf{r}_{12}'\mathbf{r}_{12} & \cdots & \mathbf{r}_{12}'\mathbf{r}_{1\bar{\iota}} \\ \vdots & \cdots & & \vdots \\ \mathbf{r}_{1\bar{\iota}}'\mathbf{r}_{11} & \mathbf{r}_{1\bar{\iota}}'\mathbf{r}_{12} & \cdots & \mathbf{r}_{1\bar{\iota}}'\mathbf{r}_{1\bar{\iota}} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{r}_{11}'\mathbf{r}_{21} & \mathbf{r}_{11}'\mathbf{r}_{22} & \cdots & \mathbf{r}_{11}'\mathbf{r}_{2\bar{\iota}} \\ \mathbf{r}_{12}'\mathbf{r}_{21} & \mathbf{r}_{12}'\mathbf{r}_{22} & \cdots & \mathbf{r}_{12}'\mathbf{r}_{2\bar{\iota}} \\ \vdots & \cdots & & \vdots \\ \mathbf{r}_{1\bar{\iota}}'\mathbf{r}_{21} & \mathbf{r}_{1\bar{\iota}}'\mathbf{r}_{22} & \cdots & \mathbf{r}_{1\bar{\iota}}'\mathbf{r}_{2\bar{\iota}} \end{bmatrix}.$$

Importantly, the form described in (SA-7.3) is usually referred to as a (tridiagonal) block Toeplitz matrix, meaning that it is a tridiagonal block matrix containing blocks repeated down the diagonals. It is well known that its asymptotic eigenvalue distribution is characterized by the Fourier transform

$$\mathscr{T}_{\bar{\iota}}(\omega) = \mathbf{A} + (\mathbf{B} + \mathbf{B}')\cos\omega, \quad \omega \in [0, 2\pi].$$

As $\kappa \to \infty$, $\lambda_{\min}(\mathbf{R}\mathbf{R}')$ converges to the minimum attained by the minimum eigenvalue of $\mathscr{T}_{\bar{\iota}}(\omega)$ as a function of ω on $[0, 2\pi]$. Similarly, the limit of $\lambda_{\max}(\mathbf{R}\mathbf{R}')$ is the maximum attained by the maximum eigenvalue of $\mathscr{T}_{\bar{\iota}}(\omega)$.

SA-7.2. *Comparison of Bias Correction Approaches.* We make some comparison of the three bias correction approaches considered in this paper.

First, the higher-order correction ($j=1$) and least squares correction ($j=2$) are closely related. Let $\mathbf{P} = (\mathbf{p}(\mathbf{x}_1), \ldots, \mathbf{p}(\mathbf{x}_n))'$ and $\tilde{\mathbf{P}} = (\tilde{\mathbf{p}}(\mathbf{x}_1), \ldots, \tilde{\mathbf{p}}(\mathbf{x}_n))'$. Clearly,

$$\begin{aligned} \widehat{\mu}_2(\mathbf{x}) &= \widehat{\mu}_1(\mathbf{x}) + \mathbf{p}(\mathbf{x})'(\mathbf{P}'\mathbf{P})^{-1}\mathbf{P}'(\mathbf{I} - \tilde{\mathbf{P}}(\tilde{\mathbf{P}}'\tilde{\mathbf{P}})^{-1}\tilde{\mathbf{P}})\mathbf{y} \\ &= \widehat{\mu}_1(\mathbf{x}) + \mathbf{p}(\mathbf{x})'(\mathbf{P}'\mathbf{P})^{-1}\mathbf{P}'\mathbf{M}_{\tilde{\mathbf{p}}}\mathbf{y} \end{aligned}$$

where $\mathbf{M}_{\tilde{\mathbf{p}}} := \mathbf{I} - \tilde{\mathbf{P}}(\tilde{\mathbf{P}}'\tilde{\mathbf{P}})^{-1}\tilde{\mathbf{P}}'$. Importantly, when $\mathbf{p}$ and $\tilde{\mathbf{p}}$ generate nested models, i.e., there exists a transformation matrix $\boldsymbol{\Upsilon}$ such that $\mathbf{p}(\cdot) = \boldsymbol{\Upsilon}\tilde{\mathbf{p}}(\cdot)$, it is easy to see that $\mathbf{M}_{\tilde{\mathbf{p}}}\mathbf{P} = \mathbf{0}$. Thus the higher-order and least squares bias correction approaches are equivalent. When $\tilde{\mathbf{p}}$ and $\mathbf{p}$ are not nested bases, the two methods will typically differ in variance and bias.

To compare their variance, we generally have

$$\begin{aligned}\mathbb{V}[\widehat{\mu}_2(\mathbf{x})|\mathbf{X}] = \mathbb{V}[\widehat{\mu}_1(\mathbf{x})|\mathbf{X}] + \mathbb{V}[\mathbf{p}(\mathbf{x})'(\mathbf{P}'\mathbf{P})^{-1}\mathbf{P}'\mathbf{M}_{\tilde{\mathbf{p}}}\mathbf{y}|\mathbf{X}]\\ + 2\mathbb{C}\mathrm{ov}[\tilde{\mathbf{p}}(\mathbf{x})'(\tilde{\mathbf{P}}'\tilde{\mathbf{P}})^{-1}\tilde{\mathbf{P}}\mathbf{y},\ \mathbf{p}(\mathbf{x})'(\mathbf{P}'\mathbf{P})^{-1}\mathbf{P}'\mathbf{M}_{\tilde{\mathbf{p}}}\mathbf{y}|\mathbf{X}].\end{aligned}$$

When $\sigma^2(\mathbf{x}) = \sigma^2$, the covariance term is 0, and thus $\widehat{\mu}_2$ has variance no less than that of $\widehat{\mu}_1$. The same conclusion is true for asymptotic variance as shown in the proof of Lemma SA-2.4.

Regarding their bias, let $\mathfrak{B}_{\tilde{m},\mathbf{0}}(\mathbf{x}) = \mathbb{E}[\widehat{\mu}_1(\mathbf{x})|\mathbf{X}] - \mu(\mathbf{x})$ denote the conditional bias of $\widehat{\mu}_1(\mathbf{x})$, and $\boldsymbol{\mathfrak{B}}_{\tilde{m},\mathbf{0}} := (\mathfrak{B}_{\tilde{m},\mathbf{0}}(\mathbf{x}_1), \ldots, \mathfrak{B}_{\tilde{m},\mathbf{0}}(\mathbf{x}_n))'$. Then $\mathbb{E}[\widehat{\mu}_2(\mathbf{x})|\mathbf{X}] - \mu(\mathbf{x}) = \mathfrak{B}_{\tilde{m},\mathbf{0}}(\mathbf{x}) - \mathbf{p}(\mathbf{x})'(\mathbf{P}'\mathbf{P})^{-1}\mathbf{P}'\boldsymbol{\mathfrak{B}}_{\tilde{m},\mathbf{0}}$. Clearly, the second term will asymptotically get close to the projection of $\mathfrak{B}_{\tilde{m},\mathbf{0}}(\mathbf{x})$ onto the space spanned by $\mathbf{p}$:

$$\mathscr{L}_{\mathbf{p}}[\mathfrak{B}_{\tilde{m},\mathbf{0}}](\mathbf{x}) := \mathbf{p}(\mathbf{x})'(\mathbb{E}[\mathbf{p}(\mathbf{x})\mathbf{p}(\mathbf{x})'])^{-1}\mathbb{E}[\mathbf{p}(\mathbf{x})\mathfrak{B}_{\tilde{m},\mathbf{0}}(\mathbf{x})]$$

where $\mathscr{L}_{\mathbf{p}}[\cdot]$ denotes the projection operator. Therefore, when $\mathbf{p}$ and $\tilde{\mathbf{p}}$ are not nested bases, we typically have the bias of $\widehat{\mu}_2$ no greater than that of $\widehat{\mu}_1$ in terms of $\|\cdot\|_{F,L_2(\mathcal{X})}$, where for a real-valued function $g(\cdot)$ on $\mathcal{X}$, $\|g\|_{F,L_2(\mathcal{X})} = (\int_{\mathcal{X}} |g(\mathbf{x})|^2 dF(\mathbf{x}))^{1/2}$.

According to the discussion above, the higher-order and least squares bias correction approaches do not dominate each other in general, and whether one is preferred to the other depends on the data generating process and the relation between the two approximation spaces (or more precisely, the approximation power of $\tilde{\mathbf{p}}$ for functions in the linear span of $\mathbf{p}$). Then a natural question follows: is there an optimal weighting scheme when $\mathbf{p}$ and $\tilde{\mathbf{p}}$ are not nested? Again, assume ε_i's are homoskedastic for simplicity, and we take a weighted average of $\widehat{\mu}_1$ and $\widehat{\mu}_2$: $\widehat{\mu}_{w,\mathtt{bc}} := w\widehat{\mu}_2 + (1-w)\widehat{\mu}_1$ where $w \in [0, 1]$. Then using a conclusion in the proof of Lemma SA-2.4, the change in the integrated asymptotic variance (weighted by the design density $f(\mathbf{x})$) is

$$w^2\sigma^2 \int_{\mathcal{X}} \mathbf{p}(\mathbf{x})'\mathbf{Q}_m^{-1}(\mathbf{Q}_m - \mathbf{Q}_{m,\tilde{m}}\mathbf{Q}_{\tilde{m}}^{-1}\mathbf{Q}_{m,\tilde{m}}')\mathbf{Q}_m^{-1}\mathbf{p}(\mathbf{x})f(\mathbf{x})d\mathbf{x} =: w^2\bar{\mathscr{V}}.$$

On the other hand, by the property of projection operator, the change in the integrated squared bias (weighted by the design density $f(\mathbf{x})$) is

$$(w^2 - 2w)\int_{\mathcal{X}} \left(\mathscr{L}_{\mathbf{p}}[\mathfrak{B}_{\tilde{m},\mathbf{0}}](\mathbf{x})\right)^2 f(\mathbf{x})d\mathbf{x} =: (w^2 - 2w)\bar{\mathscr{B}}.$$

It is easy to see the optimal weight is $w^* = \bar{\mathscr{B}}/(\bar{\mathscr{B}} + \bar{\mathscr{V}})$. Clearly, when variance is very small (e.g., σ^2 is small), w^* is close to 1 and $\widehat{\mu}_2$ is preferred, whereas when bias is small, w^* is close to 0 and one may want to use $\widehat{\mu}_1$.

Next, the comparison of plug-in bias correction with the other two is more complicated since $\widehat{\mu}_3$ generally cannot be viewed as a regression estimator with additional covariates, and the covariance between $\widehat{\mu}_0$ and the estimated bias does not vanish. For piecewise polynomials, however, all three bias correction approaches are simply equivalent under certain conditions.

To see this, suppose $\mathbf{p}$ and $\tilde{\mathbf{p}}$ are constructed on the same partitioning scheme Δ, but the order of basis increases from m to $m+1$. $\widehat{\mu}_0$ and $\widehat{\mu}_1$ are linked by (SA-7.1) since $\widehat{\mu}_0$ can be viewed as a restricted regression estimator compared with $\widehat{\mu}_1$. Specifically, one can construct a polynomial series of order $m+1$ (with degree no greater than m) within each cell $\delta \in \Delta$, and then implement a local regression restricting the coefficients of the polynomial terms of degree m to be 0. Then the restriction matrix in this case can be written as $\mathbf{R} = [\mathbf{0} \quad \mathbf{I}_\vartheta]$ where ϑ denotes the number of polynomial terms of degree m and $\mathbf{R}$ is a $\vartheta \times \tilde{K}$ matrix. Plug it in (SA-7.1), and then use the formula for matrix inverse in block form to obtain $(\tilde{\mathbf{P}}'\tilde{\mathbf{P}})^{-1}$. It is easy to see that the second term on the RHS of (SA-7.1), $\mathbf{p}(\mathbf{x})'\mathbf{M}\hat{\boldsymbol{\beta}}_{\mathtt{ur}}$, is exactly the same as the leading bias derived in [7] (see their proof of Theorem 3) with the mth derivative estimated by piecewise polynomials of order $m+1$. As explained in the proof of Lemma SA-6.3, the leading approximation error can be alternatively expressed in terms of Legendre polynomials. Asymptotically, the two expressions are equivalent since the "locally" orthogonalized polynomials of degree m will converge to the mth Legendre polynomials given by Lemma SA-6.3.

For splines or wavelets, we do not have the above equivalence in general, since bases of different orders do not generate nested spaces, and the relative performance of the three approaches depends on the relation between these approximation spaces.

SA-8. Implementation Details. In this section, we briefly discuss implementation details about choosing the IMSE-optimal tuning parameters. We restrict our attention to tensor-product partitions with the same number of knots used in every dimension. Thus the tuning parameter reduces to a scalar κ which denotes the number of subintervals used in every dimension. Also, we let the weighting function $w(\mathbf{x})$ be the density of $\mathbf{x}_i$. We offer two approaches: rule-of-thumb (ROT) and direct plug-in (DPI).

SA-8.1. *Rule-of-Thumb Choice.* The rule-of-thumb choice is based on the special case considered in Theorem 4.2 of the main paper. Specifically, assume $\mathbf{q} = \mathbf{0}$ and knots are evenly spaced. The implementation steps are summarized as follows.

- **Preliminary regression**. Implement a preliminary regression to estimate $\mu(\cdot)$ by a global polynomial of order $(m+2)$. Denote this estimate by $\hat{\mu}_{\mathtt{pre}}(\cdot)$.
- **Bias constant**: Use this preliminary regression to obtain an estimate of the mth derivatives of $\mu(\cdot)$, i.e., $\widehat{\partial^{\boldsymbol{u}}\mu}(\cdot) = \partial^{\boldsymbol{u}}\widehat{\mu}_{\mathtt{pre}}(\cdot)$, for each $\boldsymbol{u} \in \Lambda_m$. Then an estimate of the bias constant is

$$\widehat{\mathscr{B}}_{\boldsymbol{u}_1,\boldsymbol{u}_2,\mathbf{0}} = \eta_{\boldsymbol{u}_1,\boldsymbol{u}_2,\mathbf{0}} \times \frac{1}{n}\sum_{i=1}^{n} \widehat{\partial^{\boldsymbol{u}_1}\mu}(\mathbf{x}_i)\widehat{\partial^{\boldsymbol{u}_2}\mu}(\mathbf{x}_i).$$

- **Variance constant**. Implement another regression using a global polynomial of order $(m+2)$ to estimate $\mathbb{E}[y_i^2|\mathbf{x}_i = \mathbf{x}]$. Combining it with $\hat{\mu}_{\texttt{pre}}(\cdot)$, we can obtain an estimate of the conditional variance function, denoted by $\hat{\sigma}^2(\cdot)$, since $\sigma^2(\mathbf{x}) = \mathbb{E}[y_i^2|\mathbf{x}_i = \mathbf{x}] - (\mathbb{E}[y_i|\mathbf{x}_i = \mathbf{x}])^2$. Then an estimate of the variance constant is

$$\widehat{\mathscr{V}}_{\mathbf{0}} = \begin{cases} \frac{1}{n}\sum_{i=1}^n \hat{\sigma}^2(\mathbf{x}_i) & \text{for splines and wavelets,} \\ \binom{d+m-1}{m-1} \times \frac{1}{n}\sum_{i=1}^n \hat{\sigma}^2(\mathbf{x}_i) & \text{for piecewise polynomials.} \end{cases}$$

- **Rule-of-thumb** $\hat{\kappa}_{\texttt{ROT}}$. Using the above results, a simple rule-of-thumb choice of κ is

$$\hat{\kappa}_{\texttt{ROT}} = \left\lceil \left(\frac{2m \sum_{\boldsymbol{u}_1,\boldsymbol{u}_2\in\Lambda_m} \widehat{\mathscr{B}}_{\boldsymbol{u}_1,\boldsymbol{u}_2,\mathbf{0}}}{d\widehat{\mathscr{V}}_{\mathbf{0}}} \right)^{\frac{1}{2m+d}} n^{\frac{1}{2m+d}} \right\rceil.$$

Rather than assume a uniform design and evenly-spaced partition, one may use a trimmed-from-below Gaussian reference model to estimate the density of $\mathbf{x}_i$ and adjust the variance and bias constants based on Theorem 4.2. Clearly, this choice of κ is derived based on simplifying assumptions, but it still has the correct rate ($\asymp n^{\frac{1}{2m+d}}$) even in other cases.

SA-8.2. *Direct Plug-in Choice.* A direct plug-in (DPI) procedure is summarized in the following.

- **Preliminary choice of** κ: Implement the ROT procedure to obtain $\hat{\kappa}_{\texttt{ROT}}$.
- **Preliminary regression**. Given the user-specified basis (splines, wavelets, or piecewise polynomials), knot placement scheme ("uniform" or "quantile") and rule-of-thumb choice $\hat{\kappa}_{\texttt{ROT}}$, implement a series regression of order $(m+1)$ to obtain derivative estimates for every $\boldsymbol{u} \in \Lambda_m$. Denote this preliminary estimate by $\widehat{\partial^{\boldsymbol{u}}\mu}_{\texttt{pre}}(\cdot)$.
- **Bias constant**. Construct an estimate $\widehat{\mathscr{B}}_{m,\mathbf{q}}(\cdot)$ of the leading error $\mathscr{B}_{m,\mathbf{q}}(\cdot)$ simply by replacing $\partial^{\boldsymbol{u}}\mu(\cdot)$ by $\widehat{\partial^{\boldsymbol{u}}\mu}_{\texttt{pre}}(\cdot)$. $\widehat{\mathscr{B}}_{m,\mathbf{0}}(\cdot)$ can be obtained similarly. Then we use the pre-asymptotic version of the conditional bias to estimate the bias constant:

$$\widehat{\mathscr{B}}_{\kappa,\mathbf{q}} = \frac{1}{n}\sum_{i=1}^n \Big(\widehat{\mathscr{B}}_{m,\mathbf{q}}(\mathbf{x}_i) - \widehat{\boldsymbol{\gamma}}_{\mathbf{q},0}(\mathbf{x}_i)'\mathbb{E}_n[\boldsymbol{\Pi}_0(\mathbf{x}_i)\widehat{\mathscr{B}}_{m,\mathbf{0}}(\mathbf{x}_i)] \Big)^2.$$

- **Variance constant**. Implement a series regression of order m with $\kappa = \hat{\kappa}_{\texttt{ROT}}$, and then use the pre-asymptotic version of the conditional variance to obtain an estimate of the variance constant. Specifically,

$$\widehat{\mathscr{V}}_{\kappa,\mathbf{q}} = \frac{1}{n}\sum_{i=1}^n \widehat{\boldsymbol{\gamma}}_{\mathbf{q},0}(\mathbf{x}_i)'\widehat{\boldsymbol{\Sigma}}_0\widehat{\boldsymbol{\gamma}}_{\mathbf{q},0}(\mathbf{x}_i), \quad \widehat{\boldsymbol{\Sigma}}_0 = \mathbb{E}_n[\boldsymbol{\Pi}_0(\mathbf{x}_i)\boldsymbol{\Pi}_0(\mathbf{x}_i)'\widehat{\varepsilon}_{i,0}^2],$$

where $\widehat{\varepsilon}_{i,0}$'s are regression residuals. Different weighting schemes for residuals may be used, leading to various heteroskedasticity-consistent variance estimates.

- **Direct plug-in** $\hat{\kappa}_{\mathtt{DPI}}$. Combining these results, a direct plug-in choice of κ is

$$\hat{\kappa}_{\mathtt{DPI}} = \left\lceil \left(\frac{2(m-[\mathbf{q}])\kappa_{\mathtt{ROT}}^{2(m-[\mathbf{q}])}\widehat{\mathscr{B}}_{\kappa,\mathbf{q}}}{(d+2[\mathbf{q}])\kappa_{\mathtt{ROT}}^{-(d+2[\mathbf{q}])}\widehat{\mathscr{V}}_{\kappa,\mathbf{q}}} \right)^{\frac{1}{2m+d}} n^{\frac{1}{2m+d}} \right\rceil.$$

SA-9. Simulations. In this section, we present detailed simulation results. We consider the following regression functions:

$$\text{Model 1: } \mu(x) = \sin\left(\frac{\pi}{2}(2x-1)\right) / (1 + 2(2x-1)^2(\text{sign}(2x-1)+1))$$
$$\text{Model 2: } \mu(x) = \sin\left(\frac{3\pi}{2}(2x-1)\right) / (1 + 18(2x-1)^2(\text{sign}(2x-1)+1))$$
$$\text{Model 3: } \mu(x) = 2x - 1 + 5\phi(20x-10)$$
$$\text{Model 4: } \mu(\mathbf{x}) = \sin(5x_1)\sin(10x_2)$$
$$\text{Model 5: } \mu(\mathbf{x}) = (1-(4x_1-2)^2)^2\sin(5x_2)/5$$
$$\text{Model 6: } \mu(\mathbf{x}) = (1-(4x_1-2)^2)^2(2x_2-1)(x_3-0.5)$$
$$\text{Model 7: } \mu(\mathbf{x}) = \tau(x_1)\tau(x_2)\tau(x_3)$$

where $\text{sign}(x) = -1, 0,$ or 1 if $x<0$, $x=0$, or $x>0$ respectively, $\phi(\cdot)$ is the standard Normal density function, and $\tau(x) = (x-0.5) + 8(x-0.5)^2 + 6(x-0.5)^3 - 30(x-0.5)^4 - 30(x-0.5)^5$.

Given each regression function, we generate $(y_i, \mathbf{x}_i)_{i=1,\dots,n}$ by $y_i = \mu(\mathbf{x}_i) + \varepsilon_i$ where $\mathbf{x}_i \sim i.i.d.\ \mathsf{U}[0,1]^d$ and $\varepsilon_i \sim i.i.d.\ \mathsf{N}(0,1)$. For each model, we generate $5,000$ simulated datasets with $n = 1,000$.

We first present results for (tensor-product) spline regressions in Tables SA-1–SA-23. For each model, we use linear splines ($m = 2$) to form the classical point estimates. For robust inference, we use quadratic splines ($\tilde{m} = 3$) to implement bias correction. Both evenly-spaced and quantile-spaced knot placements are considered, and for simplicity point estimators and bias correction employ the same knot placements ($\Delta = \tilde{\Delta}$).

For each model, we present three sets of simulation evidence. First, for three fixed points, we calculate the (simulated) root mean squared error (RMSE), coverage rate (CR) and average confidence interval length (AL). The nominal coverage is set to be 95%. Second, to evaluate the performance of our rule-of-thumb (ROT) and direct plug-in (DPI) knot selection procedures, we show some basic summary statistics (mean, median, standard deviation, etc.) of the selected number of knots ($\hat{\kappa}_{\mathtt{ROT}}$ and $\hat{\kappa}_{\mathtt{DPI}}$). Third, for uniform confidence bands, we calculate: the proportion of values covered with probability at least 95% (CP), average coverage errors (ACE), and average width of confidence band (AW), and uniform coverage rate (UCR). The quantile estimation based on the plug-in and bootstrap methods uses 1000 random draws conditional on the data for each simulated dataset.

Next, we evaluate the performance of (Daubechies) wavelet regressions. Several distinctive features of wavelets should be noted. First, the tuning parameter of wavelets (s) is referred to as "resolution level", which relates to the number of knots κ by $\kappa = 2^s$. Thus, as s increases, the number of series terms grows very fast. This issue is even more severe for tensor-product wavelets. Given our relatively small sample size, we restrict the wavelet-based simulations to Model 1-5. Second, due to the lack of smoothness of low-order wavelet basis, the plug-in bias correction approach ($j = 3$) is not feasible unless very high-order wavelet basis is used to estimate the derivatives in the leading bias, which may not be of practical interest. Thus, we only report results based on higher-order and least-squares bias correction (as well as classical estimators).

Results for wavelets are reported in Table SA-24–SA-34, where Daubechies (father) wavelets of order 2 ($m = 2$) are used to form classical estimators, and Daubechies wavelets of order 3 ($\tilde{m} = 3$) for bias correction. In addition, as explained above, the IMSE-optimal resolution level is related to the number of knots by $s_{\texttt{IMSE}} = \log_2 \kappa_{\texttt{IMSE}}$. Similarly, for the estimated resolution levels, $\hat{s}_{\texttt{ROT}} = \log_2 \hat{\kappa}_{\texttt{ROT}}$, and $\hat{s}_{\texttt{DPI}} = \log_2 \hat{\kappa}_{\texttt{DPI}}$.

Our simulation results show that the bias correction methods perform generally well in both pointwise and uniform inference. The following are some practical guidance and caveats. First, as discussed in Section SA-7, no bias correction approach among the three dominates the others in general. When high-order bias is large, their differences may be more pronounced. Second, our simulation shows that when $d > 1$, our estimators still perform relatively well if a small number of knots are used. If not, as expected, their performance could be poor since the number of regressors explodes in such cases. Third, the ROT selection procedure used in the simulation study employs a global polynomial regression of degree $m + 2$ to form the estimates of derivatives in the leading bias, which may be too conservative for highly nonlinear models. In such cases, a higher-order global polynomial regression may offer a better initial choice of the tuning parameter and improve the performance of DPI selection.

In the end, we present a figure (see Figure SA-1) as a visual illustration, which compares confidence bands based on different estimators suggested in our paper. We use Model 1 to generate a simulated dataset, and both plug-in approximation and wild bootstrap are used in constructing bands.

SA-10. Proofs.

SA-10.1. *Proof of Lemma SA-2.1.*

PROOF. We first prove the boundedness of the eigenvalues of $\mathbf{Q}_m$. We use the fact that $\lambda_{\max}(\mathbf{Q}_m) = \max_{\mathbf{a}'\mathbf{a}=1} \mathbf{a}'\mathbf{Q}_m\mathbf{a}$ and $\lambda_{\min}(\mathbf{Q}_m) = \min_{\mathbf{a}'\mathbf{a}=1} \mathbf{a}'\mathbf{Q}_m\mathbf{a}$. By definition of $\mathbf{Q}_m$,

$$\mathbf{a}'\mathbf{Q}_m\mathbf{a} = \mathbf{a}'\mathbb{E}[\mathbf{p}(\mathbf{x}_i)\mathbf{p}(\mathbf{x}_i)']\mathbf{a} = \int_{\mathcal{X}} \Big(\sum_{k=1}^{K} a_k p_k(\mathbf{x})\Big)^2 f(\mathbf{x})\, d\mathbf{x} =: \|s(\mathbf{x})\|^2_{F,L_2(\mathcal{X})}$$

where $s(\mathbf{x}) = \sum_{k=1}^{K} a_k p_k(\mathbf{x})$. By Assumption SA-1, $\|s(\mathbf{x})\|^2_{L_2(\mathcal{X})} \lesssim \|s(\mathbf{x})\|^2_{F,L_2(\mathcal{X})} \lesssim \|s(\mathbf{x})\|^2_{L_2(\mathcal{X})}$. By Assumption SA-3(a), the number of basis functions in $\mathbf{p}(\cdot)$ which are active on a generic cell δ_l, $l = 1, \dots, \bar{\kappa}$, is bounded by a constant. Denoted them by $(\bar{p}_{l,1}, \cdots, \bar{p}_{l,M_l})'$, where M_l may vary across l. It follows from Assumption SA-3(c) that $s(\mathbf{x})^2 = (\sum_{k=1}^{M_l} a_k \bar{p}_{l,k}(\mathbf{x}))^2 \lesssim \sum_{k=1}^{M_l} a_k^2$ for all $\mathbf{x} \in \delta_l$. Taking integral and summing over all δ_l, we have $\|s(\mathbf{x})\|^2_{L_2(\mathcal{X})} \lesssim h^d$, and the upper bound on $\lambda_{\max}(\mathbf{Q}_m)$ follows. On the other hand, by Assumption SA-3(b), $\|s(\mathbf{x})\|^2_{L_2(\mathcal{H}_k)} \gtrsim a_k^2 h^d$. Then taking sum over all $\mathcal{H}_k$, $\|s(\mathbf{x})\|^2_{L_2(\mathcal{X})} \gtrsim h^d$, and the lower bound on $\lambda_{\min}(\mathbf{Q}_m)$ follows.

To derive the convergence rate of $\widehat{\mathbf{Q}}_m$, let $\alpha_{k,l} = \frac{1}{n}\sum_{i=1}^{n} \alpha_{k,l}(i)$ be the (k,l)th element of $(\widehat{\mathbf{Q}}_m - \mathbf{Q}_m)$, where $\alpha_{k,l}(i) := p_k(\mathbf{x}_i)p_l(\mathbf{x}_i) - \mathbb{E}[p_k(\mathbf{x}_i)p_l(\mathbf{x}_i)]$. It follows from Assumption SA-1 and SA-3 that $\alpha_{j,l}$ is the sum of n independent random variables with zero means, $|\alpha_{k,l}(i)| \lesssim 1$ uniformly over i, k and l, and thus $\mathbb{V}[\alpha_{k,l}(i)] \lesssim h^d/n$. By Bernstein's inequality, for every $\vartheta > 0$,

$$\mathbb{P}(|\alpha_{k,l}| > \vartheta) \le 2\exp\left(-\frac{\vartheta^2/2}{C_1 h^d/n + C_2\vartheta/(3n)}\right).$$

By Assumption SA-3, $(\widehat{\mathbf{Q}}_m - \mathbf{Q}_m)$ only has a finite number of nonzeros in any row or column, and thus for every $\vartheta > 0$,

$$\mathbb{P}(\max_{k,l} |\alpha_{k,l}| > \vartheta) \le 2CK\exp\left(-\frac{\vartheta^2/2}{C_1 h^d/n + C_2\vartheta/(3n)}\right).$$

Then $\max_{k,l} |\alpha_{k,l}| \lesssim_{\mathbb{P}} h^d\sqrt{\log n/(nh^d)}$, which suffices to show that

$$\|\widehat{\mathbf{Q}}_m - \mathbf{Q}_m\|_\infty \lesssim_{\mathbb{P}} h^d\sqrt{\log n/(nh^d)} \quad \text{and} \quad \|\widehat{\mathbf{Q}}_m - \mathbf{Q}_m\|_1 \lesssim_{\mathbb{P}} h^d\sqrt{\log n/(nh^d)}.$$

By the relation between induced operator norms, $\|\widehat{\mathbf{Q}}_m - \mathbf{Q}_m\| \lesssim_{\mathbb{P}} h^d\sqrt{\log n/(nh^d)}$. Hence when $\log n/(nh^d) = o(1)$, $\|\widehat{\mathbf{Q}}_m - \mathbf{Q}_m\| = o_{\mathbb{P}}(h^d)$. Notice that for any vector $\mathbf{a} \in \mathbb{R}^K$ such that $\mathbf{a}'\mathbf{a} = 1$,

$$\|\widehat{\mathbf{Q}}_m - \mathbf{Q}_m\| \ge |\mathbf{a}'(\widehat{\mathbf{Q}}_m - \mathbf{Q}_m)\mathbf{a}| = |\mathbf{a}'\widehat{\mathbf{Q}}_m\mathbf{a} - \mathbf{a}'\mathbf{Q}_m\mathbf{a}|.$$

Since $h^d \lesssim \mathbf{a}'\mathbf{Q}_m\mathbf{a} \lesssim h^d$, this suffices to show that $\|\widehat{\mathbf{Q}}_m\| \lesssim_{\mathbb{P}} h^d$, and with probability approaching one, $\lambda_{\min}(\widehat{\mathbf{Q}}_m) \gtrsim h^d$.

Next, we show $\|\widehat{\mathbf{Q}}_m^{-1}\|_\infty \lesssim_{\mathbb{P}} h^{-d}$. In the univariate case, it is easy to see that by Assumption SA-3, $\widehat{\mathbf{Q}}_m$ is a banded matrix with a finite bandwidth (independent of n). In the multidimensional case $\widehat{\mathbf{Q}}_m$ will take a "multi-layer" block banded matrix form. Specifically, let $\mathbf{A}_{kl}$ be the generic (k,l)th block of $\widehat{\mathbf{Q}}_m$. $\mathbf{A}_{kl} = \mathbf{0}$ if $|k-l| > L$ for some integer $L > 0$, with each nonzero $\mathbf{A}_{kl}$ banded ("two layers") or taking another block banded structure ("more than two layers"). To see this, we first arrange the ordering of basis functions appropriately by "rectangularizing" the partition Δ. By Assumption SA-2, Δ is quasi-uniform and there exists a universal measure of mesh size h. Construct an initial rectangular partition covering $\mathcal{X}$, which is formed as tensor products of intervals of length h. A generic cell in this partition is indexed by a d-tuple $(l_1, \dots, l_d)$. Then for each cell in this partition, take its intersection with $\mathcal{X}$ and exclude all cells outside of $\mathcal{X}$. Thus we construct a "trimmed" tensor-product partition $\Delta^{\mathtt{rec}} = \{\delta^{\mathtt{rec}}_{l_1\dots l_d}\}$. Clearly, each element $\delta \in \Delta$ is covered by a finite number of cells in $\Delta^{\mathtt{rec}}$. On the other hand, each $\delta^{\mathtt{rec}} \in \Delta^{\mathtt{rec}}$ is also overlapped with a finite number of cells in Δ (if not, cells in Δ overlapping with $\delta^{\mathtt{rec}}$ cannot be covered by a ball of radius $2h$). Arrange these cells by first increasing l_d with other l_ℓ's fixed at the lowest values and then increasing l_ℓ's sequentially. Then arrange the basis functions in $\mathbf{p}(\cdot)$ according to their supports. Specifically, start with basis functions that are active on the first cell, and then arrange those functions that are active on the second and have not been included yet. Continue this procedure until the functions active on the last cell have been included. According to this particular ordering, the Gram of this basis has the same banded structure as that of tensor-product local basis on tensor-product partitions. The nested banded structure involves at most d layers. The bandwidth at each layer may be different, but is bounded by a universal constant $\bar{L}$ by Assumption SA-3.

In the one-dimensional case, it follows from [13, Theorem 2.2] that $\|\widehat{\mathbf{Q}}_m^{-1}\|_\infty \lesssim_{\mathbb{P}} h^{-1}$. In the multidimensional case the original proof needs to be slightly modified. We only prove for the case when $\widehat{\mathbf{Q}}_m$ is block banded with banded blocks (two layers of banded structures). The general case follows similarly.

For universal constants C_1 and C_2, with probability approaching one, $\lambda_{\min}(\widehat{\mathbf{Q}}_m) \geq C_1 h^d$ and $\lambda_{\max}(\widehat{\mathbf{Q}}_m) \leq C_2 h^d$. Hence for any vector $\mathbf{a} \in \mathbb{R}^K$ such that $\mathbf{a}'\mathbf{a} = 1$, there are some constants C_3, C_4 and C_5 such that with probability approaching one,

$$0 < C_3 \leq \frac{\mathbf{a}'\widehat{\mathbf{Q}}_m\mathbf{a}}{C_5 h^d} \leq C_4 < 1.$$

Hence $\mathbf{\Psi} := \mathbf{I}_K - \widehat{\mathbf{Q}}_m/(C_5 h^d)$ is a block banded matrix with banded blocks satisfying $\|\mathbf{\Psi}\| < 1$ with probability approaching one. Therefore, we can write $C_5 h^d \widehat{\mathbf{Q}}_m^{-1} = \sum_{l=1}^\infty \mathbf{\Psi}^l$. The (s,t)th entry of $\widehat{\mathbf{Q}}_m^{-1}$, denoted by $\alpha_{s,t}$, is an element of $\sum_{l=1}^\infty \mathbf{\Psi}^l$.

Hence,

$$|\alpha_{s,t}| \leq \sum_{l=\chi(s,t,\bar{L})}^{\infty} \|\mathbf{\Psi}^l\| = \frac{\|\mathbf{\Psi}\|^{\chi(s,t,L)}}{1-\|\mathbf{\Psi}\|}$$

where $\chi(s,t,\bar{L})$ is a number depending on the row index s, column index t and the upper bound on bandwidths $\bar{L}$. We further denote the block row index and column index of $\alpha_{s,t}$ as r_s and r_t, and the row index and column index within the block containing it as ι_s and ι_t. $|r_s - r_t|$ and $|\iota_s - \iota_t|$ measure how far away $\alpha_{s,t}$ is from the diagonals of the entire matrix and the block it belongs to. As in the one-dimensional case, the first few products of $\mathbf{\Psi}$ do not contribute to off-diagonal blocks of the inverse matrix. As $|r_s - r_t|$ increases, $\chi(s,t,\bar{L})$ also gets larger. Meanwhile, since each block of $\widehat{\mathbf{Q}}_m$ is also banded, $\chi(s,t,\bar{L})$ also increase with $|\iota_s - \iota_t|$. By the same argument as in [13], $|\alpha_{s,t}| \leq (1-\|\mathbf{\Psi}\|)^{-1}\|\mathbf{\Psi}\|^{|r_s-r_t|/C_6+|\iota_s-\iota_t|/C_6}$ where C_6 is some constant depending on $\bar{L}$. Thus $\alpha_{s,t}$ exponentially decays when $|r_s - r_t|$ or $|\iota_s - \iota_t|$ becomes large. By convergence of geometric series, $\|\widehat{\mathbf{Q}}_m^{-1}\| \lesssim_{\mathbb{P}} h^{-d}$. $\|\mathbf{Q}_m^{-1}\|_\infty \lesssim h^{-d}$ follows similarly.

Finally, the above results immediately imply that

$$\|\widehat{\mathbf{Q}}_m^{-1} - \mathbf{Q}_m^{-1}\|_\infty \leq \|\widehat{\mathbf{Q}}_m^{-1}\|_\infty \|\widehat{\mathbf{Q}}_m - \mathbf{Q}_m\|_\infty \|\mathbf{Q}_m^{-1}\|_\infty \lesssim_{\mathbb{P}} h^{-d}\sqrt{\log n/(nh^d)}, \quad \text{and}$$
$$\|\widehat{\mathbf{Q}}_m^{-1} - \mathbf{Q}_m^{-1}\| \leq \|\widehat{\mathbf{Q}}_m^{-1}\| \|\widehat{\mathbf{Q}}_m - \mathbf{Q}_m\| \|\mathbf{Q}_m^{-1}\| \lesssim_{\mathbb{P}} h^{-d}\sqrt{\log n/(nh^d)}.$$

The bound on $\|\mathbb{E}_n[\mathbf{p}(\mathbf{x}_i)\mathbf{p}(\mathbf{x}_i)'\sigma^2(\mathbf{x}_i)] - \mathbb{E}[\mathbf{p}(\mathbf{x}_i)\mathbf{p}(\mathbf{x}_i)'\sigma^2(\mathbf{x}_i)]\|$ follows similarly, since $\sigma^2(\mathbf{x}) \lesssim 1$ uniformly over $\mathbf{x} \in \mathcal{X}$. □

SA-10.2. *Proof of Lemma SA-2.2.*

PROOF. We first prove the uniform bound on the conditional bias. By Assumption SA-4 we can find $s^* \in \mathcal{S}_{\Delta,m}$ such that $\sup_{\mathbf{x}\in\mathcal{X}} |\partial^{\mathbf{q}}\mu(\mathbf{x}) - \partial^{\mathbf{q}}s^*(\mathbf{x})| \lesssim h^{m-[\mathbf{q}]}$. Since

$$\begin{aligned}\mathbb{E}[\widehat{\partial^{\mathbf{q}}\mu}_0(\mathbf{x})|\mathbf{X}] - \partial^{\mathbf{q}}\mu(\mathbf{x}) &= \partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})'\widehat{\mathbf{Q}}_m^{-1}\mathbb{E}_n[\mathbf{p}(\mathbf{x}_i)\mu(\mathbf{x}_i)] - \partial^{\mathbf{q}}\mu(\mathbf{x}) \\ &= \partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})'\widehat{\mathbf{Q}}_m^{-1}\mathbb{E}_n[\mathbf{p}(\mathbf{x}_i)(\mu(\mathbf{x}_i) - s^*(\mathbf{x}_i))] + \partial^{\mathbf{q}}s^*(\mathbf{x}) - \partial^{\mathbf{q}}\mu(\mathbf{x}),\end{aligned}$$

it suffices to show that the first term in the second line is properly bounded uniformly over $\mathbf{x} \in \mathcal{X}$. It follows from Lemma SA-2.1 and Assumption SA-3 that

$$\sup_{\mathbf{x}\in\mathcal{X}} \left|\partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})'\widehat{\mathbf{Q}}_m^{-1}\mathbb{E}_n[\mathbf{p}(\mathbf{x}_i)(\mu(\mathbf{x}_i) - s^*(\mathbf{x}_i))]\right| \lesssim_{\mathbb{P}} h^{-[\mathbf{q}]-d}\|\mathbb{E}_n[\mathbf{p}(\mathbf{x}_i)(\mu(\mathbf{x}_i) - s^*(\mathbf{x}_i))]\|_\infty.$$

By Assumption SA-3 and SA-4, $\max_{1\leq k\leq K} |\mathbb{E}[p_k(\mathbf{x}_i)(\mu(\mathbf{x}_i) - s^*(\mathbf{x}_i))]| \lesssim h^{m+d}$, $\mathbb{V}[p_k(\mathbf{x}_i)(\mu(\mathbf{x}_i) - s^*(\mathbf{x}_i))] \lesssim h^{d+2m}$, and $\sup_{\mathbf{x}\in\mathcal{X}} |p_k(\mathbf{x})(\mu(\mathbf{x}) - s^*(\mathbf{x}))| \lesssim h^m$. By

Bernstein's inequality, for any $\vartheta > 0$,

$$\begin{aligned}
&\mathbb{P}\left(\max_{1\le k\le K}\Big|\mathbb{E}_n[p_k(\mathbf{x}_i)(\mu(\mathbf{x}_i)-s^*(\mathbf{x}_i))]-\mathbb{E}[p_k(\mathbf{x}_i)(\mu(\mathbf{x}_i)-s^*(\mathbf{x}_i))]\Big|>\vartheta\right)\\
&\le\sum_{k=1}^K\mathbb{P}\left(\Big|\mathbb{E}_n[p_k(\mathbf{x}_i)(\mu(\mathbf{x}_i)-s^*(\mathbf{x}_i))]-\mathbb{E}[p_k(\mathbf{x}_i)(\mu(\mathbf{x}_i)-s^*(\mathbf{x}_i))]\Big|>\vartheta\right)\\
&\le 2K\exp\left(\frac{-\vartheta^2/2}{C_1h^m\vartheta/(3n)+C_2h^{2m+d}/n}\right).
\end{aligned}$$

Therefore,

$$\max_{1\le k\le K}\Big|\mathbb{E}_n[p_k(\mathbf{x}_i)(\mu(\mathbf{x}_i)-s^*(\mathbf{x}_i))]-\mathbb{E}[p_k(\mathbf{x}_i)(\mu(\mathbf{x}_i)-s^*(\mathbf{x}_i))]\Big|\lesssim_{\mathbb{P}} h^{m+d}\sqrt{\frac{\log n}{nh^d}},$$

which suffices to prove the desired uniform bound since $\frac{\log n}{nh^d}=o(1)$.

