Estimation and Inference in Boundary Discontinuity Designs: Location-Based Methods Supplemental Appendix

Matias D. Cattaneo* Rocio Titiunik† Ruiqi (Rae) Yu‡

October 29, 2025

Abstract

This supplemental appendix presents more general theoretical results encompassing those reported in the paper, their theoretical proofs, and other technical results. In particular, it presents a new strong approximation result for residual-based empirical processes leveraging and extending ideas from Cattaneo and Yu [2025].

Keywords: regression discontinuity, treatment effects estimation, causal inference.

*Department of Operations Research and Financial Engineering, Princeton University.
†Department of Politics, Princeton University.
‡Department of Operations Research and Financial Engineering, Princeton University.

Contents

SA-1 Setup 2
- SA-1.1 Notation and Definitions 4
- SA-1.2 Mapping Between Paper and Supplement 5

SA-2 Preliminary Lemmas 5

SA-3 Boundary Average Treatment Effect Curve 6
- SA-3.1 Point Estimation and MSE Expansions 6
- SA-3.2 Distributional Approximation and Inference 8

SA-4 Weighted Boundary Average Treatment Effect 10

SA-5 Largest Boundary Average Treatment Effect 11

SA-6 Gaussian Strong Approximation 12
- SA-6.1 Definitions for Function Spaces 12
- SA-6.2 Residual-based Empirical Process 13

SA-7 Proofs 15
- SA-7.1 Proof of Lemma SA-1 15
- SA-7.2 Proof of Lemma SA-2 17
- SA-7.3 Proof of Lemma SA-3 18
- SA-7.4 Proof of Lemma SA-4 20
- SA-7.5 Proof of Lemma SA-5 22
- SA-7.6 Proof of Theorem SA-1 23
- SA-7.7 Proof of Theorem SA-2 24
- SA-7.8 Proof of Theorem SA-3 24
- SA-7.9 Proof of Theorem SA-4 25
- SA-7.10 Proof of Theorem SA-5 26
- SA-7.11 Proof of Theorem SA-6 29
- SA-7.12 Proof of Lemma SA-6 33
- SA-7.13 Proof of Lemma SA-7 33
- SA-7.14 Proof of Theorem SA-7 34
- SA-7.15 Proof of Theorem SA-8 34
- SA-7.16 Proof of Theorem SA-9 36
- SA-7.17 Proof of Theorem SA-10 36
- SA-7.18 Proof of Theorem SA-11 36

SA-1 Setup

This supplemental appendix considers a generalized version of the problem studied in the main paper: the location variable $\mathbf{X}_i$ is d-dimensional with $d \geq 1$ and support $\mathscr{X} \subseteq \mathbb{R}^d$, and the boundary region $\mathscr{B}$ is a low dimensional manifold with "effective dimension" $d-1$. The special case considered in the paper is $d = 2$, that is, $\mathbf{X}_i$ is bivariate and $\mathscr{B}$ is a one-dimensional (boundary) curve.

Assumption 1 from the paper is generalized to the following.

Assumption SA–1 (Data Generating Process)**.** *Let* $t \in \{0, 1\}$.

(i) $(Y_1(t), \mathbf{X}_1^\top)^\top, \ldots, (Y_n(t), \mathbf{X}_n^\top)^\top$ *are independent and identically distributed random vectors with* $\mathscr{X} = \prod_{l=1}^d [a_l, b_l]$ *for* $-\infty < a_l < b_l < \infty$ *for* $l = 1, \cdots, d$.

(ii) *The distribution of* $\mathbf{X}_i$ *has a Lebesgue density* $f_X(\mathbf{x})$ *that is continuous and bounded away from zero on* $\mathscr{X}$.

(iii) $\mu_t(\mathbf{x}) = \mathbb{E}[Y_i(t)|\mathbf{X}_i = \mathbf{x}]$ *is* $(p+1)$*-times continuously differentiable on* $\mathscr{X}$.

(iv) $\sigma_t^2(\mathbf{x}) = \mathbb{V}[Y_i(t)|\mathbf{X}_i = \mathbf{x}]$ *is bounded away from zero and continuous on* $\mathscr{X}$.

(v) $\sup_{\mathbf{x} \in \mathscr{X}} \mathbb{E}[|Y_i(t)|^{2+v}|\mathbf{X}_i = \mathbf{x}] < \infty$ *for some* $v \geq 2$.

We partition $\mathscr{X}$ into two areas, $\mathscr{A}_t \subset \mathbb{R}^d$ with $t \in \{0, 1\}$, which represent the control and treatment regions, respectively. That is, $\mathscr{X} = \mathscr{A}_0 \cup \mathscr{A}_1$, where $\mathscr{A}_0$ and $\mathscr{A}_1$ are two disjoint regions in $\mathbb{R}^d$. The observed outcome is $Y_i = \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_0)Y_i(0) + \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_1)Y_i(1)$. The "boundary" now becomes $\mathscr{B} = \mathtt{bd}(\mathscr{A}_0) \cap \mathtt{bd}(\mathscr{A}_1)$ denotes the boundary determined by the assignment regions $\mathscr{A}_t$, $t \in \{0, 1\}$, where $\mathtt{bd}(\mathscr{A}_t)$ denotes the topological boundary of $\mathscr{A}_t$. As in the paper, we assume that $\mathscr{B}$ belongs to $\mathscr{A}_1$, that is, $\mathscr{B} \subseteq \mathscr{A}_1$ and $\mathscr{B} \cup \mathscr{A}_0 = \emptyset$.

The multidimensional generalization of the three causal parameters studied in the paper are:

1. *Boundary average treatment effect curve* (BATEC):

$$\tau(\mathbf{x}) = \mathbb{E}[Y_i(1) - Y_i(0)|\mathbf{X}_i = \mathbf{x}], \qquad \mathbf{x} \in \mathscr{B} \subseteq \mathbb{R}^{d-1}.$$

2. *Weighted Boundary average treatment effect* (WBATE):

$$\tau_{\mathtt{WBATE}} = \frac{\int_{\mathscr{B}} \tau(\mathbf{x}) w(\mathbf{x}) d\mathfrak{H}^{d-1}(\mathbf{x})}{\int_{\mathscr{B}} w(\mathbf{x}) d\mathfrak{H}^{d-1}(\mathbf{x})}.$$

3. *Largest Boundary average treatment effect* (LBATE):

$$\tau(\mathbf{x}) = \sup_{\mathbf{x} \in \mathscr{B}} \tau(\mathbf{x}).$$

More generally, this supplemental appendix also considers the derivatives of the BATEC parameter:

$$\tau^{(\boldsymbol{\nu})}(\mathbf{x}) = \mu_1^{(\boldsymbol{\nu})}(\mathbf{x}) - \mu_0^{(\boldsymbol{\nu})}(\mathbf{x}), \qquad \mathbf{x} \in \mathscr{B},$$

where, using standard multi-index notation, $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_d)^\top \in \mathbb{N}_0$ with $|\boldsymbol{\nu}| = \nu_1 + \ldots + \nu_d \leq p$ and $\mu_t^{(\boldsymbol{\nu})}(\mathbf{x}) = \partial^{\nu_1} \cdots \partial^{\nu_d} \mu_t(\mathbf{x})$ for $t \in \{0, 1\}$.