Next, we prove the leading bias expansion. By Assumption SA-4, we can find $s^*\in\mathcal{S}_{\Delta,m}$ such that $\sup_{\mathbf{x}\in\mathcal{X}}|\partial^{\mathbf{q}}\mu(\mathbf{x})-\partial^{\mathbf{q}}s^*(\mathbf{x})+\mathscr{B}_{m,\mathbf{q}}(\mathbf{x})|\lesssim h^{m+\varrho-[\mathbf{q}]}$. Then the conditional bias can be expanded as follows:

$$\begin{aligned}
\mathbb{E}[\widehat{\partial^{\mathbf{q}}\mu}_0(\mathbf{x})|\mathbf{X}]-\partial^{\mathbf{q}}\mu(\mathbf{x})&=\partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})'\widehat{\mathbf{Q}}_m^{-1}\mathbb{E}_n[\mathbf{p}(\mathbf{x}_i)\mu(\mathbf{x}_i)]-\partial^{\mathbf{q}}\mu(\mathbf{x})\\
&=\partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})'\widehat{\mathbf{Q}}_m^{-1}\mathbb{E}_n[\mathbf{p}(\mathbf{x}_i)\mu(\mathbf{x}_i)]-\partial^{\mathbf{q}}s^*(\mathbf{x})+\mathscr{B}_{m,\mathbf{q}}(\mathbf{x})+O(h^{m+\varrho-[\mathbf{q}]})\\
&=\mathscr{B}_{m,\mathbf{q}}(\mathbf{x})+\partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})'\widehat{\mathbf{Q}}_m^{-1}\mathbb{E}_n[\mathbf{p}(\mathbf{x}_i)(\mu(\mathbf{x}_i)-s^*(\mathbf{x}_i))]+O(h^{m+\varrho-[\mathbf{q}]}).
\end{aligned}$$

The second term in the last line can be further written as

$$\begin{aligned}
&\partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})'\widehat{\mathbf{Q}}_m^{-1}\mathbb{E}_n[\mathbf{p}(\mathbf{x}_i)(\mu(\mathbf{x}_i)-s^*(\mathbf{x}_i))]\\
&=-\partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})'\widehat{\mathbf{Q}}_m^{-1}\mathbb{E}_n\left[\mathbf{p}(\mathbf{x}_i)\mathscr{B}_{m,\mathbf{0}}(\mathbf{x}_i)\right]\\
&\qquad+\partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})'\widehat{\mathbf{Q}}_m^{-1}\mathbb{E}_n\Big[\mathbf{p}(\mathbf{x}_i)(\mu(\mathbf{x}_i)-s^*(\mathbf{x}_i)+\mathscr{B}_{m,\mathbf{0}}(\mathbf{x}_i))\Big].
\end{aligned}$$

By the same argument used to bound $\|\mathbb{E}_n[\mathbf{p}(\mathbf{x}_i)(\mu(\mathbf{x}_i)-s^*(\mathbf{x}_i))]\|_\infty$,

$$\max_{1\le k\le K}\Big|\mathbb{E}_n\Big[p_k(\mathbf{x}_i)(\mu(\mathbf{x}_i)-s^*(\mathbf{x}_i)+\mathscr{B}_{m,\mathbf{0}}(\mathbf{x}_i))\Big]\Big|\lesssim_{\mathbb{P}} h^{m+\varrho+d}.$$

Then by Assumption SA-3 and Lemma SA-2.1,

$$\|\partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})'\widehat{\mathbf{Q}}_m^{-1}\mathbb{E}_n[\mathbf{p}(\mathbf{x}_i)(\mu(\mathbf{x}_i)-s^*(\mathbf{x}_i)+\mathscr{B}_{m,\mathbf{0}}(\mathbf{x}_i))]\|_{L_\infty(\mathcal{X})}\lesssim_{\mathbb{P}} h^{m+\varrho-[\mathbf{q}]},$$

which suffices to show the desired bias expansion.

Suppose that Equation (SA-2.6) holds. Then by Lemma SA-2.1 and Assumption SA-3, the second term in the bias expansion is $o_{\mathbb{P}}(h^{m-[\mathbf{q}]})$. Thus the leading bias is further reduced to $\mathscr{B}_{m,\mathbf{q}}(\mathbf{x})$. The proof of Lemma SA-2.2 is complete. □

SA-10.3. *Proof of Lemma SA-2.3.*

PROOF. For $j=0,1$, the results immediately follow from Assumption SA-3, SA-5 and Lemma SA-2.1. For $j=2$,

$$\begin{aligned}\sup_{\mathbf{x}\in\mathcal{X}}\|\boldsymbol{\gamma}_{\mathbf{q},2}(\mathbf{x})'\|_\infty &\leq \sup_{\mathbf{x}\in\mathcal{X}}\|\boldsymbol{\gamma}_{\mathbf{q},0}(\mathbf{x})'\|_\infty + \sup_{\mathbf{x}\in\mathcal{X}}\|\boldsymbol{\gamma}_{\mathbf{q},0}(\mathbf{x})'\mathbf{Q}_{m,\tilde{m}}\mathbf{Q}_{\tilde{m}}^{-1}\|_\infty + \sup_{\mathbf{x}\in\mathcal{X}}\|\boldsymbol{\gamma}_{\mathbf{q},1}(\mathbf{x})'\|_\infty\\ &\lesssim h^{-d-[\mathbf{q}]} + h^{-d-[\mathbf{q}]}\|\mathbf{Q}_{m,\tilde{m}}\|_\infty\|\mathbf{Q}_{\tilde{m}}^{-1}\|_\infty.\end{aligned}$$

By Assumptions SA-5, both $\mathbf{p}(\cdot)$ and $\tilde{\mathbf{p}}(\cdot)$ are local bases. Then $\mathbf{Q}_{m,\tilde{m}}$ has a finite number of nonzero elements in any row or column, and all elements in $\mathbf{Q}_{m,\tilde{m}}$ are bounded by Ch^d for some universal constant C. Thus $\|\mathbf{Q}_{m,\tilde{m}}\|_\infty \lesssim h^d$, $\|\mathbf{Q}_{m,\tilde{m}}\|_1 \lesssim h^d$ and $\|\mathbf{Q}_{m,\tilde{m}}\| \lesssim h^d$. Then $\|\boldsymbol{\gamma}_{\mathbf{q},2}(\mathbf{x})'\|_\infty \lesssim h^{-d-[\mathbf{q}]}$. Similarly, $\|\boldsymbol{\gamma}_{\mathbf{q},2}(\mathbf{x})\| \lesssim h^{-d-[\mathbf{q}]}$. The lower bound on the L_2-norm follows by Assumption SA-3. Moreover, note that

$$\begin{aligned}&\|\widehat{\boldsymbol{\gamma}}_{\mathbf{q},2}(\mathbf{x})' - \boldsymbol{\gamma}_{\mathbf{q},2}(\mathbf{x})'\|_\infty\\ &\leq \|\widehat{\boldsymbol{\gamma}}_{\mathbf{q},0}(\mathbf{x})' - \boldsymbol{\gamma}_{\mathbf{q},0}(\mathbf{x})'\|_\infty + \|\widehat{\boldsymbol{\gamma}}_{\mathbf{q},0}(\mathbf{x})'\widehat{\mathbf{Q}}_{m,\tilde{m}}\widehat{\mathbf{Q}}_{\tilde{m}}^{-1} - \boldsymbol{\gamma}_{\mathbf{q},0}\mathbf{Q}_{m,\tilde{m}}\mathbf{Q}_{\tilde{m}}^{-1}\|_\infty\\ &\quad + \|\widehat{\boldsymbol{\gamma}}_{\mathbf{q},1}(\mathbf{x})' - \boldsymbol{\gamma}_{\mathbf{q},1}(\mathbf{x})'\|_\infty\\ &\leq \|(\widehat{\boldsymbol{\gamma}}_{\mathbf{q},0}(\mathbf{x})' - \boldsymbol{\gamma}_{\mathbf{q},0}(\mathbf{x})')\widehat{\mathbf{Q}}_{m,\tilde{m}}\widehat{\mathbf{Q}}_{\tilde{m}}^{-1}\|_\infty + \|\boldsymbol{\gamma}_{\mathbf{q},0}(\mathbf{x})'(\widehat{\mathbf{Q}}_{m,\tilde{m}} - \mathbf{Q}_{m,\tilde{m}})\widehat{\mathbf{Q}}_{\tilde{m}}^{-1}\|_\infty\\ &\quad + \|\boldsymbol{\gamma}_{\mathbf{q},0}(\mathbf{x})')\mathbf{Q}_{m,\tilde{m}}(\widehat{\mathbf{Q}}_{\tilde{m}}^{-1} - \mathbf{Q}_{\tilde{m}}^{-1})\|_\infty + h^{-d-[\mathbf{q}]}\sqrt{\log n/(nh^d)},\end{aligned}$$

where the last line uses the results for $j=0,1$. Using the sparsity of $(\widehat{\mathbf{Q}}_{m,\tilde{m}} - \mathbf{Q}_{m,\tilde{m}})$ and the same argument for Lemma SA-2.1, $\|\widehat{\mathbf{Q}}_{m,\tilde{m}} - \mathbf{Q}_{m,\tilde{m}}\|_\infty \lesssim_{\mathbb{P}} h^d\sqrt{\log n/(nh^d)}$. Then the desired uniform bound follows from Lemma SA-2.1 and Assumption SA-3. The L_2-bound follows similarly.

For $j=3$, notice that

$$\begin{aligned}\|\boldsymbol{\gamma}_{\mathbf{q},3}(\mathbf{x})'\|_\infty \leq \|\boldsymbol{\gamma}_{\mathbf{q},0}(\mathbf{x})'\|_\infty &+ \sum_{\boldsymbol{u}\in\Lambda_m}\left\|\boldsymbol{\gamma}_{\boldsymbol{u},1}(\mathbf{x})'h_{\mathbf{x}}^{m-[\mathbf{q}]}B_{\boldsymbol{u},\mathbf{q}}(\mathbf{x})\right\|_\infty\\ &+ \sum_{\boldsymbol{u}\in\Lambda_m}\left\|\boldsymbol{\gamma}_{\mathbf{q},0}(\mathbf{x})'\mathbb{E}[\mathbf{p}(\mathbf{x}_i)h_{\mathbf{x}_i}^m B_{\boldsymbol{u},\mathbf{0}}(\mathbf{x}_i)\partial^{\boldsymbol{u}}\tilde{\mathbf{p}}(\mathbf{x}_i)']\mathbf{Q}_{\tilde{m}}^{-1}\right\|_\infty\end{aligned}$$

By Assumption SA-4 and SA-5, both $\mathbf{p}(\cdot)$ and $\tilde{\mathbf{p}}(\cdot)$ are locally supported and all elements in $\mathbf{p}(\mathbf{x}_i)h_{\mathbf{x}_i}^m B_{\boldsymbol{u},\mathbf{0}}(\mathbf{x}_i)\partial^{\boldsymbol{u}}\tilde{\mathbf{p}}(\mathbf{x})'$ are bounded by a universal constant. Using the argument given in Lemma SA-2.1,

$$\begin{aligned}&\|\mathbb{E}[\mathbf{p}(\mathbf{x}_i)h_{\mathbf{x}_i}^m B_{\boldsymbol{u},\mathbf{0}}(\mathbf{x}_i)\partial^{\boldsymbol{u}}\tilde{\mathbf{p}}(\mathbf{x}_i)']\| \lesssim h^d \qquad \text{and}\\ &\|\mathbb{E}_n[\mathbf{p}(\mathbf{x}_i)h_{\mathbf{x}_i}^m B_{\boldsymbol{u},\mathbf{0}}(\mathbf{x}_i)\partial^{\boldsymbol{u}}\tilde{\mathbf{p}}(\mathbf{x}_i)' - \mathbb{E}[\mathbf{p}(\mathbf{x}_i)h_{\mathbf{x}_i}^m B_{\boldsymbol{u},\mathbf{0}}(\mathbf{x}_i)\partial^{\boldsymbol{u}}\tilde{\mathbf{p}}(\mathbf{x}_i)']\| \lesssim_{\mathbb{P}} h^d\sqrt{\frac{\log n}{nh^d}}.\end{aligned}$$

Then the desired results follow from Assumption SA-3 and Lemma SA-2.1. □

SA-10.4. *Proof of Lemma SA-2.4.*

PROOF. For $j = 0, 1$, the results directly follow from Assumption SA-1, SA-3 and Lemma SA-2.1. The proof for $j = 2, 3$ is divided into three steps.

Step 1: We first establish the upper bounds on $\Omega_j(\mathbf{x})$. By Assumption SA-1,

$$\mathbf{\Sigma}_j = \mathbb{E}[\varepsilon_i^2 \mathbf{\Pi}_j(\mathbf{x}_i)\mathbf{\Pi}_j(\mathbf{x}_i)'] \lesssim \mathbb{E}[\mathbf{\Pi}_j(\mathbf{x}_i)\mathbf{\Pi}_j(\mathbf{x}_i)'].$$

To bound $\|\mathbf{\Sigma}_j\|$, it suffices to give an upper bound on $\mathbb{E}[\mathbf{\Pi}_j(\mathbf{x}_i)\mathbf{\Pi}_j(\mathbf{x}_i)']$. By Assumption SA-5 and the same argument used in the proof of Lemma SA-2.1, we have $\|\mathbb{E}[\mathbf{\Pi}_j(\mathbf{x}_i)\mathbf{\Pi}_j(\mathbf{x}_i)]\| \lesssim h^d$. By Lemma SA-2.3, $\sup_{\mathbf{x}\in\mathcal{X}} \|\boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})'\| \lesssim h^{-d-[\mathbf{q}]}$. Thus $\sup_{\mathbf{x}\in\mathcal{X}} \Omega_j(\mathbf{x}) \lesssim h^{-d-2[\mathbf{q}]}$.

Step 2: Next, we show the lower bound on Ω_2. Since $\sigma^2(\mathbf{x}) \gtrsim 1$ uniformly over $\mathbf{x} \in \mathcal{X}$, we have $\Omega_2(\mathbf{x}) \gtrsim \boldsymbol{\gamma}_{\mathbf{q},2}(\mathbf{x})'\mathbb{E}[\mathbf{\Pi}_2(\mathbf{x}_i)\mathbf{\Pi}_2(\mathbf{x}_i)']\boldsymbol{\gamma}_{\mathbf{q},2}(\mathbf{x})$. Expanding the expression on the RHS of this inequality, we have a trivial lower bound:

$$\begin{aligned}
&\boldsymbol{\gamma}_{\mathbf{q},2}(\mathbf{x})'\mathbb{E}[\mathbf{\Pi}_2(\mathbf{x}_i)\mathbf{\Pi}_2(\mathbf{x}_i)']\boldsymbol{\gamma}_{\mathbf{q},2}(\mathbf{x}) \\
=&\, \partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})'\mathbf{Q}_m^{-1}\partial^{\mathbf{q}}\mathbf{p}(\mathbf{x}) + (\boldsymbol{\gamma}_{\mathbf{q},0}(\mathbf{x})'\mathbf{Q}_{m,\tilde{m}} - \partial^{\mathbf{q}}\tilde{\mathbf{p}}(\mathbf{x})')\mathbf{Q}_{\tilde{m}}^{-1}(\boldsymbol{\gamma}_{\mathbf{q},0}(\mathbf{x})'\mathbf{Q}_{m,\tilde{m}} - \partial^{\mathbf{q}}\tilde{\mathbf{p}}(\mathbf{x})')' \\
&- 2\partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})'\mathbf{Q}_m^{-1}\mathbf{Q}_{m,\tilde{m}}\mathbf{Q}_{\tilde{m}}^{-1}(\boldsymbol{\gamma}_{\mathbf{q},0}(\mathbf{x})'\mathbf{Q}_{m,\tilde{m}} - \partial^{\mathbf{q}}\tilde{\mathbf{p}}(\mathbf{x})')' \\
=&\, \partial^{\mathbf{q}}\tilde{\mathbf{p}}(\mathbf{x})'\mathbf{Q}_{\tilde{m}}^{-1}\partial^{\mathbf{q}}\tilde{\mathbf{p}}(\mathbf{x}) + \Big[\partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})'\mathbf{Q}_m^{-1}(\mathbf{Q}_m - \mathbf{Q}_{m,\tilde{m}}\mathbf{Q}_{\tilde{m}}^{-1}\mathbf{Q}'_{m,\tilde{m}})\mathbf{Q}_m^{-1}\partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})\Big].
\end{aligned}$$

By properties of projection operator, $(\mathbf{Q}_m - \mathbf{Q}_{m,\tilde{m}}\mathbf{Q}_{\tilde{m}}^{-1}\mathbf{Q}'_{m,\tilde{m}})$ is positive semidefinite. By Assumption SA-5 and Lemma SA-2.1, $\partial^{\mathbf{q}}\tilde{\mathbf{p}}(\mathbf{x})'\mathbf{Q}_{\tilde{m}}^{-1}\partial^{\mathbf{q}}\tilde{\mathbf{p}}(\mathbf{x}) \gtrsim h^{-d-2[\mathbf{q}]}$, and thus the desired lower bound is obtained.

Step 3: Now let us bound $\Omega_3(\mathbf{x})$ from below. Suppose condition (i) in Assumption SA-5 holds. Then there exists a linear map $\mathbf{\Upsilon}$ such that $\mathbf{\Pi}_3(\cdot) = \mathbf{\Upsilon}\tilde{\mathbf{p}}(\cdot)$. By Lemma SA-2.1 and Assumption SA-1, $\Omega_3(\mathbf{x}) \gtrsim \boldsymbol{\gamma}_{\mathbf{q},3}(\mathbf{x})'\mathbf{\Upsilon}\mathbb{E}[\tilde{\mathbf{p}}(\mathbf{x}_i)\tilde{\mathbf{p}}(\mathbf{x}_i)']\mathbf{\Upsilon}'\boldsymbol{\gamma}_{\mathbf{q},3}(\mathbf{x}) \gtrsim h^d\boldsymbol{\gamma}_{\mathbf{q},3}(\mathbf{x})'\mathbf{\Upsilon}\mathbf{\Upsilon}'\boldsymbol{\gamma}_{\mathbf{q},3}(\mathbf{x})$. Define $\mathbf{v}(\mathbf{x})' := \boldsymbol{\gamma}_{\mathbf{q},3}(\mathbf{x})'\mathbf{\Upsilon}$. Notice that for any function $s(\mathbf{x}) \in \mathsf{span}(\mathbf{p}(\cdot))$, there exists some $\mathbf{c} \in \mathbb{R}^{\tilde{K}}$ such that $s(\mathbf{x}) = \tilde{\mathbf{p}}(\mathbf{x})'\mathbf{c}$. It follows that $\mathbf{v}(\mathbf{x})'\mathbb{E}[\tilde{\mathbf{p}}(\mathbf{x}_i)s(\mathbf{x}_i)] = \partial^{\mathbf{q}}s(\mathbf{x})$. Then we have

$$\|\mathbf{v}(\mathbf{x})\| \geq \frac{\|\mathbf{v}(\mathbf{x})'\mathbb{E}[\tilde{\mathbf{p}}(\mathbf{x}_i)s(\mathbf{x}_i)]\|}{\|\mathbb{E}[\tilde{\mathbf{p}}(\mathbf{x}_i)s(\mathbf{x}_i)]\|} = \frac{\|\partial^{\mathbf{q}}s(\mathbf{x})\|}{\|\mathbb{E}[\tilde{\mathbf{p}}(\mathbf{x}_i)s(\mathbf{x}_i)]\|}.$$

Since the choice of $s(\mathbf{x})$ is arbitrary within the span of $\mathbf{p}$, by Assumption SA-3(c) we can take a function in $\mathbf{p}(\mathbf{x})$ to be $s(\mathbf{x})$ such that $\|\partial^{\mathbf{q}}s(\mathbf{x})\| \geq Ch^{-[\mathbf{q}]}|$ where C is a constant independent of $\mathbf{x}$ and n. Also, $\|\mathbb{E}[\tilde{\mathbf{p}}(\mathbf{x}_i)s(\mathbf{x}_i)]\| \lesssim h^d$. Hence $h^{-[\mathbf{q}]-d} \lesssim \inf_{\mathbf{x}\in\mathcal{X}} \|\mathbf{v}(\mathbf{x})\|$. The desired bound follows.

Step 4: Now suppose that condition (ii) in Assumption SA-5 holds. Again, since $\sigma^2(\mathbf{x}) \gtrsim 1$ uniformly over $\mathbf{x} \in \mathcal{X}$, $\Omega_3(\mathbf{x}) \gtrsim \boldsymbol{\gamma}_{\mathbf{q},3}(\mathbf{x})'\mathbb{E}[\mathbf{\Pi}_3(\mathbf{x}_i)\mathbf{\Pi}_3(\mathbf{x}_i)']\boldsymbol{\gamma}_{\mathbf{q},3}(\mathbf{x})$.

Define $\mathscr{K}_h(\mathbf{x},\mathbf{x}_1) := \boldsymbol{\gamma}_{\mathbf{q},3}(\mathbf{x})'\boldsymbol{\Pi}_3(\mathbf{x}_1)$. It suffices to bound $\mathbb{E}_{\mathbf{x}_1}[\mathscr{K}_h(\mathbf{x},\mathbf{x}_1)^2]$, where $\mathbb{E}_{\mathbf{x}_1}$ denotes the expectation with respect to the distribution of $\mathbf{x}_1$. We write

$$\langle g_1, g_2\rangle_{\mathcal{U}} := \int_{\mathcal{U}} g_1(\mathbf{x}_1)g_2(\mathbf{x}_1)f(\mathbf{x}_1)d\mathbf{x}_1$$

for the inner product of $g_1(\cdot)$ and $g_2(\cdot)$ with respect to the probability measure of $\mathbf{x}_1$ on $\mathcal{U}\subset\mathcal{X}$. Clearly, $\langle g_1, g_2\rangle_{\mathcal{X}} = \mathbb{E}[g_1(\mathbf{x}_1)g_2(\mathbf{x}_1)]$. By Cauchy-Schwartz inequality, for $g\in L^2(\mathcal{U})$,

$$\langle\mathscr{K}_h(\mathbf{x},\mathbf{x}_1), g(\mathbf{x}_1)\rangle^2_{\mathcal{U}} \leq \langle\mathscr{K}_h(\mathbf{x},\mathbf{x}_1),\mathscr{K}_h(\mathbf{x},\mathbf{x}_1)\rangle_{\mathcal{U}}\cdot\langle g(\mathbf{x}_1), g(\mathbf{x}_1)\rangle_{\mathcal{U}}.$$

Given an evaluation point $\mathbf{x}\in\mathcal{X}$, choose a polynomial $\varphi_h(\mathbf{x}_1;\mathbf{x}) := \frac{(\mathbf{x}_1-\mathbf{x})^{\mathbf{q}}}{h^{[\mathbf{q}]}}$. Clearly, $\partial^{\mathbf{q}}\varphi_h(\mathbf{x}_1;\mathbf{x}) = h^{-[\mathbf{q}]}$. By Assumption SA-5, the operator $\langle\mathscr{K}_h(\mathbf{x},\mathbf{x}_1),\cdot\rangle_{\mathcal{X}}$ reproduces the $\mathbf{q}$th derivative of $\varphi_h(\mathbf{x}_1;\mathbf{x})$ at $\mathbf{x}$, i.e. $\langle\mathscr{K}_h(\mathbf{x},\mathbf{x}_1),\varphi_h(\mathbf{x}_1;\mathbf{x})\rangle_{\mathcal{X}} = h^{-[\mathbf{q}]}$.

In addition, we rectangularize Δ as described in the proof of Lemma SA-2.1. Then for each $\mathbf{x}\in\mathcal{X}$, we can choose a rectangular region that contains $\mathbf{x}$ and consists of a fixed number of subrectangles in $\Delta^{\mathtt{rec}}$. Thus, the size of this region shrinks as $n\to\infty$. Specifically, let

$$\mathcal{W}_{\mathbf{x}} := \left\{\check{\mathbf{x}} : t_{\ell,\mathbf{x}} - L \leq \check{x}_j \leq t_{\ell,\mathbf{x}} + L,\ \ell = 1,\dots,d\right\},$$

where $t_{\ell,\mathbf{x}}$ is the closest point in $\Delta^{\mathtt{rec}}$ that is no greater than x_ℓ, and L is some fixed number to be determined which only depends on d, m and $\tilde{m}$. If such a region spans outside of $\mathcal{X}$, take its intersection with $\mathcal{X}$. Then, $\langle\varphi_h(\mathbf{x}_1;\mathbf{x}),\varphi_h(\mathbf{x}_1;\mathbf{x})\rangle_{\mathcal{W}_{\mathbf{x}}} \lesssim h^d$, and

$$\langle\mathscr{K}_h(\mathbf{x},\mathbf{x}_1),\mathscr{K}_h(\mathbf{x},\mathbf{x}_1)\rangle_{\mathcal{X}} \geq \langle\mathscr{K}_h(\mathbf{x},\mathbf{x}_1),\mathscr{K}_h(\mathbf{x},\mathbf{x}_1)\rangle_{\mathcal{W}_{\mathbf{x}}} \gtrsim h^{-d}\left\langle\mathscr{K}_h(\mathbf{x},\mathbf{x}_1),\varphi_h(\mathbf{x}_1;\mathbf{x})\right\rangle^2_{\mathcal{W}_{\mathbf{x}}}.$$

It suffices to show $|\langle\mathscr{K}_h(\mathbf{x},\mathbf{x}_1),\varphi_h(\mathbf{x}_1;\mathbf{x})\rangle_{\mathcal{X}\setminus\mathcal{W}_{\mathbf{x}}}|$ can be made sufficiently small such that

$$|\langle\mathscr{K}_h(\mathbf{x},\mathbf{x}_1),\varphi_h(\mathbf{x}_1;\mathbf{x})\rangle_{\mathcal{W}_{\mathbf{x}}}| \gtrsim h^{-[\mathbf{q}]}.$$

By Lemma SA-2.1, the elements of $h^d\mathbf{Q}_m^{-1}$ and $h^d\mathbf{Q}_{\tilde{m}}^{-1}$ exponentially decays when they get far away from the (block) diagonals. In view of Assumption SA-5, with $\mathbf{x}$ fixed, $\mathscr{K}_h(\mathbf{x},\mathbf{x}_1)$ also exponentially decays as $\|\mathbf{x}_1-\mathbf{x}\|$ increases. Formally, write

$$\begin{aligned}\mathscr{K}(\mathbf{x},\mathbf{x}_1) &= \partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})'\mathbf{Q}_m^{-1}\mathbf{p}(\mathbf{x}_1) + \sum_{\boldsymbol{u}\in\Lambda_m}\partial^{\boldsymbol{u}}\tilde{\mathbf{p}}(\mathbf{x})'h_{\mathbf{x}}^{m-[\mathbf{q}]}B_{\boldsymbol{u},\mathbf{q}}(\mathbf{x})\mathbf{Q}_{\tilde{m}}^{-1}\tilde{\mathbf{p}}(\mathbf{x}_1)\\ &\quad - \partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})'\mathbf{Q}_m^{-1}\mathbb{E}\Big[\mathbf{p}(\mathbf{x}_i)h_{\mathbf{x}_i}^{\boldsymbol{u}}B_{\boldsymbol{u},\mathbf{0}}(\mathbf{x}_i)\partial^{\boldsymbol{u}}\tilde{\mathbf{p}}(\mathbf{x}_i)'\Big]\mathbf{Q}_{\tilde{m}}^{-1}\tilde{\mathbf{p}}(\mathbf{x}_1).\end{aligned}\tag{SA-10.1}$$

We temporarily change the meaning of subscripts of subrectangles in $\Delta^{\mathtt{rec}}$: let

$$\delta^{\mathtt{rec}}_{l_1\dots l_d} := \{\check{\mathbf{x}} : t_{\ell,l_\ell} \leq \check{x}_\ell \leq t_{\ell,l_\ell+1},\ \ell = 1,\dots,d\}$$

with $\delta^{\mathtt{rec}}_{\mathbf{0}}$ denoting the subrectangle where $\mathbf{x}$ is located, and index other subrectangles with $\delta^{\mathtt{rec}}_{\mathbf{0}}$ regarded as the "origin". First notice that for any given point $\mathbf{x}_0\in\mathcal{X}$,

$\mathbf{p}(\mathbf{x}_0)$ and $\tilde{\mathbf{p}}(\mathbf{x}_0)$ are two vectors containing a fixed number of nonzeros and all their elements are bounded by some universal constant. Their nonzero elements are obtained by evaluating those basis functions with local supports covering $\mathbf{x}_0$. Moreover, $h^{-d}\mathbb{E}[\mathbf{p}(\mathbf{x}_i)h_{\mathbf{x}_i}^m B_{\boldsymbol{u},\mathbf{0}}(\mathbf{x}_i)\partial^{\boldsymbol{u}}\tilde{\mathbf{p}}(\mathbf{x}_i)']$ also admits a block banded structure in the sense that only the products of basis functions with overlapping supports are nonzero, and all elements in this matrix is bounded by some universal constant. Hence for

$$\mathbf{r}_1(\mathbf{x})' = h^{[\mathbf{q}]}\partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})'\mathbf{Q}_m^{-1}\mathbb{E}\Big[\sum_{\boldsymbol{u}\in\Lambda_m}\mathbf{p}(\mathbf{x}_i)B_{\boldsymbol{u},\mathbf{0}}(\mathbf{x}_i)\partial^{\boldsymbol{u}}\tilde{\mathbf{p}}(\mathbf{x}_i)'\Big]$$

and

$$\mathbf{r}_2(\mathbf{x}_1)' = h^d\tilde{\mathbf{p}}(\mathbf{x}_1)'\mathbf{Q}_{\tilde{m}}^{-1},$$

we can find another two vectors $\bar{\mathbf{r}}_1(\mathbf{x})$ and $\bar{\mathbf{r}}_2(\mathbf{x}_1)$ with strictly positive elements that are greater than the absolute values of the corresponding elements in $\mathbf{r}_1(\mathbf{x})$ and $\mathbf{r}_2(\mathbf{x}_1)$, i.e., $\bar{\mathbf{r}}_1(\mathbf{x})$ and $\bar{\mathbf{r}}_2(\mathbf{x}_1)$ are element-wise bounds on $\mathbf{r}_1(\mathbf{x})$ and $\mathbf{r}_2(\mathbf{x}_1)$. Moreover, the elements of $\bar{\mathbf{r}}_1(\mathbf{x})$ and $\bar{\mathbf{r}}_2(\mathbf{x}_1)$ are decaying exponentially with some rates $\vartheta_1, \vartheta_2 \in (0,1)$ when they get far away from the positions of those basis functions in $\mathbf{p}(\cdot)$ and $\tilde{\mathbf{p}}(\cdot)$ whose supports are around $\mathbf{x}$ and $\mathbf{x}_1$ respectively. Notice that for some constant $\vartheta \in (0,1)$, $\{\sum_{l=i}^{\infty}\vartheta^l\}_{i=1}^{\infty}$ is also a geometric sequence. Therefore, the inner product between $\bar{\mathbf{r}}_1(\mathbf{x})$ and $\bar{\mathbf{r}}_2(\mathbf{x}_1)$, i.e., the third term in (SA-10.1), exponentially decays as $\|\mathbf{x}_1 - \mathbf{x}\|$ increases. Similarly, the inner product between $\partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})$ and $\mathbf{Q}_m^{-1}\mathbf{p}(\mathbf{x}_1)$ and that between $\sum_{\boldsymbol{u}\in\Lambda_m}\partial^{\boldsymbol{u}}\tilde{\mathbf{p}}(\mathbf{x})h_{\mathbf{x}}^{m-[\mathbf{q}]}B_{\boldsymbol{u},\mathbf{q}}(\mathbf{x})$ and $\mathbf{Q}_{\tilde{m}}^{-1}\tilde{\mathbf{p}}(\mathbf{x}_1)$ have the same property. Given these results, we have for some $\vartheta \in (0,1)$,

$$\|h^{[\mathbf{q}]+d}\mathscr{K}_h(\mathbf{x},\cdot)\|_{L_\infty(\delta_{l_1\dots l_d})} \le C\vartheta^{\sum_{\ell=1}^d |l_\ell|}.$$

Meanwhile, $\|\varphi_h(\cdot;\mathbf{x})\|_{L_\infty(\delta^{\mathtt{rec}}_{l_1\dots l_d})} \lesssim (|l_1|+1)^{q_1}\cdots(|l_d|+1)^{q_d}$. Let $\boldsymbol{l} = (l_1,\cdots,l_d)$ and $\mathbf{1} = (1,\cdots,1)$. Denote $|\boldsymbol{l}| = (|l_1|,\dots,|l_d|)$. The above results imply

$$|\langle\mathscr{K}_h(\mathbf{x},\mathbf{x}_1), \varphi_h(\mathbf{x}_1;\mathbf{x})\rangle_{\delta^{\mathtt{rec}}_{l_1\dots l_d}}| \lesssim h^{-[\mathbf{q}]}(|\boldsymbol{l}|+\mathbf{1})^{\mathbf{q}}\vartheta^{[|\boldsymbol{l}|]}.$$

Then the desired result follows from the fact that $\sum_{[|\boldsymbol{l}|]=0}^{\infty}(|\boldsymbol{l}|+\mathbf{1})^{\mathbf{q}}\vartheta^{[|\boldsymbol{l}|]}$ exists. Therefore, we can choose L large enough which is independent of $\mathbf{x}$ and n such that $|\langle\mathscr{K}_h(\mathbf{x},\mathbf{x}_1),\varphi_h(\mathbf{x}_1;\mathbf{x})\rangle_{\mathcal{W}_{\mathbf{x}}}| \gtrsim h^{-[\mathbf{q}]}$. Then the proof is complete. □

SA-10.5. *Proof of Theorem 4.1.* In this section, we provide the proof of Theorem 4.1 in the main paper.

PROOF. Regarding the integrated conditional variance,

$$\int_{\mathcal{X}} \mathbb{V}[\widehat{\partial^{\mathbf{q}}\mu}_0(\mathbf{x})|\mathbf{X}]w(\mathbf{x})\,d\mathbf{x} = \frac{1}{n}\operatorname{trace}\Big[\boldsymbol{\Sigma}_0 \int_{\mathcal{X}} \boldsymbol{\gamma}_{\mathbf{q},0}(\mathbf{x})\boldsymbol{\gamma}_{\mathbf{q},0}(\mathbf{x})'w(\mathbf{x})\,d\mathbf{x}\Big] + o_{\mathbb{P}}\Big(\frac{1}{nh^{d+2[\mathbf{q}]}}\Big)$$
$$\leq \frac{1}{n}\lambda_{\max}\Big(\mathbf{Q}_m^{-1}\mathbb{E}[\mathbf{p}(\mathbf{x}_i)\mathbf{p}(\mathbf{x}_i)'\sigma^2(\mathbf{x}_i)]\mathbf{Q}_m^{-1}\Big)\operatorname{trace}\Big[\int_{\mathcal{X}} \partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})\partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})'w(\mathbf{x})d\mathbf{x}\Big] + o_{\mathbb{P}}\Big(\frac{1}{nh^{d+2[\mathbf{q}]}}\Big)$$
$$\lesssim \frac{1}{nh^d}\operatorname{trace}\Big[\int_{\mathcal{X}} \partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})\partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})'w(\mathbf{x})d\mathbf{x}\Big] \lesssim \frac{1}{nh^{d+2[\mathbf{q}]}},$$

where the first line holds by Lemma SA-2.1, the second by Trace Inequality, and the last by the continuity of $w(\cdot)$ and Lemma SA-2.1. Since $\sigma^2(\cdot)$ and $w(\cdot)$ are bounded away from zero, the other side of the bound follows similarly.

Regarding the integrated squared bias, we have

$$\begin{aligned}
&\int_{\mathcal{X}} \Big(\mathbb{E}[\widehat{\partial^{\mathbf{q}}\mu}_0(\mathbf{x})|\mathbf{X}] - \partial^{\mathbf{q}}\mu(\mathbf{x})\Big)^2 w(\mathbf{x})d\mathbf{x}\\
&= \int_{\mathcal{X}} \Big(\mathscr{B}_{m,\mathbf{q}}(\mathbf{x}) - \boldsymbol{\gamma}_{\mathbf{q},0}(\mathbf{x})'\mathbb{E}[\mathbf{p}(\mathbf{x}_i)\mathscr{B}_{m,\mathbf{0}}(\mathbf{x}_i)]\Big)^2 w(\mathbf{x})d\mathbf{x} + o_{\mathbb{P}}(h^{2m-2[\mathbf{q}]})\\
&= \int_{\mathcal{X}} \mathscr{B}_{m,\mathbf{q}}(\mathbf{x})^2 w(\mathbf{x})d\mathbf{x} + \int_{\mathcal{X}} \Big(\boldsymbol{\gamma}_{\mathbf{q},0}(\mathbf{x})'\mathbb{E}[\mathbf{p}(\mathbf{x}_i)\mathscr{B}_{m,\mathbf{0}}(\mathbf{x}_i)]\Big)^2 w(\mathbf{x})d\mathbf{x}\\
&\quad - 2\int_{\mathcal{X}} \mathscr{B}_{m,\mathbf{q}}(\mathbf{x})\boldsymbol{\gamma}_{\mathbf{q},0}(\mathbf{x})'\mathbb{E}[\mathbf{p}(\mathbf{x}_i)\mathscr{B}_{m,\mathbf{0}}(\mathbf{x}_i)]w(\mathbf{x})d\mathbf{x} + o_{\mathbb{P}}(h^{2m-2[\mathbf{q}]})
\end{aligned}$$
$$=: \mathrm{B}_1 + \mathrm{B}_2 - 2\mathrm{B}_3 + o_{\mathbb{P}}(h^{2m-2[\mathbf{q}]}). \tag{SA-10.2}$$

Let h_δ be the diameter of δ and $\mathbf{t}_\delta^*$ be an arbitrary point in δ. Then

$$\begin{aligned}
\mathrm{B}_1 &= \sum_{\boldsymbol{u}_1,\boldsymbol{u}_2\in\Lambda_m} \int_{\mathcal{X}} \Big[\partial^{\boldsymbol{u}_1}\mu(\mathbf{x})\partial^{\boldsymbol{u}_2}\mu(\mathbf{x})h_{\mathbf{x}}^{2m-2[\mathbf{q}]}B_{\boldsymbol{u}_1,\mathbf{q}}(\mathbf{x})B_{\boldsymbol{u}_2,\mathbf{q}}(\mathbf{x})\Big]w(\mathbf{x})d\mathbf{x}\\
&= \sum_{\boldsymbol{u}_1,\boldsymbol{u}_2\in\Lambda_m}\sum_{\delta\in\Delta} \int_{\delta} \Big[h_\delta^{2m-2[\mathbf{q}]}\partial^{\boldsymbol{u}_1}\mu(\mathbf{t}_\delta^*)\partial^{\boldsymbol{u}_2}\mu(\mathbf{t}_\delta^*)B_{\boldsymbol{u}_1,\mathbf{q}}(\mathbf{x})B_{\boldsymbol{u}_2,\mathbf{q}}(\mathbf{x})\Big]w(\mathbf{t}_\delta^*)d\mathbf{x} + o(h^{2m-2[\mathbf{q}]})\\
&= \sum_{\boldsymbol{u}_1,\boldsymbol{u}_2\in\Lambda_m}\sum_{\delta\in\Delta} \Big\{h_\delta^{2m-2[\mathbf{q}]}\partial^{\boldsymbol{u}_1}\mu(\mathbf{t}_\delta^*)\partial^{\boldsymbol{u}_2}\mu(\mathbf{t}_\delta^*)w(\mathbf{t}_\delta^*)\int_{\delta} B_{\boldsymbol{u}_1,\mathbf{q}}(\mathbf{x})B_{\boldsymbol{u}_2,\mathbf{q}}(\mathbf{x})d\mathbf{x}\Big\} + o(h^{2m-2[\mathbf{q}]})\\
&\lesssim h^{2m-2[\mathbf{q}]}
\end{aligned}$$

where the third line holds by the continuity of $\partial^{\boldsymbol{u}_1}\mu(\cdot)$, $\partial^{\boldsymbol{u}_2}\mu(\cdot)$, and $w(\cdot)$, and the last by Assumption SA-2 and SA-4.

$$\begin{aligned}
\mathrm{B}_2 &= \operatorname{trace}\Big[\mathbf{Q}_m^{-1}\mathbb{E}[\mathbf{p}(\mathbf{x}_i)\mathscr{B}_{m,\mathbf{0}}(\mathbf{x}_i)]\mathbb{E}[\mathbf{p}(\mathbf{x}_i)'\mathscr{B}_{m,\mathbf{0}}(\mathbf{x}_i)]\mathbf{Q}_m^{-1}\int_{\mathcal{X}} \partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})\partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})'w(\mathbf{x})d\mathbf{x}\Big]\\
&\lesssim h^{-d-2[\mathbf{q}]}\operatorname{trace}\Big[\mathbb{E}[\mathbf{p}(\mathbf{x}_i)\mathscr{B}_{m,\mathbf{0}}(\mathbf{x}_i)]\mathbb{E}[\mathbf{p}(\mathbf{x}_i)'\mathscr{B}_{m,\mathbf{0}}(\mathbf{x}_i)]\Big] \lesssim h^{2m-2[\mathbf{q}]}
\end{aligned}$$

where the first inequality holds by Trace Inequality, Lemma SA-2.1, and Assumption SA-3, and the second by Assumption SA-3 and SA-4. Finally,

$$|\mathrm{B}_3| \leq \Big\| \int_{\mathcal{X}} \mathscr{B}_{m,\mathbf{q}}(\mathbf{x})\partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})'w(\mathbf{x})d\mathbf{x}\Big\|_\infty \Big\|\mathbf{Q}_m^{-1}\Big\|_\infty \Big\|\mathbb{E}[\mathbf{p}(\mathbf{x}_i)\mathscr{B}_{m,\mathbf{0}}(\mathbf{x}_i)]\Big\|_\infty \lesssim h^{2m-2[\mathbf{q}]}$$

where the last inequality follows from Assumption SA-3 and Lemma SA-2.1. □

SA-10.6. *Proof of Theorem SA-3.1.*

PROOF. We divide the proof into two steps.

Step 1: For the integrated variance, first define operators $\boldsymbol{\mathscr{M}}(\cdot)$ and $\boldsymbol{\mathscr{M}}_{\mathbf{q}}(\cdot)$ as follows:

$$\boldsymbol{\mathscr{M}}(\phi) := \int_{\mathcal{X}} \mathbf{p}(\mathbf{x})\mathbf{p}(\mathbf{x})'\phi(\mathbf{x})d\mathbf{x}, \quad \text{and} \quad \boldsymbol{\mathscr{M}}_{\mathbf{q}}(\phi) := \int_{\mathcal{X}} \partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})\partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})'\phi(\mathbf{x})d\mathbf{x}.$$

Then,

$$\begin{aligned}
&\int_{\mathcal{X}} \mathbb{V}[\widehat{\partial^{\mathbf{q}}\mu}_0(\mathbf{x})|\mathbf{X}]w(\mathbf{x})d\mathbf{x}\\
=&\frac{1}{n}\,\mathrm{trace}\left[\mathbf{Q}_m^{-1}\mathbb{E}[\mathbf{p}(\mathbf{x}_i)\mathbf{p}(\mathbf{x}_i)'\sigma(\mathbf{x}_i)^2]\mathbf{Q}_m^{-1}\int_{\mathcal{X}} \partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})\partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})'w(\mathbf{x})\,d\mathbf{x}\right] + o_{\mathbb{P}}\Big(\frac{1}{nh^{d+2[\mathbf{q}]}}\Big)\\
=&\frac{1}{n}\,\mathrm{trace}\left[\boldsymbol{\mathscr{M}}(f)^{-1}\boldsymbol{\mathscr{M}}(\sigma^2 f)\boldsymbol{\mathscr{M}}(f)^{-1}\boldsymbol{\mathscr{M}}_{\mathbf{q}}(w)\right] + o_{\mathbb{P}}\Big(\frac{1}{nh^{d+2[\mathbf{q}]}}\Big).
\end{aligned}$$

Moreover, define another operator $\boldsymbol{\mathscr{D}}(\cdot)$: $\boldsymbol{\mathscr{D}}(\phi) := \mathrm{diag}\{\phi(\tau_1), \phi(\tau_2), \cdots, \phi(\tau_K)\}$. Recall that $\boldsymbol{\tau}_k$ is an arbitrary point in $\mathrm{supp}(p_k)$, for $k = 1, \dots, K$. Then we can write

$$\boldsymbol{\mathscr{M}}(\phi) = \boldsymbol{\mathscr{M}}(1)\boldsymbol{\mathscr{D}}(\phi) - \boldsymbol{\mathscr{E}}(\phi) \tag{SA-10.3}$$

where $\boldsymbol{\mathscr{E}}(\phi)$ can be viewed as errors defined by Eq. (SA-10.3). Similarly, write

$$\boldsymbol{\mathscr{M}}_{\mathbf{q}}(\phi) = \boldsymbol{\mathscr{M}}_{\mathbf{q}}(1)\boldsymbol{\mathscr{D}}(\phi) - \boldsymbol{\mathscr{E}}_{\mathbf{q}}(\phi).$$

Then it directly follows that

$$\begin{aligned}
&\boldsymbol{\mathscr{M}}(f)^{-1}\boldsymbol{\mathscr{M}}(\sigma^2 f) = [\mathbf{I} - \boldsymbol{\mathscr{U}}(f)]^{-1}[\boldsymbol{\mathscr{D}}(f)^{-1}\boldsymbol{\mathscr{D}}(\sigma^2 f) - \boldsymbol{\mathscr{L}}(f, \sigma^2 f)], \quad \text{and}\\
&\boldsymbol{\mathscr{M}}(f)^{-1}\boldsymbol{\mathscr{M}}_{\mathbf{q}}(w) = [\mathbf{I} - \boldsymbol{\mathscr{U}}(f)]^{-1}[\boldsymbol{\mathscr{D}}(f)^{-1}\boldsymbol{\mathscr{M}}(1)^{-1}\boldsymbol{\mathscr{M}}_{\mathbf{q}}(1)\boldsymbol{\mathscr{D}}(w) - \boldsymbol{\mathscr{L}}_{\mathbf{q}}(f, w)]
\end{aligned}$$

where

$$\begin{aligned}
&\boldsymbol{\mathscr{U}}(\phi) := \boldsymbol{\mathscr{D}}(\phi)^{-1}\boldsymbol{\mathscr{M}}(1)^{-1}\boldsymbol{\mathscr{E}}(\phi),\\
&\boldsymbol{\mathscr{L}}(\phi, \varphi) := \boldsymbol{\mathscr{D}}(\phi)^{-1}\boldsymbol{\mathscr{M}}(1)^{-1}\boldsymbol{\mathscr{E}}(\varphi), \quad \text{and}\\
&\boldsymbol{\mathscr{L}}_{\mathbf{q}}(\phi, \varphi) := \boldsymbol{\mathscr{D}}(\phi)^{-1}\boldsymbol{\mathscr{M}}(1)^{-1}\boldsymbol{\mathscr{E}}_{\mathbf{q}}(\varphi).
\end{aligned}$$

The number of nonzeros in any row or any column of $\mathscr{E}(\phi)$ (and $\mathscr{E}_{\mathbf{q}}(\phi)$) is bounded by some constant. In fact, as explained in the proof of Lemma SA-2.1, it takes a (multi-layer) banded structure. If $\text{supp}(p_k) \cap \text{supp}(p_l) \neq \varnothing$, then by Assumption SA-3 and the continuity of f, the (k,l)th element of $\mathscr{M}(f)$ can be approximated as follows:

$$\begin{aligned}\int_{\mathcal{X}} p_k(\mathbf{x})p_l(\mathbf{x})f(\mathbf{x})d\mathbf{x} =& f(\tau_k)\int_{\mathcal{X}} p_k(\mathbf{x})p_l(\mathbf{x})d\mathbf{x} + o(h^d)\\ =& f(\tau_l)\int_{\mathcal{X}} p_k(\mathbf{x})p_l(\mathbf{x})d\mathbf{x} + o(h^d).\end{aligned} \tag{SA-10.4}$$

Moreover, since $\mathcal{X}$ is compact, f is uniformly continuous, and then $\|\mathscr{E}(f)\|_1 = o(h^d)$, $\|\mathscr{E}(f)\|_\infty = o(h^d)$, and $\|\mathscr{E}(f)\| = o(h^d)$. Since $\|\mathscr{D}(f)^{-1}\| \lesssim 1$ and $\|\mathscr{M}(1)^{-1}\| \lesssim h^{-d}$, we conclude $\|\mathscr{U}(f)\| = o(1)$. For K large enough, we can make $\|\mathscr{U}(f)\| < 1$, and thus

$$[\mathbf{I} - \mathscr{U}(f)]^{-1} = \mathbf{I} + \mathscr{U}(f) + \mathscr{U}(f)^2 + \cdots = \mathbf{I} + \mathscr{W}(f)$$

where $\mathscr{W}(f) := \sum_{l=1}^{\infty} \mathscr{U}(f)^l$. Now we can write

$$\begin{aligned}&\text{trace}\left[\mathscr{M}(f)^{-1}\mathscr{M}(\sigma^2 f)\mathscr{M}(f)^{-1}\mathscr{M}_{\mathbf{q}}(w)\right]\\ =&\text{trace}\left[\Big(\mathbf{I}+\mathscr{W}(f)\Big)\Big(\mathscr{D}(f)^{-1}\mathscr{D}(\sigma^2 f) - \mathscr{L}(f,\sigma^2 f)\Big)\Big(\mathbf{I}+\mathscr{W}(f)\Big)\right.\\ &\qquad\left.\times\Big(\mathscr{D}(f)^{-1}\mathscr{M}(1)^{-1}\mathscr{M}_{\mathbf{q}}(1)\mathscr{D}(w) - \mathscr{L}_{\mathbf{q}}(f,w)\Big)\right]\\ =&\text{trace}\left[\Big(\mathscr{D}(f)^{-1}\mathscr{D}(\sigma^2 f)+\mathbf{E}_1\Big)\Big(\mathscr{D}(f)^{-1}\mathscr{M}(1)^{-1}\mathscr{M}_{\mathbf{q}}(1)\mathscr{D}(w)+\mathbf{E}_2\Big)\right]\\ =&\text{trace}\Big[\mathscr{D}(f)^{-1}\mathscr{D}(\sigma^2 f)\mathscr{D}(f)^{-1}\mathscr{M}(1)^{-1}\mathscr{M}_{\mathbf{q}}(1)\mathscr{D}(w)+\\ &\qquad\mathbf{E}_1\mathscr{D}(f)^{-1}\mathscr{M}(1)^{-1}\mathscr{M}_{\mathbf{q}}(1)\mathscr{D}(w)+\mathscr{D}(f)^{-1}\mathscr{D}(\sigma^2 f)\mathbf{E}_2+\mathbf{E}_1\mathbf{E}_2\Big]\end{aligned}$$

where

$$\begin{aligned}\mathbf{E}_1 &= -\mathscr{L}(f,\sigma^2 f) + \mathscr{W}(f)\mathscr{D}(f)^{-1}\mathscr{D}(\sigma^2 f) - \mathscr{W}(f)\mathscr{L}(f,\sigma^2 f),\\ \mathbf{E}_2 &= -\mathscr{L}_{\mathbf{q}}(f,w) + \mathscr{W}(f)\mathscr{D}(f)^{-1}\mathscr{M}(1)^{-1}\mathscr{M}_{\mathbf{q}}(1)\mathscr{D}(w) - \mathscr{W}(f)\mathscr{L}_{\mathbf{q}}(f,w).\end{aligned}$$

By assumptions in the theorem, $\text{vol}(\delta_{\mathbf{x}}) = \prod_{\ell=1}^d b_{\mathbf{x},\ell} = \prod_{\ell=1}^d \kappa_\ell^{-1} g_\ell(\mathbf{x})^{-1} + o(h^d)$. Hence

$$\begin{aligned}&\text{trace}\left[\mathscr{D}(f)^{-1}\mathscr{D}(\sigma^2 f)\mathscr{D}(f)^{-1}\mathscr{M}(1)^{-1}\mathscr{M}_{\mathbf{q}}(1)\mathscr{D}(w)\right]\\ =&\prod_{\ell=1}^d \kappa_\ell\,\text{trace}\left[\mathscr{D}(f)^{-1}\mathscr{D}(\sigma^2 f)\mathscr{D}(f)^{-1}\mathscr{M}(1)^{-1}\mathscr{M}_{\mathbf{q}}(1)\mathscr{D}(w)\mathscr{D}\Big(\prod_{\ell=1}^d g_\ell\Big)\mathscr{D}\Big(\prod_{\ell=1}^d g_\ell\Big)^{-1}\prod_{\ell=1}^d \kappa_\ell^{-1}\right]\\ =&\ \boldsymbol{\kappa}^{\mathbf{1}+2\mathbf{q}}\left(\sum_{k=1}^K\left[\frac{\sigma^2(\boldsymbol{\tau}_k)w(\boldsymbol{\tau}_k)}{f(\boldsymbol{\tau}_k)}\prod_{\ell=1}^d g_\ell(\boldsymbol{\tau}_k)\right]\mathbf{e}_k'\mathscr{M}(1)^{-1}\boldsymbol{\kappa}^{-2\mathbf{q}}\mathscr{M}_{\mathbf{q}}(1)\mathbf{e}_k\,\text{vol}(\delta_{\boldsymbol{\tau}_k})\right) + o(\boldsymbol{\kappa}^{\mathbf{1}+2\mathbf{q}})\end{aligned}$$

By Assumptions SA-2, SA-3 and Lemma SA-2.1, the summation in parenthesis is bounded from above and below. It remains to show all other terms are of smaller order. It directly follows from the same argument as that in the proof of [1, Theorem 6.1] that the trace of the remaining terms is $o(\boldsymbol{\kappa}^{\mathbf{1}+2\mathbf{q}})$.

Step 2: For the integrated squared bias, consider the three leading terms B_1, B_2 and B_3 defined in Equation (SA-10.2). For B_1, by assumption in the theorem

$$\mathscr{B}_{m,\mathbf{q}}(\mathbf{x}) = -\sum_{\boldsymbol{u}\in\Lambda_m} \partial^{\boldsymbol{u}}\mu(\mathbf{x})\Big(\prod_{\ell=1}^{d}\kappa_\ell^{-u_\ell+q_\ell}g_\ell(\mathbf{x})^{-u_\ell+q_\ell}\Big)\mathbf{b}_{\mathbf{x}}^{-\boldsymbol{u}+\mathbf{q}}h_{\mathbf{x}}^{m-[\mathbf{q}]}B_{m,\mathbf{q}}(\mathbf{x})+o(h^{m-[\mathbf{q}]}).$$

Recall that $\boldsymbol{\kappa}=(\kappa_1,\dots,\kappa_d)$. Define $\mathbf{g}(\mathbf{x}) := (g_1(\mathbf{x}),\dots,g_d(\mathbf{x}))$. Using the above fact and the same notation as in the proof of Theorem 4.1, we have

$$\begin{aligned}
\mathrm{B}_1 &= \sum_{\boldsymbol{u}_1,\boldsymbol{u}_2\in\Lambda_m}\int_{\mathcal{X}}\left[\partial^{\boldsymbol{u}_1}\mu(\mathbf{x})\partial^{\boldsymbol{u}_2}\mu(\mathbf{x})\frac{h_{\mathbf{x}}^{2m-2[\mathbf{q}]}B_{\boldsymbol{u}_1,\mathbf{q}}(\mathbf{x})B_{\boldsymbol{u}_2,\mathbf{q}}(\mathbf{x})}{\boldsymbol{\kappa}^{\boldsymbol{u}_1+\boldsymbol{u}_2-2\mathbf{q}}\mathbf{g}(\mathbf{x})^{\boldsymbol{u}_1+\boldsymbol{u}_2-2\mathbf{q}}\mathbf{b}_{\mathbf{x}}^{\boldsymbol{u}_1+\boldsymbol{u}_2-2\mathbf{q}}}\right]w(\mathbf{x})d\mathbf{x}+o(h^{2m-2[\mathbf{q}]})\\
&= \sum_{\boldsymbol{u}_1,\boldsymbol{u}_2\in\Lambda_m}\boldsymbol{\kappa}^{-(\boldsymbol{u}_1+\boldsymbol{u}_2-2\mathbf{q})}\Bigg(\sum_{\delta\in\Delta}\left[\frac{\partial^{\boldsymbol{u}_1}g(t_\delta^*)\partial^{\boldsymbol{u}_2}g(t_\delta^*)w(t_\delta^*)}{\mathbf{g}(t_\delta^*)^{\boldsymbol{u}_1+\boldsymbol{u}_2-2\mathbf{q}}}\right]\\
&\qquad\qquad\times\left[\frac{h_{\mathbf{x}}^{2m-2[\mathbf{q}]}}{\mathbf{b}_{\mathbf{x}}^{\boldsymbol{u}_1+\boldsymbol{u}_2-2\mathbf{q}}}\int_\delta B_{\boldsymbol{u}_1,\mathbf{q}}(\mathbf{x})B_{\boldsymbol{u}_2,\mathbf{q}}(\mathbf{x})d\mathbf{x}\right]\Bigg)+o(h^{2m-2[\mathbf{q}]})\\
&= \sum_{\boldsymbol{u}_1,\boldsymbol{u}_2,\in\Lambda_m}\boldsymbol{\kappa}^{-(\boldsymbol{u}_1+\boldsymbol{u}_2-2\mathbf{q})}\eta_{\boldsymbol{u}_1,\boldsymbol{u}_2,\mathbf{q}}\int_{\mathcal{X}}\frac{\partial^{\boldsymbol{u}_1}\mu(\mathbf{x})\partial^{\boldsymbol{u}_2}\mu(\mathbf{x})w(\mathbf{x})}{\mathbf{g}(\mathbf{x})^{\boldsymbol{u}_1+\boldsymbol{u}_2-2\mathbf{q}}}d\mathbf{x}+o(h^{2m-2[\mathbf{q}]})
\end{aligned}$$

where the last line holds by the integrability of $\partial^{\boldsymbol{u}_1}\mu(\mathbf{x})\partial^{\boldsymbol{u}_2}\mu(\mathbf{x})w(\mathbf{x})/\mathbf{g}(\mathbf{x})^{\boldsymbol{u}_1+\boldsymbol{u}_2-2\mathbf{q}}$ over $\mathcal{X}$.

For B_2, first notice that

$$\left\|\mathbb{E}[\mathbf{p}(\mathbf{x}_i)\mathscr{B}_{m,\mathbf{0}}(\mathbf{x}_i)]+\sum_{\boldsymbol{u}\in\Lambda_m}\boldsymbol{\kappa}^{-\boldsymbol{u}}\int_{\mathcal{X}}\Big(\mathbf{p}(\mathbf{x})\partial^{\boldsymbol{u}}\mu(\mathbf{x})\mathbf{g}(\mathbf{x})^{-\boldsymbol{u}}\mathbf{b}_{\mathbf{x}}^{-\boldsymbol{u}}h_{\mathbf{x}}^{m}B_{m,\mathbf{0}}(\mathbf{x})f(\mathbf{x})\Big)d\mathbf{x}\right\|_\infty = o(h^{m+d}),$$

implying that the errors given rise to by approximating $\mathbf{b}_{\mathbf{x}}$ are of smaller order. The integral in this approximation is a vector with typical elements given by

$$\begin{aligned}
&\int_{\mathcal{X}}\Big(p_k(\mathbf{x})\partial^{\boldsymbol{u}}\mu(\mathbf{x})\mathbf{g}(\mathbf{x})^{-\boldsymbol{u}}\mathbf{b}_{\mathbf{x}}^{-\boldsymbol{u}}h_{\mathbf{x}}^{m}B_{m,\mathbf{0}}(\mathbf{x})f(\mathbf{x})\Big)d\mathbf{x}\\
&=\partial^{\boldsymbol{u}}\mu(\boldsymbol{\tau}_k)\mathbf{g}(\boldsymbol{\tau}_k)^{-\boldsymbol{u}}f(\boldsymbol{\tau}_k)\int_{\mathcal{X}}p_k(\mathbf{x})\mathbf{b}_{\mathbf{x}}^{-\boldsymbol{u}}h_{\mathbf{x}}^{m}B_{m,\mathbf{0}}(\mathbf{x})d\mathbf{x}+o(h^d).
\end{aligned}$$

Recall that $\mathbf{v}_{\boldsymbol{u},\mathbf{q}}$ defined in the theorem has the kth element equal to

$$\frac{\partial^{\boldsymbol{u}}\mu(\boldsymbol{\tau}_k)\sqrt{w(\boldsymbol{\tau}_k)}}{\boldsymbol{\kappa}^{\mathbf{q}}\mathbf{g}(\boldsymbol{\tau}_k)^{\boldsymbol{u}-\mathbf{q}}}\int_{\mathcal{X}}\partial^{\mathbf{q}}p_k(\mathbf{x})\mathbf{b}_{\mathbf{x}}^{-(\boldsymbol{u}-\mathbf{q})}h_{\mathbf{x}}^{m-[\mathbf{q}]}B_{\boldsymbol{u},\mathbf{q}}(\mathbf{x})d\mathbf{x}.$$

Given these results and Lemmas SA-2.1, it follows that

$$\begin{aligned}
\mathrm{B}_2 &= \operatorname{trace}\Big[\mathbf{Q}_m^{-1}\mathbb{E}[\mathbf{p}(\mathbf{x}_i)\mathscr{B}_{m,\mathbf{0}}(\mathbf{x}_i)]\mathbb{E}[\mathbf{p}(\mathbf{x}_i)'\mathscr{B}_{m,\mathbf{0}}(\mathbf{x}_i)]\mathbf{Q}_m^{-1}\int_{\mathcal{X}}\partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})\partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})'w(\mathbf{x})d\mathbf{x}\Big]\\
&= \sum_{\boldsymbol{u}_1,\boldsymbol{u}_2,\in\Lambda_m}\boldsymbol{\kappa}^{-(\boldsymbol{u}_1+\boldsymbol{u}_2-2\mathbf{q})}\operatorname{trace}\Big(\mathbf{Q}_m^{-1}\boldsymbol{\mathscr{D}}(f/\sqrt{w})\mathbf{v}_{\boldsymbol{u}_2,\mathbf{0}}\mathbf{v}'_{\boldsymbol{u}_1,\mathbf{0}}\boldsymbol{\mathscr{D}}(f/\sqrt{w})\times\\
&\qquad\qquad \mathbf{Q}_m^{-1}\boldsymbol{\kappa}^{-2\mathbf{q}}\boldsymbol{\mathscr{M}}_{\mathbf{q}}(w)\Big)+o(h^{2m-2[\mathbf{q}]})\\
&= \sum_{\boldsymbol{u}_1,\boldsymbol{u}_2,\in\Lambda_m}\boldsymbol{\kappa}^{-(\boldsymbol{u}_1+\boldsymbol{u}_2-2\mathbf{q})}\Big(\mathbf{v}'_{\boldsymbol{u}_1,\mathbf{0}}\boldsymbol{\mathscr{D}}(f/\sqrt{w})(\mathbf{I}+\boldsymbol{\mathscr{W}})\boldsymbol{\mathscr{D}}(f)^{-1}\boldsymbol{\mathscr{M}}(1)^{-1}\boldsymbol{\kappa}^{-2\mathbf{q}}\boldsymbol{\mathscr{M}}_{\mathbf{q}}(w)\\
&\qquad\qquad \times\boldsymbol{\mathscr{M}}(1)^{-1}\boldsymbol{\mathscr{D}}(f)^{-1}(\mathbf{I}+\boldsymbol{\mathscr{W}})'\boldsymbol{\mathscr{D}}(f/\sqrt{w})\mathbf{v}_{\boldsymbol{u}_2,\mathbf{0}}\Big)+o(h^{2m-2[\mathbf{q}]})\\
&= \sum_{\boldsymbol{u}_1,\boldsymbol{u}_2,\in\Lambda_m}\boldsymbol{\kappa}^{-(\boldsymbol{u}_1+\boldsymbol{u}_2-2\mathbf{q})}\Big(\mathbf{v}'_{\boldsymbol{u}_1,\mathbf{0}}\boldsymbol{\mathscr{D}}(1/\sqrt{w})\boldsymbol{\mathscr{M}}(1)^{-1}\boldsymbol{\kappa}^{-2\mathbf{q}}\boldsymbol{\mathscr{M}}_{\mathbf{q}}(1)\boldsymbol{\mathscr{D}}(w)\\
&\qquad\qquad \times\boldsymbol{\mathscr{M}}(1)^{-1}\boldsymbol{\mathscr{D}}(1/\sqrt{w})\mathbf{v}_{\boldsymbol{u}_2,\mathbf{0}}\Big)+o(h^{2m-2[\mathbf{q}]}).
\end{aligned}$$

It should be noted that (SA-10.4) implies that the approximation given by (SA-10.3) is still valid if $\boldsymbol{\mathscr{M}}(1)$ is pre-multiplied by $\boldsymbol{\mathscr{D}}(\phi)$ instead of being post-multiplied if $\phi(\cdot)$ is continuous. Therefore, we can repeat the argument in Step 1 and further write the term in parenthesis in the last line as:

$$\begin{aligned}
&\mathbf{v}'_{\boldsymbol{u}_1,\mathbf{0}}\boldsymbol{\mathscr{D}}(1/\sqrt{w})\boldsymbol{\mathscr{M}}(1)^{-1}\boldsymbol{\kappa}^{-2\mathbf{q}}\boldsymbol{\mathscr{M}}_{\mathbf{q}}(1)\boldsymbol{\mathscr{D}}(w)\boldsymbol{\mathscr{M}}(1)^{-1}\boldsymbol{\mathscr{D}}(1/\sqrt{w})\mathbf{v}_{\boldsymbol{u}_2,\mathbf{0}}\\
&=\mathbf{v}'_{\boldsymbol{u}_1,\mathbf{0}}\boldsymbol{\mathscr{D}}(1/\sqrt{w})\Big(\boldsymbol{\mathscr{D}}(1/\sqrt{w})\boldsymbol{\mathscr{M}}(1)\Big)^{-1}\boldsymbol{\kappa}^{-2\mathbf{q}}\boldsymbol{\mathscr{M}}_{\mathbf{q}}(1)\times\\
&\qquad\qquad \Big(\boldsymbol{\mathscr{M}}(1)\boldsymbol{\mathscr{D}}(1/\sqrt{w})\Big)^{-1}\boldsymbol{\mathscr{D}}(1/\sqrt{w})\mathbf{v}_{\boldsymbol{u}_2,\mathbf{0}}+o(1)\\
&=\mathbf{v}'_{\boldsymbol{u}_1,\mathbf{0}}\mathbf{H}_{\mathbf{0}}^{-1}\mathbf{H}_{\mathbf{q}}\mathbf{H}_{\mathbf{0}}^{-1}\mathbf{v}_{\boldsymbol{u}_2,\mathbf{0}}+o(1)
\end{aligned}$$

where $|\mathbf{v}'_{\boldsymbol{u}_1,\mathbf{0}}\mathbf{H}_{\mathbf{0}}^{-1}\mathbf{H}_{\mathbf{q}}\mathbf{H}_{\mathbf{0}}^{-1}\mathbf{v}_{\boldsymbol{u}_2,\mathbf{0}}|\lesssim 1$ by Assumption SA-2, SA-3 and Lemma SA-2.1.

Finally, for B_3, notice that

$$\begin{aligned}
&\Big\|\int_{\mathcal{X}}\mathscr{B}_{m,\mathbf{q}}(\mathbf{x})\partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})'w(\mathbf{x})d\mathbf{x}+\\
&\qquad \sum_{\boldsymbol{u}\in\Lambda_m}\int_{\mathcal{X}}\left(\frac{\partial^{\boldsymbol{u}}\mu(\mathbf{x})w(\mathbf{x})}{\boldsymbol{\kappa}^{\boldsymbol{u}-\mathbf{q}}\mathbf{g}(\mathbf{x})^{\boldsymbol{u}-\mathbf{q}}}\cdot\frac{h_{\mathbf{x}}^{m-[\mathbf{q}]}\partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})'B_{\boldsymbol{u},\mathbf{q}}(\mathbf{x})}{\mathbf{b}_{\mathbf{x}}^{\boldsymbol{u}-\mathbf{q}}}\right)d\mathbf{x}\Big\|_{\infty}=o(h^{m+d-2[\mathbf{q}]}).
\end{aligned}$$

Thus repeating the argument for B_2, we have

$$\begin{aligned}
\mathrm{B}_3 &= \left(\int_{\mathcal{X}}\mathscr{B}_{m,\mathbf{q}}(\mathbf{x})\partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})'w(\mathbf{x})d\mathbf{x}\right)\mathbf{Q}_m^{-1}\mathbb{E}[\mathbf{p}(\mathbf{x}_i)\mathscr{B}_{m,\mathbf{0}}(\mathbf{x}_i)]\\
&= \sum_{\boldsymbol{u}_1,\boldsymbol{u}_2\in\Lambda_m}\boldsymbol{\kappa}^{-(\boldsymbol{u}_1+\boldsymbol{u}_2-2\mathbf{q})}\mathbf{v}'_{\boldsymbol{u}_1,\mathbf{q}}\boldsymbol{\mathscr{D}}(\sqrt{w})\mathbf{H}_{\mathbf{0}}^{-1}\boldsymbol{\mathscr{D}}(f)^{-1}\boldsymbol{\mathscr{D}}(f/\sqrt{w})\mathbf{v}_{\boldsymbol{u}_2,\mathbf{0}}+o(h^{2m-2[\mathbf{q}]})\\
&= \sum_{\boldsymbol{u}_1,\boldsymbol{u}_2\in\Lambda_m}\boldsymbol{\kappa}^{-(\boldsymbol{u}_1+\boldsymbol{u}_2-2\mathbf{q})}\mathbf{v}'_{\boldsymbol{u}_1,\mathbf{q}}\mathbf{H}_{\mathbf{0}}^{-1}\mathbf{v}_{\boldsymbol{u}_2,\mathbf{0}}+o(h^{2m-2[\mathbf{q}]})
\end{aligned}$$

where $|\mathbf{v}'_{\boldsymbol{u}_1,\mathbf{q}}\mathbf{H}_{\mathbf{0}}^{-1}\mathbf{v}_{\boldsymbol{u}_2,\mathbf{0}}| \lesssim 1$ by Assumption SA-2, SA-3 and Lemma SA-2.1. Then the proof is complete. □

SA-10.7. *Proof of Theorem 4.2.*

PROOF. Continue the calculation in the proof of Theorem SA-3.1. For the integrated variance, when $\mathbf{q} = \mathbf{0}$ and $\mathbf{p}$ generates J complete covers, $\mathscr{M}(1)^{-1}\mathscr{M}_{\mathbf{q}}(1) = \mathbf{I}_K$ and hence

$$\operatorname{trace}\Big[\mathscr{D}(f)^{-1}\mathscr{D}(\sigma^2 f)\mathscr{D}(f)^{-1}\mathscr{D}(w)\Big] = \prod_{\ell=1}^{d}\kappa_\ell \times J\int_{\mathcal{X}}\frac{\sigma^2(\mathbf{x})w(\mathbf{x})}{f(\mathbf{x})}\prod_{\ell=1}^{d}g_\ell(\mathbf{x})\,d\mathbf{x} + o(\prod_{\ell=1}^{d}\kappa_\ell).$$

For the integrated squared bias, since the approximate orthogonality condition holds, both B_2 and B_3 are of smaller order, and the leading term in the integrated squared bias reduces to B_1 only. □

SA-10.8. *Proof of Lemma SA-4.1.*

PROOF. For $j = 0, 1$, the results directly follow from Assumption SA-5, Lemma SA-2.2, and [4, Lemma 4.1]. For $j = 2, 3$, conditional on $\mathbf{X}$, $R_{1n,\mathbf{q}}(\mathbf{x})$ has mean zero, and its variance can be bounded as follows:

$$\begin{aligned}
\mathbb{V}[R_{1n,\mathbf{q}}(\mathbf{x})|\mathbf{X}] &\lesssim \frac{1}{n}\left[\widehat{\boldsymbol{\gamma}}_{\mathbf{q},j}(\mathbf{x})' - \boldsymbol{\gamma}_{\mathbf{q},1}(\mathbf{x})'\right]\mathbb{E}[\boldsymbol{\Pi}_j(\mathbf{x}_i)\boldsymbol{\Pi}_j(\mathbf{x}_i)']\left[\widehat{\boldsymbol{\gamma}}_{\mathbf{q},j}(\mathbf{x}) - \boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})\right]\\
&\lesssim \frac{1}{n}\|\widehat{\boldsymbol{\gamma}}_{\mathbf{q},j}(\mathbf{x})' - \boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})'\|^2\|\mathbb{E}[\boldsymbol{\Pi}_j(\mathbf{x}_i)\boldsymbol{\Pi}_j(\mathbf{x}_i)']\|\\
&\lesssim_{\mathbb{P}} \frac{1}{n}h^{-2[\mathbf{q}]-2d}\frac{\log n}{nh^d}h^d = \frac{\log n}{n^2h^{2d+2[\mathbf{q}]}},
\end{aligned}$$

where the third line holds by Lemma SA-2.3 and the fact that $\mathbb{E}[\boldsymbol{\Pi}_j(\mathbf{x}_i)\boldsymbol{\Pi}_j(\mathbf{x}_i)'] \lesssim h^d$ shown in the proof of Lemma SA-2.4. Then by Chebyshev's inequality we conclude that $R_{1n,\mathbf{q}}(\mathbf{x}) \lesssim_{\mathbb{P}} \frac{\sqrt{\log n}}{nh^{d+[\mathbf{q}]}}$.

For the conditional bias $R_{2n,\mathbf{q}}(\mathbf{x})$, we analyze the least squares bias correction and plug-in bias correction separately. For $j = 2$, by construction,

$$\begin{aligned}
&\mathbb{E}[\widehat{\partial^{\mathbf{q}}\mu}_2(\mathbf{x})|\mathbf{X}] - \partial^{\mathbf{q}}\mu(\mathbf{x})\\
&= \Big(\mathbb{E}[\widehat{\partial^{\mathbf{q}}\mu}_1(\mathbf{x})|\mathbf{X}] - \partial^{\mathbf{q}}\mu(\mathbf{x})\Big) - \partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})'\widehat{\mathbf{Q}}_m^{-1}\mathbb{E}_n[\mathbf{p}(\mathbf{x}_i)\mathfrak{B}_{\tilde{m},\mathbf{0}}(\mathbf{x}_i)]\\
&= O_{\mathbb{P}}(h^{m+\varrho-[\mathbf{q}]}) - \partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})'\widehat{\mathbf{Q}}_m^{-1}\mathbb{E}_n[\mathbf{p}(\mathbf{x}_i)\mathfrak{B}_{\tilde{m},\mathbf{0}}(\mathbf{x}_i)]
\end{aligned}$$

where $\mathfrak{B}_{\tilde{m},\mathbf{0}}(\mathbf{x}_i) = \mathbb{E}[\widehat{\mu}_1(\mathbf{x}_i)|\mathbf{X}] - \mu(\mathbf{x}_i)$ is the conditional bias of $\widehat{\mu}_1(\mathbf{x}_i)$ and the last line follows from Lemma SA-2.2.

Since $\|\mathbb{E}_n[\mathbf{p}(\mathbf{x}_i)\mathfrak{B}_{\tilde{m},\mathbf{0}}(\mathbf{x}_i)]\|_\infty \leq \sup_{\mathbf{x}\in\mathcal{X}}|\mathfrak{B}_{\tilde{m},\mathbf{0}}(\mathbf{x})|\,\|\mathbb{E}_n[|\mathbf{p}(\mathbf{x}_i)|]\|_\infty$, using the same proof strategy as that for Lemma SA-2.2, we have $\|\mathbb{E}_n[|\mathbf{p}(\mathbf{x}_i)|]\|_\infty \lesssim_{\mathbb{P}} h^d$. Also, by

Lemma SA-2.2, $\sup_{\mathbf{x}\in\mathcal{X}}|\mathfrak{B}_{\tilde{m},\mathbf{0}}(\mathbf{x})| \lesssim_{\mathbb{P}} h^{m+\varrho}$. Then by Lemma SA-2.3, the conditional bias of $\widehat{\partial^{\mathbf{q}}\mu}_2(\mathbf{x})$ is $O_{\mathbb{P}}(h^{m+\varrho-[\mathbf{q}]})$.

Next, for $j=3$, using Lemma SA-2.2, we have

$$\begin{aligned}
&\mathbb{E}[\widehat{\partial^{\mathbf{q}}\mu}_3(\mathbf{x})|\mathbf{X}] - \partial^{\mathbf{q}}\mu(\mathbf{x})\\
=&\mathbb{E}\Big[\widehat{\partial^{\mathbf{q}}\mu}_0(\mathbf{x}) + \sum_{\boldsymbol{u}\in\Lambda_m}\widehat{\partial^{\boldsymbol{u}}\mu}_1(\mathbf{x})h_{\mathbf{x}}^{m-[\mathbf{q}]}B_{\boldsymbol{u},\mathbf{q}}(\mathbf{x})\\
&\qquad\qquad + \partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})'\widehat{\mathbf{Q}}_m^{-1}\mathbb{E}_n[\mathbf{p}(\mathbf{x}_i)\widehat{\mathscr{B}}_{m,\mathbf{0}}(\mathbf{x}_i)\Big|\mathbf{X}\Big] - \partial^{\mathbf{q}}\mu(\mathbf{x})\\
=&\sum_{\boldsymbol{u}\in\Lambda_m} h_{\mathbf{x}}^{m-[\mathbf{q}]}B_{\boldsymbol{u},\mathbf{q}}(\mathbf{x})\mathbb{E}[\widehat{\partial^{\boldsymbol{u}}\mu}_1(\mathbf{x}) - \partial^{\boldsymbol{u}}\mu(\mathbf{x})|\mathbf{X}]\\
&+ \partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})'\widehat{\mathbf{Q}}_m^{-1}\mathbb{E}_n\Big[\mathbf{p}(\mathbf{x}_i)\Big(\mathbb{E}[\widehat{\mathscr{B}}_{m,\mathbf{0}}(\mathbf{x}_i)|\mathbf{X}] - \mathscr{B}_{m,\mathbf{0}}(\mathbf{x}_i)\Big)\Big] + O_{\mathbb{P}}(h^{m+\varrho-[\mathbf{q}]})\\
=&\partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})'\widehat{\mathbf{Q}}_m^{-1}\mathbb{E}_n\Big[\mathbf{p}(\mathbf{x}_i)\Big(\mathbb{E}[\widehat{\mathscr{B}}_{m,\mathbf{0}}(\mathbf{x}_i)|\mathbf{X}] - \mathscr{B}_{m,\mathbf{0}}(\mathbf{x}_i)\Big)\Big] + O_{\mathbb{P}}(h^{m+\varrho-[\mathbf{q}]}),
\end{aligned}$$

where $\widehat{\mathscr{B}}_{m,\mathbf{0}}(\mathbf{x}) = -\sum_{\boldsymbol{u}\in\Lambda_m}\widehat{\partial^{\boldsymbol{u}}\mu}_1(\mathbf{x})h_{\mathbf{x}}^m B_{\boldsymbol{u},\mathbf{0}}(\mathbf{x})$, and the last line follows from Assumption SA-4, SA-5 and Lemma SA-2.2. Also by Lemma SA-2.2 and the fact that $B_{\boldsymbol{u},\mathbf{0}}(\cdot)$ is bounded, $\sup_{\mathbf{x}\in\mathcal{X}}|\mathbb{E}[\widehat{\mathscr{B}}_{m,\mathbf{0}}(\mathbf{x})|\mathbf{X}] - \mathscr{B}_{m,\mathbf{0}}(\mathbf{x})| \lesssim_{\mathbb{P}} h^{m+\varrho}$. The desired result immediately follows by the same argument as for $j=2$. □

SA-10.9. *Proof of Theorem 5.1.*

PROOF. For $j=0$, the result directly follows from [4, Theorem 4.2] combined with our Lemma SA-4.1. For $j=1,2,3$, by Lemma SA-2.4 and Lemma SA-4.1, it suffices to show that $\boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})'\mathbb{G}_n[\boldsymbol{\Pi}_j(\mathbf{x}_i)\varepsilon_i]/\sqrt{\Omega_j(\mathbf{x})}$ weakly converges to the standard Normal distribution. First, by construction

$$\mathbb{V}\Big[\frac{\boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})'}{\sqrt{\Omega_j(\mathbf{x})}}\mathbb{G}_n[\boldsymbol{\Pi}_j(\mathbf{x}_i)\varepsilon_i]\Big] = 1.$$

Next, we write $a_{ni} := \frac{\boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})'\boldsymbol{\Pi}_j(\mathbf{x}_i)}{\sqrt{\Omega_j(\mathbf{x})}}$. For all $\vartheta>0$,

$$\begin{aligned}
\sum_{i=1}^n \mathbb{E}\left[a_{ni}^2\frac{\varepsilon_i^2}{n}\mathbb{1}\{|a_{ni}\varepsilon_i/\sqrt{n}|>\vartheta\}\right] &\leq \mathbb{E}\left[\mathbb{E}\left[a_{ni}^2\varepsilon_i^2\mathbb{1}\{|\varepsilon_i|>\vartheta\sqrt{n}/|a_{ni}|\}\Big|\mathbf{x}_i\right]\right]\\
&\leq \mathbb{E}[a_{ni}^2]\cdot\sup_{\mathbf{x}\in\mathcal{X}}\mathbb{E}\left[\varepsilon_i^2\mathbb{1}\{|\varepsilon_i|>\vartheta\sqrt{n}/|a_{ni}|\}\Big|\mathbf{x}_i=\mathbf{x}\right]\\
&\lesssim \frac{h^{-2[\mathbf{q}]-2d}h^d}{h^{-d-2[\mathbf{q}]}}\sup_{\mathbf{x}\in\mathcal{X}}\mathbb{E}\left[\varepsilon_i^2\mathbb{1}\Big\{|\varepsilon_i|>\frac{\vartheta\sqrt{n}}{|a_{ni}|}\Big\}\Big|\mathbf{x}_i=\mathbf{x}\right]\\
&\lesssim \sup_{\mathbf{x}\in\mathcal{X}}\mathbb{E}\left[\varepsilon_i^2\mathbb{1}\{|\varepsilon_i|>\vartheta\sqrt{n}/|a_{ni}|\}\Big|\mathbf{x}_i=\mathbf{x}\right],
\end{aligned}\tag{SA-10.5}$$

where the third line follows from Lemma SA-2.3 and Lemma SA-2.4. Since $|a_{ni}| \lesssim h^{-d/2}$ and $\log n/(nh^d) = o(1)$, it follows that $\sqrt{n}/|a_{ni}| \to \infty$ as $n \to \infty$. By the moment condition in the theorem, the upper bound in Eq. (SA-10.5) goes to 0 as $n \to \infty$, which completes the proof. □

SA-10.10. *Proof of Lemma SA-4.2.*

PROOF. Consider the conditions in (i) hold. The proof is divided into two steps.