The treatment effect estimator process along the boundary (submanifold) is

$$\Big(\widehat{\tau}^{(\boldsymbol{\nu})}(\mathbf{x}) = \widehat{\mu}_1^{(\boldsymbol{\nu})}(\mathbf{x}) - \widehat{\mu}_0^{(\boldsymbol{\nu})}(\mathbf{x}) : \mathbf{x} \in \mathscr{B}\Big),$$

where $\widehat{\mu}_t^{(\boldsymbol{\nu})}(\mathbf{x}) = \mathbf{e}_{1+\boldsymbol{\nu}}^\top \widehat{\boldsymbol{\beta}}_t(\mathbf{x})$ for $t \in \{0,1\}$ with

$$\widehat{\boldsymbol{\beta}}_t(\mathbf{x}) = \operatorname*{arg\,min}_{\boldsymbol{\beta} \in \mathbb{R}^{\mathfrak{p}_p+1}} \mathbb{E}_n\Big[\big(Y_i - \mathbf{r}_p(\mathbf{X}_i - \mathbf{x})^\top \boldsymbol{\beta}\big)^2 K_h(\mathbf{X}_i - \mathbf{x}) \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)\Big], \qquad \mathbf{x} \in \mathscr{B},$$

with $\mathfrak{p}_p = \frac{(d+p)!}{d!p!}$, $\mathbf{r}_p(\mathbf{u})$ denotes the pth order polynomial expansion of the d-variate vector $\mathbf{u} = (u_1, \cdots, u_d)^\top$, $K_h(\mathbf{u}) = K(u_1/h, \cdots, u_d/h)/h^d$ for a d-variate kernel function $K(\cdot)$ and a bandwidth parameter h.

We impose the following assumption on the d-variate kernel function and assignment boundary (submanifold) $\mathscr{B}$.

Assumption SA–2 (Kernel and Boundary). *Let $t \in \{0,1\}$.*

(i) *$\mathscr{B}$ is compact $(d-1)$-rectifiable, with $\mathfrak{H}^{d-1}(\mathscr{B})$ positive and finite.*

(ii) *$K : \mathbb{R}^d \to [0, \infty)$ is compact supported and Lipschitz continuous or $K(\mathbf{u}) = \mathbb{1}(\mathbf{u} \in [-1,1]^d)$.*

(iii) *There exists a set $U \subseteq \mathbb{R}^d$, such that $K(\mathbf{u}) \geq \kappa > 0$ for all $\mathbf{u} \in U$, $\lambda_{\min}(\int_U \mathbf{r}_p(\mathbf{z})\mathbf{r}_p(\mathbf{z})^\top d\mathbf{z}) > 0$, and $\liminf_{h \downarrow 0} \inf_{\mathbf{x} \in \mathscr{B}} \int_U K(\mathbf{u}) \mathbb{1}(\mathbf{x} + h\mathbf{u} \in \mathscr{A}_t) d\mathbf{u} \gtrsim 1$.*

Note that in case $d = 2$, if we assume $\mathscr{B}$ is a rectifiable curve, then Assumption SA–2 (i) holds.

Under the assumptions imposed,

$$\widehat{\boldsymbol{\beta}}_t(\mathbf{x}) = \mathbf{H}^{-1} \widehat{\boldsymbol{\Gamma}}_{t,\mathbf{x}}^{-1} \mathbb{E}_n\Big[\mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{x}}{h}\Big) K_h(\mathbf{X}_i - \mathbf{x}) Y_i \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)\Big],$$

where $\mathbf{H} = \operatorname{diag}((h^{|\mathbf{v}|})_{0 \leq |\mathbf{v}| \leq p})$ with $\mathbf{v}$ running through all $\frac{d+\mathfrak{p}}{d!\mathfrak{p}!}$ multi-indices such that $|\mathbf{v}| \leq p$, and

$$\widehat{\boldsymbol{\Gamma}}_{t,\mathbf{x}} = \mathbb{E}_n\Big[\mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{x}}{h}\Big) \mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{x}}{h}\Big)^\top K_h(\mathbf{X}_i - \mathbf{x}) \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)\Big],$$

where its population analogue is

$$\boldsymbol{\Gamma}_{t,\mathbf{x}} = \mathbb{E}\Big[\mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{x}}{h}\Big) \mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{x}}{h}\Big)^\top K_h(\mathbf{X}_i - \mathbf{x}) \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)\Big].$$

Note that $\left\|\mathbf{e}_{1+\boldsymbol{\nu}}^\top \mathbf{H}^{-1}\right\|_2 = \left\|\mathbf{e}_{1+\boldsymbol{\nu}}^\top \mathbf{H}^{-1}\right\|_\infty = h^{-|\boldsymbol{\nu}|}$. In addition, define

$$\mathbf{Q}_{t,\mathbf{x}} = \mathbb{E}_n\left[\mathbf{r}_p\left(\frac{\mathbf{X}_i - \mathbf{x}}{h}\right) K_h(\mathbf{X}_i - \mathbf{x}) \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t) u_i\right],$$

where $u_i = Y_i - [\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_0)\mu_0(\mathbf{X}_i) + \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_1)\mu_1(\mathbf{X}_i)] = Y_i - \mathbb{E}[Y_i|\mathbf{X}_i]$.

For $\mathbf{x}_1, \mathbf{x}_2 \in \mathscr{B}$ and $t \in \{0,1\}$, we introduce the following quantities:

$$\begin{aligned}
\widehat{\boldsymbol{\Sigma}}_{t,\mathbf{x}_1,\mathbf{x}_2} &= h^d \mathbb{E}_n\Big[\mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{x}_1}{h}\Big) \mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{x}_2}{h}\Big)^\top K_h(\mathbf{X}_i - \mathbf{x}_1) K_h(\mathbf{X}_i - \mathbf{x}_2) \widehat{\varepsilon}_i(\mathbf{x}_1) \widehat{\varepsilon}_i(\mathbf{x}_2) \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)\Big], \\
\boldsymbol{\Sigma}_{t,\mathbf{x}_1,\mathbf{x}_2} &= h^d \mathbb{E}\Big[\mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{x}_1}{h}\Big) \mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{x}_2}{h}\Big)^\top K_h(\mathbf{X}_i - \mathbf{x}_1) K_h(\mathbf{X}_i - \mathbf{x}_2) \sigma_t^2(\mathbf{X}_i) \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)\Big], \\
\widehat{\Omega}_{t,\mathbf{x}_1,\mathbf{x}_2}^{(\boldsymbol{\nu})} &= \frac{1}{nh^{d+2|\boldsymbol{\nu}|}} \mathbf{e}_{1+\boldsymbol{\nu}}^\top \widehat{\boldsymbol{\Gamma}}_{t,\mathbf{x}_1}^{-1} \widehat{\boldsymbol{\Sigma}}_{t,\mathbf{x}_1,\mathbf{x}_2} \widehat{\boldsymbol{\Gamma}}_{t,\mathbf{x}_2}^{-1} \mathbf{e}_{1+\boldsymbol{\nu}}, \qquad \widehat{\Omega}_{\mathbf{x}_1,\mathbf{x}_2}^{(\boldsymbol{\nu})} = \widehat{\Omega}_{0,\mathbf{x}_1,\mathbf{x}_2}^{(\boldsymbol{\nu})} + \widehat{\Omega}_{1,\mathbf{x}_1,\mathbf{x}_2}^{(\boldsymbol{\nu})}, \\
\Omega_{t,\mathbf{x}_1,\mathbf{x}_2}^{(\boldsymbol{\nu})} &= \frac{1}{nh^{d+2|\boldsymbol{\nu}|}} \mathbf{e}_{1+\boldsymbol{\nu}}^\top \boldsymbol{\Gamma}_{t,\mathbf{x}_1}^{-1} \boldsymbol{\Sigma}_{t,\mathbf{x}_1,\mathbf{x}_2} \boldsymbol{\Gamma}_{t,\mathbf{x}_2}^{-1} \mathbf{e}_{1+\boldsymbol{\nu}}, \qquad \Omega_{\mathbf{x}_1,\mathbf{x}_2}^{(\boldsymbol{\nu})} = \Omega_{0,\mathbf{x}_1,\mathbf{x}_2}^{(\boldsymbol{\nu})} + \Omega_{1,\mathbf{x}_1,\mathbf{x}_2}^{(\boldsymbol{\nu})},
\end{aligned}$$

where $\widehat{\varepsilon}_i(\mathbf{x}) = Y_i - \mathbf{r}_p(\mathbf{X}_i - \mathbf{x})^\top [\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_0)\widehat{\boldsymbol{\beta}}_0(\mathbf{x}) + \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_1)\widehat{\boldsymbol{\beta}}_1(\mathbf{x})]$.