Step 1: We first bound $\sup_{\mathbf{x}\in\mathcal{X}} |R_{1n,\mathbf{q}}(\mathbf{x})|$ for $j = 0, 1, 2, 3$. To simplify notation, we write $\mathbf{\Pi}_j(\mathbf{x}_i) = (\pi_1(\mathbf{x}_i), \ldots, \pi_{K_j}(\mathbf{x}_i))'$ where $K_j = \dim(\mathbf{\Pi}_j(\cdot))$. We truncate the errors by an increasing sequence of constants $\{\vartheta_n : n \geq 1\}$ such that $\vartheta_n \asymp \sqrt{nh^d/\log n}$. Let $H_{ik} = \pi_k(\mathbf{x}_i)(\varepsilon_i \mathbb{1}\{|\varepsilon_i| \leq \vartheta_n\} - \mathbb{E}[\varepsilon_i \mathbb{1}\{|\varepsilon_i| \leq \vartheta_n\}|\mathbf{x}_i])$ and $T_{ik} = \pi_k(\mathbf{x}_i)(\varepsilon_i \mathbb{1}\{|\varepsilon_i| > \vartheta_n\} - \mathbb{E}[\varepsilon_i \mathbb{1}\{|\varepsilon_i| > \vartheta_n\}|\mathbf{x}_i])$. Regarding the truncated term, it follows from the truncation strategy, Assumption SA-3, and SA-5 that $|H_{ik}| \leq \vartheta_n$ and $\mathbb{E}[H_{ik}^2] \lesssim h^d$. By Bernstein's inequality, for every $t > 0$,

$$
\begin{aligned}
&\mathbb{P}\left(\max_{1\leq k\leq K_j} |\mathbb{E}_n[H_{ik}]| > h^d \sqrt{\log n/(nh^d)}\, t\right) \\
&\leq 2\sum_{k=1}^{K_j} \exp\left\{-\frac{n^2 h^{2d} h^{-d} \log n t^2/n}{C_1 nh^d + C_2 \vartheta_n nh^d \sqrt{\log n/(nh^d)} t}\right\} \\
&\leq C \exp\left\{\log n \Big(1 - \frac{t^2}{C_1 + C_2 \vartheta_n \sqrt{\log n/(nh^d)}\, t}\Big)\right\}, \qquad \text{(SA-10.6)}
\end{aligned}
$$

which is arbitrarily small for t large enough by the truncation strategy. By Lemma SA-2.3,

$$
\begin{aligned}
&\sup_{\mathbf{x}\in\mathcal{X}} \left|(\widehat{\boldsymbol{\gamma}}_{\mathbf{q},j}(\mathbf{x})' - \boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})')\mathbb{E}_n[\mathbf{\Pi}_j(\mathbf{x}_i)(\varepsilon_i \mathbb{1}\{|\varepsilon_i| \leq \vartheta_n\} - \mathbb{E}[\varepsilon_i \mathbb{1}\{|\varepsilon_i| \leq \vartheta_n\}|\mathbf{x}_i])]\right| \\
&\leq \sup_{\mathbf{x}\in\mathcal{X}} \|\widehat{\boldsymbol{\gamma}}_{\mathbf{q},j}(\mathbf{x})' - \boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})'\|_\infty \|\mathbb{E}_n[\mathbf{\Pi}_j(\mathbf{x}_i)(\varepsilon_i \mathbb{1}\{|\varepsilon_i| \leq \vartheta_n\} - \mathbb{E}[\varepsilon_i \mathbb{1}\{|\varepsilon_i| \leq \vartheta_n\}|\mathbf{x}_i])]\|_\infty \\
&\lesssim_{\mathbb{P}} h^{-[\mathbf{q}]-d} \sqrt{\log n/(nh^d)} h^d \sqrt{\log n/(nh^d)} = h^{-[\mathbf{q}]} \log n/(nh^d).
\end{aligned}
$$

Regarding the tails, let $\mathscr{K}_{ji}(\mathbf{x}) := (\widehat{\boldsymbol{\gamma}}_{\mathbf{q},j}(\mathbf{x})' - \boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})')\mathbf{\Pi}_j(\mathbf{x}_i)$. By Lemma SA-2.3 and Assumption SA-3, we have $\sup_{\mathbf{x}\in\mathcal{X}} |\mathscr{K}_{ji}(\mathbf{x})| \lesssim_{\mathbb{P}} h^{-d-[\mathbf{q}]} \sqrt{\log n/(nh^d)}$. Let $\mathcal{A}_n(M)$ denote the event on which $\sup_{\mathbf{x}\in\mathcal{X}} |\mathscr{K}_{ji}(\mathbf{x})| \leq M h^{-d-[\mathbf{q}]} \sqrt{\log n/(nh^d)}$ for some $M > 0$, and $\mathbb{1}_{\mathcal{A}_n(M)}$ be an indicator function of $\mathcal{A}_n(M)$. Then by Markov's inequality, for any $t > 0$,

$$\mathbb{P}\Big(\sup_{\mathbf{x}\in\mathcal{X}}\Big|\mathbb{E}_n[\mathbb{1}_{\mathcal{A}_n(M)}\mathscr{K}_{ji}(\varepsilon_i\mathbb{1}\{|\varepsilon_i|>\vartheta_n\}-\mathbb{E}[\varepsilon_i\mathbb{1}\{|\varepsilon_i|>\vartheta_n\}|\mathbf{x}_i])]\Big| > \frac{t\log n}{nh^{d+[\mathbf{q}]}}\Big)$$

$$\lesssim \frac{Mh^{-d-[\mathbf{q}]}\sqrt{\log n/(nh^d)}\mathbb{E}[|\varepsilon_i|\mathbb{1}\{|\varepsilon_i|>\vartheta_n\}]}{th^{-[\mathbf{q}]}\log n/(nh^d)}$$

(SA-10.7)

$$\le \frac{M\sqrt{n}}{t\sqrt{h^d\log n}}\frac{\mathbb{E}[|\varepsilon_i|^{2+\nu}]}{\vartheta_n^{1+\nu}}$$

which is arbitrarily small for t/M large enough by the additional moment condition specified in the lemma and the rate restriction. Since $\mathbb{P}(\mathcal{A}_n(M)^c) = o(1)$ as $M\to\infty$, simply let $t = M^2$ and $M\to\infty$, then the desired conclusion immediately follows.

Step 2: Next, we bound $\sup_{\mathbf{x}\in\mathcal{X}}|R_{2n,\mathbf{q}}(\mathbf{x})|$. For $j = 0, 1$, the result directly follows from Lemma SA-2.2. For $j = 2$, notice that the proof of Lemma SA-4.1 essentially establishes a bound on the uniform norm of $\partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})'\mathbf{Q}_m^{-1}\mathbb{E}_n[\mathbf{p}(\mathbf{x}_i)\mathfrak{B}_{\tilde{m},\mathbf{0}}(\mathbf{x}_i)]$. The bound on $(\mathbb{E}[\widehat{\partial^{\mathbf{q}}\mu}(\mathbf{x})|\mathbf{X}] - \partial^{\mathbf{q}}\mu(\mathbf{x}))$ follows from Lemma SA-2.2. Then the desired bound on $R_{2n,\mathbf{q}}(\mathbf{x})$ is obtained.

For $j = 3$, notice that by Assumption SA-4 we can write $R_{2n,\mathbf{q}}(\mathbf{x})$ explicitly as

$$\widehat{\gamma}_{\mathbf{q},0}(\mathbf{x})\Big\{\mathbb{E}_n[\mathbf{p}(\mathbf{x}_i)(\mu(\mathbf{x}_i)-s^*(\mathbf{x}_i))] - \mathbb{E}_n\Big[\mathbf{p}(\mathbf{x}_i)\sum_{\boldsymbol{u}\in\Lambda_m}h_{\mathbf{x}_i}^m B_{\boldsymbol{u},\mathbf{0}}(\mathbf{x}_i)\mathbb{E}\Big[\widehat{\partial^{\boldsymbol{u}}\mu}_1(\mathbf{x}_i)\Big|\mathbf{X}\Big]\Big]\Big\}$$
$$+\,\partial^{\mathbf{q}}s^*(\mathbf{x}) - \partial^{\mathbf{q}}\mu(\mathbf{x}) - \mathscr{B}_{m,\mathbf{q}}(\mathbf{x})$$
$$+\,\mathscr{B}_{m,\mathbf{q}}(\mathbf{x}) + \sum_{\boldsymbol{u}\in\Lambda_m}h_{\mathbf{x}}^{m-[\mathbf{q}]}B_{\boldsymbol{u},\mathbf{q}}(\mathbf{x})\partial^{\boldsymbol{u}}\tilde{\mathbf{p}}(\mathbf{x})'\widehat{\mathbf{Q}}_{\tilde{m}}^{-1}\mathbb{E}_n[\tilde{\mathbf{p}}(\mathbf{x}_i)\mu(\mathbf{x}_i)].$$

The second line is uniformly bounded by Assumption SA-4: $\sup_{\mathbf{x}\in\mathcal{X}}|\partial^{\mathbf{q}}s^*(\mathbf{x}) - \partial^{\mathbf{q}}\mu(\mathbf{x}) - \mathscr{B}_{m,\mathbf{q}}(\mathbf{x})| \lesssim h^{m+\varrho-[\mathbf{q}]}$. The third line can be written as

$$\mathscr{B}_{m,\mathbf{q}}(\mathbf{x}) + \sum_{\boldsymbol{u}\in\Lambda_m}h_{\mathbf{x}}^{m-[\mathbf{q}]}B_{\boldsymbol{u},\mathbf{q}}(\mathbf{x})\partial^{\boldsymbol{u}}\tilde{\mathbf{p}}(\mathbf{x})'\widehat{\mathbf{Q}}_{\tilde{m}}^{-1}\mathbb{E}_n[\tilde{\mathbf{p}}(\mathbf{x}_i)\mu(\mathbf{x}_i)]$$
$$= \sum_{\boldsymbol{u}\in\Lambda_m}h_{\mathbf{x}}^{m-[\mathbf{q}]}B_{\boldsymbol{u},\mathbf{q}}(\mathbf{x})\Big(\partial^{\boldsymbol{u}}\tilde{\mathbf{p}}(\mathbf{x})'\widehat{\mathbf{Q}}_{\tilde{m}}^{-1}\mathbb{E}_n[\tilde{\mathbf{p}}(\mathbf{x}_i)\mu(\mathbf{x}_i)] - \partial^{\boldsymbol{u}}\mu(\mathbf{x})\Big).$$

By Assumption SA-4, $\sup_{\mathbf{x}\in\mathcal{X}}|B_{\boldsymbol{u},\mathbf{q}}(\mathbf{x})| \lesssim 1$. Moreover, as we have shown in the proof of Lemma SA-2.2, $\sup_{\mathbf{x}\in\mathcal{X}}|\partial^{\boldsymbol{u}}\tilde{\mathbf{p}}(\mathbf{x})'\widehat{\mathbf{Q}}_{\tilde{m}}^{-1}\mathbb{E}_n[\tilde{\mathbf{p}}(\mathbf{x}_i)\mu(\mathbf{x}_i)] - \partial^{\boldsymbol{u}}\mu(\mathbf{x})| \lesssim_{\mathbb{P}} h^{\tilde{m}-[\boldsymbol{u}]}$ which suffices to show that the third line is $O_{\mathbb{P}}(h^{m+\varrho-[\mathbf{q}]})$.

In addition, using the proof strategy for Lemma SA-2.2, we have

$$\sup_{\mathbf{x}\in\mathcal{X}}\Big|\widehat{\gamma}_{\mathbf{q},0}(\mathbf{x})'\mathbb{E}_n[\mathbf{p}(\mathbf{x}_i)(\mu(\mathbf{x}_i)-s^*(\mathbf{x}_i)+\mathscr{B}_{m,\mathbf{0}}(\mathbf{x}_i))]\Big| \lesssim_{\mathbb{P}} h^{m+\varrho-[\mathbf{q}]}, \quad \text{and}$$
$$\sup_{\mathbf{x}\in\mathcal{X}}\Big|\widehat{\gamma}_{\mathbf{q},0}(\mathbf{x})'\mathbb{E}_n\Big[\mathbf{p}(\mathbf{x}_i)h_{\mathbf{x}_i}^m B_{\boldsymbol{u},\mathbf{0}}(\mathbf{x}_i)\Big(\mathbb{E}[\widehat{\partial^{\boldsymbol{u}}\mu}_1(\mathbf{x}_i)|\mathbf{X}] - \partial^{\boldsymbol{u}}\mu(\mathbf{x}_i)\Big)\Big]\Big| \lesssim_{\mathbb{P}} h^{m+\varrho-[\mathbf{q}]}.$$

Finally, if the conditions in (ii) hold, then it suffices to adjust the proof for $R_{1n,\mathbf{q}}(\mathbf{x})$. We still use the same proof strategy, but let $\vartheta_n = \log n$. For the truncated term, it can be seen from Bernstein's inequality (Equation SA-10.6) that the upper bound can be made arbitrarily small for t large enough when $(\log n)^3/(nh^d) \lesssim 1$. On the other hand, when applying Markov's inequality to control the tail (Equation (SA-10.7)), we employ the exponential moment condition:

$$\mathbb{P}\left(\sup_{\mathbf{x}\in\mathcal{X}}\left|\mathbb{E}_n[\mathbb{1}_{\mathcal{A}_n(M)}\mathscr{K}_{ji}(\varepsilon_i\mathbb{1}\{|\varepsilon_i|>\vartheta_n\}-\mathbb{E}[\varepsilon_i\mathbb{1}\{|\varepsilon_i|>\vartheta_n\}|\mathbf{x}_i])]\right|>\frac{t\log n}{nh^{d+[\mathbf{q}]}}\right)$$
$$\lesssim \frac{M\sqrt{n}}{t\sqrt{h^d\log n}}\mathbb{E}[|\varepsilon_i|\mathbb{1}\{|\varepsilon_i|>\vartheta_n\}] \leq \frac{M\sqrt{n}}{t\sqrt{h^d\log n}}\frac{\mathbb{E}[|\varepsilon_i|^3\exp(|\varepsilon_i|)]}{\vartheta_n^2\exp(\vartheta_n)}$$
$$\leq \frac{M}{t(\log n)^{5/2}\sqrt{nh^d}}\mathbb{E}[|\varepsilon_i|^3\exp(|\varepsilon_i|)]$$

which is arbitrarily small for t/M large enough. Thus the same bound on $R_{1n,\mathbf{q}}$ is established. Then the proof is complete. □

SA-10.11. *Proof of Theorem SA-4.1.*

PROOF. Regarding the L_2 convergence, by Lemma SA-2.2,

$$\int_{\mathcal{X}}\left(\left(\widehat{\partial^{\mathbf{q}}\mu}_0(\mathbf{x})-\partial^{\mathbf{q}}\mu(\mathbf{x})\right)^2\right)w(\mathbf{x})d\mathbf{x}$$
$$=\Big(\mathbb{E}_n[\mathbf{\Pi}_0(\mathbf{x}_i)\varepsilon_i]'\Big)\Big(\int_{\mathcal{X}}\widehat{\boldsymbol{\gamma}}_{\mathbf{q},0}(\mathbf{x})\widehat{\boldsymbol{\gamma}}_{\mathbf{q},0}(\mathbf{x})'w(\mathbf{x})d\mathbf{x}\Big)\Big(\mathbb{E}_n[\mathbf{\Pi}_0(\mathbf{x}_i)\varepsilon_i]\Big)+O_{\mathbb{P}}(h^{2(m-[\mathbf{q}])}).$$

Notice that in the proof of Lemma SA-2.2, the uniform bound on the conditional bias does not require explicit expression of leading approximation error. Then by Lemma SA-2.1, we have $\int_{\mathcal{X}}\widehat{\boldsymbol{\gamma}}_{\mathbf{q},0}(\mathbf{x})\widehat{\boldsymbol{\gamma}}_{\mathbf{q},0}(\mathbf{x})'w(\mathbf{x})d\mathbf{x}\lesssim h^{-d-2[\mathbf{q}]}$. Also, $\mathbb{E}[\|\mathbb{E}_n[\mathbf{\Pi}_0(\mathbf{x}_i)\varepsilon_i]\|^2]\lesssim \mathbb{E}[\mathbf{\Pi}_0(\mathbf{x}_i)'\mathbf{\Pi}_0(\mathbf{x}_i)/n]\lesssim 1/n$. The desired L_2-convergence rate follows.

Regarding the uniform convergence, consider the case when the conditions of Lemma SA-4.2 hold. We use the same truncation strategy. Specifically, separate ε_i into

$$\varepsilon_i\mathbb{1}\{|\varepsilon_i|\leq\vartheta_n\}-\mathbb{E}[\varepsilon_i\mathbb{1}\{|\varepsilon_i|\leq\vartheta_n\}|\mathbf{x}_i]\quad\text{and}\quad\varepsilon_i\mathbb{1}\{|\varepsilon_i|>\vartheta_n\}-\mathbb{E}[\varepsilon_i\mathbb{1}\{|\varepsilon_i|>\vartheta_n\}|\mathbf{x}_i],$$

where $\vartheta_n \asymp \sqrt{nh^d/\log n}$. By Lemmas SA-2.3, $\sup_{\mathbf{x}\in\mathcal{X}}|\boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})'\mathbf{\Pi}_j(\mathbf{x}_i)|\lesssim_{\mathbb{P}} h^{-d-[\mathbf{q}]}$. Then repeating the argument given in the proof of Lemma SA-4.2 for the truncated and tails respectively, we have

$$\sup_{\mathbf{x}\in\mathcal{X}}|\boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})'\mathbb{E}_n[\mathbf{\Pi}_j(\mathbf{x}_i)\varepsilon_i]|\lesssim_{\mathbb{P}} h^{-[\mathbf{q}]}\sqrt{\log n/(nh^d)}.$$

Moreover, $\bar{R}_{1n,\mathbf{q}} = o(h^{-d/2-[\mathbf{q}]}\sqrt{\log n/n})$ since $\log n/(nh^d) = o(1)$. In view of these bounds and Lemma SA-4.2, the desired rate of uniform convergence follows.

Finally, the same results can be proved under the conditions in (ii) of Lemma SA-4.2 if we let $\vartheta_n = \log n$ and assume $(\log n)^3/(nh^d)\lesssim 1$. □

SA-10.12. *Proof of Theorem SA-4.2.*

PROOF. First consider the conditions in part (i) hold. Notice that for $j = 0, 1, 2, 3$,

$$\widehat{\mathbf{\Sigma}}_j - \mathbf{\Sigma}_j = \mathbb{E}_n[(\widehat{\varepsilon}^2_{i,j} - \varepsilon^2_i)\mathbf{\Pi}_j(\mathbf{x}_i)\mathbf{\Pi}_j(\mathbf{x}_i)'] + \Big(\mathbb{E}_n[\varepsilon^2_i\mathbf{\Pi}_j(\mathbf{x}_i)\mathbf{\Pi}_j(\mathbf{x}_i)'] - \mathbf{\Sigma}_j\Big). \quad \text{(SA-10.8)}$$

We then divide our proof into four steps.

Step 1: For $\|\mathbb{E}_n[\mathbf{\Pi}_j(\mathbf{x}_i)\mathbf{\Pi}_j(\mathbf{x}_i)']\|$, by Assumption SA-3, SA-5, and the same argument used in the proof of Lemma SA-2.1, $\|\mathbb{E}[\mathbf{\Pi}_j(\mathbf{x}_i)'\mathbf{\Pi}_j(\mathbf{x}_i)']\| \lesssim h^d$, and

$$\|\mathbb{E}_n[\mathbf{\Pi}_j(\mathbf{x}_i)\mathbf{\Pi}_j(\mathbf{x}_i)'] - \mathbb{E}[\mathbf{\Pi}_j(\mathbf{x}_i)\mathbf{\Pi}_j(\mathbf{x}_i)']\| \lesssim_{\mathbb{P}} h^d\sqrt{\log n/(nh^d)}.$$

By the triangle inequality and the fact that $\frac{\log n}{nh^d} = o(1)$, $\|\mathbb{E}_n[\mathbf{\Pi}_j(\mathbf{x}_i)\mathbf{\Pi}_j(\mathbf{x}_i)'] \lesssim_{\mathbb{P}} h^d$.

Step 2: Next, we bound the second term in Equation (SA-10.8). To simplify our notations, let $\mathbf{L}_j(\mathbf{x}_i) := \mathbf{W}_j^{-1/2}\mathbf{\Pi}_j(\mathbf{x}_i)$ be the normalized basis, where $\mathbf{W}_j = \mathbf{Q}_m$ for $j = 0$, $\mathbf{W}_j = \mathbf{Q}_{\tilde{m}}$ for $j = 1$, and $\mathbf{W}_j = \mathrm{diag}\{\mathbf{Q}_m, \mathbf{Q}_{\tilde{m}}\}$ for $j = 2, 3$. Introduce a sequence of positive numbers: $M_n^2 \asymp \frac{K^{1+1/\nu}n^{1/(2+\nu)}}{(\log n)^{1/(2+\nu)}}$. Then, we write

$$\mathbf{H}_j(\mathbf{x}_i) = \varepsilon^2_i\mathbf{L}_j(\mathbf{x}_i)\mathbf{L}_j(\mathbf{x}_i)'\mathbb{1}\{\|\varepsilon^2_i\mathbf{L}_j(\mathbf{x}_i)\mathbf{L}_j(\mathbf{x}_i)'\| \le M_n^2\},$$

and

$$\mathbf{T}_j(\mathbf{x}_i) = \varepsilon^2_i\mathbf{L}_j(\mathbf{x}_i)\mathbf{L}_j(\mathbf{x}_i)'\mathbb{1}\{\|\varepsilon^2_i\mathbf{L}_j(\mathbf{x}_i)\mathbf{L}_j(\mathbf{x}_i)'\| > M_n^2\}.$$

Clearly,

$$\begin{aligned}&\mathbb{E}_n[\mathbf{L}_j(\mathbf{x}_i)\mathbf{L}_j(\mathbf{x}_i)'\varepsilon^2_i] - \mathbb{E}[\mathbf{L}_j(\mathbf{x}_i)\mathbf{L}_j(\mathbf{x}_i)'\varepsilon^2_i]\\ &= (\mathbb{E}_n[\mathbf{H}_j(\mathbf{x}_i)] - \mathbb{E}[\mathbf{H}_j(\mathbf{x}_i)]) + (\mathbb{E}_n[\mathbf{T}_j(\mathbf{x}_i)] - \mathbb{E}[\mathbf{T}_j(\mathbf{x}_i)]).\end{aligned}$$

For the truncated terms, by definition, $\|\mathbf{H}_j(\mathbf{x}_i)\| \le M_n^2$. By the triangle inequality and Jensen's inequality, $\|\mathbf{H}_j(\mathbf{x}_i) - \mathbb{E}[\mathbf{H}_j(\mathbf{x}_i)]\| \le 2M_n^2$. Also, by Assumption SA-1,

$$\begin{aligned}\mathbb{E}[(\mathbf{H}_j(\mathbf{x}_i) - \mathbb{E}[\mathbf{H}_j(\mathbf{x}_i)])^2] &\le \mathbb{E}[\varepsilon^4_i\|\mathbf{L}_j(\mathbf{x}_i)\|^2\mathbf{L}_j(\mathbf{x}_i)\mathbf{L}_j(\mathbf{x}_i)'\mathbb{1}\{\|\varepsilon^2_i\mathbf{L}_j(\mathbf{x}_i)\mathbf{L}_j(\mathbf{x}_i)'\| \le M_n^2\}]\\ &\le M_n^2\mathbb{E}[\varepsilon^2_i\mathbf{L}_j(\mathbf{x}_i)\mathbf{L}_j(\mathbf{x}_i)'\mathbb{1}\{\|\varepsilon^2_i\mathbf{L}_j(\mathbf{x}_i)\mathbf{L}_j(\mathbf{x}_i)'\| \le M_n^2\}]\\ &\lesssim M_n^2\mathbb{E}[\mathbf{L}_j(\mathbf{x}_i)\mathbf{L}_j(\mathbf{x}_i)'],\end{aligned}$$

where the inequalities are understood in the sense of semi-definite matrices. Thus $\|\mathbb{E}[(\mathbf{H}_j(\mathbf{x}_i) - \mathbb{E}[\mathbf{H}_j(\mathbf{x}_i)])^2]\| \lesssim M_n^2$. Let $\vartheta_n = \sqrt{(\log n)^{\frac{\nu}{2+\nu}}/(n^{\frac{\nu}{2+\nu}}h^d)}$. By an inequality of [25] for independent matrices, for all $t > 0$,

$$\begin{aligned}&\mathbb{P}\Big(\Big\|\mathbb{E}_n[\mathbf{H}_j(\mathbf{x}_i)] - \mathbb{E}[\mathbf{H}_j(\mathbf{x}_i)]\Big\| > \vartheta_n t\Big)\\ &\le \exp\Big(\log n - \frac{\vartheta_n^2t^2/2}{M_n^2/n + M_n^2\vartheta_n t/(3n)}\Big)\\ &\le \exp\Big\{\log n\Big(1 - \frac{t^2/2}{M_n^2\log n\,\vartheta_n^{-2}n^{-1}(1 + \vartheta_n t/3)}\Big)\Big\}\end{aligned}$$

where $M_n^2 \log n\, \vartheta_n^{-2} n^{-1} \asymp (\log n)^{\frac{1}{2+\nu}} / (n^{\frac{1}{2+\nu}} h^{d/\nu}) = o(1)$ and $\vartheta_n = o(1)$. Hence

$$\left\| \mathbb{E}_n[\mathbf{H}_j(\mathbf{x}_i)] - \mathbb{E}[\mathbf{H}_j(\mathbf{x}_i)] \right\| \lesssim_{\mathbb{P}} \vartheta_n = o(1).$$

Regarding the tails, it directly follows from Lemma SA-2.1 that $\|\mathbf{T}_j(\mathbf{x}_i)\| \lesssim h^{-d} \varepsilon_i^2 \mathbb{1}\{\varepsilon_i^2 \gtrsim M_n^2 h^d\}$. Then by the triangle inequality, Jensen's inequality, and the assumption that $(2+\nu)$th moment of ε_i is bounded,

$$\begin{aligned}\mathbb{E}\left[\left\| \mathbb{E}_n[\mathbf{T}_j(\mathbf{x}_i)] - \mathbb{E}[\mathbf{T}_j(\mathbf{x}_i)] \right\| \right] &\lesssim 2h^{-d} \mathbb{E}[\varepsilon_i^2 \mathbb{1}\{|\varepsilon_i| \gtrsim M_n \sqrt{h^d}\}] \\ &\lesssim \frac{2h^{-d(1+\nu/2)} \mathbb{E}[|\varepsilon_i|^{2+\nu} \mathbb{1}\{|\varepsilon_i| \gtrsim M_n \sqrt{h^d}\}]}{M_n^\nu} \lesssim \vartheta_n.\end{aligned}$$

By Markov's inequality, $\|\mathbb{E}_n[\mathbf{T}_j(\mathbf{x}_i) - \mathbb{E}[\mathbf{T}_j(\mathbf{x}_i)]]\| \lesssim_{\mathbb{P}} \vartheta_n$. Since $\|\mathbf{W}_j^{1/2}\| \lesssim h^{d/2}$ and $\|\mathbf{W}_j^{-1/2}\| \lesssim h^{-d/2}$, we conclude that $\|\mathbb{E}_n[\mathbf{\Pi}_j(\mathbf{x}_i)\mathbf{\Pi}_j(\mathbf{x}_i)'\varepsilon_i^2] - \mathbf{\Sigma}_j\| \lesssim_{\mathbb{P}} h^d \vartheta_n = o_{\mathbb{P}}(h^d)$.

Step 3: The first term in Equation (SA-10.8) satisfies

$$\begin{aligned}&\|\mathbb{E}_n[(\widehat{\varepsilon}_{i,j}^2 - \varepsilon_i^2)\mathbf{\Pi}_j(\mathbf{x}_i)\mathbf{\Pi}_j(\mathbf{x}_i)']\| \\ \leq\, &\|\mathbb{E}_n[(\mu(\mathbf{x}_i) - \widehat{\mu}_j(\mathbf{x}_i))^2 \mathbf{\Pi}_j(\mathbf{x}_i)\mathbf{\Pi}_j(\mathbf{x}_i)']\| + 2\|\mathbb{E}_n[(\mu(\mathbf{x}_i) - \widehat{\mu}_j(\mathbf{x}_i))\varepsilon_i \mathbf{\Pi}_j(\mathbf{x}_i)\mathbf{\Pi}_j(\mathbf{x}_i)']\| \\ \leq\, &\max_{1\leq i\leq n} |\mu(\mathbf{x}_i) - \widehat{\mu}_j(\mathbf{x}_i)|^2 \|\mathbb{E}_n[\mathbf{\Pi}_j(\mathbf{x}_i)\mathbf{\Pi}_j(\mathbf{x}_i)']\| + \\ &\max_{1\leq i\leq n} |\mu(\mathbf{x}_i) - \widehat{\mu}_j(\mathbf{x}_i)| \left(\|\mathbb{E}_n[\mathbf{\Pi}_j(\mathbf{x}_i)\mathbf{\Pi}_j(\mathbf{x}_i)']\| + \|\mathbb{E}_n[\mathbf{\Pi}_j(\mathbf{x}_i)\mathbf{\Pi}_j(\mathbf{x}_i)'\varepsilon_i^2]\| \right)\end{aligned}$$

where the last line follows from the fact that $2|a| \leq 1 + a^2$. By Theorem SA-4.1 and the results proved in Step 1 and 2, we have $\max_{1\leq i\leq n} |\mu(\mathbf{x}_i) - \widehat{\mu}_j(\mathbf{x}_i)| = R_{\mathbf{0},j}^{\mathtt{uc}} = o_{\mathbb{P}}(1)$, $\|\mathbb{E}_n[\mathbf{\Pi}_j(\mathbf{x}_i)\mathbf{\Pi}_j(\mathbf{x}_i)']\| \lesssim_{\mathbb{P}} h^d$ and $\|\mathbb{E}_n[\mathbf{\Pi}_j(\mathbf{x}_i)\mathbf{\Pi}_j(\mathbf{x}_i)'\varepsilon_i^2]\| \lesssim_{\mathbb{P}} h^d$. Hence

$$\|\widehat{\mathbf{\Sigma}}_j - \mathbf{\Sigma}_j\| \lesssim_{\mathbb{P}} h^d (R_{\mathbf{0},j}^{\mathtt{uc}} + \vartheta_n) = o_{\mathbb{P}}(h^d). \tag{SA-10.9}$$

Step 4: Using all above results, we have

$$\begin{aligned}\left| \widehat{\Omega}_j(\mathbf{x}) - \Omega_j(\mathbf{x}) \right| =\, &\|\widehat{\boldsymbol{\gamma}}_{\mathbf{q},j}(\mathbf{x})' \widehat{\mathbf{\Sigma}}_j \widehat{\boldsymbol{\gamma}}_{\mathbf{q},j}(\mathbf{x}) - \boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})' \mathbf{\Sigma}_j \boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})\| \\ \leq\, &\|(\widehat{\boldsymbol{\gamma}}_{\mathbf{q},j}(\mathbf{x})' - \boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x}))' \widehat{\mathbf{\Sigma}}_j \widehat{\boldsymbol{\gamma}}_{\mathbf{q},j}(\mathbf{x})\| + \|\boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})' (\widehat{\mathbf{\Sigma}}_j - \mathbf{\Sigma}_j) \widehat{\boldsymbol{\gamma}}_{\mathbf{q},j}(\mathbf{x})\| \\ &+ \|\boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})' \mathbf{\Sigma}_j (\widehat{\boldsymbol{\gamma}}_{\mathbf{q},j}(\mathbf{x}) - \boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x}))\|.\end{aligned}$$

By Lemma SA-2.3 and Equation (SA-10.9), we have

$$\sup_{\mathbf{x}\in\mathcal{X}} \left| \widehat{\Omega}_j(\mathbf{x}) - \Omega_j(\mathbf{x}) \right| \lesssim_{\mathbb{P}} h^{-d-2[\mathbf{q}]} (R_{\mathbf{0},j}^{\mathtt{uc}} + \vartheta_n) = o_{\mathbb{P}}(h^{-d-2[\mathbf{q}]}).$$

Finally, when the conditions in part (ii) hold, we only need to adjust the proof in Step 2. Apply the same proof strategy with $M_n = \sqrt{Ch^{-d}}\log n$ and $\vartheta_n = \sqrt{\frac{(\log n)^3}{nh^d}}$. For the truncated term, since $M_n^2 \log n/(\vartheta_n^2 n) \lesssim 1$, by Bernstein's inequality, $\|\mathbb{E}_n[\mathbf{H}_j(\mathbf{x}_i)] - \mathbb{E}[\mathbf{H}_j(\mathbf{x}_i)]\| \lesssim_{\mathbb{P}} \vartheta_n = o_{\mathbb{P}}(1)$. On the other hand, when applying Markov's inequality to bound the tail, we employ the stronger moment condition:

$$\begin{aligned}\mathbb{E}\Big[\|\mathbb{E}_n[\mathbf{T}_j(\mathbf{x}_i)] - \mathbb{E}[\mathbf{T}_j(\mathbf{x}_i)]\|\Big] &\lesssim 2h^{-d}\mathbb{E}[\varepsilon_i^2 \mathbb{1}\{|\varepsilon_i| \geq M_n/\sqrt{Ch^{-d}}\}] \\ &\lesssim \frac{h^{-3d/2}}{M_n \exp(M_n/\sqrt{Ch^{-d}})}\mathbb{E}[|\varepsilon_i|^3 \exp(|\varepsilon_i|)] \lesssim \vartheta_n = o(1).\end{aligned}$$

Then the proof is complete. □

SA-10.13. *Proof of Theorem SA-5.1.*

Proof. By Lemma SA-5.1, the remainders in the linearization of $\widehat{T}_j(\cdot)$ is $o_{\mathbb{P}}(r_n^{-1})$ in $\mathcal{L}^\infty(\mathcal{X})$, and hence we only need to show that $Z_j(\cdot)$ approximates $t_j(\cdot)$.

Suppose that the conditions in (i) hold. As a first step, we define $\mathscr{K}(\mathbf{x}, \mathbf{x}_i) := \frac{\boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})'\boldsymbol{\Pi}_j(\mathbf{x}_i)}{\sqrt{\Omega_j(\mathbf{x})}}$ for each $j = 0, 1, 2, 3$, and will construct conditional coupling for $t_j(\cdot)$. One can verify conditions in Lemma 8.2 of the main paper (see Section 8.3) by using the properties of local bases considered in this paper, but here we directly follow the strategy given in the proof of Lemma 8.2, which will unify the proofs for $d = 1$ and Haar basis with $d > 1$.

We need to construct a rearranged sequence $\{\mathbf{x}_{i,n}\}_{i=1}^n$ of $\{\mathbf{x}_i\}_{i=1}^n$. If $d = 1$, we order them from the smallest to the largest: $x_{1,n} \leq \cdots \leq x_{n,n}$ (which are simply order statistics as in the proof of Lemma 8.2 of the main paper). For Haar basis with $d > 1$ and $j = 0$, we first classify $\{\mathbf{x}_i\}_{i=1}^n$ into $\bar{\kappa}$ groups so that the points in the same group belong to the same cell (recall that $\bar{\kappa}$ is the number of cells in Δ). Number the cells in an arbitrary way. Under the new ordering, the points in the first group is followed by the second, then the third, and so on. Ordering within each group is arbitrary. $\{\sigma_i\}_{i=1}^n$ and $\{\varepsilon_i\}_{i=1}^n$ are rearranged as $\{\sigma_{i,n}\}_{i=1}^n$ and $\{\varepsilon_{i,n}\}_{i=1}^n$ accordingly. Again, the ordering strategies described above only depend on the values of $\{\mathbf{x}_i\}_{i=1}^n$, and thus conditional on $\mathbf{X}$, the new sequence $\{\varepsilon_{i,n}\}_{i=1}^n$ is still independent. As explained in the proof of Lemma 8.2, we can find a sequence of i.i.d standard normal random variables $\{\zeta_{i,n}\}_{i=1}^n$ such that $\max_{1\leq l\leq n}|S_{l,n}| \lesssim_{\mathbb{P}} n^{\frac{1}{2+\bar{\nu}}}$ where $S_{l,n} = \sum_{i=1}^l(\varepsilon_{i,n} - \sigma_{i,n}\zeta_{i,n})$. We will write $\{\zeta_i\}_{i=1}^n$ for the sequence of $\zeta_{i,n}$'s rearranged based on the original ordering of $\{\mathbf{x}_i\}_{i=1}^n$. Using summation by parts,

$$\begin{aligned}&\sup_{\mathbf{x}\in\mathcal{X}}\Big|\sum_{i=1}^n \mathscr{K}(\mathbf{x}, \mathbf{x}_{i,n})(\varepsilon_{i,n} - \sigma_{i,n}\zeta_{i,n})\Big| \\ =& \sup_{\mathbf{x}\in\mathcal{X}}\Big|\mathscr{K}(\mathbf{x}, \mathbf{x}_{n,n})S_{n,n} - \sum_{i=1}^{n-1} S_{i,n}\left(\mathscr{K}(\mathbf{x}, \mathbf{x}_{i+1,n}) - \mathscr{K}(\mathbf{x}, \mathbf{x}_{i,n})\right)\Big|\end{aligned}$$

$$\leq \sup_{\mathbf{x}\in\mathcal{X}} \max_{1\leq i\leq n} |\mathscr{K}(\mathbf{x},\mathbf{x}_i)||S_{n,n}| + \sup_{\mathbf{x}\in\mathcal{X}} \Big| \frac{\boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})'}{\sqrt{\Omega_j(\mathbf{x})}} \sum_{i=1}^{n-1} S_{i,n}(\boldsymbol{\Pi}_j(\mathbf{x}_{i+1,n}) - \boldsymbol{\Pi}_j(\mathbf{x}_{i,n}))\Big|$$

$$\leq \sup_{\mathbf{x}\in\mathcal{X}} \max_{1\leq i\leq n} |\mathscr{K}(\mathbf{x},\mathbf{x}_i)||S_{n,n}| + \sup_{\mathbf{x}\in\mathcal{X}} \Big\| \frac{\boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})'}{\sqrt{\Omega_j(\mathbf{x})}} \Big\|_\infty \Big\| \sum_{i=1}^{n-1} S_{i,n}(\boldsymbol{\Pi}_j(\mathbf{x}_{i+1,n}) - \boldsymbol{\Pi}_j(\mathbf{x}_{i,n}))\Big\|_\infty.$$

By Assumption SA-3, Lemma SA-2.3, and SA-2.4, $\sup_{\mathbf{x}\in\mathcal{X}} \max_{1\leq i\leq n} |\mathscr{K}(\mathbf{x},\mathbf{x}_i)| \lesssim h^{-d/2}$ and

$$\sup_{\mathbf{x}\in\mathcal{X}} \Big\| \frac{\boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})'}{\sqrt{\Omega_j(\mathbf{x})}} \Big\|_\infty \lesssim h^{-d/2}. \tag{SA-10.10}$$

Also, if we write the lth element of $\boldsymbol{\Pi}_j(\cdot)$ as $\pi_{j,l}(\cdot)$, then

$$\begin{aligned}
&\max_{1\leq l\leq K_j} \Big| \sum_{i=1}^{n-1} \Big(\pi_{j,l}(\mathbf{x}_{i+1,n}) - \pi_{j,l}(\mathbf{x}_{i,n})\Big) S_{l,n}\Big| \\
&\leq \max_{1\leq l\leq K_j} \sum_{i=1}^{n-1} \Big|\pi_{j,l}(\mathbf{x}_{i+1,n}) - \pi_{j,l}(\mathbf{x}_{i,n})\Big| \max_{1\leq \ell\leq n} \Big|S_{\ell,n}\Big| \lesssim_{\mathbb{P}} n^{\frac{1}{2+\bar{\nu}}},
\end{aligned}$$

where the last inequality follows from [22, Corollary 5], the ordering of $\{\mathbf{x}_{i,n}\}_{i=1}^n$, Assumption SA-3 and the fact that when $d>1$, each $\pi_{j,l}(\cdot)$ is a Haar basis function. This shows that there exists independent standard Normal random variables $\{\zeta_i\}_{i=1}^n$ such that

$$\mathbb{G}_n[\mathscr{K}(\mathbf{x},\mathbf{x}_i)\varepsilon_i] =_d z_j(\mathbf{x}) + o_{\mathbb{P}}(r_n^{-1}) \quad \text{where} \quad z_j(\mathbf{x}) := \mathbb{G}_n[\mathscr{K}(\mathbf{x},\mathbf{x}_i)\sigma_i\zeta_i].$$

Next, we convert the conditional coupling to an unconditional one. Notice that

$$z_j(\mathbf{x}) =_{d|\mathbf{X}} \frac{\boldsymbol{\gamma}_{\mathbf{q},j}(x)'}{\sqrt{\Omega_j(x)}} \bar{\boldsymbol{\Sigma}}_j^{1/2}\mathbf{N}_{K_j} = \frac{\boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})'}{\sqrt{\Omega_j(\mathbf{x})}} \boldsymbol{\Sigma}_j^{1/2}\mathbf{N}_{K_j} + \frac{\boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})'}{\sqrt{\Omega_j(\mathbf{x})}} \left(\bar{\boldsymbol{\Sigma}}_j^{1/2} - \boldsymbol{\Sigma}_j^{1/2}\right)\mathbf{N}_{K_j}$$

where $\bar{\boldsymbol{\Sigma}}_j := \mathbb{E}_n[\boldsymbol{\Pi}_j(\mathbf{x}_i)\boldsymbol{\Pi}_j(\mathbf{x}_i)'\sigma^2(\mathbf{x}_i)]$, $\mathbf{N}_{K_j}$ is a K_j-dimensional standard Normal vector (independent of $\mathbf{X}$) and "$=_{d|\mathbf{X}}$" denotes that two processes have the same conditional distribution given $\mathbf{X}$. For the second term, we already have Equation (SA-10.10), and by the Gaussian maximal inequality (see [10, Lemma 13]),

$$\mathbb{E}\left[\Big\|\left(\bar{\boldsymbol{\Sigma}}_j^{1/2} - \boldsymbol{\Sigma}_j^{1/2}\right)\mathbf{N}_{K_j}\Big\|_\infty \Big|\mathbf{X}\right] \lesssim \sqrt{\log n}\,\Big\|\bar{\boldsymbol{\Sigma}}_j^{1/2} - \boldsymbol{\Sigma}_j^{1/2}\Big\|.$$

By the same argument given for Theorem SA-4.2, $\Big\|\bar{\boldsymbol{\Sigma}}_j - \boldsymbol{\Sigma}_j\Big\| \lesssim_{\mathbb{P}} h^d\sqrt{\log n/(nh^d)}$. Then it follows from [5, Theorem X.1.1] that

$$\Big\|\bar{\boldsymbol{\Sigma}}_j^{1/2} - \boldsymbol{\Sigma}_j^{1/2}\Big\| \lesssim_{\mathbb{P}} h^{d/2}\left(\frac{\log n}{nh^d}\right)^{1/4}.$$

For $j = 0, 1$, a sharper bound is available: by Theorem X.3.8 of [5] and Lemma SA-2.1,

$$\|\bar{\mathbf{\Sigma}}_j^{1/2} - \mathbf{\Sigma}_j^{1/2}\| \leq \frac{1}{\lambda_{\min}(\mathbf{\Sigma}_j)^{1/2}}\|\bar{\mathbf{\Sigma}}_j - \mathbf{\Sigma}_j\| \lesssim_{\mathbb{P}} h^{d/2}\sqrt{\log n/(nh^d)}.$$

Thus using all these results, we have

$$\mathbb{E}\Big[\sup_{\mathbf{x}\in\mathcal{X}}\Big|\frac{\boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})'}{\sqrt{\Omega_j(\mathbf{x})}}\Big(\bar{\mathbf{\Sigma}}_j^{1/2} - \mathbf{\Sigma}_j^{1/2}\Big)\mathbf{N}_{K_j}\Big|\;\Big|\mathbf{X}\Big] \lesssim_{\mathbb{P}} h^{-d/2}\sqrt{\log n}\,\Big\|\bar{\mathbf{\Sigma}}_j^{1/2} - \mathbf{\Sigma}_j^{1/2}\Big\| = o_{\mathbb{P}}(r_n^{-1}),$$

where the last equality holds by the additional rate restriction given in the theorem (for $j = 0, 1$, no additional restriction is needed). By Markov inequality, this suffices to show that for any $\vartheta > 0$,

$$\mathbb{P}\left(\sup_{x\in\mathcal{X}}\Big|z_j(\mathbf{x}) - Z_j(\mathbf{x})\Big| > \vartheta\,\Big|\mathbf{X}\right) = o_{\mathbb{P}}(r_n^{-1}).$$

Since the conditional probability is bounded, by dominated convergence theorem, the desired result immediately follows.

When the conditions in (ii) hold, the proof remains the same except that we employ [21, Theorem 1] to construct strong approximation to the partial sum process of $\{\varepsilon_{i,n}\}_{i=1}^n$. □

SA-10.14. *Proof of Theorem SA-5.4*.

PROOF. Suppose that the conditions in (i) of Lemma SA-5.1 hold. The other case follows similarly. Let $z_j(\mathbf{x}) := \mathbb{G}_n[\mathscr{K}(\mathbf{x}, \mathbf{x}_i)\sigma_i\zeta_i]$. First notice that for $j = 0, 1, 2, 3$,

$$\begin{aligned}\sup_{\mathbf{x}\in\mathcal{X}}|\widehat{z}_j(\mathbf{x}) - z_j(\mathbf{x})| &= \sup_{\mathbf{x}\in\mathcal{X}}\left|\frac{\widehat{\boldsymbol{\gamma}}_{\mathbf{q},j}(\mathbf{x})'}{\widehat{\Omega}_j(\mathbf{x})^{1/2}}\mathbb{G}_n[\mathbf{\Pi}_j(\mathbf{x}_i)\zeta_i\hat{\sigma}(\mathbf{x}_i)] - \frac{\boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})'}{\Omega_j(\mathbf{x})^{1/2}}\mathbb{G}_n[\mathbf{\Pi}_j(\mathbf{x}_i)\zeta_i\sigma(\mathbf{x}_i)]\right| \\ &=: \sup_{\mathbf{x}\in\mathcal{X}}|D_n(\mathbf{x})|.\end{aligned}$$

Conditional on the data, $\{D_n(\mathbf{x}), \mathbf{x} \in \mathcal{X}\}$ is a Gaussian process with zero means. Let $\mathbb{E}^*$ denote the expectation with respect to the distribution of $\{\zeta_i\}_{i=1}^n$. For notational simplicity, define a norm $\|\cdot\|_{n,2}$ by $\|\mathbf{a}\|_{n,2}^2 = n^{-1}\sum_{i=1}^n a_i^2$ for $\mathbf{a} \in \mathbb{R}^n$. Then by

Cauchy-Schwarz inequality,

$$\begin{aligned}\mathbb{E}^*[D_n(\mathbf{x})^2] &= \frac{1}{n}\sum_{i=1}^n \Big[\frac{\widehat{\boldsymbol{\gamma}}_{\mathbf{q},j}(\mathbf{x})'}{\widehat{\Omega}_j(\mathbf{x})^{1/2}}\boldsymbol{\Pi}_j(\mathbf{x}_i)\hat{\sigma}(\mathbf{x}_i) - \frac{\boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})'}{\Omega_j(\mathbf{x})^{1/2}}\boldsymbol{\Pi}_j(\mathbf{x}_i)\sigma(\mathbf{x}_i)\Big]^2 \\
&= \frac{1}{n}\sum_{i=1}^n \Big\{\frac{\widehat{\boldsymbol{\gamma}}_{\mathbf{q},j}(\mathbf{x})' - \boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})'}{\widehat{\Omega}_j(\mathbf{x})^{1/2}}\boldsymbol{\Pi}_j(\mathbf{x}_i)\hat{\sigma}(\mathbf{x}_i) \\
&\quad + \boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})'(\widehat{\Omega}_j(\mathbf{x})^{-1/2} - \Omega_j(\mathbf{x})^{-1/2})\boldsymbol{\Pi}_j(\mathbf{x}_i)\hat{\sigma}(\mathbf{x}_i) \\
&\quad + \frac{\boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})'}{\Omega_j(\mathbf{x})^{1/2}}\boldsymbol{\Pi}_j(\mathbf{x}_i)(\hat{\sigma}(\mathbf{x}_i) - \sigma(\mathbf{x}_i))\Big\}^2 \\
&\lesssim \Big\|\frac{\widehat{\boldsymbol{\gamma}}_{\mathbf{q},j}(\mathbf{x})' - \boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})'}{\widehat{\Omega}_j(\mathbf{x})^{1/2}}\boldsymbol{\Pi}_j(\mathbf{x}_i)\hat{\sigma}(\mathbf{x}_i)\Big\|_{n,2}^2 \\
&\quad + \Big\|\boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})'(\widehat{\Omega}_j(\mathbf{x})^{-1/2} - \Omega_j(\mathbf{x})^{-1/2})\boldsymbol{\Pi}_j(\mathbf{x}_i)\hat{\sigma}(\mathbf{x}_i)\Big\|_{n,2}^2 \\
&\quad + \Big\|\frac{\boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})'}{\Omega_j(\mathbf{x})^{1/2}}\boldsymbol{\Pi}_j(\mathbf{x}_i)(\hat{\sigma}(\mathbf{x}_i) - \sigma(\mathbf{x}_i))\Big\|_{n,2}^2.\end{aligned}$$

By Lemmas SA-2.3 and SA-2.4, Theorem SA-4.2, and $\max_{1\leq i\leq n}|\hat{\sigma}^2(\mathbf{x}_i) - \sigma^2(\mathbf{x}_i)| = o_{\mathbb{P}}(1/(r_n\sqrt{\log n}))$,

$$\begin{aligned}\sup_{\mathbf{x}\in\mathcal{X}}\mathbb{E}^*[D_n(\mathbf{x})^2] &\lesssim_{\mathbb{P}} \frac{\log n}{nh^d} + \Big(\frac{n^{\frac{2}{2+\nu}}(\log n)^{\frac{\nu}{2+\nu}}}{nh^d} + a_n\Big) + (\max_{1\leq i\leq n}|\hat{\sigma}(\mathbf{x}_i) - \sigma(\mathbf{x}_i)|)^2 \\
&= o_{\mathbb{P}}\Big(\frac{1}{r_n^2\log n}\Big)\end{aligned}$$

where $a_n = h^{2m}$ for $j = 0$ and $a_n = h^{2m+2\varrho}$ for $j = 1, 2, 3$. Moreover, using the fact that for $\mathbf{x}, \check{\mathbf{x}} \in \mathcal{X}$,

$$\begin{aligned}&(\mathbb{E}^*[(D_n(\mathbf{x}) - D_n(\check{\mathbf{x}}))^2])^{1/2} \\
&\lesssim \Big\|\Big(\frac{\widehat{\boldsymbol{\gamma}}_{\mathbf{q},j}(\mathbf{x})'}{\widehat{\Omega}_j(\mathbf{x})^{1/2}} - \frac{\widehat{\boldsymbol{\gamma}}_{\mathbf{q},j}(\check{\mathbf{x}})'}{\widehat{\Omega}_j(\check{\mathbf{x}})^{1/2}}\Big)\boldsymbol{\Pi}_j(\mathbf{x}_i)\hat{\sigma}(\mathbf{x}_i)\Big\|_{n,2} \\
&\quad + \Big\|\Big(\frac{\boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})'}{\Omega_j(\mathbf{x})^{1/2}} - \frac{\boldsymbol{\gamma}_{\mathbf{q},j}(\check{\mathbf{x}})'}{\Omega_j(\check{\mathbf{x}})^{1/2}}\Big)\boldsymbol{\Pi}_j(\mathbf{x}_i)\sigma(\mathbf{x}_i)\Big\|_{n,2},\end{aligned}$$

we will show later in the proof that $\sup_{\delta\in\Delta}\sup_{\mathbf{x},\check{\mathbf{x}}\in\mathrm{clo}(\delta)}(\mathbb{E}^*[(D_n(\mathbf{x}) - D_n(\check{\mathbf{x}}))^2])^{1/2} \lesssim_{\mathbb{P}} h^{-1}\|\mathbf{x} - \check{\mathbf{x}}\|$. Since $\sup_{\delta\in\Delta}\mathrm{vol}(\delta) \lesssim h^d$, and the number of cells in Δ is $\bar{\kappa} \asymp h^{-d}$, we apply Dudley's inequality (see the proof of [4, Lemma 4.2] for more details of this inequality) to obtain $\mathbb{E}^*\Big[\sup_{\mathbf{x}\in\mathcal{X}}|D_n(\mathbf{x})|\Big] = o_{\mathbb{P}}(r_n^{-1})$, which suffices to show for any $t > 0$, $\mathbb{P}^*(\sup_{\mathbf{x}\in\mathcal{X}}|D_n| > tr_n^{-1}) = o_{\mathbb{P}}(1)$ by Markov inequality. Then, to show $z_j(\cdot)$ is approximated by $Z_j(\cdot)$, we only need to repeat the argument given in the proof of Theorem SA-5.1.

In the end, we establish the desired Lipschitz bound on the basis and variance. For any $\mathbf{x}, \check{\mathbf{x}} \in \delta$, by Assumption SA-3, there are only a finite number of basis functions in $\mathbf{p}$ and $\tilde{\mathbf{p}}$ which are active on δ, and hence it follows from Lemma SA-2.1 that there exists some universal constant $C_1 > 0$ such that $\|\boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x}) - \boldsymbol{\gamma}_{\mathbf{q},j}(\check{\mathbf{x}})\| \leq C_1 h^{-[\mathbf{q}]-1} h^{-d}\|\mathbf{x}-\check{\mathbf{x}}\|$. Similarly, by Lemma SA-2.3 and Theorem SA-4.2, $\|\widehat{\boldsymbol{\gamma}}_{\mathbf{q},j}(\mathbf{x}) - \widehat{\boldsymbol{\gamma}}_{\mathbf{q}.j}(\check{\mathbf{x}})\| \lesssim_{\mathbb{P}} h^{-[\mathbf{q}]-1}h^{-d}\|\mathbf{x}-\check{\mathbf{x}}\|$, $|\Omega_j(\mathbf{x}) - \Omega_j(\check{\mathbf{x}})| \lesssim h^{-d-2[\mathbf{q}]-1}\|\mathbf{x}-\check{\mathbf{x}}\|$, and $|\widehat{\Omega}_j(\mathbf{x}) - \widehat{\Omega}_j(\check{\mathbf{x}})| \lesssim_{\mathbb{P}} h^{-d-2[\mathbf{q}]-1}\|\mathbf{x}-\check{\mathbf{x}}\|$. Then the proof is complete. □

SA-10.15. *Proof of Theorem SA-5.5.*

PROOF. Let $\mathbb{E}^*$ be the expectation conditional on the data. Note that $\widehat{\varepsilon}_{i,j} = y_i - \widehat{\mu}_j(\mathbf{x}_i) = \mu(\mathbf{x}_i) - \widehat{\mu}_j(\mathbf{x}_i) + \varepsilon_i$. Then

$$\begin{aligned}
\widehat{z}_j^*(\mathbf{x}) =& \frac{\boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})'}{\Omega_j(\mathbf{x})^{1/2}}\mathbb{G}_n[\boldsymbol{\Pi}_j(\mathbf{x}_i)\widehat{\varepsilon}_{i,j}\omega_i] + o_{\mathbb{P}}(r_n^{-1})\\
=& \frac{\boldsymbol{\gamma}_{\mathbf{q}.j}(\mathbf{x})'}{\Omega_j(\mathbf{x})^{1/2}}\mathbb{G}_n[\boldsymbol{\Pi}_j(\mathbf{x}_i)\varepsilon_i\omega_i] + \\
& \frac{\boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})'}{\Omega_j(\mathbf{x})^{1/2}}\mathbb{G}_n[\boldsymbol{\Pi}_j(\mathbf{x}_i)(\mu(\mathbf{x}_i) - \widehat{\mu}_j(\mathbf{x}_i))\omega_i] + o_{\mathbb{P}}(r_n^{-1})\\
=& \frac{\boldsymbol{\gamma}_{\mathbf{q},j}(\mathbf{x})'}{\Omega_j(\mathbf{x})^{1/2}}\mathbb{G}_n[\boldsymbol{\Pi}_j(\mathbf{x}_i)\varepsilon_i\omega_i] + o_{\mathbb{P}}(r_n^{-1}) \quad \text{in} \quad \mathcal{L}^\infty(\mathcal{X}), \qquad \text{(SA-10.11)}
\end{aligned}$$

where the first and second equalities follow from a similar argument for Theorem SA-5.4, and the last line uses the rate of uniform convergence given in Theorem SA-4.1. Note that we employ different rate restrictions given in the theorem: for $j = 1, 2, 3$, $h^{m+\varrho}\sqrt{\log n} \lesssim \sqrt{n}h^{d/2+m+\varrho} = o(r_n^{-1})$, whereas for $j = 0$, $h^m\sqrt{\log n} \lesssim \sqrt{n}h^{d/2+m} = o(r_n^{-1})$. Moreover, either $\frac{\log n}{\sqrt{nh^d}} \lesssim \frac{n^{\frac{1}{2+\nu}}(\log n)^{\frac{1+\nu}{2+\nu}}}{\sqrt{nh^d}} = o(r_n^{-1})$, or $\frac{\log n}{\sqrt{nh^d}} \lesssim \frac{(\log n)^2}{\sqrt{nh^d}} = o(r_n^{-1})$. Simply notice that $M_n \lesssim_{\mathbb{P}} 1$ means $\mathbb{P}(|M_n| > \vartheta_n) = o(1)$ for any $\vartheta_n \to \infty$. By Markov inequality, it immediately follows that $\mathbb{P}^*(|M_n| > \vartheta_n) = o_{\mathbb{P}}(1)$. Thus the above derivation still holds in P-probability if we replace $\mathbb{P}$ by $\mathbb{P}^*$.

Then we only need to construct strong approximation to the leading term in Equation (SA-10.11). Repeat the argument for Theorem SA-5.1 conditional on the data. Regarding the conditional coupling step, we still rearrange terms according to the values of $\{\mathbf{x}_i\}_{i=1}^n$ as described in the proof of Theorem SA-5.1. Now if $(2+\nu)$th moment of ε_i is bounded, then conditional on the data, $\{\varepsilon_{i,n}\omega_{i,n}\}_{i=1}^n$ is independent and $\mathbb{E}_n[\mathbb{E}^*[|\varepsilon_i\omega_i|^{2+\nu}]] \lesssim_{\mathbb{P}} 1$. Thus, strong approximation to the partial sum process $\{\sum_{i=1}^l \varepsilon_{i,n}\omega_{i,n} : 1 \leq l \leq n\}$ can be obtained by using [22, Corollary 5]. When the conditions in (ii) of the theorem hold, [21, Theorem 1] can be employed to construct the strong approximation to this partial sum process. Thus, conditional on the data, there exists a K_j-dimensional standard Normal vector $\mathbf{N}_{K_j}$ such that

$$\widehat{z}^*(\cdot) =_{d^*} \frac{\boldsymbol{\gamma}_{\mathbf{q},j}(\cdot)'}{\Omega_j(\cdot)^{1/2}}\left(\mathbb{E}_n[\boldsymbol{\Pi}_j(\mathbf{x}_i)\boldsymbol{\Pi}_j(\mathbf{x}_i)'\varepsilon_i^2]\right)^{1/2}\mathbf{N}_{K_j} + o_{\mathbb{P}^*}(r_n^{-1}).$$

By Theorem SA-4.2 and the argument given in the proof of Theorem SA-5.1, we can further show that the conditional Gaussian process on the RHS is further approximated by $Z_j(\cdot)$. Then the proof is complete. □

SA-10.16. *Proof of Theorem SA-5.6.*

PROOF. The proof is exactly the same as that for Theorem SA-5.5 except that we apply the improved Yurinskii's inequality as in Theorem SA-5.2 conditional on the data. The conditions needed for coupling can be easily verified since $\boldsymbol{\omega}$ is assumed to be independent of the data and bounded. □

SA-10.17. *Proof of Theorem SA-5.7.*

PROOF. The proof is similar to that given in Section 8.6 of the main paper, and thus omitted here.

□

SA-10.18. *Proof of Lemma SA-6.1.*

PROOF. (a): Assumption SA-3(a) directly follows from the construction of tensor-product B-splines. Assumption SA-3(b) follows from [23, Theorem 12.5]. To show Assumption SA-3(c), note that for a univariate B-spline basis $\breve{\mathbf{p}}_\ell(x_\ell) = (\breve{p}_1(x_\ell), \cdots, \breve{p}_{K_\ell}(x_\ell))'$, there exists a universal constant $C > 0$ such that for any $\varsigma_\ell \leq m$, $1 \leq l_\ell \leq K_\ell$, and $\delta \in \Delta$, $\|d^{\varsigma_\ell}\breve{p}_{l_\ell}(x_\ell)/dx_\ell^{\varsigma_\ell}\|_{L_\infty(\mathrm{clo}(\delta))} \lesssim h^{-\varsigma_\ell}$. Since there are only a fixed number of nonzero elements in $\mathbf{p}$, we have for $[\boldsymbol{\varsigma}] \leq m$, $\sup_{\delta\in\Delta}\sup_{\mathbf{x}\in\mathrm{clo}(\delta)}\|\partial^{\boldsymbol{\varsigma}}\mathbf{p}(\mathbf{x})\| \lesssim h^{-[\boldsymbol{\varsigma}]}$. To derive the other side of the bound, notice that the proof of [28, Lemma 5.4] shows that for the univariate B-spline basis $\breve{\mathbf{p}}_\ell(x_\ell)$, for any $\varsigma_\ell \leq m-1$, $x_\ell \in \mathcal{X}_\ell$, $\|d^{\varsigma_\ell}\breve{\mathbf{p}}_\ell(x_\ell)/dx_\ell^{\varsigma_\ell}\| \gtrsim h^{-\varsigma_\ell}$. Since for any x_ℓ, there are only m nonzero elements in $\breve{\mathbf{p}}_\ell(x_\ell)$, this suffices to show that for any $x_\ell \in \mathcal{X}_\ell$, there exists some $\breve{p}_{l_\ell}(x_\ell)$ such that $|d^{\varsigma_\ell}p_{l_\ell}(x_\ell)/dx_\ell^{\varsigma_\ell}| \gtrsim h^{-\varsigma_\ell}$. Then the lower bound directly follows from the construction of tensor-product B-splines.

(b): The proof of orthogonality between the constructed leading error and B-splines can be found in [2]. Regarding the bias expansion, we first consider $\boldsymbol{\varsigma} = \mathbf{0}$. Noticing that

$$\mathscr{B}_{m,\mathbf{0}}(\mathbf{x}) = -\sum_{\ell=1}^{d}\frac{\partial^m\mu(\mathbf{t}_{\mathbf{x}}^L)}{\partial x_\ell^m}\frac{b_{\ell,l_\ell}^m}{m!}B_m\Big(\frac{x_\ell - t_{\ell,l_\ell}}{b_{\ell,l_\ell}}\Big) + O(h^{m+\varrho}) \text{ for } \mathbf{x}\in\delta_{l_1\dots l_d}$$

where $B_m(\cdot)$ is the mth Bernoulli polynomial, we only need to focus on the first term on the RHS, denoted by $\bar{\mathscr{B}}_m(\mathbf{x})$. By construction, $\bar{\mathscr{B}}_m(\mathbf{x})$ is continuous on the interior of each subrectangle $\delta_{l_1\dots l_d}$, and the discontinuity only takes place at boundaries of subrectangles. Let $J_{\mathbf{0}}$ denote the magnitude of a (generic) jump of $\bar{\mathscr{B}}_m(\mathbf{x})$. By Assumption SA-1, $J_{\mathbf{0}}$ is also a jump of $\bar{\mu} := \mu + \bar{\mathscr{B}}_m$. We first bound it as [2] did in their proof. We introduce the following notation:

i. $\boldsymbol{\tau} := (\tau_1, \cdots, \tau_d)$ is a point on the boundary of a generic rectangle $\delta_{l_1 \dots l_d}$;
ii. $\boldsymbol{\tau}^- := (\tau_1^-, \cdots, \tau_d^-)$ and $\boldsymbol{\tau}^+ := (\tau_1^+, \cdots, \tau_d^+)$ are two points close to $\boldsymbol{\tau}$ but belong to two different subrectangles $\delta_{l_1 \dots l_d}^- := \{\mathbf{x}: t_{\ell,l_\ell}^- \leq x_\ell < t_{\ell,l_\ell+1}^-\}$ and $\delta_{l_1 \dots l_d}^+ := \{\mathbf{x}: t_{\ell,l_\ell}^+ \leq x_\ell < t_{\ell,l_\ell+1}^+\}$;
iii. $\mathbf{t}_-^L$ and $\mathbf{t}_+^L$ are the starting points of $\delta_{l_1 \dots l_d}^-$ and $\delta_{l_1 \dots l_d}^+$;
iv. $(b_{1,-}, \cdots, b_{d,-})$ and $(b_{1,+}, \cdots, b_{d,+})$ are the corresponding mesh widths of $\delta_{l_1 \dots l_d}^-$ and $\delta_{l_1 \dots l_d}^+$;
v. $\Xi := \{\ell\colon\ \tau_\ell^- - \tau_\ell \text{ and } \tau_\ell^+ - \tau_\ell \text{ differ in signs}\}$.

In words, the index set Ξ indicates the directions in which we cross boundaries when we move from $\boldsymbol{\tau}_\ell^-$ to $\boldsymbol{\tau}_\ell^+$. To further simplify notation, we write $\bar{\mu}(\boldsymbol{\tau}^-) := \lim_{\mathbf{x}\to\boldsymbol{\tau}, \mathbf{x}\in\delta_{l_1 \dots l_d}^-} \bar{\mu}(\mathbf{x})$ and $\bar{\mu}(\boldsymbol{\tau}^+) := \lim_{\mathbf{x}\to\boldsymbol{\tau}, \mathbf{x}\in\delta_{l_1 \dots l_d}^+} \bar{\mu}(\mathbf{x})$. Then we have

$$\begin{aligned}
J_{\mathbf{0}} &= |\bar{\mu}(\boldsymbol{\tau}^+) - \bar{\mu}(\boldsymbol{\tau}^-)| = \Big|\bar{\mathscr{B}}_m(\boldsymbol{\tau}^+) - \bar{\mathscr{B}}_m(\boldsymbol{\tau}^-))\Big| \\
&= \sum_{\ell\in\Xi} (B_m(0)|/m!) \Big| \frac{\partial^m \mu(\mathbf{t}_+^L)}{\partial x_\ell^m} b_{\ell,+}^m - \frac{\partial^m \mu(\mathbf{t}_-^L)}{\partial x_\ell^m} b_{\ell,-}^m \Big| \\
&= \sum_{\ell\in\Xi} (B_m(0)|/m!) \Big| \Big(\frac{\partial^m \mu(\mathbf{t}_+^L)}{\partial x_\ell^m} - \frac{\partial^m \mu(\mathbf{t}_-^L)}{\partial x_\ell^m}\Big) b_{\ell,+}^m + \frac{\partial^m \mu(\mathbf{t}_-^L)}{\partial x_\ell^m}(b_{\ell,+}^m - b_{\ell,-}^m) \Big| \\
&\leq \sum_{\ell\in\Xi} (B_m(0)|/m!) \Big[O(h^{m+\varrho}) + Ch^{m-1}|b_{\ell,+} - b_{\ell,-}|\Big] \\
&\leq \sum_{\ell\in\Xi} (B_m(0)|/m!) \Big[O(h^{m+\varrho}) + Ch^{m-1}O(h^{1+\varrho})\Big],
\end{aligned}$$

where the fourth line follows from Assumption SA-1, and the last from the stronger quasi-uniformity condition specified in the Lemma. This suffices to show that $J_{\mathbf{0}}$ is $O(h^{m+\varrho})$.

Then we mimic the proof strategy used in [23, Theorem 12.7]. By [23, Theorem 12.6], we can construct a bounded linear operator $\mathscr{L}[\cdot]$ mapping $\mathcal{L}_1(\mathcal{X})$ onto $\mathcal{S}_{\Delta,m}$ with $\mathscr{L}[s] = s$ for all $s \in \mathcal{S}_{\Delta,m}$. Specifically, $\mathscr{L}[\cdot]$ is defined as

$$\mathscr{L}[\mu](\mathbf{x}) := \sum_{l_1=1}^{K_1} \cdots \sum_{l_d=1}^{K_d} (\psi_{l_1 \dots l_d}\mu) p_{l_1 \dots l_d}(\mathbf{x})$$

where $\{\psi_{l_1 \dots l_d}\}_{l_1=1,\dots,l_d=1}^{K_1,\dots,K_d}$ is the dual basis defined in [23, Equation (12.24)]. By multi-dimensional Taylor expansion, there exists a polynomial $\varphi_{l_1 \dots l_d}$ such that $\|\bar{\mu} - \varphi_{l_1 \dots l_d}\|_{L_\infty(\delta_{l_1 \dots l_d})} \lesssim h^{m+\varrho}$, and the degree of $\varphi_{l_1 \dots l_d}$ is no greater than $m-1$. Since $\mathscr{L}$ reproduces polynomials, we have

$$\begin{aligned}
\|\bar{\mu} - \mathscr{L}[\bar{\mu}]\|_{L_\infty(\delta_{l_1 \dots l_d})} &\leq \|\bar{\mu} - \varphi_{l_1 \dots l_d}\|_{L_\infty(\delta_{l_1 \dots l_d})} + \|\mathscr{L}[\bar{\mu} - \varphi_{l_1 \dots l_d}]\|_{L_\infty(\delta_{l_1 \dots l_d})} \\
&\leq C\|\bar{\mu} - \varphi_{l_1 \dots l_d}\|_{L_\infty(\delta_{l_1 \dots l_d})} \lesssim h^{m+\varrho}.
\end{aligned}$$

Taking account of the jumps of $\bar{\mu}$ along boundaries, the approximation error of $\mathscr{L}[\bar{\mu}]$ is still $O(h^{m+\varrho})$. Evaluating the L_∞ norm on all subrectangles, we conclude that there exists some $s^* \in \mathcal{S}_{\Delta,m}$ such that $\|\mu + \bar{\mathscr{B}}_m - s^*\|_{L_\infty(\mathcal{X})} \lesssim h^{m+\varrho}$.

For other $\boldsymbol{\varsigma}$, we only need to show that the desired result holds for $s^* = \mathscr{L}[\bar{\mu}]$. By construction of $\mathscr{L}$,

(SA-10.12)
$$|\partial^{\boldsymbol{\varsigma}}(\mathscr{L}[\bar{\mu}])| \leq \sum_{l_1=1}^{m+\kappa_1} \cdots \sum_{l_d=1}^{m+\kappa_d} |\psi_{l_1\dots l_d}\bar{\mu}||\partial^{\boldsymbol{\varsigma}} p_{l_1\dots l_d}(\mathbf{x})| \leq Ch^{-[\boldsymbol{\varsigma}]}\|\bar{\mu}\|_{L_\infty(\delta_{l_1\dots l_d})}$$

where the last line follows from [23, Theorem 12.5]. Then we have

$$\begin{aligned}
\|\partial^{\boldsymbol{\varsigma}}\mu + \partial^{\boldsymbol{\varsigma}}\bar{\mathscr{B}}_m - \partial^{\boldsymbol{\varsigma}}(\mathscr{L}[\bar{\mu}])\|_{L_\infty(\delta_{l_1\dots l_d})} &\leq \|\partial^{\boldsymbol{\varsigma}}\mu + \partial^{\boldsymbol{\varsigma}}\mathscr{B}_m^* - \partial^{\boldsymbol{\varsigma}}\varphi_{l_1\dots l_d}\|_{L_\infty(\delta_{l_1\dots l_d})} \\
&\quad + \|\partial^{\boldsymbol{\varsigma}}(\mathscr{L}[\bar{\mu} - \varphi_{l_1\dots l_d}])\|_{L_\infty(\delta_{l_1\dots l_d})} \\
&\leq O(h^{m+\varrho-[\boldsymbol{\varsigma}]}) + Ch^{-[\boldsymbol{\varsigma}]}\|\bar{\mu} - \varphi_{l_1\dots l_d}\|_{L_\infty(\delta_{l_1\dots l_d})} \\
&\lesssim h^{m+\varrho-[\boldsymbol{\varsigma}]}
\end{aligned}$$

where the second inequality follows from Taylor expansion and Equation (SA-10.12). Moreover, by an argument similar to that for $J_{\mathbf{0}}$, the jump of $\partial^{\boldsymbol{\varsigma}}\bar{\mathscr{B}}_m$ is $O(h^{m+\varrho-[\boldsymbol{\varsigma}]})$.

(c): By construction of $\tilde{\mathbf{p}}$, $\rho = 1$. By the same argument as in part (a) and (b), $\tilde{\mathbf{p}}$ satisfies Assumption SA-3 and SA-4. Finally, by definition of tensor-product splines, both $\mathbf{p}$ and $\tilde{\mathbf{p}}$ reproduce polynomials of degree no greater than $m-1$. Then the proof is complete. □

SA-10.19. *Proof of Lemma SA-6.2.*

PROOF. (a): Assumption SA-3(a) holds by the fact that the father wavelet is compactly supported and $\{\phi_{sl}\}$ is generated by translation and dilation. Assumption SA-1(b) follows from the fact that $\{\phi_{sl}\}$ is an orthonormal basis with respect to the Lebesgue measure. For Assumption SA-3(c), notice that

$$\frac{d^{\varsigma_\ell}\phi(2^s x_\ell - l_\ell)}{dx_\ell^{\varsigma_\ell}} = 2^{s\varsigma_\ell}\frac{d^{\varsigma_\ell}\phi(z)}{dz^{\varsigma_\ell}}\Big|_{z=2^s x_\ell - l_\ell} = b^{-\varsigma_\ell}\frac{d^{\varsigma_\ell}\phi(z)}{dz^{\varsigma_\ell}}\Big|_{z=2^s x_\ell - l_\ell}.$$

Since the wavelet basis reproduces polynomials of degree no greater than $m-1$ and ϕ is assumed to have $q+1$ continuous derivatives, the desired bounds follow.

(b): We employ the strategy used in [24], but extend their proof to the multidimensional case. First, we denote by V_s^ℓ the closure of the level-s subspace spanned by $\{\phi_{sl}(x_\ell)\}$, and let W_s^ℓ be the orthogonal complement of V_s^ℓ in V_{s+1}^ℓ. Then we write $\mathcal{V}_s := \otimes_{\ell=1}^d V_s^\ell$ for the space spanned by the tensor-product level-s father wavelets, and $\mathcal{W}_s$ denotes the orthogonal complement of $\mathcal{V}_s$ in $\mathcal{V}_{s+1}$. We use the fact that

$\mathcal{W}_s = \oplus_{i=1}^{2^d-1}\mathcal{W}_{s,i}$, where $\oplus$ denotes "direct sum", and each $\mathcal{W}_{s,i}$ takes the following form: $\mathcal{W}_{s,i} = \otimes_{\ell=1}^{d} Z_s^\ell$. Each Z_s^ℓ is either V_s^ℓ or W_s^ℓ, but $\{Z_s^\ell\}_{\ell=1}^d$ cannot be identical to $\{V_s^\ell\}_{\ell=1}^d$. There are in total (2^d-1) such subspaces. Accordingly, a typical element of a basis vector for $\mathcal{W}_s$ can be written as

$$\bar{\psi}_{s\mathbf{l}\boldsymbol{\alpha}}(\mathbf{x}) = \prod_{\ell=1}^{d}\left[\alpha_\ell \phi_{sl_\ell}(x_\ell) + (1-\alpha_\ell)\psi_{sl_\ell}(x_\ell)\right]$$

where $\mathbf{l} = (l_1, \dots, l_d)$ and $\alpha_\ell = 0$ or 1, but $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d) \neq (1, \dots, 1)$. Then it directly follows from the properties of wavelet basis that for $\bar{\psi}_{s\mathbf{l}\boldsymbol{\alpha}}$, $s \geq m$,
(SA-10.13)

$$\langle \mathbf{x}^{\boldsymbol{\varsigma}}, \bar{\psi}_{s\mathbf{l}\boldsymbol{\alpha}}(\mathbf{x})\rangle := \int_{\mathcal{X}} \mathbf{x}^{\boldsymbol{\varsigma}}\bar{\psi}_{s\mathbf{l}\boldsymbol{\alpha}}(\mathbf{x})d\mathbf{x} = 0, \text{ for } \boldsymbol{\varsigma} \text{ such that } [\boldsymbol{\varsigma}] \leq m, \text{ and } \varsigma_\ell \neq m \ \forall \ell.$$

Denote by $\mathscr{L}_s[\cdot]$ the orthogonal projection operator onto $\mathcal{W}_s$. Then the approximation error of the tensor-product wavelet space $\mathcal{V}_{s_n}$ can be written as

$$\begin{aligned}
\sum_{s=s_n}^{\infty} \mathscr{L}_s[\mu](\mathbf{x}) &= \sum_{s=s_n}^{\infty}\sum_{\boldsymbol{\alpha}}\sum_{\mathbf{l}} \langle \mu(\check{\mathbf{x}}), \bar{\psi}_{s\mathbf{l}\boldsymbol{\alpha}}(\check{\mathbf{x}})\rangle \bar{\psi}_{s\mathbf{l}\boldsymbol{\alpha}}(\mathbf{x}) \\
&= \sum_{s=s_n}^{\infty}\sum_{\boldsymbol{\alpha}}\sum_{\mathbf{l}} \Big\langle \sum_{[\boldsymbol{\varsigma}]\leq m} \partial^{\boldsymbol{\varsigma}}\mu(\mathbf{x})\frac{(\check{\mathbf{x}}-\mathbf{x})^{\boldsymbol{\varsigma}}}{\boldsymbol{\varsigma}!} + \vartheta_n(\check{\mathbf{x}}, \mathbf{x}),\ \bar{\psi}_{s\mathbf{l}\boldsymbol{\alpha}}(\check{\mathbf{x}})\Big\rangle \bar{\psi}_{s\mathbf{l}\boldsymbol{\alpha}}(\mathbf{x})
\end{aligned}$$

where $\vartheta_n(\check{\mathbf{x}}, \mathbf{x}) \lesssim \|\check{\mathbf{x}} - \mathbf{x}\|^{m+\varrho}$, and the inner product in the above equations are taken with respect to $\check{\mathbf{x}}$ in terms of the Lebesgue measure. The index sets where $\boldsymbol{\alpha}$ and $\mathbf{l}$ live are described in Section SA-6 and the proof, and omitted in the above derivation for simplicity. By Assumption SA-1 and Assumption SA-3,

$$\begin{aligned}
&\sup_{\mathbf{x}\in\mathcal{X}} \Big| \sum_{s=s_n}^{\infty}\sum_{\boldsymbol{\alpha}}\sum_{\mathbf{l}} \Big\langle \vartheta_n(\check{\mathbf{x}}, \mathbf{x}),\ \bar{\psi}_{s\mathbf{l}\boldsymbol{\alpha}}(\check{\mathbf{x}})\Big\rangle \bar{\psi}_{s\mathbf{l}\boldsymbol{\alpha}}(\mathbf{x})\Big| \\
&= \sup_{\mathbf{x}\in\mathcal{X}} \Big| \sum_{s=s_n}^{\infty} \Big(\frac{b}{2^{s-s_n}}\Big)^{m+\varrho} \sum_{\boldsymbol{\alpha}}\sum_{\mathbf{l}} \Big\langle \vartheta_n(\check{\mathbf{x}}, \mathbf{x})2^{s(m+\varrho)},\ \bar{\psi}_{s\mathbf{l}\boldsymbol{\alpha}}(\check{\mathbf{x}})\Big\rangle \bar{\psi}_{s\mathbf{l}\boldsymbol{\alpha}}(\mathbf{x})\Big| \lesssim b^{m+\varrho}.
\end{aligned}$$

Recall that $b = 2^{-s_n}$.

Regarding the leading terms

$$\sum_{s=s_n}^{\infty}\sum_{\boldsymbol{\alpha}}\sum_{\mathbf{l}} \Big\langle \sum_{[\boldsymbol{\varsigma}]\leq m} \partial^{\boldsymbol{\varsigma}}\mu(\mathbf{x})\frac{(\check{\mathbf{x}}-\mathbf{x})^{\boldsymbol{\varsigma}}}{\boldsymbol{\varsigma}!},\ \bar{\psi}_{s\mathbf{l}\boldsymbol{\alpha}}(\check{\mathbf{x}})\Big\rangle \bar{\psi}_{s\mathbf{l}\boldsymbol{\alpha}}(\mathbf{x}),$$

it is clear that the coefficients of the wavelet basis can be viewed as a linear combination of the inner products of monomials and the mother wavelets themselves, and thus by Equation (SA-10.13), the leading error is of order b^m and can be characterized as

$$\mathfrak{B}_{m,\mathbf{0}}(\mathbf{x}) = -\sum_{\boldsymbol{u}\in\Lambda_m} \frac{b^m}{\boldsymbol{u}!}\partial^{\boldsymbol{u}}\mu(\mathbf{x})B^{\mathtt{W}}_{\boldsymbol{u},\mathbf{0}}(\mathbf{x}/b).$$

$B^{\mathtt{W}}_{\boldsymbol{u},\mathbf{0}}$ is referred to as "monowavelet" in [24]. Here we extend it to the multidimensional case. Specifically, define a mapping

$$\begin{aligned} \varphi:\ \Lambda_m &\to \{1,\dots,d\} \\ \boldsymbol{u} &\mapsto \ell \end{aligned}$$

such that $\varphi(\boldsymbol{u})$th element of $\boldsymbol{u}$ is nonzero. We denote $\mathbf{l}_{-\ell} := (l_1,\dots,l_{\ell-1},l_{\ell+1},\dots,l_d)$ and $\mathcal{L}_s^{-\ell} := \Big\{\mathbf{l}_{-\ell} : l_{\ell'} \in \mathcal{L}_s, j' = \{1,\cdots,d\}\setminus\{\ell\}\Big\}$. Then define

$$\varpi_{\boldsymbol{u},s}(\mathbf{x}) = \sum_{l_{\varphi(\boldsymbol{u})}\in\mathcal{L}_s}\ \sum_{\mathbf{l}_{-\varphi(\boldsymbol{u})}\in\mathcal{L}_s^{-\varphi(\boldsymbol{u})}} c_m\psi(2^s x_{\varphi(\boldsymbol{u})} - l_{\varphi(\boldsymbol{u})}) \prod_{\substack{\ell=1,\dots d\\ \ell\neq\varphi(\boldsymbol{u})}} \phi(2^s x_\ell - l_\ell)$$

where $c_m := \int_0^1 x^m\psi(x)\,dx$. Then $B^{\mathtt{W}}_{\boldsymbol{u},\mathbf{0}}(\cdot)$ can be expressed as

$$B^{\mathtt{W}}_{\boldsymbol{u},\mathbf{0}}(\mathbf{x}) = \sum_{s=0}^{\infty} 2^{-sm}\varpi_{\boldsymbol{u},s}(\mathbf{x}) =: \sum_{s=0}^{\infty}\xi_{\boldsymbol{u},s}(\mathbf{x}). \tag{SA-10.14}$$

Moreover, since the series in Equation (SA-10.14) converges uniformly and for $s \geq s_n$, $\varpi^*_{\boldsymbol{u},s}(\mathbf{x})$ is orthogonal to the tensor-product wavelet basis $\mathbf{p}$ with respect to the Lebesgue measure, by Dominated Convergence Theorem, the approximate orthogonality condition holds.

For other $\boldsymbol{\varsigma}$, let

$$\mathscr{B}_{m,\boldsymbol{\varsigma}}(\mathbf{x}) = -\sum_{\boldsymbol{u}\in\Lambda_m} \frac{b^{m-[\boldsymbol{\varsigma}]}}{\boldsymbol{u}!}\partial^{\boldsymbol{u}}\mu(\mathbf{x})B^{\mathtt{W}}_{\boldsymbol{u},\boldsymbol{\varsigma}}(\mathbf{x}/b)$$

where $B^{\mathtt{W}}_{\boldsymbol{u},\boldsymbol{\varsigma}}(\mathbf{x}) = \partial^{\boldsymbol{\varsigma}}B^{\mathtt{W}}_{\boldsymbol{u},\mathbf{0}}(\mathbf{x})$. Since for $\boldsymbol{\varsigma}$ such that $[\boldsymbol{\varsigma}] \leq \varsigma$, $\partial^{\boldsymbol{\varsigma}}\phi$ and $\partial^{\boldsymbol{\varsigma}}\psi$ are continuously differentiable, $\sum_{s=0}^{\infty} 2^{-sm}\partial^{\boldsymbol{\varsigma}}\varpi_{\boldsymbol{u},s}(\mathbf{x})$ converges uniformly, and hence we can interchange the differentiation and infinite summation. Therefore, $B^{\mathtt{W}}_{\boldsymbol{u},\boldsymbol{\varsigma}}(\cdot)$ is well defined and continuously differentiable. Then the lipschitz condition on $B^{\mathtt{W}}_{\boldsymbol{u},\boldsymbol{\varsigma}}(\cdot)$ in Assumption SA-4 holds.

Let s^* be the orthogonal projection of μ onto $\mathcal{V}_{s_n}$. To complete the proof of part (b), it suffices to show $\|\partial^{\boldsymbol{\varsigma}}\mu - \partial^{\boldsymbol{\varsigma}}s^* + \mathscr{B}_{m,\boldsymbol{\varsigma}}\|_{L_\infty(\mathcal{X})} \lesssim b^{m+\varrho-[\boldsymbol{\varsigma}]}$. Given a resolution level s_n,

$$\begin{aligned}
&\sum_{s=s_n}^{\infty}\sum_{\boldsymbol{\alpha}}\sum_{\mathbf{l}}\langle\mu(\check{\mathbf{x}}),\bar{\psi}_{s\mathbf{l}\boldsymbol{\alpha}}(\check{\mathbf{x}})\rangle\partial^{\boldsymbol{\varsigma}}\bar{\psi}_{s\mathbf{l}\boldsymbol{\alpha}}(\mathbf{x})\\
&=\sum_{s=s_n}^{\infty}\sum_{\boldsymbol{\alpha}}\sum_{\mathbf{l}}\Big\langle\sum_{[\boldsymbol{u}]\leq m}\partial^{\boldsymbol{u}}\mu(\mathbf{x})\frac{(\check{\mathbf{x}}-\mathbf{x})^{\boldsymbol{u}}}{\boldsymbol{u}!}+\vartheta_n(\check{\mathbf{x}},\mathbf{x}),\ \bar{\psi}_{s\mathbf{l}\boldsymbol{\alpha}}(\check{\mathbf{x}})\Big\rangle\partial^{\boldsymbol{\varsigma}}\bar{\psi}_{s\mathbf{l}\boldsymbol{\alpha}}(\mathbf{x})\\
&=b^{m-[\boldsymbol{\varsigma}]}\sum_{s=s_n}^{\infty}\frac{2^{[\boldsymbol{\varsigma}](s_n-s)}}{2^{m(s-s_n)}}\sum_{\boldsymbol{\alpha}}\sum_{\mathbf{l}}\frac{2^{sd}}{2^{-sm}}\Big\langle\sum_{[\boldsymbol{u}]\leq m}\partial^{\boldsymbol{u}}\mu(\mathbf{x})\frac{(\check{\mathbf{x}}-\mathbf{x})^{\boldsymbol{u}}}{\boldsymbol{u}!}+\vartheta_n(\check{\mathbf{x}},\mathbf{x}),\\
&\qquad\qquad 2^{-sd/2}\bar{\psi}_{s\mathbf{l}\boldsymbol{\alpha}}(\check{\mathbf{x}})\Big\rangle\ \partial^{\boldsymbol{\varsigma}}\Big(2^{-sd/2}\bar{\psi}_{s\mathbf{l}\boldsymbol{\alpha}}(\mathbf{x})\Big).
\end{aligned}$$

By changing variables, the vanishing moments of the wavelet function, and the fact that geometric series converges, the last line uniformly converges to the $\boldsymbol{\varsigma}$th derivative of the approximation error of $\mathcal{V}_{s_n}$, $\mathscr{B}_{m,\boldsymbol{\varsigma}}(\cdot)$ is the leading error and the remainder is $O(b^{m+\varrho-[\boldsymbol{\varsigma}]})$.