Finally, to ensure that various weighted integral functionals on submanifolds (over the assignment boundary $\mathscr{B}$) are well-defined, we impose the following conditions on the weight function.

Assumption SA–3 (Weight Function and Boundary)**.** *Let $w : \mathscr{B} \mapsto \mathbb{R}$ with $\sup_{\mathbf{x}\in\mathscr{B}} |w(\mathbf{x})| < \infty$, $\inf_{\mathbf{x}\in\mathscr{B}} |w(\mathbf{x})| > 0$, and $\int_{\mathscr{B}} |w(\mathbf{x})| d\mathfrak{H}^{d-1}(\mathbf{x}) < \infty$.*

SA-1.1 Notation and Definitions

For textbook references on empirical process, see van der Vaart and Wellner [1996], Dudley [2014], and Giné and Nickl [2016]. For textbook reference on geometric measure theory, see Simon et al. [1984], Federer [2014], and Folland [2002].

(i) *Multi-index Notations.* For a multi-index $\mathbf{u} = (u_1, \ldots, u_d) \in \mathbb{N}^d$, denote $|\mathbf{u}| = \sum_{i=1}^d u_d$, $\mathbf{u}! = \Pi_{i=1}^d u_d$. Denote $\mathbf{r}_p(\mathbf{u}) = (1, u_1, \ldots, u_d, u_1^2, \ldots, u_d^2, \ldots, u_1^p, \ldots, u_d^p)$, that is, all monomials $u_1^{\alpha_1} \cdots u_d^{\alpha_d}$ such that $\alpha_i \in \mathbb{N}$ and $\sum_{i=1}^d \alpha_i \leq p$. Define $\mathbf{e}_{1+\boldsymbol{\nu}}$ to be the $p_d = \frac{(d+p)!}{d!p!}$-dimensional vector such that $\mathbf{e}_{1+\boldsymbol{\nu}}^\top \mathbf{r}_p(\mathbf{u}) = \mathbf{u}^{\boldsymbol{\nu}}$ for all $\mathbf{u} \in \mathbb{R}^d$ and $|\boldsymbol{\nu}| \leq p$.

(ii) *Norms.* For a vector $\mathbf{v} \in \mathbb{R}^k$, $\|\mathbf{v}\| = (\sum_{i=1}^k \mathbf{v}_i^2)^{1/2}$, $\|\mathbf{v}\|_\infty = \max_{1\leq i\leq k} |\mathbf{v}_i|$. For a matrix $A \in \mathbb{R}^{m\times n}$, $\|A\|_p = \sup_{\|\mathbf{x}\|_p=1} \|A\mathbf{x}\|_p$, $p \in \mathbb{N} \cup \{\infty\}$, and $\lambda_{\min}(A)$ denotes its minimum eigenvalue. For a function f on a metric space (S, d), $\|f\|_\infty = \sup_{\mathbf{x}\in\mathcal{X}} |f(\mathbf{x})|$. For a probability measure Q on $(\mathcal{S}, \mathscr{S})$ and $p \geq 1$, define $\|f\|_{Q,p} = (\int_{\mathcal{S}} |f|^p dQ)^{1/p}$. For a set $E \subseteq \mathbb{R}^d$, denote by $\mathfrak{m}(E)$ the Lebesgue measure of E.

(iii) *Empirical Process.* We use standard empirical process notations: $\mathbb{E}_n[g(\mathbf{v}_i)] = \frac{1}{n}\sum_{i=1}^n g(\mathbf{v}_i)$ and $\mathbb{G}_n[g(\mathbf{v}_i)] = \frac{1}{\sqrt{n}}\sum_{i=1}^n (g(\mathbf{v}_i) - \mathbb{E}[g(\mathbf{v}_i)])$. Let $(\mathcal{S}, d)$ be a semi-metric space. The covering number $N(\mathcal{S}, d, \varepsilon)$ is the minimal number of balls $B_s(\varepsilon) = \{t : d(t, s) < \varepsilon\}$ needed to cover $\mathcal{S}$. A $\mathbb{P}$-*Brownian bridge* is a mean-zero Gaussian random function $W_n(f), f \in L_2(\mathcal{X}, \mathbb{P})$ with the covariance $\mathbb{E}[W_{\mathbb{P}}(f)W_{\mathbb{P}}(g)] = \mathbb{P}(fg) - \mathbb{P}(f)\mathbb{P}(g)$, for $f, g \in L_2(\mathcal{X}, \mathbb{P})$. A class $\mathscr{F} \subseteq L_2(\mathcal{X}, \mathbb{P})$ is $\mathbb{P}$-*pregaussian* if there is a version of $\mathbb{P}$-Brownian bridge $W_{\mathbb{P}}$ such that $W_{\mathbb{P}} \in C(\mathscr{F}; \rho_{\mathbb{P}})$ almost surely, where $\rho_{\mathbb{P}}$ is the semi-metric on $L_2(\mathcal{X}, \mathbb{P})$ is defined by $\rho_{\mathbb{P}}(f, g) = (\|f - g\|_{\mathbb{P},2}^2 - (\int f\, d\mathbb{P} - \int g\, d\mathbb{P})^2)^{1/2}$, for $f, g \in L_2(\mathcal{X}, \mathbb{P})$.

(iv) *Geometric Measure Theory.* For a set $E \subseteq \mathcal{X}$, the *De Giorgi perimeter of E related to $\mathcal{X}$* is $\mathscr{L}(E) = \mathtt{TV}_{\{\mathbb{1}_E\},\mathcal{X}}$. For $d \in \mathbb{N}$ and $0 \leq m \leq d$, the m-dimensional Hausdorff (outer) measure is given by $\mathfrak{H}^m(A) = \lim_{\delta\downarrow 0} \mathfrak{H}_\delta^m(A)$, $A \subseteq \mathbb{R}^d$, where for each $\delta > 0$, $\mathfrak{H}_\delta^m(A)$ is defined by taking $\mathfrak{H}_\delta^m(\emptyset) = 0$, and for any non-empty $A \subseteq \mathbb{R}^d$, $\mathfrak{H}_\delta^m(A) = \frac{\pi^{m/2}}{\Gamma(m/2+1)} \inf \sum_{j=1}^\infty (\mathrm{diam}(C_j)/2)^m$, and the infimum is taken over all countable collections $C_1, C_2, \cdots$ of subsets of $\mathbb{R}^d$ such that $\mathrm{diam}(C_j) < \delta$ and $A \subseteq \cup_{j=1}^\infty C_j$. Integration against $\mathfrak{H}^m$ is defined via Carathéodory's Theorem following the classical measure-theoretic literature. The Hausdorff dimension $\dim_{\mathfrak{H}}(A)$ of A is defined by $\dim_{\mathfrak{H}}(A) = \inf\{t \geq 0 : \mathfrak{H}^t(A) = 0\}$. A set $A \subseteq \mathbb{R}^d$ is said to be k-rectifiable if A is of Hausdorff dimension k, and there exist a countable collection $\{f_i\}$ of continuously differentiable maps $f_i : \mathbb{R}^k \to \mathbb{R}^d$ such that $\mathfrak{H}^k(E \setminus \cup_{i=0}^\infty f_i(\mathbb{R}^k)) = 0$. B is a *rectifiable curve* if there exists a Lipschitz continuous function $\gamma : [0, 1] \to \mathbb{R}$ such that $B = \gamma([0, 1])$. We define the curve length function of B to be $\mathfrak{L}(B) = \sup_{\pi\in\Pi} s(\pi, \gamma)$, where $\Pi = \{(t_0, t_1, \ldots, t_N) : N \in \mathbb{N}, 0 \leq t_0 < t_1 < \ldots \leq t_N \leq 1\}$ and $s(\pi, \gamma) = \sum_{i=0}^N \|\gamma(t_i) - \gamma(t_{i+1})\|_2$ for $\pi = (t_0, t_1, \ldots, t_N)$.

(v) *Bounds and Asymptotics.* For reals sequences $a_n = o(b_n)$ if $\limsup \frac{|a_n|}{|b_n|} = 0$, $a_n \lesssim b_n$ if there exists some constant C and $N > 0$ such that $n > N$ implies $|a_n| \leq C|b_n|$. For sequences of random variables $a_n = o_{\mathbb{P}}(b_n)$ if $\operatorname{plim}_{n\to\infty} \frac{a_n}{b_n} = 0$, $|a_n| \lesssim_{\mathbb{P}} |b_n|$ if $\limsup_{M\to\infty} \limsup_{n\to\infty} \mathbb{P}[|\frac{a_n}{b_n}| \geq M] = 0$.

All limits are taken such that $h \to 0$ as $n \to \infty$. Most of our results hold with h fixed but small enough, but we do not make this distinction explicit to avoid overly-complex statements.

SA-1.2 Mapping Between Paper and Supplement

The results in the paper are special cases of the results in this supplemental appendix as follows.

- Theorem 1 in the paper corresponds to Theorem SA-1 with $d = 2$.
- Theorem 2 in the paper corresponds to Theorem SA-2 with $d = 2$.
- Theorem 3 in the paper corresponds to Theorems SA-3 and SA-6 with $d = 2$.
- Theorem 4 in the paper corresponds to Theorem SA-7 with $d = 2$.
- Theorem 5 in the paper corresponds to Theorem SA-8 with $d = 2$.
- Theorem 6 in the paper corresponds to Theorem SA-10 with $d = 2$.

SA-2 Preliminary Lemmas

Let $\mathbf{X} = (\mathbf{X}_1^\top, \cdots, \mathbf{X}_n^\top)$, and recall that $t \in \{0, 1\}$.

Lemma SA-1 (Invertibility)**.** *Suppose Assumption SA–1(i,ii) and Assumption SA–2 hold. Then for $t = 0, 1$,*

$$\liminf_{h \to 0} \inf_{\mathbf{x} \in \mathscr{B}} \lambda_{\min}(\boldsymbol{\Gamma}_{t,\mathbf{x}}) > 0.$$

Lemma SA-2 (Gram)**.** *Suppose Assumption SA–1(i,ii) and Assumption SA–2 hold. If $\frac{\log(1/h)}{nh^d} = o(1)$, then*

$$\sup_{\mathbf{x} \in \mathscr{B}} \left\|\widehat{\boldsymbol{\Gamma}}_{t,\mathbf{x}} - \boldsymbol{\Gamma}_{t,\mathbf{x}}\right\| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}}, \qquad \sup_{\mathbf{x} \in \mathscr{B}} \left\|\widehat{\boldsymbol{\Gamma}}_{t,\mathbf{x}}^{-1} - \boldsymbol{\Gamma}_{t,\mathbf{x}}^{-1}\right\| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}},$$

and if further $h = o(1)$, then $1 \lesssim_{\mathbb{P}} \inf_{\mathbf{x} \in \mathscr{B}} \left\|\widehat{\boldsymbol{\Gamma}}_{t,\mathbf{x}}\right\| \lesssim \sup_{\mathbf{x} \in \mathscr{B}} \left\|\widehat{\boldsymbol{\Gamma}}_{t,\mathbf{x}}\right\| \lesssim_{\mathbb{P}} 1$.

Lemma SA-3 (Stochastic Linear Approximation)**.** *Suppose Assumption SA–1(i,ii,iv,v) and Assumption SA–2 hold. Suppose $\frac{\log(1/h)}{nh^d} = o(1)$, then*

$$\sup_{\mathbf{x} \in \mathscr{B}} \left|\mathbf{Q}_{t,\mathbf{x}}\right| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{1+v}{2+v}} h^d},$$

and if further $h = o(1)$,

$$\sup_{\mathbf{x} \in \mathscr{B}} \left|\widehat{\mu}_t^{(\boldsymbol{\nu})}(\mathbf{x}) - \mathbb{E}\left[\widehat{\mu}_t^{(\boldsymbol{\nu})}(\mathbf{x}) \middle| \mathbf{X}\right] - \mathbf{e}_{1+\boldsymbol{\nu}}^\top \mathbf{H}^{-1} \boldsymbol{\Gamma}_{t,\mathbf{x}}^{-1} \mathbf{Q}_{t,\mathbf{x}}\right| \lesssim_{\mathbb{P}} h^{-|\boldsymbol{\nu}|} \sqrt{\frac{\log(1/h)}{nh^d}} \left(\sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{1+v}{2+v}} h^d}\right).$$

Lemma SA-4 (Covariance)**.** *Suppose Assumptions SA–1 and SA–2 hold. If $\frac{\log(1/h)}{nh^d} = o(1)$, then*

$$\sup_{\mathbf{x}_1, \mathbf{x}_2 \in \mathscr{B}} \left\|\widehat{\boldsymbol{\Sigma}}_{t,\mathbf{x}_1,\mathbf{x}_2} - \boldsymbol{\Sigma}_{t,\mathbf{x}_1,\mathbf{x}_2}\right\| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{v}{2+v}} h^d} + h^{p+1},$$

$$\sup_{\mathbf{x}_1,\mathbf{x}_2\in\mathscr{B}} \big|\widehat{\Omega}^{(\boldsymbol{\nu})}_{\mathbf{x}_1,\mathbf{x}_2} - \Omega^{(\boldsymbol{\nu})}_{\mathbf{x}_1,\mathbf{x}_2}\big| \lesssim_{\mathbb{P}} (nh^{d+2|\boldsymbol{\nu}|})^{-1}\Big(\sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{v}{2+v}}h^d} + h^{p+1}\Big),$$

and

$$\sup_{\mathbf{x}\in\mathscr{B}} \big|(\widehat{\Omega}^{(\boldsymbol{\nu})}_{\mathbf{x},\mathbf{x}})^{-1/2} - (\Omega^{(\boldsymbol{\nu})}_{\mathbf{x},\mathbf{x}})^{-1/2}\big| \lesssim_{\mathbb{P}} \sqrt{nh^{d+2|\boldsymbol{\nu}|}}\Big(\sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{v}{2+v}}h^d} + h^{p+1}\Big).$$

Lemma SA-5 (Bias)**.** *Suppose Assumption SA–1(i,ii,iii) and Assumption SA–2 hold. If* $\frac{\log(1/h)}{nh^d} = o(1)$ *and* $h = o(1)$*, then*

$$\sup_{\mathbf{x}\in\mathscr{B}} \big|\mathbb{E}[\widehat{\mu}^{(\boldsymbol{\nu})}_t(\mathbf{x})|\mathbf{X}] - \mu^{(\boldsymbol{\nu})}_t(\mathbf{x})\big| \lesssim_{\mathbb{P}} h^{p+1-|\boldsymbol{\nu}|},$$

implying

$$\sup_{\mathbf{x}\in\mathscr{B}} \big|\mathbb{E}[\widehat{\mu}^{(\boldsymbol{\nu})}_t(\mathbf{x})|\mathbf{X}] - \mu^{(\boldsymbol{\nu})}_t(\mathbf{x}) - h^{p+1-|\boldsymbol{\nu}|}\widehat{B}^{(\boldsymbol{\nu})}_{t,\mathbf{x}}\big| = o_{\mathbb{P}}(h^{p+1-|\boldsymbol{\nu}|}),$$

with $\sup_{\mathbf{x}\in\mathscr{B}} \big|\widehat{B}^{(\boldsymbol{\nu})}_{t,\mathbf{x}} - B^{(\boldsymbol{\nu})}_{t,\mathbf{x}}\big| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}}$*, and hence* $\sup_{\mathbf{x}\in\mathscr{B}} |\widehat{B}^{(\boldsymbol{\nu})}_{t,\mathbf{x}}| \lesssim_{\mathbb{P}} 1$.