(c): By construction of $\tilde{\mathbf{p}}$, $\rho=1$. By the same argument as that for part (a) and (b), $\tilde{\mathbf{p}}$ satisfies Assumption SA-3 and SA-4. Finally, both $\mathbf{p}$ and $\tilde{\mathbf{p}}$ reproduce polynomials of degree no greater than $m-1$. Thus Assumption SA-5 holds. The proof is complete. □

SA-10.20. *Proof of Lemma SA-6.3.*

Proof. (a): By construction, each basis function $p_k(\mathbf{x})$ is supported by only one subrectangle, and there are only a fixed number of $p_k(\mathbf{x})$'s which are not identically zero on each subrectangle. Thus Assumption SA-3(a) is satisfied. In addition, given a generic subrectangle $\delta_{l_1\dots l_d}$, store all basis functions supported on $\delta_{l_1\dots l_d}$ in a vector $\mathbf{p}_{l_1\dots l_d}$. By [7, Lemma A.3], $\mathbf{Q}_{l_1\dots l_d}:=\mathbb{E}[\mathbf{p}_{l_1\dots l_d}(\mathbf{x}_i)\mathbf{p}_{l_1\dots l_d}(\mathbf{x}_i)']\asymp\mathbf{I}_{\dim(\mathbf{R}(\cdot))}$, where $\mathbf{I}_{\dim(\mathbf{R}(\cdot))}$ is an identity matrix of size $\dim(\mathbf{R}(\cdot))$. In fact, $\int_{\delta_{l_1\dots l_d}}\mathbf{p}_{l_1\dots l_d}(\mathbf{x})\mathbf{p}_{l_1\dots l_d}(\mathbf{x})'\,d\mathbf{x}$ is a finite-dimensional matrix with the minimum eigenvalue bounded from below by Ch^d for some $C>0$. Hence for any $\mathbf{a}\in\mathbb{R}^{\dim(\mathbf{R}(\cdot))}$,

$$\mathbf{a}'\int_{\delta_{l_1\dots l_d}}\mathbf{p}_{l_1\dots l_d}(\mathbf{x})\mathbf{p}_{l_1\dots l_d}(\mathbf{x})'\,d\mathbf{x}\,\mathbf{a}\geq Ch^d\mathbf{a}'\mathbf{a}$$

which suffices to show Assumption SA-3(b).

To show Assumption SA-3(c), simply notice that given any $\mathbf{x}\in\mathcal{X}$, there are only a fixed number of nonzero elements in $\partial^{\mathbf{q}}\mathbf{p}(\mathbf{x})$, and for any $k=1,\dots,K$,

$$\sup_{\delta\in\Delta}\sup_{\mathbf{x}\in\mathrm{clo}(\delta)}|\partial^{\boldsymbol{\varsigma}}p_k(\mathbf{x})|\lesssim h^{-[\boldsymbol{\varsigma}]}\max_{[\boldsymbol{\alpha}]=m-1}\frac{\boldsymbol{\alpha}!}{(\boldsymbol{\alpha}-\boldsymbol{\varsigma})!}.$$

Moreover, for any $\mathbf{x}\in\mathcal{X}$, there exists some p_k in $\mathbf{p}$ such that for $[\boldsymbol{\varsigma}]\leq m-1$, $|\partial^{\boldsymbol{\varsigma}}p_k(\mathbf{x})|\gtrsim h^{-[\boldsymbol{\varsigma}]}$.

(b): The result directly follows from the proofs of Lemma A.2 and Theorem 3 in [7]. The only difference is that we use shifted Legendre polynomials to re-express the approximating function $s^*(\mathbf{x}) = \mathbf{p}(\mathbf{x})'\boldsymbol{\beta}^*$ and the leading error. Clearly, $\boldsymbol{\beta}^*$ is just a linear combination of coefficients of power series basis defined in their paper. The orthogonality between the approximation basis and the leading error holds by the property of Legendre polynomials and the fact that every basis function is locally supported on only one cell.

(c): By construction of $\tilde{\mathbf{p}}$, $\rho = 1$. By the same argument as that for part (a) and (b), $\tilde{\mathbf{p}}$ satisfies Assumption SA-3 and SA-4. Finally, when the degree of piecewise polynomials is increased, $\tilde{\mathbf{p}}$ spans a larger space containing the span of $\mathbf{p}$, and both bases reproduce polynomials of degree no greater than $m-1$. Thus Assumption SA-5 holds. □

SA-10.21. *Proof of Lemma SA-6.4.*

PROOF. Assumption SA-3(a), SA-3(c) and SA-4 directly follow from the construction of this basis and Taylor expansion restricted to a particular cell. For Assumption SA-3(b), given a generic cell δ, by Assumption SA-2, we can find an inscribed ball of diameter ℓ_1 that is proportional to $h_{\mathbf{x}}$. Thus we can further find an inscribed rectangle with lengths equal to ℓ_2 that is also proportional to $h_{\mathbf{x}}$. Thus, by changing variables and the same argument as that for the basis defined on rectangular cells, we have Assumption SA-3(b) holds. The properties of $\tilde{\mathbf{p}}$ follow similarly as in Lemma SA-6.3. □

SA-10.22. *Proof of Lemma SA-7.1.*

PROOF. For the upper bound on the maximum eigenvalue, simply notice that all elements in $\mathbf{R}$ is bounded by some constant C, and the number of nonzeros in any row or column of $\mathbf{R}$ is bounded by some constant L. Then for any $\boldsymbol{\alpha} \in \mathbb{R}^{\vartheta_\iota}$ such that $\|\boldsymbol{\alpha}\| = 1$, $\boldsymbol{\alpha}'\mathbf{R}\mathbf{R}'\boldsymbol{\alpha} \le L^2C^2\|\boldsymbol{\alpha}\|^2 \lesssim 1$.

For the other side of the bound, since $\mathbf{R}\mathbf{R}'$ is a symmetric block Toeplitz matrix, Szegö's theorem and its extensions state that the asymptotic behavior of Toeplitz or block Toeplitz matrices is characterized by the corresponding Fourier transformation of their entries. See [16] for more details. Specifically, $\mathbf{R}\mathbf{R}'$ is transformed into the following matrix

$$\mathscr{T}_{\bar{\iota}}(\omega) = \begin{bmatrix} \mathbf{r}'_{11}\mathbf{r}_{11} + 2\mathbf{r}'_{11}\mathbf{r}_{21}\cos\omega & \cdots & \mathbf{r}'_{11}\mathbf{r}_{1\bar{\iota}} + (\mathbf{r}'_{11}\mathbf{r}_{2\bar{\iota}} + \mathbf{r}'_{1\bar{\iota}}\mathbf{r}_{21})\cos\omega \\ \vdots & \ddots & \vdots \\ \mathbf{r}'_{1\bar{\iota}}\mathbf{r}_{11} + (\mathbf{r}'_{1\bar{\iota}}\mathbf{r}_{21} + \mathbf{r}'_{11}\mathbf{r}_{2\bar{\iota}})\cos\omega & \cdots & \mathbf{r}'_{1\bar{\iota}}\mathbf{r}_{1\bar{\iota}} + 2\mathbf{r}'_{1\bar{\iota}}\mathbf{r}_{2\bar{\iota}}\cos\omega \end{bmatrix},$$

Using the representation of $\mathbf{R}\mathbf{R}'$ given in the discussion after the lemma, we can concisely write $\mathscr{T}_{\bar{\iota}}(\omega) = \mathbf{A} + (\mathbf{B} + \mathbf{B}')\cos\omega$. By [15, Equation (7)], we have

$$\lambda_{\min}(\mathbf{R}\mathbf{R}') \to \min_{\omega\in[0,2\pi]} \lambda_{\min}(\mathscr{T}_{\bar{\iota}}(\omega)) \quad \text{as} \quad \kappa \to \infty.$$

The minimum of the minimum eigenvalue function of $\mathscr{T}_{\bar{\iota}}(\omega)$ is attainable since each entry of $\mathscr{T}_{\bar{\iota}}(\omega)$ is a linear function of $\cos\omega$, and thus each coefficient of the corresponding characteristic polynomial is a continuous function of $\cos\omega$. By Theorem 3.9.1 of [26], there exist $\bar{\iota}$ continuous functions of $\cos\omega$ such that they are the roots of the characteristic polynomial, and thus the minimum eigenvalue is a continuous function of $\cos\omega$. In addition, since $\mathbf{RR}'$ is positive semidefinite, this function is nonnegative over $[-1,1]$.

By construction, $\mathbf{RR}'$ is a real symmetric positive semidefinite matrix, and thus its eigenvalues are real and nonnegative. Moreover, given any fixed κ, $\mathbf{RR}'$ is positive definite since the restrictions specified in $\mathbf{R}$ are non-redundant. Therefore, it suffices to show that the limit of the minimum eigenvalue sequence is bounded away from zero. The original problem is transformed into showing that the minimum eigenvalue of a finite-dimensional matrix $\mathscr{T}_{\bar{\iota}}(\omega)$ is strictly positive for any $\omega\in[0,2\pi]$.

The next critical fact we employ is that the smallest eigenvalue as a function of a real symmetric matrix is concave [see 18, Property 2.1]. In our case, each entry is a linear function of $\cos\omega$, and thus $\lambda_{\min}(\mathscr{T}_{\bar{\iota}}(\omega))$ is concave with respect to $\cos\omega$. Therefore, the minimum of the smallest eigenvalue function could be attained only at two endpoints, i.e., when $\cos\omega=1$ or $\cos\omega=-1$.

We start with the case in which $(m-1)$ continuity constraints are imposed at each knot, i.e., $\bar{\iota}=m-1$. Since each knot is treated the same way, the restriction matrix $\mathbf{R}$ can be fully characterized by $\bar{\iota}$ row vectors. A typical restriction that the ςth derivative $(0\le\varsigma\le\bar{\iota}-1)$ is continuous at a knot can be represented by the following vector

$$\Big(\underbrace{\bar{P}_0^{(\varsigma)}(1),\dots,\bar{P}_{m-1}^{(\varsigma)}(1)}_{\text{left interval}},\ \underbrace{-\bar{P}_0^{(\varsigma)}(-1),\dots,-\bar{P}_{m-1}^{(\varsigma)}(-1)}_{\text{right interval}}\Big)$$

where we omit all zero entries and $\bar{P}_l^{(\varsigma)}(x)$, $0\le l\le m-1$, denotes the ςth derivative of the *normalized* Legendre polynomial of degree l. Generally, Legendre polynomial $P_l(x)$ can be written as

$$P_l(x)=\frac{1}{2^l}\sum_{i=0}^{l}\binom{l}{i}^2(x-1)^{l-i}(x+1)^i,$$

and they have the following properties: for any $l,l'\in\mathbb{Z}_+$

$$P_l(1)=1,\quad P_l(-x)=(-1)^lP_l(x),\quad \int_{-1}^{1}P_l(x)P_{l'}(x)\,dx=\frac{2}{2l+1}\delta_{ll'}$$

where $\delta_{ll'}$ is the Kronecker delta. Therefore, $\bar{P}_l(x)=\frac{\sqrt{2l+1}}{\sqrt{2}}P_l(x)$.

Using these formulas,

$$P_l^{(\varsigma)}(1)=2^{-\varsigma}\varsigma!\binom{l}{\varsigma}\binom{l+\varsigma}{\varsigma}=2^{-\varsigma}\frac{(l+\varsigma)(l+\varsigma-1)\cdots(l-\varsigma+1)}{\varsigma!}.$$

Therefore, $\bar{P}_l^{(\varsigma)}(1) = \frac{\sqrt{2l+1}}{\sqrt{2}} 2^{-\varsigma} \varsigma! \binom{l}{\varsigma}\binom{l+\varsigma}{\varsigma}$. In addition, since Legendre polynomials are symmetric or antisymmetric, we have $\bar{P}_l^{(\varsigma)}(-1) = (-1)^{l+\varsigma}\bar{P}_l^{(\varsigma)}(1)$. Thus we obtain an explicit expression for $\mathbf{R}$.

Next, $\mathbf{R}\mathbf{R}'$ can be fully characterized by two matrices $\mathbf{A}$ and $\mathbf{B}$ as in (SA-7.3). In what follows we use $A[\varsigma,\ell]$ to denote the (ς,ℓ)th element of $\mathbf{A}$. The same notation is used for $\mathbf{B}$ and $\boldsymbol{\mathscr{T}}_{\bar{\iota}}(\omega)$. If we arrange restrictions by increasing ς (here we allow row and column indices to start from 0), then we have

$$A[\varsigma,\varsigma] = \sum_{u=\varsigma}^{\bar{\iota}} (2u+1)2^{-2\varsigma}\left[\varsigma!\binom{u}{\varsigma}\binom{u+\varsigma}{\varsigma}\right]^2,$$

and for $\varsigma > \ell$

$$A[\varsigma,\ell] = \begin{cases} 0 & \varsigma+\ell \text{ is odd} \\ \sum_{u=\varsigma}^{\bar{\iota}} (2u+1)2^{-\varsigma-\ell}\varsigma!\ell!\binom{u}{\varsigma}\binom{u+\varsigma}{\varsigma}\binom{u}{\ell}\binom{u+\ell}{\ell} & \varsigma+\ell \text{ is even}\end{cases}.$$

$\mathbf{B}$ can be expressed explicitly as well:

$$B[\varsigma,\varsigma] = \sum_{u=\varsigma}^{\bar{\iota}} (-1)^{u+\varsigma+1}\frac{2u+1}{2}2^{-2\varsigma}\left[\varsigma!\binom{u}{\varsigma}\binom{u+\varsigma}{\varsigma}\right]^2, \quad \text{and}$$

$$B[\varsigma,\ell] = \begin{cases} \sum_{u=\varsigma}^{\bar{\iota}} (-1)^{u+\varsigma+1}\frac{2u+1}{2}2^{-\varsigma-\ell}\varsigma!\ell!\binom{u}{\varsigma}\binom{u+\varsigma}{\varsigma}\binom{u}{\ell}\binom{u+\ell}{\ell} & \varsigma > \ell \\ (-1)^{\ell+\varsigma}B[\ell,\varsigma] & \varsigma < \ell \end{cases}.$$

Therefore, (ς,ℓ)th element of $\boldsymbol{\mathscr{T}}_{\bar{\iota}}(\omega)$ is

$$\mathscr{T}_{\bar{\iota}}(\omega)[\varsigma,\ell] = \begin{cases} 0 & \ell+\varsigma \text{ is odd} \\ A[\varsigma,\ell] + 2B[\varsigma,\ell]\cos\omega & \ell+\varsigma \text{ is even}\end{cases}.$$

When $\ell+\varsigma$ is even, the corresponding entry of $\boldsymbol{\mathscr{T}}_{\bar{\iota}}(\omega)$ is nonzero. In addition, the summands in $A[\varsigma,\ell]$ and $2B[\varsigma,\ell]$ are the same in terms of absolute values and only differ in signs. Consider the case when $\cos\omega = 1$ and $\varsigma \geq \ell$. There are several cases:

(i) $\bar{\iota}$ is even, ς is even

$$\mathscr{T}_{\bar{\iota}}(\omega)[\varsigma,\ell] = \sum_{u=\varsigma/2}^{(\bar{\iota}-2)/2} \Big(2(2u+1)+1\Big)2^{-\varsigma-\ell+1}\varsigma!\ell!\binom{2u+1}{\varsigma}\binom{2u+1+\varsigma}{\varsigma}\binom{2u+1}{\ell}\binom{2u+1+\ell}{\ell};$$

(ii) $\bar{\iota}$ is even, ς is odd

$$\mathscr{T}_{\bar{\iota}}(\omega)[\varsigma,\ell] = \sum_{u=(\varsigma+1)/2}^{\bar{\iota}/2} \Big(2(2u)+1\Big)2^{-\varsigma-\ell+1}\varsigma!\ell!\binom{2u}{\varsigma}\binom{2u+\varsigma}{\varsigma}\binom{2u}{\ell}\binom{2u+\ell}{\ell};$$

(iii) $\bar{\iota}$ is odd, ς is even

$$\mathscr{T}_{\bar{\iota}}(\omega)[\varsigma,\ell] = \sum_{u=\varsigma/2}^{(\bar{\iota}-1)/2} \Big(2(2u+1)+1\Big)2^{-\varsigma-\ell+1}\varsigma!\ell!\binom{2u+1}{\varsigma}\binom{2u+1+\varsigma}{\varsigma}\binom{2u+1}{\ell}\binom{2u+1+\ell}{\ell};$$

(iv) $\bar{\iota}$ is odd, ς is odd

$$\mathscr{T}_{\bar{\iota}}(\omega)[\varsigma,\ell]=\sum_{u=(\varsigma+1)/2}^{(\bar{\iota}-1)/2}\Big(2(2u)+1\Big)2^{-\varsigma-\ell+1}\varsigma!\ell!\binom{2u}{\varsigma}\binom{2u+\varsigma}{\varsigma}\binom{2u}{\ell}\binom{2u+\ell}{\ell}.$$

When $\bar{\iota}$ is odd, $\mathscr{T}_{\bar{\iota}}(\omega)$ can be written as a Gram matrix $\mathscr{T}_{\bar{\iota}}(\omega)=\mathbf{G}\mathbf{G}'$ where

$$\mathbf{G}=\begin{bmatrix}\bar{P}_1(1) & 0 & \bar{P}_3(1) & \cdots & 0 & \bar{P}_{m-1}(1)\\ 0 & \bar{P}_2^{(1)}(1) & 0 & \cdots & \bar{P}_{m-2}^{(1)}(1) & 0\\ 0 & 0 & \bar{P}_3^{(2)}(1) & \cdots & 0 & \bar{P}_{m-1}^{(2)}(1)\\ \vdots & \vdots & \vdots & & \vdots & \vdots\\ 0 & 0 & 0 & \cdots & 0 & \bar{P}_{m-1}^{(m-2)}(1)\end{bmatrix}.$$

Clearly, it is a row echelon matrix and has full row rank. Thus, the minimum eigenvalue of $\mathscr{T}_{\bar{\iota}}(\omega)$ is strictly positive.

When $\bar{\iota}$ is even, $\mathscr{T}_{\bar{\iota}}(\omega)$ can be written as $\mathbf{G}\mathbf{G}'$ where

$$\mathbf{G}=\begin{bmatrix}\bar{P}_1(1) & 0 & \bar{P}_3(1) & \cdots & \bar{P}_{m-2}(1) & 0\\ 0 & \bar{P}_2^{(1)}(1) & 0 & \cdots & 0 & \bar{P}_{m-1}^{(1)}(1)\\ 0 & 0 & \bar{P}_3^{(2)}(1) & \cdots & \bar{P}_{m-2}^{(2)}(1) & 0\\ \vdots & \vdots & \vdots & & \vdots & \vdots\\ 0 & 0 & 0 & \cdots & 0 & \bar{P}_{m-1}^{(m-2)}(1)\end{bmatrix}.$$

The case of $\cos\omega=-1$ can be proved the same way. This suffices to show that the minimum eigenvalue of $\mathscr{T}_{\bar{\iota}}(\omega)$ as a function of $\cos\omega$ is strictly positive at the two endpoints, 1 and -1, thus completing the proof for $\bar{\iota}=m-1$.

To complete the proof of the lemma, it remains to extend this result to the case in which fewer constraints are imposed. Compared with the case when $(m-1)$ constraints are imposed, some rows in the bigger restriction matrix are removed, and accordingly, $\mathbf{R}\mathbf{R}'$ is a principle submatrix of the original one. By Cauchy Interlacing Theorem, the smallest eigenvalue of the principle submatrix must be no less than the smallest eigenvalue of the original matrix. Combining this fact with the results proved for $\bar{\iota}=m-1$, we have the minimum eigenvalue of $\mathbf{R}\mathbf{R}'$ uniformly bounded away from 0 when fewer restrictions are imposed, and then the proof is complete. □

Table SA-1
Pointwise Results, Model 1, Spline, Evenly-spaced, $n = 1000$, 5000 Replications

	κ	$x = 0.2$ RMSE	CR	AL	$x = 0.5$ RMSE	CR	AL	$x = 0.8$ RMSE	CR	AL
$j = 0$										
	1.0	0.084	56.4	0.185	0.057	84.9	0.249	0.069	69.3	0.185
	2.0	0.055	79.0	0.190	0.045	82.3	0.175	0.002	94.5	0.190
$\kappa_{\texttt{IMSE}}$	3.0	0.006	94.9	0.251	0.046	91.6	0.328	0.036	90.9	0.251
	4.0	0.028	93.8	0.376	0.004	94.7	0.222	0.022	94.0	0.376
	5.0	0.003	94.9	0.320	0.018	94.4	0.399	0.002	94.8	0.320
$\hat{\kappa}_{\texttt{ROT}}$	4.1	0.023	94.5	0.359	0.002	94.9	0.254	0.019	95.0	0.359
$\hat{\kappa}_{\texttt{DPI}}$	4.7	0.010	94.9	0.337	0.008	93.8	0.311	0.007	94.7	0.337
$j = 1$										
	1.0	0.032	90.8	0.205	0.030	90.3	0.186	0.056	80.8	0.205
	2.0	0.026	93.3	0.269	0.013	94.1	0.251	0.033	92.4	0.269
$\kappa_{\texttt{IMSE}}$	3.0	0.002	95.3	0.305	0.003	94.8	0.226	0.001	94.8	0.304
	4.0	0.003	95.2	0.279	0.008	94.7	0.309	0.010	94.4	0.278
	5.0	0.001	94.7	0.296	0.008	94.8	0.268	0.009	94.7	0.296
$\hat{\kappa}_{\texttt{ROT}}$	4.1	0.002	95.2	0.283	0.008	94.8	0.297	0.009	94.5	0.283
$\hat{\kappa}_{\texttt{DPI}}$	4.7	0.001	95.2	0.297	0.007	94.9	0.294	0.009	94.6	0.296
$j = 2$										
	1.0	0.029	91.4	0.205	0.018	93.4	0.194	0.054	82.1	0.205
	2.0	0.032	92.4	0.270	0.013	94.2	0.251	0.038	91.3	0.270
$\kappa_{\texttt{IMSE}}$	3.0	0.002	95.1	0.315	0.004	94.7	0.268	0.002	94.8	0.315
	4.0	0.004	94.5	0.314	0.009	94.8	0.311	0.001	94.9	0.314
	5.0	0.000	94.4	0.311	0.001	94.5	0.340	0.006	94.8	0.311
$\hat{\kappa}_{\texttt{ROT}}$	4.1	0.003	94.7	0.313	0.008	94.8	0.312	0.000	94.7	0.313
$\hat{\kappa}_{\texttt{DPI}}$	4.7	0.002	94.6	0.318	0.004	94.8	0.330	0.002	95.1	0.318
$j = 3$										
	1.0	0.045	85.1	0.198	0.018	33.9	0.268	0.037	88.7	0.197
	2.0	0.005	94.9	0.223	0.012	94.1	0.198	0.042	87.9	0.223
$\kappa_{\texttt{IMSE}}$	3.0	0.005	95.0	0.248	0.016	90.0	0.320	0.036	90.7	0.248
	4.0	0.003	94.8	0.351	0.008	94.4	0.250	0.021	93.6	0.351
	5.0	0.002	95.0	0.319	0.003	94.0	0.390	0.003	94.8	0.320
$\hat{\kappa}_{\texttt{ROT}}$	4.1	0.004	94.7	0.339	0.007	94.2	0.275	0.019	94.1	0.339
$\hat{\kappa}_{\texttt{DPI}}$	4.7	0.002	94.9	0.329	0.005	94.2	0.322	0.003	94.8	0.329

Table SA-2

Pointwise Results, Model 2, Spline, Evenly-spaced, $n = 1000$, 5000 Replications

	κ	$x=0.2$ RMSE	CR	AL	$x=0.5$ RMSE	CR	AL	$x=0.8$ RMSE	CR	AL
$j=0$										
	5.0	0.065	86.9	0.321	0.082	87.1	0.402	0.001	94.9	0.320
	6.0	0.043	90.6	0.278	0.095	69.9	0.263	0.004	94.8	0.277
$\kappa_{\texttt{IMSE}}$	7.0	0.008	94.8	0.290	0.077	89.6	0.462	0.006	94.8	0.290
	8.0	0.018	93.8	0.375	0.046	90.4	0.297	0.002	94.8	0.375
	9.0	0.026	93.9	0.513	0.060	91.9	0.515	0.004	94.8	0.514
$\hat{\kappa}_{\texttt{ROT}}$	6.1	0.040	90.4	0.279	0.083	71.8	0.277	0.003	95.0	0.278
$\hat{\kappa}_{\texttt{DPI}}$	7.8	0.014	93.6	0.361	0.008	93.2	0.360	0.003	94.3	0.361
$j=1$										
	5.0	0.065	85.7	0.297	0.098	69.9	0.268	0.048	89.9	0.297
	6.0	0.001	94.0	0.360	0.020	94.1	0.363	0.027	93.7	0.360
$\kappa_{\texttt{IMSE}}$	7.0	0.009	94.2	0.396	0.032	92.9	0.307	0.005	94.5	0.396
	8.0	0.011	94.2	0.382	0.019	94.2	0.411	0.001	94.8	0.382
	9.0	0.002	94.4	0.351	0.007	94.6	0.343	0.000	94.9	0.351
$\hat{\kappa}_{\texttt{ROT}}$	6.1	0.000	94.2	0.362	0.016	94.0	0.359	0.025	93.3	0.362
$\hat{\kappa}_{\texttt{DPI}}$	7.8	0.011	94.3	0.384	0.006	92.6	0.376	0.000	95.0	0.384
$j=2$										
	5.0	0.045	91.2	0.311	0.014	94.3	0.340	0.027	93.5	0.311
	6.0	0.005	94.3	0.362	0.024	93.9	0.367	0.031	93.5	0.362
$\kappa_{\texttt{IMSE}}$	7.0	0.014	94.3	0.403	0.001	94.4	0.398	0.013	94.5	0.403
	8.0	0.001	94.5	0.413	0.022	94.4	0.416	0.015	94.3	0.414
	9.0	0.001	94.3	0.434	0.002	94.3	0.446	0.002	94.7	0.435
$\hat{\kappa}_{\texttt{ROT}}$	6.1	0.005	94.4	0.365	0.022	93.9	0.369	0.027	93.1	0.365
$\hat{\kappa}_{\texttt{DPI}}$	7.8	0.004	94.7	0.412	0.015	94.3	0.413	0.012	94.4	0.412
$j=3$										
	5.0	0.074	84.8	0.321	0.044	67.9	0.393	0.002	94.7	0.320
	6.0	0.019	93.9	0.302	0.041	91.3	0.298	0.013	94.7	0.302
$\kappa_{\texttt{IMSE}}$	7.0	0.001	94.4	0.325	0.035	86.0	0.452	0.007	94.8	0.325
	8.0	0.018	93.9	0.374	0.009	94.5	0.337	0.002	94.8	0.374
	9.0	0.024	92.9	0.492	0.023	90.5	0.504	0.003	94.9	0.493
$\hat{\kappa}_{\texttt{ROT}}$	6.1	0.017	93.6	0.304	0.035	90.9	0.309	0.013	94.7	0.304
$\hat{\kappa}_{\texttt{DPI}}$	7.8	0.016	93.7	0.368	0.003	93.0	0.382	0.003	94.6	0.369

TABLE SA-3

Pointwise Results, Model 3, Spline, Evenly-spaced, $n = 1000$, 5000 Replications

	κ	$x=0.2$ RMSE	$x=0.2$ CR	$x=0.2$ AL	$x=0.5$ RMSE	$x=0.5$ CR	$x=0.5$ AL	$x=0.8$ RMSE	$x=0.8$ CR	$x=0.8$ AL
$j=0$										
	10.0	0.002	94.5	0.423	0.487	0.04	0.343	0.002	94.4	0.423
	11.0	0.003	94.5	0.357	0.233	63.40	0.566	0.003	94.5	0.357
$\kappa_{\texttt{IMSE}}$	12.0	0.004	94.1	0.368	0.349	3.88	0.366	0.003	94.9	0.368
	13.0	0.005	94.6	0.466	0.229	68.30	0.611	0.004	94.6	0.467
	14.0	0.003	94.1	0.625	0.251	28.10	0.389	0.001	94.5	0.627
$\hat{\kappa}_{\texttt{ROT}}$	3.7	0.209	31.9	0.342	1.010	0.00	0.294	0.207	32.0	0.342
$\hat{\kappa}_{\texttt{DPI}}$	7.6	0.011	92.8	0.375	0.467	19.70	0.371	0.011	93.1	0.376
$j=1$										
	10.0	0.000	94.3	0.398	0.035	93.40	0.454	0.000	94.8	0.399
	11.0	0.029	93.5	0.466	0.281	17.90	0.383	0.028	94.1	0.467
$\kappa_{\texttt{IMSE}}$	12.0	0.005	94.0	0.490	0.010	93.90	0.493	0.004	94.7	0.491
	13.0	0.002	93.9	0.459	0.176	60.80	0.410	0.001	94.5	0.460
	14.0	0.004	94.2	0.423	0.025	93.70	0.530	0.004	94.8	0.423
$\hat{\kappa}_{\texttt{ROT}}$	3.7	0.101	71.5	0.290	0.823	0.00	0.310	0.100	73.4	0.290
$\hat{\kappa}_{\texttt{DPI}}$	7.6	0.032	91.8	0.371	0.304	31.90	0.382	0.032	90.8	0.371
$j=2$										
	10.0	0.001	94.1	0.433	0.037	93.40	0.460	0.000	94.8	0.434
	11.0	0.036	93.4	0.474	0.046	92.50	0.490	0.035	93.6	0.475
$\kappa_{\texttt{IMSE}}$	12.0	0.009	94.1	0.500	0.019	93.90	0.500	0.008	94.7	0.501
	13.0	0.012	94.0	0.503	0.030	93.70	0.528	0.014	94.4	0.505
	14.0	0.002	94.2	0.540	0.037	93.40	0.538	0.004	94.1	0.541
$\hat{\kappa}_{\texttt{ROT}}$	3.7	0.104	64.7	0.320	0.733	0.00	0.326	0.102	65.2	0.320
$\hat{\kappa}_{\texttt{DPI}}$	7.6	0.007	90.5	0.405	0.204	53.90	0.412	0.006	90.5	0.405
$j=3$										
	10.0	0.002	94.3	0.420	0.280	19.30	0.386	0.002	94.3	0.421
	11.0	0.007	94.3	0.393	0.066	91.30	0.551	0.007	94.2	0.393
$\kappa_{\texttt{IMSE}}$	12.0	0.006	94.2	0.409	0.171	63.40	0.413	0.005	94.7	0.410
	13.0	0.005	94.7	0.465	0.069	91.10	0.595	0.004	94.7	0.467
	14.0	0.005	94.4	0.608	0.103	83.30	0.440	0.001	94.6	0.610
$\hat{\kappa}_{\texttt{ROT}}$	3.7	0.249	20.0	0.327	0.902	0.00	0.312	0.250	19.5	0.326
$\hat{\kappa}_{\texttt{DPI}}$	7.6	0.024	93.1	0.379	0.366	28.40	0.392	0.024	93.6	0.380

TABLE SA-4
Pointwise Results, Model 4, Spline, Evenly-spaced, $n = 1000$, 5000 Replications

		$\mathbf{x} = (0.5, 0.5)$			$\mathbf{x} = (0.1, 0.5)$			$\mathbf{x} = (0.1, 0.1)$		
	κ	RMSE	CR	AL	RMSE	CR	AL	RMSE	CR	AL
$j = 0$										
	2.0	0.515	0.0	0.267	0.405	1.68	0.384	0.087	89.4	0.574
	3.0	0.320	69.4	0.877	0.133	87.80	0.694	0.116	85.3	0.556
$\kappa_{\texttt{IMSE}}$	4.0	0.182	57.4	0.402	0.141	78.30	0.470	0.074	90.3	0.550
	5.0	0.187	90.0	1.300	0.107	91.40	0.882	0.034	93.4	0.596
	6.0	0.070	91.5	0.564	0.045	93.60	0.648	0.003	93.6	0.740
$\hat{\kappa}_{\texttt{ROT}}$	2.9	0.266	65.1	0.839	0.100	81.80	0.675	0.115	85.7	0.558
$\hat{\kappa}_{\texttt{DPI}}$	4.0	0.182	57.4	0.402	0.141	78.30	0.470	0.074	90.3	0.550
$j = 1$										
	2.0	0.026	93.8	0.510	0.000	95.10	0.516	0.006	94.5	0.526
	3.0	0.164	66.4	0.417	0.119	85.70	0.516	0.077	92.3	0.636
$\kappa_{\texttt{IMSE}}$	4.0	0.042	93.2	0.774	0.035	94.20	0.820	0.039	93.5	0.869
	5.0	0.024	93.9	0.585	0.012	94.30	0.828	0.013	93.5	1.170
	6.0	0.013	93.8	1.070	0.023	93.40	1.240	0.006	93.0	1.440
$\hat{\kappa}_{\texttt{ROT}}$	2.9	0.155	67.9	0.423	0.111	87.10	0.516	0.072	92.2	0.628
$\hat{\kappa}_{\texttt{DPI}}$	4.0	0.042	93.2	0.774	0.035	94.20	0.820	0.039	93.5	0.869
$j = 2$										
	2.0	0.027	93.7	0.510	0.001	95.20	0.520	0.000	94.4	0.538
	3.0	0.029	93.8	0.658	0.035	94.60	0.597	0.072	92.3	0.641
$\kappa_{\texttt{IMSE}}$	4.0	0.052	93.2	0.780	0.041	94.30	0.826	0.040	93.6	0.873
	5.0	0.017	94.0	1.060	0.004	94.10	0.955	0.014	93.6	1.180
	6.0	0.014	94.4	1.090	0.025	93.90	1.260	0.007	93.4	1.480
$\hat{\kappa}_{\texttt{ROT}}$	2.9	0.030	93.7	0.648	0.032	94.80	0.592	0.068	92.2	0.634
$\hat{\kappa}_{\texttt{DPI}}$	4.0	0.051	93.2	0.780	0.041	94.30	0.826	0.040	93.6	0.873
$j = 3$										
	2.0	0.254	27.8	0.390	0.190	59.60	0.434	0.028	94.3	0.549
	3.0	0.146	86.5	0.834	0.054	92.10	0.693	0.063	93.2	0.625
$\kappa_{\texttt{IMSE}}$	4.0	0.044	93.1	0.564	0.028	95.00	0.685	0.039	93.8	0.815
	5.0	0.057	93.9	1.260	0.043	94.10	0.962	0.023	93.6	0.998
	6.0	0.006	94.1	0.790	0.007	94.20	0.901	0.020	93.4	1.030
$\hat{\kappa}_{\texttt{ROT}}$	2.9	0.120	82.9	0.806	0.039	88.10	0.677	0.061	93.1	0.620
$\hat{\kappa}_{\texttt{DPI}}$	4.0	0.044	93.1	0.564	0.028	95.00	0.685	0.039	93.8	0.815

Table SA-5
Pointwise Results, Model 5, Spline, Evenly-spaced, $n = 1000$, 5000 Replications

	κ	$\mathbf{x} = (0.5, 0.5)$			$\mathbf{x} = (0.1, 0.5)$			$\mathbf{x} = (0.1, 0.1)$		
		RMSE	CR	AL	RMSE	CR	AL	RMSE	CR	AL
$j = 0$										
	2.0	0.129	47.8	0.251	0.035	93.1	0.386	0.068	91.3	0.561
	3.0	0.086	92.2	0.870	0.178	82.7	0.701	0.083	89.5	0.555
$\kappa_{\mathtt{IMSE}}$	4.0	0.021	94.0	0.399	0.082	89.4	0.471	0.074	90.4	0.551
	5.0	0.022	94.2	1.300	0.087	92.3	0.882	0.055	92.5	0.596
	6.0	0.014	93.9	0.563	0.025	94.4	0.648	0.034	93.6	0.740
$\hat{\kappa}_{\mathtt{ROT}}$	3.0	0.086	92.0	0.868	0.177	82.7	0.700	0.083	89.5	0.555
$\hat{\kappa}_{\mathtt{DPI}}$	3.9	0.015	94.0	0.434	0.088	88.9	0.488	0.074	90.4	0.551
$j = 1$										
	2.0	0.028	93.7	0.509	0.062	92.2	0.517	0.066	91.9	0.526
	3.0	0.016	94.1	0.414	0.015	95.3	0.512	0.026	94.5	0.634
$\kappa_{\mathtt{IMSE}}$	4.0	0.001	93.9	0.773	0.000	95.0	0.820	0.010	93.7	0.868
	5.0	0.006	94.4	0.585	0.013	94.3	0.828	0.003	93.6	1.170
	6.0	0.000	93.9	1.070	0.019	93.4	1.240	0.004	93.0	1.440
$\hat{\kappa}_{\mathtt{ROT}}$	3.0	0.016	94.1	0.415	0.015	95.3	0.513	0.026	94.5	0.635
$\hat{\kappa}_{\mathtt{DPI}}$	3.9	0.001	94.0	0.746	0.001	95.0	0.798	0.010	93.7	0.851
$j = 2$										
	2.0	0.031	93.7	0.510	0.060	92.7	0.520	0.060	92.4	0.537
	3.0	0.015	94.0	0.657	0.016	94.7	0.596	0.025	94.4	0.639
$\kappa_{\mathtt{IMSE}}$	4.0	0.001	94.0	0.780	0.000	95.2	0.826	0.009	93.9	0.872
	5.0	0.004	94.0	1.060	0.014	93.9	0.955	0.004	93.6	1.180
	6.0	0.000	94.5	1.090	0.020	94.0	1.260	0.005	93.4	1.480
$\hat{\kappa}_{\mathtt{ROT}}$	3.0	0.015	94.0	0.657	0.016	94.7	0.597	0.025	94.4	0.639
$\hat{\kappa}_{\mathtt{DPI}}$	3.9	0.000	93.9	0.771	0.002	95.2	0.809	0.010	93.9	0.855
$j = 3$										
	2.0	0.064	88.8	0.373	0.011	95.1	0.422	0.045	93.3	0.541
	3.0	0.056	93.6	0.831	0.053	93.6	0.693	0.030	94.4	0.623
$\kappa_{\mathtt{IMSE}}$	4.0	0.002	94.1	0.562	0.011	95.3	0.683	0.024	94.1	0.815
	5.0	0.002	94.2	1.260	0.012	94.1	0.962	0.019	93.6	0.998
	6.0	0.003	93.9	0.790	0.001	94.1	0.901	0.018	93.3	1.030
$\hat{\kappa}_{\mathtt{ROT}}$	3.0	0.056	93.6	0.830	0.053	93.6	0.693	0.030	94.3	0.623
$\hat{\kappa}_{\mathtt{DPI}}$	3.9	0.001	94.2	0.581	0.014	95.4	0.684	0.024	94.0	0.801

Table SA-6
Pointwise Results, Model 6, Spline, Evenly-spaced, $n = 1000$, 5000 Replications

	κ	$\mathbf{x} = (0.5, 0.5, 0.5)$ RMSE	CR	AL	$\mathbf{x} = (0.1, 0.1, 0.5)$ RMSE	CR	AL	$\mathbf{x} = (0.1, 0.1, 0.1)$ RMSE	CR	AL
$j = 0$										
	1.0	0.002	94.9	1.000	0.003	93.4	1.030	0.152	88.0	1.16
	2.0	0.001	94.7	0.363	0.005	91.8	0.816	0.231	79.8	1.33
$\kappa_{\texttt{IMSE}}$	3.0	0.004	93.8	2.410	0.014	89.4	1.500	0.274	73.7	1.28
$\hat{\kappa}_{\texttt{ROT}}$	2.1	0.003	94.8	0.570	0.006	91.5	0.887	0.236	79.3	1.33
$\hat{\kappa}_{\texttt{DPI}}$	3.0	0.003	93.8	2.400	0.014	89.5	1.500	0.274	73.7	1.28
$j = 1$										
	1.0	0.001	94.7	0.433	0.005	91.4	0.807	0.195	78.9	1.19
	2.0	0.000	95.5	1.090	0.010	91.7	1.110	0.130	84.1	1.25
$\kappa_{\texttt{IMSE}}$	3.0	0.002	96.3	0.842	0.014	91.7	1.300	0.081	85.5	2.09
$\hat{\kappa}_{\texttt{ROT}}$	2.1	0.000	95.3	1.070	0.010	91.9	1.130	0.121	84.1	1.33
$\hat{\kappa}_{\texttt{DPI}}$	3.0	0.002	96.3	0.845	0.014	91.7	1.300	0.081	85.5	2.09
$j = 2$										
	1.0	0.000	95.0	0.563	0.006	92.1	0.856	0.118	85.4	1.27
	2.0	0.001	95.4	1.090	0.010	92.0	1.130	0.124	86.4	1.35
$\kappa_{\texttt{IMSE}}$	3.0	0.008	94.8	1.850	0.014	91.7	1.440	0.083	87.4	2.22
$\hat{\kappa}_{\texttt{ROT}}$	2.1	0.000	95.2	1.170	0.009	92.2	1.170	0.116	86.4	1.43
$\hat{\kappa}_{\texttt{DPI}}$	3.0	0.008	94.8	1.850	0.015	91.7	1.440	0.084	87.4	2.22
$j = 3$										
	1.0	0.001	95.7	1.020	0.008	92.7	1.210	0.132	87.0	1.18
	2.0	0.001	95.3	0.860	0.004	93.3	0.952	0.077	89.3	1.31
$\kappa_{\texttt{IMSE}}$	3.0	0.006	94.7	2.380	0.014	93.5	1.860	0.033	91.5	2.28
$\hat{\kappa}_{\texttt{ROT}}$	2.1	0.000	95.4	1.010	0.005	93.3	1.050	0.076	89.6	1.38
$\hat{\kappa}_{\texttt{DPI}}$	3.0	0.006	94.7	2.380	0.014	93.5	1.900	0.033	91.5	2.31