SA-3 Boundary Average Treatment Effect Curve

SA-3.1 Point Estimation and MSE Expansions

Theorem SA-1 (Convergence Rates)**.** *Suppose Assumptions SA–1 and SA–2 hold. If* $\frac{\log(1/h)}{nh^d} = o(1)$ *and* $h = o(1)$*, then*

$$\big|\widehat{\tau}^{(\boldsymbol{\nu})}(\mathbf{x}) - \tau^{(\boldsymbol{\nu})}(\mathbf{x})\big| \lesssim_{\mathbb{P}} h^{-|\boldsymbol{\nu}|}\Big(\frac{1}{\sqrt{nh^d}} + \frac{1}{n^{\frac{1+v}{2+v}}h^d} + h^{p+1}\Big)$$

for $\mathbf{x}\in\mathscr{B}$*, and*

$$\sup_{\mathbf{x}\in\mathscr{B}} \big|\widehat{\tau}^{(\boldsymbol{\nu})}(\mathbf{x}) - \tau^{(\boldsymbol{\nu})}(\mathbf{x})\big| \lesssim_{\mathbb{P}} h^{-|\boldsymbol{\nu}|}\Big(\sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{1+v}{2+v}}h^d} + h^{p+1}\Big).$$

The conditional mean-squared error (MSE) is

$$\mathrm{MSE}_{\boldsymbol{\nu}}(\mathbf{x}) = \mathbb{E}\Big[(\widehat{\tau}^{(\boldsymbol{\nu})}(\mathbf{x}) - \tau^{(\boldsymbol{\nu})}(\mathbf{x}))^2\Big|\mathbf{X}\Big]$$

for $\mathbf{x}\in\mathscr{B}$, and the conditional integrated MSE (IMSE) is

$$\mathrm{IMSE}_{\boldsymbol{\nu}} = \int_{\mathscr{B}} \mathrm{MSE}_{\boldsymbol{\nu}}(\mathbf{x})w(\mathbf{x})d\mathfrak{H}^{d-1}(\mathbf{x}),$$

where $w(\mathbf{x})$ satisfies Assumption SA–3. To state the MSE expansions, we introduce some more notation for the leading bias and variance:

$$B^{(\boldsymbol{\nu})}_{\mathbf{x}} = B^{(\boldsymbol{\nu})}_{1,\mathbf{x}} - B^{(\boldsymbol{\nu})}_{0,\mathbf{x}}, \qquad B^{(\boldsymbol{\nu})}_{t,\mathbf{x}} = \mathbf{e}^{\top}_{1+\boldsymbol{\nu}}\boldsymbol{\Gamma}^{-1}_{t,\mathbf{x}} \sum_{|\boldsymbol{\omega}|=p+1} \frac{\mu^{(\boldsymbol{\omega})}_t(\mathbf{x})}{\boldsymbol{\omega}!}\mathbb{E}\Big[\mathbf{r}_p\Big(\frac{\mathbf{X}_i-\mathbf{x}}{h}\Big)\Big(\frac{\mathbf{X}_i-\mathbf{x}}{h}\Big)^{\boldsymbol{\omega}}K_h(\mathbf{X}_i-\mathbf{x})\Big],$$

where

$$V_{\mathbf{x}}^{(\boldsymbol{\nu})} = V_{0,\mathbf{x}}^{(\boldsymbol{\nu})} + V_{1,\mathbf{x}}^{(\boldsymbol{\nu})}, \qquad V_{t,\mathbf{x}}^{(\boldsymbol{\nu})} = \mathbf{e}_{1+\boldsymbol{\nu}}^{\top}\boldsymbol{\Gamma}_{t,\mathbf{x}}^{-1}\boldsymbol{\Sigma}_{t,\mathbf{x},\mathbf{x}}\boldsymbol{\Gamma}_{t,\mathbf{x}}^{-1}\mathbf{e}_{1+\boldsymbol{\nu}} = nh^{d+2|\boldsymbol{\nu}|}\Omega_{t,\mathbf{x},\mathbf{x}}^{(\boldsymbol{\nu})}.$$

Theorem SA-2 (MSE Expansions)**.** *Suppose Assumptions SA–1, SA–2 and SA–3 hold. If* $\frac{\log(1/h)}{nh^d} = o(1)$ *and* $h = o(1)$, *then*

$$\mathrm{MSE}_{\boldsymbol{\nu}}(\mathbf{x}) = \big(h^{p+1-|\boldsymbol{\nu}|}B_{\mathbf{x}}^{(\boldsymbol{\nu})}\big)^2 + \frac{1}{nh^{d+2|\boldsymbol{\nu}|}}V_{\mathbf{x}}^{(\boldsymbol{\nu})} + o_{\mathbb{P}}\big(h^{2p+2-2|\boldsymbol{\nu}|} + n^{-1}h^{-d-2|\boldsymbol{\nu}|}\big)$$

for $\mathbf{x} \in \mathscr{B}$, *and*

$$\mathrm{IMSE}_{\boldsymbol{\nu}} = \int_{\mathscr{B}} \Big[\big(h^{p+1-|\boldsymbol{\nu}}B_{\mathbf{x}}^{(\boldsymbol{\nu})}\big)^2 + \frac{1}{nh^{d+2|\boldsymbol{\nu}|}}V_{\mathbf{x}}^{(\boldsymbol{\nu})}\Big] w(\mathbf{x})d\mathfrak{H}^{d-1}(\mathbf{x}) + o_{\mathbb{P}}\big(h^{2p+2-2|\boldsymbol{\nu}|} + n^{-1}h^{-d-2|\boldsymbol{\nu}|}\big).$$

Theorem SA-2 can be used to develop (feasible) bandwidth selectors. If $\widehat{B}_{\mathbf{x}}^{(\boldsymbol{\nu})} \neq 0$, the asymptotic MSE-optimal bandwidth is

$$h_{\mathrm{MSE},\boldsymbol{\nu},p}(\mathbf{x}) = \left(\frac{(d+2|\boldsymbol{\nu}|)V_{\mathbf{x}}^{(\boldsymbol{\nu})}}{(2p+2-2|\boldsymbol{\nu}|)(B_{\mathbf{x}}^{(\boldsymbol{\nu})})^2}\frac{1}{n}\right)^{\frac{1}{2p+d+2}}$$

for $\mathbf{x} \in \mathscr{B}$. Similarly, if $\int_{\mathscr{B}}(B_{\mathbf{x}}^{(\boldsymbol{\nu})})^2 w(\mathbf{x})dH^{d-1}(\mathbf{x}) \neq 0$, the asymptotic IMSE-optimal bandwidth is

$$h_{\mathrm{IMSE},\boldsymbol{\nu},p} = \left(\frac{(d+2|\boldsymbol{\nu}|)\int_{\mathscr{B}} V_{\mathbf{x}}^{(\boldsymbol{\nu})}w(\mathbf{x})d\mathfrak{H}^{d-1}(\mathbf{x})}{(2p+2-2|\boldsymbol{\nu}|)\int_{\mathscr{B}}(B_{\mathbf{x}}^{(\boldsymbol{\nu})})^2 w(\mathbf{x})d\mathfrak{H}^{d-1}(\mathbf{x})}\frac{1}{n}\right)^{\frac{1}{2p+d+2}}.$$