Table SA-7

Pointwise Results, Model 7, Spline, Evenly-spaced, $n = 1000$, 5000 Replications

	κ	$\mathbf{x} = (0.5, 0.5, 0.5)$ RMSE	CR	AL	$\mathbf{x} = (0.1, 0.1, 0.5)$ RMSE	CR	AL	$\mathbf{x} = (0.1, 0.1, 0.1)$ RMSE	CR	AL
$j = 0$										
	1.0	0.002	95.0	0.999	0.004	93.6	0.992	0.003	92.8	0.99
$\kappa_{\mathtt{IMSE}}$	2.0	0.004	94.6	0.362	0.004	92.4	0.789	0.003	89.5	1.15
	3.0	0.003	93.8	2.410	0.014	89.1	1.490	0.007	85.8	1.21
$\hat{\kappa}_{\mathtt{ROT}}$	2.1	0.003	94.6	0.512	0.004	92.3	0.839	0.003	89.1	1.15
$\hat{\kappa}_{\mathtt{DPI}}$	3.0	0.002	93.8	2.400	0.014	89.1	1.490	0.002	85.8	1.23
$j = 1$										
	1.0	0.004	94.6	0.432	0.005	91.8	0.783	0.001	88.8	1.06
$\kappa_{\mathtt{IMSE}}$	2.0	0.000	95.4	1.090	0.010	91.5	1.100	0.006	86.5	1.24
	3.0	0.002	96.3	0.842	0.014	91.7	1.300	0.038	85.7	2.08
$\hat{\kappa}_{\mathtt{ROT}}$	2.1	0.000	95.4	1.070	0.010	91.5	1.120	0.001	86.2	1.31
$\hat{\kappa}_{\mathtt{DPI}}$	3.0	0.002	96.3	0.844	0.015	91.7	1.300	0.018	85.8	2.28
$j = 2$										
	1.0	0.004	95.1	0.562	0.005	92.7	0.838	0.001	90.8	1.12
$\kappa_{\mathtt{IMSE}}$	2.0	0.001	95.4	1.090	0.010	92.1	1.130	0.004	89.0	1.34
	3.0	0.008	94.8	1.850	0.014	91.7	1.440	0.039	87.6	2.22
$\hat{\kappa}_{\mathtt{ROT}}$	2.1	0.001	95.4	1.150	0.010	92.0	1.160	0.003	88.9	1.41
$\hat{\kappa}_{\mathtt{DPI}}$	3.0	0.008	94.8	1.850	0.015	91.6	1.440	0.018	87.6	2.44
$j = 3$										
	1.0	0.006	95.6	1.020	0.007	93.1	1.160	0.001	92.0	1.07
$\kappa_{\mathtt{IMSE}}$	2.0	0.003	95.2	0.859	0.004	93.5	0.946	0.000	90.7	1.27
	3.0	0.005	94.6	2.380	0.014	93.5	1.860	0.023	91.7	2.27
$\hat{\kappa}_{\mathtt{ROT}}$	2.1	0.005	95.3	0.970	0.004	93.5	1.020	0.004	90.6	1.33
$\hat{\kappa}_{\mathtt{DPI}}$	3.0	0.005	94.6	2.370	0.012	93.5	2.030	0.224	91.7	21.30

Table SA-8

Pointwise Results, Model 1, Spline, Quantile-spaced, $n = 1000$, 5000 Replications

	κ	$x = 0.2$ RMSE	CR	AL	$x = 0.5$ RMSE	CR	AL	$x = 0.8$ RMSE	CR	AL
$j = 0$										
	1.0	0.084	56.5	0.185	0.046	87.6	0.240	0.068	69.6	0.185
	2.0	0.054	79.0	0.190	0.045	81.9	0.176	0.003	94.2	0.190
$\kappa_{\mathtt{IMSE}}$	3.0	0.006	94.8	0.252	0.036	92.1	0.308	0.036	90.8	0.253
	4.0	0.021	94.3	0.352	0.003	94.5	0.225	0.018	94.0	0.352
	5.0	0.003	94.6	0.322	0.013	94.5	0.364	0.001	94.9	0.321
$\hat{\kappa}_{\mathtt{ROT}}$	4.1	0.018	94.5	0.341	0.001	94.6	0.250	0.016	95.0	0.340
$\hat{\kappa}_{\mathtt{DPI}}$	4.7	0.008	94.9	0.328	0.006	94.0	0.299	0.006	94.6	0.327
$j = 1$										
	1.0	0.032	90.7	0.205	0.030	89.5	0.187	0.056	80.7	0.205
	2.0	0.025	93.3	0.269	0.012	94.2	0.251	0.033	92.3	0.269
$\kappa_{\mathtt{IMSE}}$	3.0	0.002	95.4	0.303	0.004	94.6	0.228	0.001	94.8	0.303
	4.0	0.003	95.1	0.281	0.008	94.6	0.307	0.010	94.4	0.280
	5.0	0.001	94.7	0.298	0.008	94.7	0.274	0.008	94.7	0.298
$\hat{\kappa}_{\mathtt{ROT}}$	4.1	0.002	95.1	0.285	0.008	94.6	0.297	0.009	94.5	0.285
$\hat{\kappa}_{\mathtt{DPI}}$	4.7	0.001	95.1	0.298	0.007	94.7	0.295	0.009	94.6	0.299
$j = 2$										
	1.0	0.029	91.2	0.205	0.019	93.1	0.194	0.054	82.0	0.206
	2.0	0.031	92.4	0.270	0.013	94.2	0.251	0.038	91.2	0.270
$\kappa_{\mathtt{IMSE}}$	3.0	0.002	95.0	0.314	0.004	94.5	0.264	0.001	94.8	0.313
	4.0	0.003	94.6	0.310	0.009	94.6	0.310	0.000	95.0	0.310
	5.0	0.000	94.3	0.314	0.002	94.5	0.329	0.005	94.8	0.314
$\hat{\kappa}_{\mathtt{ROT}}$	4.1	0.002	94.7	0.311	0.008	94.7	0.309	0.001	95.0	0.311
$\hat{\kappa}_{\mathtt{DPI}}$	4.7	0.001	94.4	0.318	0.004	94.6	0.324	0.002	95.2	0.318
$j = 3$										
	1.0	0.046	85.0	0.198	0.022	35.5	0.253	0.037	88.5	0.197
	2.0	0.005	94.9	0.223	0.012	93.9	0.199	0.043	87.6	0.223
$\kappa_{\mathtt{IMSE}}$	3.0	0.005	94.9	0.249	0.012	89.9	0.299	0.037	90.7	0.249
	4.0	0.000	94.8	0.340	0.008	94.7	0.251	0.005	93.7	0.341
	5.0	0.002	94.8	0.324	0.003	93.8	0.355	0.004	94.9	0.324
$\hat{\kappa}_{\mathtt{ROT}}$	4.1	0.002	94.5	0.332	0.007	94.4	0.269	0.006	94.3	0.331
$\hat{\kappa}_{\mathtt{DPI}}$	4.7	0.001	94.8	0.327	0.005	94.1	0.307	0.001	94.7	0.326

Table SA-9

Pointwise Results, Model 2, Spline, Quantile-spaced, $n = 1000$, 5000 Replications

	κ	$x = 0.2$			$x = 0.5$			$x = 0.8$		
		RMSE	CR	AL	RMSE	CR	AL	RMSE	CR	AL
$j = 0$										
	5.0	0.058	87.7	0.323	0.039	88.5	0.367	0.003	94.3	0.321
	6.0	0.041	90.7	0.282	0.087	72.0	0.271	0.003	94.8	0.281
$\kappa_{\texttt{IMSE}}$	7.0	0.009	94.2	0.296	0.036	91.1	0.409	0.006	94.6	0.297
	8.0	0.015	94.2	0.377	0.037	90.7	0.313	0.001	94.6	0.380
	9.0	0.014	94.0	0.456	0.026	92.5	0.443	0.001	94.7	0.456
$\hat{\kappa}_{\texttt{ROT}}$	6.1	0.039	90.5	0.284	0.080	73.2	0.281	0.002	95.1	0.282
$\hat{\kappa}_{\texttt{DPI}}$	7.8	0.010	94.0	0.365	0.017	91.6	0.350	0.002	94.4	0.367
$j = 1$										
	5.0	0.060	86.6	0.298	0.091	71.2	0.274	0.046	90.3	0.299
	6.0	0.001	94.1	0.357	0.014	93.4	0.358	0.026	93.8	0.358
$\kappa_{\texttt{IMSE}}$	7.0	0.008	94.1	0.391	0.026	92.6	0.319	0.004	94.7	0.391
	8.0	0.009	94.3	0.380	0.014	93.6	0.401	0.001	94.8	0.379
	9.0	0.002	94.3	0.364	0.002	94.2	0.362	0.000	94.7	0.364
$\hat{\kappa}_{\texttt{ROT}}$	6.1	0.001	94.2	0.359	0.012	93.3	0.355	0.023	93.4	0.360
$\hat{\kappa}_{\texttt{DPI}}$	7.8	0.009	94.4	0.381	0.006	93.0	0.377	0.000	94.9	0.381
$j = 2$										
	5.0	0.040	91.2	0.314	0.020	93.0	0.329	0.027	93.2	0.314
	6.0	0.004	94.0	0.361	0.022	93.9	0.365	0.028	93.5	0.362
$\kappa_{\texttt{IMSE}}$	7.0	0.012	94.2	0.400	0.002	94.4	0.382	0.012	94.3	0.400
	8.0	0.000	94.4	0.413	0.020	94.3	0.412	0.013	94.2	0.414
	9.0	0.001	94.2	0.422	0.004	93.9	0.427	0.001	94.7	0.423
$\hat{\kappa}_{\texttt{ROT}}$	6.1	0.004	94.3	0.364	0.020	94.0	0.366	0.025	93.0	0.365
$\hat{\kappa}_{\texttt{DPI}}$	7.8	0.002	94.7	0.411	0.014	94.4	0.406	0.010	94.4	0.411
$j = 3$										
	5.0	0.077	83.1	0.326	0.010	71.0	0.358	0.001	94.7	0.324
	6.0	0.019	93.7	0.305	0.040	90.1	0.300	0.012	94.5	0.305
$\kappa_{\texttt{IMSE}}$	7.0	0.002	94.4	0.327	0.012	87.7	0.400	0.008	94.7	0.327
	8.0	0.016	93.9	0.379	0.008	93.7	0.344	0.001	94.4	0.381
	9.0	0.004	93.6	0.445	0.010	91.4	0.437	0.002	95.0	0.446
$\hat{\kappa}_{\texttt{ROT}}$	6.1	0.018	93.5	0.307	0.038	89.6	0.308	0.012	94.5	0.307
$\hat{\kappa}_{\texttt{DPI}}$	7.8	0.013	94.0	0.373	0.004	92.4	0.367	0.003	94.6	0.374

TABLE SA-10
Pointwise Results, Model 3, Spline, Quantile-spaced, $n = 1000$, 5000 Replications

	κ	$x = 0.2$ RMSE	CR	AL	$x = 0.5$ RMSE	CR	AL	$x = 0.8$ RMSE	CR	AL
$j = 0$										
	10.0	0.002	94.1	0.426	0.396	10.8	0.367	0.001	94.6	0.424
	11.0	0.002	94.4	0.376	0.028	70.6	0.477	0.002	94.5	0.375
$\kappa_{\texttt{IMSE}}$	12.0	0.003	94.5	0.389	0.250	32.2	0.403	0.003	94.5	0.391
	13.0	0.004	94.5	0.466	0.017	74.7	0.502	0.003	94.5	0.470
	14.0	0.003	93.9	0.526	0.154	56.7	0.440	0.001	94.8	0.527
$\hat{\kappa}_{\texttt{ROT}}$	3.7	0.186	38.8	0.326	1.020	0.0	0.290	0.186	38.7	0.327
$\hat{\kappa}_{\texttt{DPI}}$	7.7	0.010	92.6	0.364	0.487	23.1	0.363	0.011	92.9	0.364
$j = 1$										
	10.0	0.001	94.1	0.400	0.096	80.2	0.441	0.001	94.4	0.402
	11.0	0.022	93.7	0.455	0.218	43.2	0.408	0.022	94.5	0.457
$\kappa_{\texttt{IMSE}}$	12.0	0.007	94.1	0.476	0.049	87.3	0.472	0.006	94.6	0.476
	13.0	0.004	93.9	0.459	0.120	75.7	0.443	0.003	94.3	0.459
	14.0	0.003	94.4	0.450	0.027	90.7	0.501	0.002	94.4	0.452
$\hat{\kappa}_{\texttt{ROT}}$	3.7	0.102	68.6	0.292	0.826	0.0	0.311	0.100	69.7	0.291
$\hat{\kappa}_{\texttt{DPI}}$	7.7	0.032	91.7	0.374	0.322	26.1	0.383	0.032	90.6	0.374
$j = 2$										
	10.0	0.007	94.0	0.439	0.053	90.5	0.455	0.007	94.6	0.440
	11.0	0.029	93.6	0.471	0.065	89.7	0.469	0.028	94.1	0.472
$\kappa_{\texttt{IMSE}}$	12.0	0.009	94.1	0.495	0.004	93.5	0.494	0.007	94.4	0.496
	13.0	0.007	94.3	0.507	0.030	93.1	0.506	0.009	94.5	0.508
	14.0	0.000	94.0	0.520	0.010	93.3	0.530	0.003	94.5	0.521
$\hat{\kappa}_{\texttt{ROT}}$	3.7	0.111	62.8	0.317	0.740	0.0	0.324	0.109	63.4	0.317
$\hat{\kappa}_{\texttt{DPI}}$	7.7	0.016	89.3	0.403	0.220	50.5	0.406	0.016	89.3	0.404
$j = 3$										
	10.0	0.005	94.4	0.429	0.259	29.3	0.395	0.004	94.4	0.428
	11.0	0.005	94.3	0.404	0.083	79.5	0.471	0.005	94.3	0.404
$\kappa_{\texttt{IMSE}}$	12.0	0.005	94.6	0.420	0.150	68.2	0.430	0.005	94.6	0.421
	13.0	0.004	94.3	0.474	0.051	87.2	0.499	0.002	94.5	0.478
	14.0	0.004	94.3	0.522	0.085	85.5	0.465	0.001	95.1	0.523
$\hat{\kappa}_{\texttt{ROT}}$	3.7	0.156	36.5	0.318	0.923	0.0	0.306	0.152	37.4	0.319
$\hat{\kappa}_{\texttt{DPI}}$	7.7	0.021	93.1	0.371	0.399	22.5	0.379	0.020	93.7	0.372

Table SA-11

Pointwise Results, Model 4, Spline, Quantile-spaced, $n = 1000$, 5000 Replications

		$\mathbf{x} = (0.5, 0.5)$			$\mathbf{x} = (0.1, 0.5)$			$\mathbf{x} = (0.1, 0.1)$		
	κ	RMSE	CR	AL	RMSE	CR	AL	RMSE	CR	AL
$j = 0$										
	2.0	0.511	0.0	0.269	0.402	2.16	0.385	0.087	89.2	0.573
	3.0	0.214	79.5	0.774	0.081	91.00	0.652	0.115	85.4	0.556
$\kappa_{\texttt{IMSE}}$	4.0	0.174	61.3	0.414	0.136	79.40	0.477	0.073	90.2	0.552
	5.0	0.113	91.8	1.080	0.062	93.30	0.807	0.034	93.5	0.602
	6.0	0.062	92.0	0.600	0.039	93.70	0.673	0.003	93.9	0.755
$\hat{\kappa}_{\texttt{ROT}}$	2.9	0.168	74.4	0.742	0.051	85.00	0.635	0.114	85.8	0.558
$\hat{\kappa}_{\texttt{DPI}}$	4.0	0.174	61.3	0.414	0.135	79.40	0.478	0.073	90.2	0.552
$j = 1$										
	2.0	0.030	93.5	0.509	0.002	95.00	0.516	0.005	94.2	0.527
	3.0	0.158	68.5	0.426	0.115	86.20	0.522	0.077	92.0	0.637
$\kappa_{\texttt{IMSE}}$	4.0	0.038	93.6	0.764	0.031	94.30	0.814	0.039	93.7	0.870
	5.0	0.020	93.7	0.611	0.008	94.40	0.844	0.013	93.5	1.170
	6.0	0.011	93.6	1.040	0.020	93.90	1.220	0.006	93.1	1.430
$\hat{\kappa}_{\texttt{ROT}}$	2.9	0.150	69.8	0.431	0.107	87.50	0.521	0.072	91.9	0.630
$\hat{\kappa}_{\texttt{DPI}}$	4.0	0.038	93.6	0.764	0.031	94.30	0.814	0.039	93.7	0.870
$j = 2$										
	2.0	0.030	93.6	0.509	0.003	95.20	0.519	0.000	94.5	0.538
	3.0	0.041	93.2	0.611	0.040	94.30	0.589	0.073	92.2	0.642
$\kappa_{\texttt{IMSE}}$	4.0	0.048	93.4	0.773	0.040	94.30	0.822	0.040	93.8	0.874
	5.0	0.013	94.0	0.938	0.001	94.30	0.943	0.013	93.7	1.180
	6.0	0.013	94.3	1.070	0.023	94.10	1.240	0.006	93.1	1.460
$\hat{\kappa}_{\texttt{ROT}}$	2.9	0.041	93.2	0.604	0.037	94.50	0.585	0.068	92.1	0.636
$\hat{\kappa}_{\texttt{DPI}}$	4.0	0.048	93.4	0.773	0.039	94.30	0.822	0.040	93.8	0.874
$j = 3$										
	2.0	0.257	26.6	0.388	0.191	58.70	0.435	0.029	94.5	0.550
	3.0	0.070	90.0	0.741	0.022	93.00	0.646	0.064	92.9	0.625
$\kappa_{\texttt{IMSE}}$	4.0	0.045	93.0	0.560	0.028	95.10	0.680	0.040	94.0	0.812
	5.0	0.026	93.9	1.050	0.024	94.20	0.869	0.024	93.5	0.985
	6.0	0.005	93.9	0.783	0.008	94.50	0.891	0.020	93.4	1.020
$\hat{\kappa}_{\texttt{ROT}}$	2.9	0.049	85.8	0.719	0.009	88.90	0.633	0.062	92.8	0.621
$\hat{\kappa}_{\texttt{DPI}}$	4.0	0.045	93.0	0.561	0.028	95.10	0.681	0.040	94.0	0.812

Table SA-12

Pointwise Results, Model 5, Spline, Quantile-spaced, $n = 1000$, 5000 Replications

	κ	$\mathbf{x} = (0.5, 0.5)$ RMSE	CR	AL	$\mathbf{x} = (0.1, 0.5)$ RMSE	CR	AL	$\mathbf{x} = (0.1, 0.1)$ RMSE	CR	AL
$j = 0$										
	2.0	0.128	48.4	0.253	0.036	93.0	0.387	0.068	91.2	0.561
	3.0	0.063	92.8	0.768	0.162	83.0	0.658	0.083	89.4	0.555
$\kappa_{\texttt{IMSE}}$	4.0	0.020	94.1	0.411	0.082	89.1	0.479	0.074	90.6	0.553
	5.0	0.010	94.1	1.080	0.078	92.7	0.808	0.055	92.2	0.603
	6.0	0.012	94.0	0.599	0.025	94.2	0.673	0.034	93.6	0.755
$\hat{\kappa}_{\texttt{ROT}}$	3.0	0.063	92.7	0.766	0.161	83.0	0.658	0.083	89.4	0.555
$\hat{\kappa}_{\texttt{DPI}}$	3.9	0.016	94.3	0.437	0.087	88.7	0.491	0.074	90.6	0.553
$j = 1$										
	2.0	0.027	93.7	0.507	0.061	92.3	0.516	0.066	91.8	0.527
	3.0	0.015	94.3	0.423	0.015	95.4	0.518	0.026	94.4	0.636
$\kappa_{\texttt{IMSE}}$	4.0	0.000	93.7	0.764	0.000	95.1	0.814	0.010	93.8	0.869
	5.0	0.005	94.2	0.611	0.014	94.3	0.844	0.004	93.6	1.170
	6.0	0.001	93.7	1.040	0.018	93.8	1.220	0.004	93.1	1.430
$\hat{\kappa}_{\texttt{ROT}}$	3.0	0.015	94.3	0.424	0.015	95.4	0.519	0.026	94.4	0.636
$\hat{\kappa}_{\texttt{DPI}}$	3.9	0.000	93.7	0.739	0.002	95.2	0.793	0.011	93.9	0.852
$j = 2$										
	2.0	0.030	93.7	0.508	0.059	92.8	0.520	0.060	92.3	0.538
	3.0	0.015	94.1	0.610	0.016	95.0	0.589	0.025	94.5	0.640
$\kappa_{\texttt{IMSE}}$	4.0	0.000	94.0	0.772	0.000	95.4	0.821	0.010	93.9	0.873
	5.0	0.004	94.2	0.938	0.014	94.2	0.943	0.004	93.6	1.180
	6.0	0.001	94.2	1.070	0.019	94.2	1.240	0.004	93.1	1.460
$\hat{\kappa}_{\texttt{ROT}}$	3.0	0.015	94.1	0.610	0.016	95.0	0.589	0.025	94.5	0.641
$\hat{\kappa}_{\texttt{DPI}}$	3.9	0.000	94.0	0.761	0.002	95.3	0.805	0.011	93.9	0.856
$j = 3$										
	2.0	0.065	88.9	0.371	0.011	95.0	0.423	0.045	93.4	0.541
	3.0	0.038	93.6	0.738	0.044	94.1	0.646	0.031	94.2	0.623
$\kappa_{\texttt{IMSE}}$	4.0	0.002	94.3	0.558	0.011	95.2	0.679	0.025	94.2	0.811
	5.0	0.003	94.4	1.050	0.010	94.4	0.869	0.019	93.7	0.985
	6.0	0.004	94.3	0.782	0.001	94.5	0.891	0.017	93.4	1.020
$\hat{\kappa}_{\texttt{ROT}}$	3.0	0.038	93.5	0.737	0.044	94.1	0.646	0.031	94.2	0.623
$\hat{\kappa}_{\texttt{DPI}}$	3.9	0.001	94.3	0.571	0.014	95.2	0.676	0.025	94.2	0.797

Table SA-13
Pointwise Results, Model 6, Spline, Quantile-spaced, $n = 1000$, 5000 Replications

	κ	$\mathbf{x} = (0.5, 0.5, 0.5)$			$\mathbf{x} = (0.1, 0.1, 0.5)$			$\mathbf{x} = (0.1, 0.1, 0.1)$		
		RMSE	CR	AL	RMSE	CR	AL	RMSE	CR	AL
$j = 0$										
	1.0	0.001	95.0	0.899	0.004	93.4	0.995	0.153	88.1	1.16
	2.0	0.001	94.5	0.367	0.005	91.8	0.818	0.230	79.6	1.33
$\kappa_{\mathtt{IMSE}}$	3.0	0.001	93.9	1.990	0.014	89.4	1.410	0.270	73.9	1.27
$\hat{\kappa}_{\mathtt{ROT}}$	2.1	0.002	94.6	0.534	0.006	91.4	0.878	0.235	79.1	1.32
$\hat{\kappa}_{\mathtt{DPI}}$	3.0	0.001	93.9	1.990	0.014	89.4	1.410	0.270	73.9	1.27
$j = 1$										
	1.0	0.001	94.5	0.438	0.005	91.4	0.809	0.195	79.2	1.19
	2.0	0.000	95.4	1.080	0.010	91.6	1.110	0.129	84.2	1.25
$\kappa_{\mathtt{IMSE}}$	3.0	0.001	96.3	0.869	0.016	91.9	1.310	0.075	85.5	2.03
$\hat{\kappa}_{\mathtt{ROT}}$	2.1	0.000	95.4	1.060	0.010	92.0	1.130	0.120	84.2	1.33
$\hat{\kappa}_{\mathtt{DPI}}$	3.0	0.001	96.3	0.871	0.016	91.9	1.310	0.076	85.5	2.04
$j = 2$										
	1.0	0.000	95.1	0.542	0.006	92.2	0.855	0.118	85.3	1.27
	2.0	0.000	95.5	1.090	0.010	92.1	1.130	0.122	86.6	1.35
$\kappa_{\mathtt{IMSE}}$	3.0	0.005	94.5	1.600	0.016	91.9	1.440	0.077	87.5	2.16
$\hat{\kappa}_{\mathtt{ROT}}$	2.1	0.000	95.2	1.140	0.009	92.2	1.170	0.115	86.6	1.43
$\hat{\kappa}_{\mathtt{DPI}}$	3.0	0.005	94.5	1.600	0.016	91.8	1.440	0.078	87.5	2.17
$j = 3$										
	1.0	0.001	95.4	0.948	0.000	92.6	1.140	0.133	87.1	1.18
	2.0	0.001	95.2	0.849	0.004	93.2	0.954	0.076	89.4	1.31
$\kappa_{\mathtt{IMSE}}$	3.0	0.000	94.5	1.980	0.015	92.7	1.730	0.043	91.7	2.10
$\hat{\kappa}_{\mathtt{ROT}}$	2.1	0.003	95.2	0.966	0.006	93.2	1.040	0.075	89.7	1.38
$\hat{\kappa}_{\mathtt{DPI}}$	3.0	0.000	94.5	2.000	0.018	92.8	2.160	0.043	91.7	2.13

Table SA-14

Pointwise Results, Model 7, Spline, Quantile-spaced, $n = 1000$, 5000 Replications

	κ	$\mathbf{x} = (0.5, 0.5, 0.5)$			$\mathbf{x} = (0.1, 0.1, 0.5)$			$\mathbf{x} = (0.1, 0.1, 0.1)$		
		RMSE	CR	AL	RMSE	CR	AL	RMSE	CR	AL
$j = 0$										
	1.0	0.001	95.1	0.895	0.004	93.6	0.956	0.003	92.9	0.989
$\kappa_{\texttt{IMSE}}$	2.0	0.004	94.5	0.367	0.004	92.3	0.790	0.003	89.6	1.150
	3.0	0.001	93.7	1.990	0.014	89.4	1.400	0.005	85.6	1.210
$\hat{\kappa}_{\texttt{ROT}}$	2.1	0.003	94.6	0.486	0.004	92.1	0.833	0.003	89.2	1.150
$\hat{\kappa}_{\texttt{DPI}}$	3.0	0.001	93.7	1.990	0.014	89.4	1.400	0.003	85.7	1.220
$j = 1$										
	1.0	0.003	94.6	0.437	0.005	91.9	0.785	0.001	88.9	1.060
$\kappa_{\texttt{IMSE}}$	2.0	0.000	95.4	1.080	0.010	91.4	1.100	0.006	86.9	1.240
	3.0	0.001	96.3	0.869	0.016	91.8	1.310	0.033	85.8	2.030
$\hat{\kappa}_{\texttt{ROT}}$	2.1	0.000	95.4	1.070	0.010	91.5	1.120	0.000	86.6	1.300
$\hat{\kappa}_{\texttt{DPI}}$	3.0	0.001	96.3	0.870	0.016	91.7	1.310	0.012	85.8	2.080
$j = 2$										
	1.0	0.003	95.1	0.541	0.005	92.7	0.837	0.001	91.0	1.120
$\kappa_{\texttt{IMSE}}$	2.0	0.001	95.5	1.090	0.010	91.9	1.130	0.003	88.9	1.340
	3.0	0.005	94.4	1.600	0.016	91.9	1.440	0.034	87.7	2.160
$\hat{\kappa}_{\texttt{ROT}}$	2.1	0.001	95.3	1.120	0.010	92.0	1.160	0.002	88.7	1.400
$\hat{\kappa}_{\texttt{DPI}}$	3.0	0.005	94.5	1.600	0.016	91.9	1.440	0.012	87.7	2.220
$j = 3$										
	1.0	0.004	95.4	0.943	0.002	93.3	1.090	0.001	92.1	1.070
$\kappa_{\texttt{IMSE}}$	2.0	0.003	95.2	0.848	0.003	93.4	0.948	0.000	90.6	1.270
	3.0	0.000	94.4	1.980	0.015	92.8	1.730	0.012	91.9	2.090
$\hat{\kappa}_{\texttt{ROT}}$	2.1	0.003	95.2	0.931	0.005	93.4	1.000	0.002	90.6	1.320
$\hat{\kappa}_{\texttt{DPI}}$	3.0	0.000	94.4	1.980	0.014	92.8	1.750	0.014	91.9	2.790

Table SA-15

Tuning parameters, Spline, Evenly-spaced, $n = 1000$, 5000 Replications

	Min.	1st qu.	Median	Mean	3rd qu.	Max.	Std. dev.
Model 1, $\kappa_{\texttt{IMSE}} = 3.4$							
$\hat{\kappa}_{\texttt{ROT}}$	3	4	4	4.07	4	5	0.45
$\hat{\kappa}_{\texttt{DPI}}$	3	4	5	4.67	5	7	0.72
Model 2, $\kappa_{\texttt{IMSE}} = 7.2$							
$\hat{\kappa}_{\texttt{ROT}}$	5	6	6	6.07	6	7	0.26
$\hat{\kappa}_{\texttt{DPI}}$	6	7	8	7.79	8	10	0.57
Model 3, $\kappa_{\texttt{IMSE}} = 11.9$							
$\hat{\kappa}_{\texttt{ROT}}$	3	3	4	3.68	4	4	0.47
$\hat{\kappa}_{\texttt{DPI}}$	5	7	8	7.65	8	9	1.09
Model 4, $\kappa_{\texttt{IMSE}} = 4.4$							
$\hat{\kappa}_{\texttt{ROT}}$	2	3	3	2.94	3	3	0.24
$\hat{\kappa}_{\texttt{DPI}}$	4	4	4	4.00	4	5	0.02
Model 5, $\kappa_{\texttt{IMSE}} = 4.2$							
$\hat{\kappa}_{\texttt{ROT}}$	2	3	3	3.00	3	4	0.06
$\hat{\kappa}_{\texttt{DPI}}$	3	4	4	3.93	4	4	0.26
Model 6, $\kappa_{\texttt{IMSE}} = 3.3$							
$\hat{\kappa}_{\texttt{ROT}}$	2	2	2	2.10	2	3	0.31
$\hat{\kappa}_{\texttt{DPI}}$	3	3	3	3.00	3	4	0.04
Model 7, $\kappa_{\texttt{IMSE}} = 2.3$							
$\hat{\kappa}_{\texttt{ROT}}$	2	2	2	2.07	2	3	0.26
$\hat{\kappa}_{\texttt{DPI}}$	2	3	3	3.00	3	4	0.05

TABLE SA-16
Tuning parameters, Spline, Quantile-spaced, $n = 1000$, 5000 Replications

	Min.	1st qu.	Median	Mean	3rd qu.	Max.	Std. dev.
Model 1, $\kappa_{\mathtt{IMSE}} = 3.4$							
$\hat{\kappa}_{\mathtt{ROT}}$	3	4	4	4.07	4	5	0.45
$\hat{\kappa}_{\mathtt{DPI}}$	3	4	5	4.69	5	7	0.72
Model 2, $\kappa_{\mathtt{IMSE}} = 7.2$							
$\hat{\kappa}_{\mathtt{ROT}}$	5	6	6	6.07	6	7	0.26
$\hat{\kappa}_{\mathtt{DPI}}$	6	7	8	7.84	8	10	0.58
Model 3, $\kappa_{\mathtt{IMSE}} = 11.9$							
$\hat{\kappa}_{\mathtt{ROT}}$	3	3	4	3.68	4	4	0.47
$\hat{\kappa}_{\mathtt{DPI}}$	5	7	8	7.66	8	9	1.05
Model 4, $\kappa_{\mathtt{IMSE}} = 4.4$							
$\hat{\kappa}_{\mathtt{ROT}}$	2	3	3	2.94	3	3	0.24
$\hat{\kappa}_{\mathtt{DPI}}$	4	4	4	4.00	4	5	0.03
Model 5, $\kappa_{\mathtt{IMSE}} = 4.2$							
$\hat{\kappa}_{\mathtt{ROT}}$	2	3	3	3.00	3	4	0.06
$\hat{\kappa}_{\mathtt{DPI}}$	3	4	4	3.93	4	4	0.26
Model 6, $\kappa_{\mathtt{IMSE}} = 3.3$							
$\hat{\kappa}_{\mathtt{ROT}}$	2	2	2	2.10	2	3	0.31
$\hat{\kappa}_{\mathtt{DPI}}$	3	3	3	3.00	3	4	0.04
Model 7, $\kappa_{\mathtt{IMSE}} = 2.3$							
$\hat{\kappa}_{\mathtt{ROT}}$	2	2	2	2.07	2	3	0.26
$\hat{\kappa}_{\mathtt{DPI}}$	2	3	3	3.00	3	4	0.04

Table SA-17

Uniform Results, Model 1, Spline, $n = 1000$, 5000 Replications.

	Evenly-spaced				Quantile-spaced			
	CP	ACE	AW	UCR	CP	ACE	AW	UCR
Plug-in, $j = 0$								
$\kappa_{\texttt{IMSE}}$	0.92	0.017	0.384	79.9	0.92	0.017	0.383	80.0
$\hat{\kappa}_{\texttt{ROT}}$	1.00	0.008	0.433	90.1	1.00	0.008	0.433	89.9
$\hat{\kappa}_{\texttt{DPI}}$	1.00	0.006	0.460	91.6	1.00	0.006	0.460	91.6
Plug-in, $j = 1$								
$\kappa_{\texttt{IMSE}}$	1.00	0.005	0.426	93.8	1.00	0.005	0.426	93.5
$\hat{\kappa}_{\texttt{ROT}}$	1.00	0.005	0.473	93.5	1.00	0.005	0.473	93.4
$\hat{\kappa}_{\texttt{DPI}}$	1.00	0.004	0.497	93.2	1.00	0.004	0.498	93.2
Plug-in, $j = 2$								
$\kappa_{\texttt{IMSE}}$	1.00	0.005	0.443	94.1	1.00	0.005	0.443	94.0
$\hat{\kappa}_{\texttt{ROT}}$	1.00	0.004	0.497	93.4	1.00	0.004	0.497	93.5
$\hat{\kappa}_{\texttt{DPI}}$	1.00	0.004	0.526	93.6	1.00	0.004	0.527	93.6
Plug-in, $j = 3$								
$\kappa_{\texttt{IMSE}}$	1.00	0.008	0.413	89.0	1.00	0.008	0.413	88.8
$\hat{\kappa}_{\texttt{ROT}}$	1.00	0.005	0.463	93.0	1.00	0.005	0.463	92.7
$\hat{\kappa}_{\texttt{DPI}}$	1.00	0.004	0.490	93.2	1.00	0.004	0.490	93.3
Bootstrap, $j = 0$								
$\kappa_{\texttt{IMSE}}$	0.92	0.017	0.382	79.7	0.92	0.017	0.382	79.6
$\hat{\kappa}_{\texttt{ROT}}$	1.00	0.008	0.431	89.7	1.00	0.008	0.430	89.6
$\hat{\kappa}_{\texttt{DPI}}$	1.00	0.006	0.457	91.1	1.00	0.006	0.456	91.0
Bootstrap, $j = 1$								
$\kappa_{\texttt{IMSE}}$	1.00	0.005	0.424	93.7	1.00	0.005	0.423	93.5
$\hat{\kappa}_{\texttt{ROT}}$	1.00	0.005	0.469	92.9	1.00	0.005	0.469	92.9
$\hat{\kappa}_{\texttt{DPI}}$	1.00	0.004	0.493	92.6	1.00	0.004	0.493	92.5
Bootstrap, $j = 2$								
$\kappa_{\texttt{IMSE}}$	1.00	0.005	0.440	94.0	1.00	0.005	0.440	93.8
$\hat{\kappa}_{\texttt{ROT}}$	1.00	0.005	0.494	93.3	1.00	0.005	0.493	93.3
$\hat{\kappa}_{\texttt{DPI}}$	1.00	0.004	0.522	93.0	1.00	0.004	0.522	93.2
Bootstrap, $j = 3$								
$\kappa_{\texttt{IMSE}}$	1.00	0.008	0.411	88.5	1.00	0.008	0.411	88.5
$\hat{\kappa}_{\texttt{ROT}}$	1.00	0.005	0.460	92.5	1.00	0.005	0.460	92.3
$\hat{\kappa}_{\texttt{DPI}}$	1.00	0.004	0.486	92.9	1.00	0.004	0.487	92.8

Table SA-18
Uniform Results, Model 2, Spline, $n = 1000$, 5000 Replications.

	Evenly-spaced				Quantile-spaced			
	CP	ACE	AW	UCR	CP	ACE	AW	UCR
Plug-in, $j = 0$								
$\kappa_{\texttt{IMSE}}$	0.95	0.013	0.556	68.8	0.95	0.014	0.554	68.7
$\hat{\kappa}_{\texttt{ROT}}$	0.92	0.016	0.520	64.9	0.89	0.020	0.518	60.0
$\hat{\kappa}_{\texttt{DPI}}$	0.97	0.007	0.585	81.8	0.98	0.008	0.585	81.0
Plug-in, $j = 1$								
$\kappa_{\texttt{IMSE}}$	1.00	0.004	0.588	91.9	1.00	0.004	0.588	90.8
$\hat{\kappa}_{\texttt{ROT}}$	1.00	0.008	0.553	82.1	0.98	0.008	0.553	81.5
$\hat{\kappa}_{\texttt{DPI}}$	1.00	0.004	0.616	90.4	1.00	0.004	0.617	90.4
Plug-in, $j = 2$								
$\kappa_{\texttt{IMSE}}$	1.00	0.003	0.632	92.4	1.00	0.003	0.631	92.2
$\hat{\kappa}_{\texttt{ROT}}$	1.00	0.005	0.592	89.4	1.00	0.005	0.591	89.6
$\hat{\kappa}_{\texttt{DPI}}$	1.00	0.003	0.664	92.3	1.00	0.003	0.665	92.3
Plug-in, $j = 3$								
$\kappa_{\texttt{IMSE}}$	1.00	0.006	0.588	84.6	1.00	0.006	0.586	84.6
$\hat{\kappa}_{\texttt{ROT}}$	0.99	0.007	0.551	79.9	1.00	0.008	0.550	78.9
$\hat{\kappa}_{\texttt{DPI}}$	1.00	0.004	0.617	89.9	1.00	0.004	0.618	89.0
Bootstrap, $j = 0$								
$\kappa_{\texttt{IMSE}}$	0.95	0.014	0.551	67.5	0.95	0.015	0.549	67.2
$\hat{\kappa}_{\texttt{ROT}}$	0.91	0.017	0.515	63.4	0.88	0.021	0.514	58.6
$\hat{\kappa}_{\texttt{DPI}}$	0.97	0.008	0.579	80.6	0.98	0.008	0.578	79.6
Bootstrap, $j = 1$								
$\kappa_{\texttt{IMSE}}$	1.00	0.004	0.581	91.2	1.00	0.005	0.581	89.9
$\hat{\kappa}_{\texttt{ROT}}$	0.99	0.009	0.548	80.7	0.98	0.009	0.547	80.5
$\hat{\kappa}_{\texttt{DPI}}$	1.00	0.005	0.608	89.5	1.00	0.004	0.610	89.3
Bootstrap, $j = 2$								
$\kappa_{\texttt{IMSE}}$	1.00	0.004	0.624	91.6	1.00	0.004	0.623	91.4
$\hat{\kappa}_{\texttt{ROT}}$	1.00	0.005	0.585	88.6	1.00	0.005	0.585	88.7
$\hat{\kappa}_{\texttt{DPI}}$	1.00	0.003	0.655	91.6	1.00	0.003	0.656	91.3
Bootstrap, $j = 3$								
$\kappa_{\texttt{IMSE}}$	1.00	0.007	0.581	83.4	1.00	0.007	0.580	83.6
$\hat{\kappa}_{\texttt{ROT}}$	0.99	0.008	0.546	79.1	1.00	0.009	0.544	77.9
$\hat{\kappa}_{\texttt{DPI}}$	1.00	0.004	0.609	88.8	1.00	0.004	0.610	88.2

TABLE SA-19
Uniform Results, Model 3, Spline, $n = 1000$, 5000 Replications.

	Evenly-spaced				Quantile-spaced			
	CP	ACE	AW	UCR	CP	ACE	AW	UCR
Plug-in, $j = 0$								
$\kappa_{\texttt{IMSE}}$	0.88	0.033	0.731	9.3	0.87	0.027	0.726	26.2
$\hat{\kappa}_{\texttt{ROT}}$	0.24	0.445	0.437	0.0	0.22	0.439	0.436	0.0
$\hat{\kappa}_{\texttt{DPI}}$	0.61	0.157	0.588	24.3	0.59	0.155	0.585	11.5
Plug-in, $j = 1$								
$\kappa_{\texttt{IMSE}}$	1.00	0.003	0.755	91.0	1.00	0.006	0.754	81.4
$\hat{\kappa}_{\texttt{ROT}}$	0.18	0.365	0.472	0.0	0.15	0.365	0.472	0.0
$\hat{\kappa}_{\texttt{DPI}}$	0.63	0.075	0.614	36.3	0.58	0.089	0.615	19.4
Plug-in, $j = 2$								
$\kappa_{\texttt{IMSE}}$	1.00	0.002	0.822	91.6	1.00	0.003	0.820	90.3
$\hat{\kappa}_{\texttt{ROT}}$	0.12	0.360	0.495	0.0	0.12	0.363	0.495	0.0
$\hat{\kappa}_{\texttt{DPI}}$	0.84	0.033	0.660	46.8	0.79	0.037	0.660	43.7
Plug-in, $j = 3$								
$\kappa_{\texttt{IMSE}}$	0.92	0.011	0.766	61.3	0.94	0.010	0.761	65.5
$\hat{\kappa}_{\texttt{ROT}}$	0.24	0.401	0.469	0.0	0.16	0.397	0.468	0.0
$\hat{\kappa}_{\texttt{DPI}}$	0.63	0.140	0.619	18.5	0.63	0.133	0.619	11.4
Bootstrap, $j = 0$								
$\kappa_{\texttt{IMSE}}$	0.88	0.035	0.719	7.7	0.85	0.029	0.713	24.2
$\hat{\kappa}_{\texttt{ROT}}$	0.24	0.446	0.435	0.0	0.22	0.440	0.434	0.0
$\hat{\kappa}_{\texttt{DPI}}$	0.61	0.159	0.582	23.8	0.57	0.157	0.579	11.2
Bootstrap, $j = 1$								
$\kappa_{\texttt{IMSE}}$	1.00	0.003	0.741	89.8	1.00	0.006	0.741	79.3
$\hat{\kappa}_{\texttt{ROT}}$	0.18	0.368	0.469	0.0	0.15	0.367	0.469	0.0
$\hat{\kappa}_{\texttt{DPI}}$	0.62	0.077	0.607	35.5	0.56	0.092	0.608	18.6
Bootstrap, $j = 2$								
$\kappa_{\texttt{IMSE}}$	1.00	0.003	0.806	90.2	1.00	0.003	0.804	88.9
$\hat{\kappa}_{\texttt{ROT}}$	0.12	0.363	0.492	0.0	0.12	0.365	0.492	0.0
$\hat{\kappa}_{\texttt{DPI}}$	0.82	0.035	0.651	45.4	0.77	0.039	0.651	42.4
Bootstrap, $j = 3$								
$\kappa_{\texttt{IMSE}}$	0.91	0.013	0.751	57.0	0.90	0.011	0.747	62.4
$\hat{\kappa}_{\texttt{ROT}}$	0.24	0.403	0.466	0.0	0.14	0.399	0.465	0.0
$\hat{\kappa}_{\texttt{DPI}}$	0.62	0.142	0.612	17.9	0.61	0.136	0.611	10.6

Table SA-20
Uniform Results, Model 4, Spline, $n = 1000$, 5000 Replications.