In practice, the the unknown bias and variance quantities can be replaced with (consistent) estimators thereof. For example, $\widehat{B}_{\mathbf{x}}^{(\boldsymbol{\nu})} = \widehat{B}_{1,\mathbf{x}}^{(\boldsymbol{\nu})} - \widehat{B}_{0,\mathbf{x}}^{(\boldsymbol{\nu})}$ with

$$\widehat{B}_{t,\mathbf{x}}^{(\boldsymbol{\nu})} = \mathbf{e}_{1+\boldsymbol{\nu}}^{\top}\widehat{\boldsymbol{\Gamma}}_{t,\mathbf{x}}^{-1}\sum_{|\boldsymbol{\omega}|=p+1}\frac{\mu_t^{(\boldsymbol{\omega})}(\mathbf{x})}{\boldsymbol{\omega}!}\mathbb{E}_n\Big[\mathbf{r}_p\Big(\frac{\mathbf{X}_i-\mathbf{x}}{h}\Big)\Big(\frac{\mathbf{X}_i-\mathbf{x}}{h}\Big)^{\boldsymbol{\omega}}K_h(\mathbf{X}_i-\mathbf{x})\Big],$$

where the unknown functions $\mu_t^{(\boldsymbol{\omega})}(\mathbf{x})$ can be estimated using higher-order local polynomial estimators, and $\widehat{V}_{\mathbf{x}}^{(\boldsymbol{\nu})} = \widehat{V}_{0,\mathbf{x}}^{(\boldsymbol{\nu})} + \widehat{V}_{1,\mathbf{x}}^{(\boldsymbol{\nu})}$ with

$$\widehat{V}_{t,\mathbf{x}}^{(\boldsymbol{\nu})} = \mathbf{e}_{1+\boldsymbol{\nu}}^{\top}\widehat{\boldsymbol{\Gamma}}_{t,\mathbf{x}}^{-1}\widehat{\boldsymbol{\Sigma}}_{t,\mathbf{x},\mathbf{x}}\widehat{\boldsymbol{\Gamma}}_{t,\mathbf{x}}^{-1}\mathbf{e}_{1+\boldsymbol{\nu}},$$

which corresponds to a standard variance estimator (which is also used for asymptotic inference as discussed below).

Finally, notice that the pointwise convergence rate and MSE expansion can be obtained under the slightly weaker side rate condition $nh^d \to \infty$. We do not make this distinction explicit to simplify the exposition.

SA-3.2 Distributional Approximation and Inference

Let $\mathbf{W} = ((\mathbf{X}_1^\top, Y_1), \cdots, (\mathbf{X}_n^\top, Y_n))$, and recall that $t \in \{0,1\}$. For $|\boldsymbol{\nu}| \le p$, define the feasible t-statistic

$$\widehat{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x}) = \frac{\widehat{\tau}^{(\boldsymbol{\nu})}(\mathbf{x}) - \tau^{(\boldsymbol{\nu})}(\mathbf{x})}{\sqrt{\widehat{\Omega}^{(\boldsymbol{\nu})}_{\mathbf{x},\mathbf{x}}}}, \qquad \mathbf{x} \in \mathscr{B}.$$

The associated $100(1-\alpha)\%$ confidence interval estimator is

$$\widehat{\mathrm{I}}^{(\boldsymbol{\nu})}_{\alpha}(\mathbf{x}) = \Big[\, \widehat{\tau}^{(\boldsymbol{\nu})}(\mathbf{x}) - \mathscr{q}_\alpha \sqrt{\widehat{\Omega}^{(\boldsymbol{\nu})}_{\mathbf{x},\mathbf{x}}}\,,\ \widehat{\tau}^{(\boldsymbol{\nu})}(\mathbf{x}) + \mathscr{q}_\alpha \sqrt{\widehat{\Omega}^{(\boldsymbol{\nu})}_{\mathbf{x},\mathbf{x}}}\,\Big],$$

where $\mathscr{q}_\alpha$ denotes an appropriate quantile depending on the desired confidence level $\alpha \in (0,1)$, and coverage objective (pointwise vs. uniform over $\mathscr{B}$). The following theorem establishes pointwise asymptotic normality and validity of confidence intervals. Let $\Phi(\cdot)$ be the cumulative distribution function of a standard univariate Gaussian random variable.

Theorem SA-3 (Confidence Intervals)**.** *Suppose Assumptions SA–1 and SA–2 hold. If $nh^d \to \infty$ and $nh^d h^{2(p+1)} \to 0$, then*

$$\sup_{u \in \mathbb{R}} \Big| \mathbb{P}\big(\widehat{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x}) \le u\big) - \Phi(u) \Big| = o(1), \qquad \mathbf{x} \in \mathscr{B},$$

and

$$\mathbb{P}\big(\tau^{(\boldsymbol{\nu})}(\mathbf{x}) \in \widehat{\mathrm{I}}^{(\boldsymbol{\nu})}_{\alpha}(\mathbf{x})\big) = 1 - \alpha + o(1), \qquad \mathbf{x} \in \mathscr{B},$$

provided that $\mathscr{q}_\alpha = \inf\{c > 0 : \mathbb{P}(|\widehat{Z}| \ge c|\mathbf{W}) \le \alpha\}$ *with* $\widehat{Z}|\mathbf{W} \sim \mathsf{Normal}(0, \widehat{\Omega}^{(\boldsymbol{\nu})}_{\mathbf{x},\mathbf{x}})$.

For uniform inference, we rely on a new strong approximation result established in Section SA-6. First, we simplify the statistic $\widehat{\mathrm{T}}^{(\boldsymbol{\nu})}$, which is not directly a sum of independent random variables. Let

$$\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x}) = \mathbb{E}_n\Big[(\Omega^{(\boldsymbol{\nu})}_{\mathbf{x},\mathbf{x}})^{-1/2} \mathbf{e}_{1+\boldsymbol{\nu}}^\top \mathbf{H}^{-1} \Big[\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_1)\boldsymbol{\Gamma}_{1,\mathbf{x}}^{-1} - \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_0)\boldsymbol{\Gamma}_{0,\mathbf{x}}^{-1}\Big] \mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{x}}{h}\Big) K_h(\mathbf{X}_i - \mathbf{x}) u_i\Big],$$

where recall that $u_i = Y_i - \sum_{t \in \{0,1\}} \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)\mu_t(\mathbf{X}_i) = \mathbb{E}[Y_i|\mathbf{X}_i]$.

Theorem SA-4 (Stochastic Linearization)**.** *Suppose Assumptions SA–1 and SA–2 hold. If $\frac{\log(1/h)}{n^{\frac{v}{2+v}} h^d} = o(1)$ and $h = o(1)$, then*

$$\sup_{\mathbf{x} \in \mathscr{B}} \Big| \widehat{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x}) - \overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x}) \Big| \lesssim_{\mathbb{P}} h^{p+1}\sqrt{nh^d} + \sqrt{\log(1/h)}\Big(\sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{v}{2+v}} h^d}\Big).$$

We can now exploit the linear structure of $(\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x}) : \mathbf{x} \in \mathscr{B})$, that is, an average of i.n.i.d. random vectors. Define the following functions indexed by $\mathbf{x} \in \mathscr{B}$:

$$g_{\mathbf{x}}(\mathbf{u}) = \mathbb{1}(\mathbf{u} \in \mathscr{A}_1)\mathscr{K}_1^{(\boldsymbol{\nu})}(\mathbf{u}; \mathbf{x}) - \mathbb{1}(\mathbf{u} \in \mathscr{A}_0)\mathscr{K}_0^{(\boldsymbol{\nu})}(\mathbf{u}; \mathbf{x}), \qquad \mathbf{u} \in \mathscr{X},$$

and

$$\mathscr{K}_t^{(\boldsymbol{\nu})}(\mathbf{u};\mathbf{x}) = n^{-1/2}(\Omega_{\mathbf{x},\mathbf{x}}^{(\boldsymbol{\nu})})^{-1/2}\mathbf{e}_{1+\boldsymbol{\nu}}^\top\mathbf{H}^{-1}\boldsymbol{\Gamma}_{t,\mathbf{x}}^{-1}\mathbf{r}_p\left(\frac{\mathbf{u}-\mathbf{x}}{h}\right)K_h(\mathbf{u}-\mathbf{x}), \qquad \mathbf{u}\in\mathscr{X}, \quad t\in\{0,1\}.$$