	Evenly-spaced				**Quantile-spaced**			
	CP	ACE	AW	UCR	CP	ACE	AW	UCR
Plug-in, $j = 0$								
$\kappa_{\texttt{IMSE}}$	1.00	0.007	1.085	58.3	0.98	0.008	1.096	52.3
$\hat{\kappa}_{\texttt{ROT}}$	0.52	0.055	0.890	5.0	0.53	0.059	0.895	7.6
$\hat{\kappa}_{\texttt{DPI}}$	1.00	0.007	1.085	58.3	0.98	0.008	1.096	52.4
Plug-in, $j = 1$								
$\kappa_{\texttt{IMSE}}$	1.00	0.002	1.257	86.5	1.00	0.002	1.253	87.0
$\hat{\kappa}_{\texttt{ROT}}$	0.96	0.011	1.028	44.1	0.95	0.012	1.028	44.4
$\hat{\kappa}_{\texttt{DPI}}$	1.00	0.002	1.257	86.5	1.00	0.002	1.253	87.0
Plug-in, $j = 2$								
$\kappa_{\texttt{IMSE}}$	1.00	0.002	1.356	87.9	1.00	0.002	1.357	88.0
$\hat{\kappa}_{\texttt{ROT}}$	1.00	0.004	1.092	76.3	1.00	0.004	1.094	75.2
$\hat{\kappa}_{\texttt{DPI}}$	1.00	0.002	1.356	87.9	1.00	0.002	1.357	88.1
Plug-in, $j = 3$								
$\kappa_{\texttt{IMSE}}$	1.00	0.003	1.131	80.4	1.00	0.003	1.157	78.8
$\hat{\kappa}_{\texttt{ROT}}$	0.82	0.023	0.972	58.7	0.79	0.025	0.977	51.0
$\hat{\kappa}_{\texttt{DPI}}$	1.00	0.003	1.131	80.5	1.00	0.003	1.157	78.8
Bootstrap, $j = 0$								
$\kappa_{\texttt{IMSE}}$	0.98	0.009	1.041	48.8	0.95	0.011	1.052	43.1
$\hat{\kappa}_{\texttt{ROT}}$	0.50	0.059	0.866	3.4	0.50	0.064	0.870	5.9
$\hat{\kappa}_{\texttt{DPI}}$	0.98	0.009	1.041	48.8	0.95	0.011	1.053	43.2
Bootstrap, $j = 1$								
$\kappa_{\texttt{IMSE}}$	1.00	0.003	1.213	82.1	1.00	0.003	1.209	82.6
$\hat{\kappa}_{\texttt{ROT}}$	0.95	0.014	1.004	38.1	0.94	0.014	1.003	38.8
$\hat{\kappa}_{\texttt{DPI}}$	1.00	0.003	1.213	82.1	1.00	0.003	1.210	82.5
Bootstrap, $j = 2$								
$\kappa_{\texttt{IMSE}}$	1.00	0.002	1.305	83.0	1.00	0.002	1.305	83.9
$\hat{\kappa}_{\texttt{ROT}}$	1.00	0.005	1.061	71.1	1.00	0.005	1.063	69.5
$\hat{\kappa}_{\texttt{DPI}}$	1.00	0.002	1.305	83.0	1.00	0.002	1.305	83.8
Bootstrap, $j = 3$								
$\kappa_{\texttt{IMSE}}$	1.00	0.004	1.092	74.7	1.00	0.004	1.117	72.3
$\hat{\kappa}_{\texttt{ROT}}$	0.79	0.025	0.948	53.7	0.79	0.028	0.953	46.1
$\hat{\kappa}_{\texttt{DPI}}$	1.00	0.004	1.092	74.8	1.00	0.004	1.117	72.3

Table SA-21

Uniform Results, Model 5, Spline, $n = 1000$, 5000 Replications.

	Evenly-spaced				**Quantile-spaced**			
	CP	ACE	AW	UCR	CP	ACE	AW	UCR
Plug-in, $j = 0$								
$\kappa_{\texttt{IMSE}}$	1	0.003	1.081	80.9	1	0.003	1.092	80.6
$\hat{\kappa}_{\texttt{ROT}}$	1	0.006	0.896	69.5	1	0.006	0.900	66.9
$\hat{\kappa}_{\texttt{DPI}}$	1	0.003	1.068	79.8	1	0.003	1.078	79.0
Plug-in, $j = 1$								
$\kappa_{\texttt{IMSE}}$	1	0.001	1.256	90.1	1	0.001	1.251	90.2
$\hat{\kappa}_{\texttt{ROT}}$	1	0.001	1.033	91.4	1	0.001	1.033	91.3
$\hat{\kappa}_{\texttt{DPI}}$	1	0.001	1.240	90.2	1	0.001	1.235	90.4
Plug-in, $j = 2$								
$\kappa_{\texttt{IMSE}}$	1	0.001	1.355	90.5	1	0.001	1.356	90.2
$\hat{\kappa}_{\texttt{ROT}}$	1	0.001	1.101	91.1	1	0.001	1.103	90.9
$\hat{\kappa}_{\texttt{DPI}}$	1	0.001	1.336	90.5	1	0.001	1.337	90.3
Plug-in, $j = 3$								
$\kappa_{\texttt{IMSE}}$	1	0.001	1.127	89.4	1	0.001	1.153	89.3
$\hat{\kappa}_{\texttt{ROT}}$	1	0.002	0.976	85.1	1	0.002	0.981	84.3
$\hat{\kappa}_{\texttt{DPI}}$	1	0.001	1.116	89.3	1	0.001	1.140	89.3
Bootstrap, $j = 0$								
$\kappa_{\texttt{IMSE}}$	1	0.004	1.038	75.1	1	0.004	1.049	74.5
$\hat{\kappa}_{\texttt{ROT}}$	1	0.007	0.871	63.4	1	0.008	0.875	61.2
$\hat{\kappa}_{\texttt{DPI}}$	1	0.004	1.025	73.9	1	0.004	1.036	72.9
Bootstrap, $j = 1$								
$\kappa_{\texttt{IMSE}}$	1	0.002	1.212	86.9	1	0.002	1.208	86.8
$\hat{\kappa}_{\texttt{ROT}}$	1	0.002	1.009	89.0	1	0.002	1.008	88.9
$\hat{\kappa}_{\texttt{DPI}}$	1	0.002	1.197	87.0	1	0.002	1.194	87.1
Bootstrap, $j = 2$								
$\kappa_{\texttt{IMSE}}$	1	0.002	1.304	86.9	1	0.002	1.304	86.3
$\hat{\kappa}_{\texttt{ROT}}$	1	0.002	1.070	88.3	1	0.002	1.071	88.3
$\hat{\kappa}_{\texttt{DPI}}$	1	0.002	1.287	87.1	1	0.002	1.287	86.5
Bootstrap, $j = 3$								
$\kappa_{\texttt{IMSE}}$	1	0.002	1.089	85.7	1	0.002	1.114	85.6
$\hat{\kappa}_{\texttt{ROT}}$	1	0.003	0.951	81.8	1	0.003	0.956	80.7
$\hat{\kappa}_{\texttt{DPI}}$	1	0.002	1.078	85.6	1	0.002	1.102	85.7

TABLE SA-22
Uniform Results, Model 6, Spline, $n = 1000$, 5000 Replications.

	Evenly-spaced				Quantile-spaced			
	CP	ACE	AW	UCR	CP	ACE	AW	UCR
Plug-in, $j = 0$								
$\kappa_{\texttt{IMSE}}$	1.00	0.003	1.723	75.4	1.00	0.003	1.745	75.0
$\hat{\kappa}_{\texttt{ROT}}$	0.95	0.010	1.227	38.8	0.95	0.010	1.236	39.3
$\hat{\kappa}_{\texttt{DPI}}$	1.00	0.003	1.726	75.3	1.00	0.003	1.747	75.0
Plug-in, $j = 1$								
$\kappa_{\texttt{IMSE}}$	1.00	0.002	2.715	81.1	1.00	0.002	2.698	81.4
$\hat{\kappa}_{\texttt{ROT}}$	1.00	0.002	1.865	87.1	1.00	0.002	1.861	87.4
$\hat{\kappa}_{\texttt{DPI}}$	1.00	0.002	2.716	81.1	1.00	0.002	2.699	81.4
Plug-in, $j = 2$								
$\kappa_{\texttt{IMSE}}$	1.00	0.002	2.860	84.1	1.00	0.002	2.851	84.1
$\hat{\kappa}_{\texttt{ROT}}$	1.00	0.001	1.939	87.9	1.00	0.001	1.937	87.9
$\hat{\kappa}_{\texttt{DPI}}$	1.00	0.002	2.861	84.1	1.00	0.002	2.853	84.0
Plug-in, $j = 3$								
$\kappa_{\texttt{IMSE}}$	1.00	0.001	2.172	89.3	1.00	0.001	2.225	89.1
$\hat{\kappa}_{\texttt{ROT}}$	1.00	0.002	1.571	84.3	1.00	0.002	1.592	84.5
$\hat{\kappa}_{\texttt{DPI}}$	1.00	0.001	2.579	89.3	1.00	0.001	2.387	89.1
Bootstrap, $j = 0$								
$\kappa_{\texttt{IMSE}}$	1.00	0.005	1.618	63.9	1.00	0.005	1.639	64.4
$\hat{\kappa}_{\texttt{ROT}}$	0.94	0.013	1.175	29.6	0.94	0.013	1.183	30.3
$\hat{\kappa}_{\texttt{DPI}}$	1.00	0.005	1.621	63.8	1.00	0.005	1.641	64.4
Bootstrap, $j = 1$								
$\kappa_{\texttt{IMSE}}$	1.00	0.004	2.503	66.6	1.00	0.004	2.489	67.6
$\hat{\kappa}_{\texttt{ROT}}$	1.00	0.003	1.772	80.3	1.00	0.002	1.768	80.9
$\hat{\kappa}_{\texttt{DPI}}$	1.00	0.004	2.504	66.6	1.00	0.004	2.490	67.6
Bootstrap, $j = 2$								
$\kappa_{\texttt{IMSE}}$	1.00	0.003	2.645	72.1	1.00	0.003	2.637	72.2
$\hat{\kappa}_{\texttt{ROT}}$	1.00	0.002	1.836	81.8	1.00	0.002	1.835	81.4
$\hat{\kappa}_{\texttt{DPI}}$	1.00	0.003	2.647	72.1	1.00	0.003	2.639	72.2
Bootstrap, $j = 3$								
$\kappa_{\texttt{IMSE}}$	1.00	0.002	2.042	81.9	1.00	0.002	2.092	82.0
$\hat{\kappa}_{\texttt{ROT}}$	1.00	0.003	1.504	77.8	1.00	0.003	1.525	78.0
$\hat{\kappa}_{\texttt{DPI}}$	1.00	0.002	2.427	81.9	1.00	0.002	2.242	82.0

TABLE SA-23
Uniform Results, Model 7, Spline, $n = 1000$, 5000 Replications.

	Evenly-spaced				Quantile-spaced			
	CP	ACE	AW	UCR	CP	ACE	AW	UCR
Plug-in, $j = 0$								
$\kappa_{\texttt{IMSE}}$	0.99	0.004	1.153	69.2	0.99	0.004	1.160	69.7
$\hat{\kappa}_{\texttt{ROT}}$	0.99	0.004	1.195	70.1	0.99	0.004	1.203	70.5
$\hat{\kappa}_{\texttt{DPI}}$	1.00	0.003	1.721	78.6	1.00	0.003	1.741	78.9
Plug-in, $j = 1$								
$\kappa_{\texttt{IMSE}}$	1.00	0.001	1.766	88.6	1.00	0.001	1.763	88.5
$\hat{\kappa}_{\texttt{ROT}}$	1.00	0.001	1.836	88.4	1.00	0.001	1.832	88.3
$\hat{\kappa}_{\texttt{DPI}}$	1.00	0.002	2.714	81.2	1.00	0.002	2.697	81.1
Plug-in, $j = 2$								
$\kappa_{\texttt{IMSE}}$	1.00	0.001	1.831	88.6	1.00	0.001	1.830	88.6
$\hat{\kappa}_{\texttt{ROT}}$	1.00	0.001	1.907	88.4	1.00	0.001	1.906	88.4
$\hat{\kappa}_{\texttt{DPI}}$	1.00	0.002	2.860	84.1	1.00	0.002	2.851	84.1
Plug-in, $j = 3$								
$\kappa_{\texttt{IMSE}}$	1.00	0.001	1.491	88.8	1.00	0.001	1.508	89.4
$\hat{\kappa}_{\texttt{ROT}}$	1.00	0.001	1.541	89.2	1.00	0.001	1.560	89.6
$\hat{\kappa}_{\texttt{DPI}}$	1.00	0.001	2.199	89.7	1.00	0.001	2.227	89.7
Bootstrap, $j = 0$								
$\kappa_{\texttt{IMSE}}$	0.99	0.005	1.108	62.0	0.99	0.005	1.115	62.8
$\hat{\kappa}_{\texttt{ROT}}$	0.99	0.005	1.145	62.7	0.99	0.005	1.153	63.5
$\hat{\kappa}_{\texttt{DPI}}$	0.99	0.004	1.615	68.4	0.99	0.004	1.635	68.2
Bootstrap, $j = 1$								
$\kappa_{\texttt{IMSE}}$	1.00	0.002	1.685	82.8	1.00	0.002	1.683	82.9
$\hat{\kappa}_{\texttt{ROT}}$	1.00	0.002	1.746	81.9	1.00	0.002	1.743	82.1
$\hat{\kappa}_{\texttt{DPI}}$	1.00	0.004	2.503	66.9	1.00	0.004	2.488	67.6
Bootstrap, $j = 2$								
$\kappa_{\texttt{IMSE}}$	1.00	0.002	1.741	83.0	1.00	0.002	1.741	83.0
$\hat{\kappa}_{\texttt{ROT}}$	1.00	0.002	1.808	82.3	1.00	0.002	1.807	82.1
$\hat{\kappa}_{\texttt{DPI}}$	1.00	0.003	2.646	72.1	1.00	0.003	2.638	72.1
Bootstrap, $j = 3$								
$\kappa_{\texttt{IMSE}}$	1.00	0.002	1.433	84.6	1.00	0.002	1.449	85.0
$\hat{\kappa}_{\texttt{ROT}}$	1.00	0.002	1.478	84.8	1.00	0.002	1.496	85.2
$\hat{\kappa}_{\texttt{DPI}}$	1.00	0.002	2.066	82.9	1.00	0.002	2.093	82.9

Table SA-24
Pointwise Results, Model 1, Wavelet, $n = 1000$, 5000 Replications

	κ	$x=0.2$ RMSE	CR	AL	$x=0.5$ RMSE	CR	AL	$x=0.8$ RMSE	CR	AL
$j=0$										
$s_{\mathtt{IMSE}}$	3.0	0.003	94.3	0.339	0.002	94.2	0.476	0.022	92.7	0.203
	4.0	0.002	94.3	0.365	0.003	94.1	0.662	0.002	94.3	0.592
	5.0	0.001	94.3	0.563	0.002	93.7	0.911	0.001	94.3	0.698
	6.0	0.003	92.8	1.140	0.004	92.5	1.240	0.001	94.4	0.742
$\hat{s}_{\mathtt{ROT}}$	2.0	0.003	94.3	0.339	0.002	94.2	0.476	0.022	92.7	0.203
$\hat{s}_{\mathtt{DPI}}$	2.8	0.003	94.3	0.339	0.002	94.2	0.476	0.022	92.7	0.203
$j=1$										
$s_{\mathtt{IMSE}}$	3.0	0.002	95.0	0.370	0.036	93.6	0.449	0.034	90.5	0.230
	4.0	0.002	94.0	0.534	0.001	94.1	0.664	0.002	94.6	0.312
	5.0	0.003	93.8	0.595	0.003	93.9	0.933	0.001	94.2	0.569
	6.0	0.001	93.1	0.938	0.003	92.7	1.300	0.001	93.2	1.000
$\hat{s}_{\mathtt{ROT}}$	2.0	0.002	95.0	0.370	0.036	93.6	0.449	0.034	90.5	0.230
$\hat{s}_{\mathtt{DPI}}$	2.8	0.002	95.0	0.370	0.036	93.6	0.449	0.034	90.5	0.230
$j=2$										
$s_{\mathtt{IMSE}}$	3.0	0.002	94.5	0.393	0.009	94.2	0.523	0.031	91.4	0.237
	4.0	0.002	94.2	0.542	0.003	94.0	0.733	0.000	94.3	0.581
	5.0	0.002	94.0	0.625	0.001	93.8	1.030	0.000	94.2	0.635
	6.0	0.002	92.5	1.060	0.005	92.8	1.430	0.001	93.4	1.030
$\hat{s}_{\mathtt{ROT}}$	2.0	0.002	94.5	0.393	0.009	94.2	0.523	0.031	91.4	0.237
$\hat{s}_{\mathtt{DPI}}$	2.8	0.002	94.5	0.393	0.009	94.2	0.523	0.031	91.4	0.237

Table SA-25
Pointwise Results, Model 2, Wavelet, $n = 1000$, 5000 Replications

	κ	$x=0.2$ RMSE	$x=0.2$ CR	$x=0.2$ AL	$x=0.5$ RMSE	$x=0.5$ CR	$x=0.5$ AL	$x=0.8$ RMSE	$x=0.8$ CR	$x=0.8$ AL
$j=0$										
$s_{\mathtt{IMSE}}$	3	0.029	93.0	0.340	0.109	84.8	0.479	0.038	88.6	0.203
	4	0.008	94.3	0.365	0.012	93.9	0.662	0.003	94.4	0.592
	5	0.001	94.3	0.563	0.003	93.7	0.911	0.001	94.3	0.698
	6	0.002	92.8	1.140	0.004	92.5	1.240	0.001	94.4	0.742
$\hat{s}_{\mathtt{ROT}}$	3	0.029	93.0	0.340	0.109	84.8	0.479	0.038	88.6	0.203
$\hat{s}_{\mathtt{DPI}}$	3	0.029	93.1	0.340	0.106	85.0	0.484	0.037	88.6	0.213
$j=1$										
$s_{\mathtt{IMSE}}$	3	0.030	93.7	0.370	0.250	42.6	0.454	0.116	50.2	0.231
	4	0.005	94.1	0.534	0.039	93.3	0.664	0.004	94.8	0.312
	5	0.002	93.9	0.595	0.003	93.8	0.933	0.001	94.2	0.569
	6	0.001	93.1	0.938	0.002	92.7	1.300	0.001	93.2	1.000
$\hat{s}_{\mathtt{ROT}}$	3	0.030	93.7	0.370	0.250	42.6	0.454	0.116	50.2	0.231
$\hat{s}_{\mathtt{DPI}}$	3	0.027	93.6	0.374	0.245	43.8	0.460	0.113	51.8	0.234
$j=2$										
$s_{\mathtt{IMSE}}$	3	0.037	92.9	0.394	0.100	87.8	0.524	0.087	70.2	0.237
	4	0.004	94.2	0.542	0.022	94.1	0.734	0.000	94.3	0.581
	5	0.001	93.9	0.625	0.003	93.8	1.030	0.000	94.2	0.635
	6	0.002	92.5	1.060	0.004	92.8	1.430	0.001	93.4	1.030
$\hat{s}_{\mathtt{ROT}}$	3	0.037	92.9	0.394	0.100	87.8	0.524	0.087	70.2	0.237
$\hat{s}_{\mathtt{DPI}}$	3	0.035	92.7	0.398	0.098	87.9	0.529	0.084	71.0	0.246

Table SA-26

Pointwise Results, Model 3, Wavelet, $n = 1000$, 5000 Replications

		$x = 0.2$			$x = 0.5$			$x = 0.8$		
	κ	RMSE	CR	AL	RMSE	CR	AL	RMSE	CR	AL
$j = 0$										
	3.0	0.002	94.3	0.339	0.056	86.2	0.497	0.023	92.3	0.204
$s_{\mathtt{IMSE}}$	4.0	0.001	94.3	0.365	0.218	73.6	0.669	0.000	94.3	0.592
	5.0	0.002	94.3	0.563	0.109	90.8	0.914	0.001	94.3	0.698
	6.0	0.003	92.8	1.140	0.026	92.4	1.240	0.001	94.4	0.742
$\hat{s}_{\mathtt{ROT}}$	2.0	0.002	94.3	0.339	0.056	86.2	0.497	0.023	92.3	0.204
$\hat{s}_{\mathtt{DPI}}$	3.1	0.002	94.3	0.343	0.017	82.0	0.521	0.023	92.5	0.260
$j = 1$										
	3.0	0.000	95.1	0.370	0.009	94.2	0.453	0.153	26.5	0.232
$s_{\mathtt{IMSE}}$	4.0	0.002	94.1	0.534	0.196	77.3	0.668	0.001	94.6	0.312
	5.0	0.003	93.8	0.595	0.036	93.6	0.933	0.001	94.2	0.569
	6.0	0.001	93.1	0.938	0.001	92.7	1.300	0.001	93.2	1.000
$\hat{s}_{\mathtt{ROT}}$	2.0	0.000	95.1	0.370	0.009	94.2	0.453	0.153	26.5	0.232
$\hat{s}_{\mathtt{DPI}}$	3.1	0.002	95.0	0.393	0.021	90.1	0.483	0.129	35.9	0.244
$j = 2$										
	3.0	0.006	94.5	0.393	0.115	85.5	0.525	0.135	39.8	0.238
$s_{\mathtt{IMSE}}$	4.0	0.002	94.2	0.542	0.217	77.5	0.739	0.001	94.3	0.581
	5.0	0.002	94.0	0.625	0.040	93.5	1.030	0.000	94.2	0.635
	6.0	0.002	92.5	1.060	0.002	92.8	1.430	0.001	93.4	1.030
$\hat{s}_{\mathtt{ROT}}$	2.0	0.006	94.5	0.393	0.115	85.5	0.525	0.135	39.8	0.238
$\hat{s}_{\mathtt{DPI}}$	3.1	0.004	94.7	0.414	0.129	84.3	0.555	0.113	47.2	0.288

Table SA-27
Pointwise Results, Model 4, Wavelet, $n = 1000$, 5000 Replications

	κ	$\mathbf{x} = (0.5, 0.5)$ RMSE	CR	AL	$\mathbf{x} = (0.1, 0.5)$ RMSE	CR	AL	$\mathbf{x} = (0.1, 0.1)$ RMSE	CR	AL
$j = 0$										
$s_{\texttt{IMSE}}$	3.0	0.252	89.9	1.88	0.167	90.3	1.79	0.051	90.6	1.72
	4.0	0.056	87.1	4.06	0.064	88.0	3.01	0.013	88.0	2.19
$\hat{s}_{\texttt{ROT}}$	1.9	0.252	89.9	1.88	0.167	90.3	1.79	0.051	90.6	1.72
$\hat{s}_{\texttt{DPI}}$	2.0	0.252	89.9	1.88	0.167	90.3	1.79	0.051	90.6	1.72
$j = 1$										
$s_{\texttt{IMSE}}$	3.0	0.202	90.5	1.66	0.148	91.3	1.31	0.015	92.8	1.05
	4.0	0.017	88.9	4.16	0.001	90.0	4.04	0.030	88.8	5.08
$\hat{s}_{\texttt{ROT}}$	1.9	0.202	90.5	1.66	0.148	91.3	1.31	0.015	92.8	1.05
$\hat{s}_{\texttt{DPI}}$	2.0	0.202	90.5	1.66	0.148	91.3	1.31	0.015	92.8	1.05
$j = 2$										
$s_{\texttt{IMSE}}$	3.0	0.020	93.5	2.17	0.066	92.4	1.78	0.005	92.0	1.46
	4.0	0.025	89.5	5.01	0.002	89.7	4.40	0.026	88.7	5.07
$\hat{s}_{\texttt{ROT}}$	1.9	0.020	93.5	2.17	0.066	92.4	1.78	0.005	92.0	1.46
$\hat{s}_{\texttt{DPI}}$	2.0	0.020	93.5	2.17	0.066	92.4	1.78	0.005	92.0	1.46

TABLE SA-28

Pointwise Results, Model 5, Wavelet, $n = 1000$, 5000 Replications

	κ	$\mathbf{x} = (0.5, 0.5)$ RMSE	CR	AL	$\mathbf{x} = (0.1, 0.5)$ RMSE	CR	AL	$\mathbf{x} = (0.1, 0.1)$ RMSE	CR	AL
$j = 0$										
$s_{\texttt{IMSE}}$	3	0.034	93.3	1.87	0.015	91.8	1.79	0.007	90.6	1.72
	4	0.021	87.0	4.06	0.011	88.2	3.01	0.015	88.1	2.19
$\hat{s}_{\texttt{ROT}}$	2	0.034	93.3	1.87	0.015	91.8	1.79	0.007	90.6	1.72
$\hat{s}_{\texttt{DPI}}$	2	0.034	93.3	1.87	0.015	91.8	1.79	0.007	90.6	1.72
$j = 1$										
$s_{\texttt{IMSE}}$	3	0.099	92.6	1.65	0.017	93.8	1.30	0.014	92.8	1.05
	4	0.018	88.9	4.16	0.001	90.0	4.04	0.029	88.9	5.08
$\hat{s}_{\texttt{ROT}}$	2	0.099	92.6	1.65	0.017	93.8	1.30	0.014	92.8	1.05
$\hat{s}_{\texttt{DPI}}$	2	0.099	92.6	1.65	0.017	93.8	1.30	0.014	92.8	1.05
$j = 2$										
$s_{\texttt{IMSE}}$	3	0.009	93.6	2.16	0.016	92.3	1.78	0.009	92.0	1.46
	4	0.026	89.5	5.01	0.002	89.7	4.40	0.025	88.7	5.07
$\hat{s}_{\texttt{ROT}}$	2	0.009	93.6	2.16	0.016	92.3	1.78	0.009	92.0	1.46
$\hat{s}_{\texttt{DPI}}$	2	0.009	93.6	2.16	0.016	92.3	1.78	0.009	92.0	1.46

Table SA-29
Tuning parameters, Wavelets, $n = 1000$, 5000 Replications

	Min.	1st qu.	Median	Mean	3rd qu.	Max.	Std. dev.
Model 1, $s_{\mathtt{IMSE}} = 2.4$							
$\hat{s}_{\mathtt{ROT}}$	2	2	2	2.00	2	3	0.05
$\hat{s}_{\mathtt{DPI}}$	2	3	3	2.81	3	3	0.39
Model 2, $s_{\mathtt{IMSE}} = 3.5$							
$\hat{s}_{\mathtt{ROT}}$	3	3	3	3.00	3	3	0.00
$\hat{s}_{\mathtt{DPI}}$	3	3	3	3.03	3	4	0.16
Model 3, $s_{\mathtt{IMSE}} = 4.2$							
$\hat{s}_{\mathtt{ROT}}$	2	2	2	2.00	2	2	0.00
$\hat{s}_{\mathtt{DPI}}$	3	3	3	3.15	3	4	0.35
Model 4, $s_{\mathtt{IMSE}} = 2.7$							
$\hat{s}_{\mathtt{ROT}}$	1	2	2	1.85	2	2	0.35
$\hat{s}_{\mathtt{DPI}}$	2	2	2	2.00	2	2	0.00
Model 5, $s_{\mathtt{IMSE}} = 2.6$							
$\hat{s}_{\mathtt{ROT}}$	1	2	2	1.99	2	2	0.09
$\hat{s}_{\mathtt{DPI}}$	2	2	2	2.00	2	2	0.00

Table SA-30
Uniform Results, Model 1, Wavelet, $n = 1000$, 5000 Replications.

	Evenly-spaced			
	CP	ACE	AW	UCR
Plug-in, $j = 0$				
$s_{\texttt{IMSE}}$	1	0.005	0.509	91.1
$\hat{s}_{\texttt{ROT}}$	1	0.005	0.509	91.1
$\hat{s}_{\texttt{DPI}}$	1	0.005	0.509	91.1
Plug-in, $j = 1$				
$s_{\texttt{IMSE}}$	1	0.006	0.504	89.9
$\hat{s}_{\texttt{ROT}}$	1	0.006	0.504	89.9
$\hat{s}_{\texttt{DPI}}$	1	0.006	0.504	89.9
Plug-in, $j = 2$				
$s_{\texttt{IMSE}}$	1	0.004	0.576	91.4
$\hat{s}_{\texttt{ROT}}$	1	0.004	0.576	91.4
$\hat{s}_{\texttt{DPI}}$	1	0.004	0.576	91.4
Bootstrap, $j = 0$				
$s_{\texttt{IMSE}}$	1	0.005	0.505	90.6
$\hat{s}_{\texttt{ROT}}$	1	0.005	0.505	90.6
$\hat{s}_{\texttt{DPI}}$	1	0.005	0.505	90.6
Bootstrap, $j = 1$				
$s_{\texttt{IMSE}}$	1	0.006	0.500	89.4
$\hat{s}_{\texttt{ROT}}$	1	0.006	0.500	89.4
$\hat{s}_{\texttt{DPI}}$	1	0.006	0.500	89.4
Bootstrap, $j = 2$				
$s_{\texttt{IMSE}}$	1	0.005	0.569	90.4
$\hat{s}_{\texttt{ROT}}$	1	0.005	0.569	90.4
$\hat{s}_{\texttt{DPI}}$	1	0.005	0.569	90.4

Table SA-31
Uniform Results, Model 2, Wavelet, $n = 1000$, 5000 Replications.

	Evenly-spaced			
	CP	ACE	AW	UCR
Plug-in, $j = 0$				
$s_{\mathtt{IMSE}}$	0.77	0.054	0.511	4.1
$\hat{s}_{\mathtt{ROT}}$	0.77	0.054	0.511	4.1
$\hat{s}_{\mathtt{DPI}}$	0.77	0.053	0.518	6.0
Plug-in, $j = 1$				
$s_{\mathtt{IMSE}}$	0.69	0.095	0.507	0.5
$\hat{s}_{\mathtt{ROT}}$	0.69	0.095	0.507	0.5
$\hat{s}_{\mathtt{DPI}}$	0.70	0.093	0.513	2.5
Plug-in, $j = 2$				
$s_{\mathtt{IMSE}}$	0.97	0.012	0.576	71.2
$\hat{s}_{\mathtt{ROT}}$	0.97	0.012	0.576	71.2
$\hat{s}_{\mathtt{DPI}}$	0.97	0.012	0.584	71.7
Bootstrap, $j = 0$				
$s_{\mathtt{IMSE}}$	0.77	0.056	0.507	3.4
$\hat{s}_{\mathtt{ROT}}$	0.77	0.056	0.507	3.4
$\hat{s}_{\mathtt{DPI}}$	0.77	0.054	0.513	5.3
Bootstrap, $j = 1$				
$s_{\mathtt{IMSE}}$	0.67	0.098	0.502	0.3
$\hat{s}_{\mathtt{ROT}}$	0.67	0.098	0.502	0.3
$\hat{s}_{\mathtt{DPI}}$	0.69	0.095	0.508	2.3
Bootstrap, $j = 2$				
$s_{\mathtt{IMSE}}$	0.96	0.013	0.569	69.3
$\hat{s}_{\mathtt{ROT}}$	0.96	0.013	0.569	69.3
$\hat{s}_{\mathtt{DPI}}$	0.96	0.013	0.576	69.7

Table SA-32
Uniform Results, Model 3, Wavelet, $n = 1000$, 5000 Replications.

	Evenly-spaced			
	CP	ACE	AW	UCR
Plug-in, $j = 0$				
$s_{\texttt{IMSE}}$	0.96	0.019	0.781	13.9
$\hat{s}_{\texttt{ROT}}$	0.72	0.175	0.517	0.0
$\hat{s}_{\texttt{DPI}}$	0.73	0.150	0.555	2.6
Plug-in, $j = 1$				
$s_{\texttt{IMSE}}$	0.99	0.006	0.777	70.4
$\hat{s}_{\texttt{ROT}}$	0.66	0.122	0.507	0.0
$\hat{s}_{\texttt{DPI}}$	0.67	0.102	0.546	8.9
Plug-in, $j = 2$				
$s_{\texttt{IMSE}}$	0.99	0.004	0.868	80.1
$\hat{s}_{\texttt{ROT}}$	0.70	0.052	0.577	11.0
$\hat{s}_{\texttt{DPI}}$	0.73	0.042	0.619	21.2
Bootstrap, $j = 0$				
$s_{\texttt{IMSE}}$	0.95	0.021	0.765	11.6
$\hat{s}_{\texttt{ROT}}$	0.71	0.176	0.513	0.0
$\hat{s}_{\texttt{DPI}}$	0.73	0.152	0.549	2.3
Bootstrap, $j = 1$				
$s_{\texttt{IMSE}}$	0.99	0.007	0.761	67.0
$\hat{s}_{\texttt{ROT}}$	0.65	0.125	0.502	0.0
$\hat{s}_{\texttt{DPI}}$	0.67	0.105	0.540	8.5
Bootstrap, $j = 2$				
$s_{\texttt{IMSE}}$	0.99	0.005	0.846	77.3
$\hat{s}_{\texttt{ROT}}$	0.69	0.055	0.570	9.8
$\hat{s}_{\texttt{DPI}}$	0.71	0.045	0.610	19.5

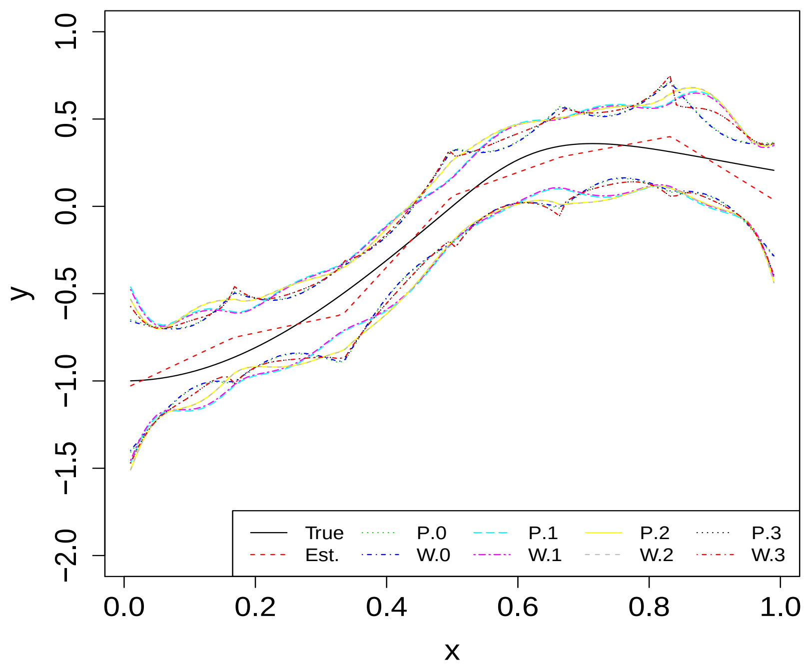
TABLE SA-33
Uniform Results, Model 4, Wavelet, $n = 1000$, 5000 Replications.

	Evenly-spaced			
	CP	ACE	AW	UCR
Plug-in, $j = 0$				
$s_{\texttt{IMSE}}$	0.93	0.032	1.499	2.0
$\hat{s}_{\texttt{ROT}}$	0.93	0.032	1.499	2.0
$\hat{s}_{\texttt{DPI}}$	0.93	0.032	1.499	2.0
Plug-in, $j = 1$				
$s_{\texttt{IMSE}}$	0.92	0.019	1.392	17.1
$\hat{s}_{\texttt{ROT}}$	0.92	0.019	1.392	17.1
$\hat{s}_{\texttt{DPI}}$	0.92	0.019	1.392	17.1
Plug-in, $j = 2$				
$s_{\texttt{IMSE}}$	0.98	0.006	1.726	60.4
$\hat{s}_{\texttt{ROT}}$	0.98	0.006	1.726	60.4
$\hat{s}_{\texttt{DPI}}$	0.98	0.006	1.726	60.4
Bootstrap, $j = 0$				
$s_{\texttt{IMSE}}$	0.91	0.040	1.394	0.6
$\hat{s}_{\texttt{ROT}}$	0.91	0.040	1.394	0.6
$\hat{s}_{\texttt{DPI}}$	0.91	0.040	1.394	0.6
Bootstrap, $j = 1$				
$s_{\texttt{IMSE}}$	0.91	0.024	1.320	9.8
$\hat{s}_{\texttt{ROT}}$	0.91	0.024	1.320	9.8
$\hat{s}_{\texttt{DPI}}$	0.91	0.024	1.320	9.8
Bootstrap, $j = 2$				
$s_{\texttt{IMSE}}$	0.96	0.009	1.612	44.3
$\hat{s}_{\texttt{ROT}}$	0.96	0.009	1.612	44.3
$\hat{s}_{\texttt{DPI}}$	0.96	0.009	1.612	44.3

TABLE SA-34
Uniform Results, Model 5, Wavelet, $n = 1000$, 5000 Replications.

	Evenly-spaced			
	CP	ACE	AW	UCR
Plug-in, $j = 0$				
$s_{\texttt{IMSE}}$	0.97	0.009	1.491	50.5
$\hat{s}_{\texttt{ROT}}$	0.97	0.009	1.491	50.5
$\hat{s}_{\texttt{DPI}}$	0.97	0.009	1.491	50.5
Plug-in, $j = 1$				
$s_{\texttt{IMSE}}$	0.99	0.005	1.387	69.2
$\hat{s}_{\texttt{ROT}}$	0.99	0.005	1.387	69.2
$\hat{s}_{\texttt{DPI}}$	0.99	0.005	1.387	69.2
Plug-in, $j = 2$				
$s_{\texttt{IMSE}}$	1.00	0.003	1.725	75.8
$\hat{s}_{\texttt{ROT}}$	1.00	0.003	1.725	75.8
$\hat{s}_{\texttt{DPI}}$	1.00	0.003	1.725	75.8
Bootstrap, $j = 0$				
$s_{\texttt{IMSE}}$	0.95	0.013	1.387	35.7
$\hat{s}_{\texttt{ROT}}$	0.95	0.013	1.387	35.7
$\hat{s}_{\texttt{DPI}}$	0.95	0.013	1.387	35.7
Bootstrap, $j = 1$				
$s_{\texttt{IMSE}}$	0.98	0.007	1.315	59.1
$\hat{s}_{\texttt{ROT}}$	0.98	0.007	1.315	59.1
$\hat{s}_{\texttt{DPI}}$	0.98	0.007	1.315	59.1
Bootstrap, $j = 2$				
$s_{\texttt{IMSE}}$	1.00	0.005	1.610	62.3
$\hat{s}_{\texttt{ROT}}$	1.00	0.005	1.610	62.3
$\hat{s}_{\texttt{DPI}}$	1.00	0.005	1.610	62.3

Fig SA-1: Confidence Bands, Model 1, Spline



Note: "P.j" denotes the confidence band based on plug-in approximation using the jth estimator, and "W.j" based on wild bootstrap. $j = 0$: the classical estimator; $j = 1$: higher-order bias correction; $j = 2$: least-squares bias correction; $j = 3$: plug-in bias correction.

References.

[1] Agarwal, G. G. and Studden, W. (1980). Asymptotic Integrated Mean Square Error Using Least Squares and Bias Minimizing Splines. *Annals of Statistics* **8** 1307–1325.

[2] Barrow, D. L. and Smith, P. W. (1978). Asymptotic Properties of Best $L_2[0,1]$ Approximation by Splines with Variable Knots. *Quarterly of Applied Mathematics* **36** 293–304.

[3] Belloni, A., Chernozhukov, V., Chetverikov, D. and Fernandez-Val, I. (2019). Conditional Quantile Processes based on Series or Many Regressors. *Journal of Econometrics*, forthcoming.

[4] Belloni, A., Chernozhukov, V., Chetverikov, D. and Kato, K. (2015). Some New Asymptotic Theory for Least Squares Series: Pointwise and Uniform Results. *Journal of Econometrics* **186** 345–366.

[5] Bhatia, R. (2013). *Matrix Analysis*. Springer.

[6] Buse, A. and Lim, L. (1977). Cubic splines as a special case of restricted least squares. *Journal of the American Statistical Association* **72** 64–68.

[7] Cattaneo, M. D. and Farrell, M. H. (2013). Optimal Convergence Rates, Bahadur Representation, and Asymptotic Normality of Partitioning Estimators. *Journal of Econometrics* **174** 127–143.

[8] Cattaneo, M. D., Farrell, M. H. and Feng, Y. (2019). `lspartition`: Partitioning-Based Least Squares Regression. *R Journal*, to appear.

[9] Chernozhukov, V., Chetverikov, D. and Kato, K. (2014). Anti-Concentration and Honest Adaptive Confidence Bands. *Annals of Statistics* **42** 1787–1818.

[10] Chernozhukov, V., Lee, S. and Rosen, A. M. (2013). Intersection bounds: estimation and inference. *Econometrica* **81** 667–737.

[11] Chui, C. K. (2016). *An introduction to wavelets*. Elsevier.

[12] Cohen, A., Daubechies, I. and Vial, P. (1993). Wavelets on the interval and fast wavelet transforms. *Applied and Computational Harmonic Analysis* **1** 54–81.

[13] Demko, S. (1977). Inverses of Band Matrices and Local Convergence of Spline Projections. *SIAM Journal on Numerical Analysis* **14** 616–619.

[14] Eggermont, P. P. B. and LaRiccia, V. N. (2009). *Maximum Penalized Likelihood Estimation: Regression*. Springer, New York, NY.

[15] Gazzah, H., Regalia, P. A. and Delmas, J.-P. (2001). Asymptotic eigenvalue distribution of block Toeplitz matrices and application to blind SIMO channel identification. *IEEE Transactions on Information Theory* **47** 1243–1251.

[16] Grenander, U. and Szegö, G. (2001). *Toeplitz forms and their applications* **321**. University of California Press.

[17] Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (2012). *Wavelets, Approximation, and Statistical Applications*. Springer.

[18] Hiriart-Urruty, J.-B. and Ye, D. (1995). Sensitivity analysis of all eigenvalues of a symmetric matrix. *Numerische Mathematik* **70** 45–72.

[19] Meyer, Y. (1995). *Wavelets and Operators*. Cambridge university press.

[20] Micula, G. and Micula, S. (1999). *Handbook of Splines*. Springer Netherlands, Dordrecht.

[21] Sakhanenko, A. (1985). Convergence Rate in the Invariance Principle for Non-identically Distributed Variables with Exponential Moments. *Advances in Probability Theory: Limit Theorems for Sums of Random Variables* 2–73.

[22] Sakhanenko, A. (1991). On the Accuracy of Normal Approximation in the Invariance Principle. *Siberian Advances in Mathematics* **1** 58–91.

[23] Schumaker, L. (2007). *Spline functions: basic theory*. Cambridge University Press.

[24] Sweldens, W. and Piessens, R. (1994). Asymptotic error expansion of wavelet approximations of smooth functions II. *Numerische Mathematik* **68** 377–401.

[25] Tropp, J. A. (2012). User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics* **12** 389–434.

[26] Tyrtyshnikov, E. E. (1997). *A brief introduction to numerical analysis*. Birkhuser Boston,

Boston, MA.

[27] Zhou, S., Shen, X. and Wolfe, D. (1998). Local Asymptotics for Regression Splines and Confidence Regions. *Annals of Statistics* **26** 1760–1782.

[28] Zhou, S. and Wolfe, D. A. (2000). On Derivative Estimation in Spline Regression. *Statistica Sinica* **10** 93–108.

Matias D. Cattaneo
Department of Operations Research
and Financial Engineering
Princeton University
Princeton, NJ 08544
E-mail: cattaneo@princeton.edu

Max H. Farrell
Booth School of Business
University of Chicago
Chicago, IL 60637
E-mail: max.farrell@chicagobooth.edu

Yingjie Feng
Department of Politics
Princeton University
Princeton, NJ 08544
E-mail: yingjief@princeton.edu