Define the associated class of functions $\mathscr{G} = \{g_{\mathbf{x}} : \mathbf{x}\in\mathscr{B}\}$ and $\mathscr{R} = \{\mathrm{Id}\}$, where $\mathrm{Id}(x) = x$, for all $x\in\mathbb{R}$. Then, the *residual-based empirical process* is

$$R_n(g,r) = n^{-1/2}\sum_{i=1}^n\Big[g(\mathbf{X}_i)r(Y_i) - g(\mathbf{X}_i)\mathbb{E}[r(Y_i)|\mathbf{X}_i]\Big], \qquad g\in\mathscr{G}, r\in\mathscr{R},$$

and therefore

$$\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x}) = R_n(g_{\mathbf{x}},\mathrm{Id}), \qquad \mathbf{x}\in\mathscr{B}.$$

Leveraging ideas in Cattaneo and Yu [2025], Theorem SA-11 gives a new Gaussian strong approximation that can be applied to our current setup. Specifically, our new theorem allows for polynomial moment bound on the conditional distribution of $Y_i|\mathbf{X}_i$.

Theorem SA-5 (Gaussian Strong Approximation: $\overline{\mathrm{T}}^{(\boldsymbol{\nu})}$)**.** *Suppose Assumptions SA–1 and SA–2 hold, and that there exists a constant $C>0$ such that for $t\in\{0,1\}$ and for any $\mathbf{x}\in\mathscr{B}$, the De Giorgi perimeter of the set $E_{t,\mathbf{x}} = \{\mathbf{y}\in\mathscr{A}_t : (\mathbf{y}-\mathbf{x})/h\in\mathrm{Supp}(K)\}$ satisfies $\mathscr{L}(E_{t,\mathbf{x}})\leq Ch^{d-1}$. If $\liminf_{n\to\infty}\frac{\log h}{\log n} > -\infty$ and $nh^d\to\infty$ as $n\to\infty$, then (on a possibly enlarged probability space) there exists a mean-zero Gaussian process $Z^{(\boldsymbol{\nu})}$ indexed by $\mathscr{B}$ with almost surely continuous sample path such that*

$$\mathbb{E}\Big[\sup_{\mathbf{x}\in\mathscr{B}}\big|\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x}) - Z^{(\boldsymbol{\nu})}(\mathbf{x})\big|\Big] \lesssim (\log n)^{\frac{3}{2}}\left(\frac{1}{nh^d}\right)^{\frac{1}{2d+2}\cdot\frac{v}{v+2}} + \log(n)\left(\frac{1}{n^{\frac{v}{2+v}}h^d}\right)^{\frac{1}{2}},$$

where $\lesssim$ is up to a universal constant, and $Z^{(\boldsymbol{\nu})}$ has the same covariance structure as $\overline{\mathrm{T}}^{(\boldsymbol{\nu})}$; that is, $\mathbb{C}\mathrm{ov}[\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x}_1),\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x}_2)] = \mathbb{C}\mathrm{ov}[Z^{(\boldsymbol{\nu})}(\mathbf{x}_1),Z^{(\boldsymbol{\nu})}(\mathbf{x}_2)]$ for all $\mathbf{x}_1,\mathbf{x}_2\in\mathscr{B}$.

Theorem SA-5 can be used to construct confidence bands for $(\tau^{(\boldsymbol{\nu})}(\mathbf{x}) : \mathbf{x}\in\mathscr{B})$. Let $(\widehat{Z}^{(\boldsymbol{\nu})}(\mathbf{x}) : \mathbf{x}\in\mathscr{B})$ be a (conditionally on $\mathbf{W}$) mean-zero Gaussian process with feasible (conditional) covariance function

$$\mathbb{C}\mathrm{ov}\Big[\widehat{Z}^{(\boldsymbol{\nu})}(\mathbf{x}_1),\widehat{Z}^{(\boldsymbol{\nu})}(\mathbf{x}_2)\Big|\mathbf{W}\Big] = \frac{\widehat{\Omega}_{\mathbf{x}_1,\mathbf{x}_2}^{(\boldsymbol{\nu})}}{\sqrt{\widehat{\Omega}_{\mathbf{x}_1,\mathbf{x}_1}^{(\boldsymbol{\nu})}\widehat{\Omega}_{\mathbf{x}_2,\mathbf{x}_2}^{(\boldsymbol{\nu})}}}, \qquad \mathbf{x}_1,\mathbf{x}_2\in\mathscr{B}.$$

Theorem SA-6 (Confidence Bands)**.** *Suppose the assumptions and conditions in Theorem SA-5 hold. If $\liminf_{n\to\infty}\frac{\log h}{\log n} > -\infty$, $\frac{(\log n)^3}{n^{\frac{v}{2+v}}h^d} = o(1)$ and $h^{p+1}\sqrt{nh^d} = o(1)$, then*

$$\sup_{u\in\mathbb{R}}\Big|\mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}}\big|\widehat{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})\big|\leq u\Big) - \mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}}\big|\widehat{Z}^{(\boldsymbol{\nu})}(\mathbf{x})\big|\leq u\Big|\mathbf{W}\Big)\Big| = o_{\mathbb{P}}(1)$$

and

$$\mathbb{P}\Big[\tau^{(\boldsymbol{\nu})}(\mathbf{x})\in\widehat{\mathrm{I}}_\alpha^{(\boldsymbol{\nu})}(\mathbf{x}),\ \textit{for all } \mathbf{x}\in\mathscr{B}\Big] = 1-\alpha+o(1),$$

provided that $\mathscr{q}_\alpha = \inf\big\{c>0 : \mathbb{P}\big(\sup_{\mathbf{x}\in\mathscr{B}}\big|\widehat{Z}^{(\boldsymbol{\nu})}(\mathbf{x})\big|\geq c\big|\mathbf{W}\big)\leq\alpha\big\}$.

SA-4 Weighted Boundary Average Treatment Effect

Without loss of generality, we set $\int_{\mathscr{B}} w(\mathbf{b})d\mathfrak{H}^{d-1}(\mathbf{b}) = 1$, and the parameter of interest is the (weighted) average treatment effect along the boundary:

$$\tau_{\texttt{WBATE}} = \int_{\mathscr{B}} \tau(\mathbf{b})w(\mathbf{b})d\mathfrak{H}^{d-1}(\mathbf{b}),$$

where the weight function $w : \mathcal{X} \mapsto \mathbb{R}$ satisfies Assumption SA–3.

The (weighted) boundary average treatment effect estimator along the boundary is

$$\widehat{\tau}_{\texttt{WBATE}} = \int_{\mathscr{B}} \widehat{\tau}(\mathbf{b})w(\mathbf{b})\, d\mathfrak{H}^{d-1}(\mathbf{b}),$$

Our first lemma in this section studies the conditional bias of $\tau_{\texttt{WBATE}}$. Let

$$B_{\texttt{WBATE}} = B_{1,\texttt{WBATE}} - B_{0,\texttt{WBATE}}, \qquad B_{t,\texttt{WBATE}} = \int_{\mathscr{B}} B_{t,\mathbf{b}}^{(\mathbf{0})} w(\mathbf{b})d\mathfrak{H}^{d-1}(\mathbf{b}),$$

for $t \in \{0, 1\}$.

Lemma SA-6 (Bias: WBATE)**.** *Suppose Assumption SA–1(i)-(iii), SA–2 and SA–3 hold. If* $\frac{\log(1/h)}{nh^d} = o(1)$ *and* $h = o(1)$*, then*

$$\mathbb{E}[\widehat{\tau}_{\texttt{WBATE}}|\mathbf{X}] - \tau_{\texttt{WBATE}} = h^{p+1}B_{\texttt{WBATE}} + o_{\mathbb{P}}(h^{p+1}).$$

The next lemma studies the conditional variance of $\tau_{\texttt{WBATE}}$, and a plug-in estimator thereof. Let

$$\Omega_{\texttt{WBATE}} = \Omega_{1,\texttt{WBATE}} + \Omega_{0,\texttt{WBATE}}, \qquad \Omega_{t,\texttt{WBATE}} = \int_{\mathscr{B}}\int_{\mathscr{B}} \Omega_{t,\mathbf{b}_1,\mathbf{b}_2}^{(\mathbf{0})} w(\mathbf{b}_1)w(\mathbf{b}_2)d\mathfrak{H}^{d-1}(\mathbf{b}_1)d\mathfrak{H}^{d-1}(\mathbf{b}_2)$$

and

$$\widehat{\Omega}_{\texttt{WBATE}} = \widehat{\Omega}_{1,\texttt{WBATE}} + \widehat{\Omega}_{0,\texttt{WBATE}}, \qquad \widehat{\Omega}_{t,\texttt{WBATE}} = \int_{\mathscr{B}}\int_{\mathscr{B}} \widehat{\Omega}_{t,\mathbf{b}_1,\mathbf{b}_2}^{(\mathbf{0})} w(\mathbf{b}_1)w(\mathbf{b}_2)d\mathfrak{H}^{d-1}(\mathbf{b}_1)d\mathfrak{H}^{d-1}(\mathbf{b}_2),$$

for $t \in \{0, 1\}$.

Lemma SA-7 (Variance: WBATE)**.** *Suppose Assumptions SA–1, SA–2 and SA–3 hold. If* $\frac{\log(1/h)}{nh^d} = o(1)$ *and* $h = o(1)$*, then*

$$\mathbb{V}[\widehat{\tau}_{\texttt{WBATE}}|\mathbf{X}] = \Omega_{\texttt{WBATE}} + O_{\mathbb{P}}\Big(h^{d-1}\frac{\log(1/h)^{1/2}}{(nh^d)^{3/2}}\Big) = \Omega_{\texttt{WBATE}} + o_{\mathbb{P}}((nh)^{-1}),$$

where

$$(nh)^{-1} \lesssim \Omega_{\texttt{WBATE}} \lesssim (nh)^{-1}.$$

If, in addition, $\frac{\log(1/h)}{n^{\frac{v}{2+v}}h^d} = o(1)$*, then*

$$\mathbb{V}[\widehat{\tau}_{\texttt{WBATE}}|\mathbf{X}] = \widehat{\Omega}_{\texttt{WBATE}} + o_{\mathbb{P}}((nh)^{-1}).$$

Theorem SA-7 (MSE Expansion: WBATE)**.** *Suppose Assumptions SA–1, SA–2 and SA–3 hold. If* $\frac{\log(1/h)}{n^{\frac{v}{2+v}}h^d} = o(1)$ *and* $h = o(1)$*, then*

$$\mathbb{E}[(\widehat{\tau}_{\texttt{WBATE}} - \tau_{\texttt{WBATE}})^2|\mathbf{X}] = \Omega_{\texttt{WBATE}} + h^{2p+2}B_{\texttt{WBATE}}^2 + o_{\mathbb{P}}((nh)^{-1}) + o_{\mathbb{P}}(h^{2p+2}).$$

MSE-optimal bandwidth selection follows directly from Theorem SA-7.

For inference, we consider the feasible t-statistics

$$\widehat{\mathrm{T}}_{\texttt{WBATE}} = \frac{\widehat{\tau}_{\texttt{WBATE}} - \tau_{\texttt{WBATE}}}{\sqrt{\widehat{\Omega}_{\texttt{WBATE}}}}.$$

Theorem SA-8 (Asymptotic Normality: WBATE)**.** *Suppose Assumptions SA–1, SA–2 and SA–3 hold. If* $\frac{\log(1/h)}{n^{\frac{v}{2+v}}h^d} = o(1)$ *and* $nh^{2p+3} = o(1)$*, then*

$$\sup_{u\in\mathbb{R}} \left|\mathbb{P}(\widehat{\mathrm{T}}_{\texttt{WBATE}} \leq u) - \Phi(u)\right| = o(1).$$

SA-5 Largest Boundary Average Treatment Effect

Consider the maximum treatment effect over the boundary, defined by

$$\tau_{\texttt{LBATE}} = \sup_{\mathbf{b}\in\mathscr{B}} \tau(\mathbf{b}).$$

Theorem SA-9 (Convergence Rate: LBATE)**.** *Suppose Assumptions SA–1 and SA–2 hold. If* $\frac{\log(1/h)}{nh^d} = o(1)$ *and* $h = o(1)$*, then*

$$\left|\widehat{\tau}_{\texttt{LBATE}} - \tau_{\texttt{LBATE}}\right| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{1+v}{2+v}}h^d} + h^{p+1}.$$

Recall from Section SA-3.2 that $(\widehat{Z}(\mathbf{x}) : \mathbf{x} \in \mathscr{B})$ is a (conditionally on $\mathbf{W}$) mean-zero Gaussian process with feasible (conditional) covariance function

$$\mathbb{C}\mathrm{ov}\Big[\widehat{Z}(\mathbf{x}_1), \widehat{Z}(\mathbf{x}_2)\Big|\mathbf{W}\Big] = \frac{\widehat{\Omega}_{\mathbf{x}_1,\mathbf{x}_2}}{\sqrt{\widehat{\Omega}_{\mathbf{x}_1,\mathbf{x}_1}\widehat{\Omega}_{\mathbf{x}_2,\mathbf{x}_2}}}, \qquad \mathbf{x}_1, \mathbf{x}_2 \in \mathscr{B}.$$

Consider the confidence interval given by

$$\widehat{\mathrm{I}}_{\alpha,\texttt{LBATE}} = \left[\sup_{\mathbf{b}\in\mathscr{B}} \left(\widehat{\tau}(\mathbf{b}) - \mathscr{q}_\alpha\sqrt{\widehat{\Omega}_{\mathbf{b},\mathbf{b}}}\right),\ \sup_{\mathbf{b}\in\mathscr{B}} \left(\widehat{\tau}(\mathbf{b}) + \mathscr{q}_\alpha\sqrt{\widehat{\Omega}_{\mathbf{b},\mathbf{b}}}\right)\right],$$

where $\mathscr{q}_\alpha = \inf\left\{c > 0 : \mathbb{P}\big(\sup_{\mathbf{x}\in\mathscr{B}}\left|\widehat{Z}(\mathbf{x})\right| \geq c\big|\mathbf{W}\big) \leq \alpha\right\}$.

Theorem SA-10 (Confidence Interval: LBATE)**.** *Suppose the assumptions and conditions in Theorem SA-5 hold. If* $\liminf_{n\to\infty}\frac{\log h}{\log n} > -\infty$, $\frac{(\log n)^3}{n^{\frac{v}{2+v}}h^d} = o(1)$ *and* $h^{p+1}\sqrt{nh^d} = o(1)$*, then*

$$\mathbb{P}\Big[\tau_{\texttt{LBATE}} \in \widehat{\mathrm{I}}_{\alpha,\texttt{LBATE}}\Big] \geq 1 - \alpha + o(1).$$

SA-6 Gaussian Strong Approximation

We present a Gaussian strong approximation theorem, which is the key technical tool behind Theorem SA-5. The theorem builds on and generalizes the results in Cattaneo and Yu [2025]. Consider the *residual-based empirical process* given by

$$R_n[g, r] = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big[g(\mathbf{x}_i) r(y_i) - \mathbb{E}[g(\mathbf{x}_i) r(y_i) | \mathbf{x}_i] \Big], \qquad g \in \mathscr{G}, r \in \mathscr{R},$$

where $\mathscr{G}$ and $\mathscr{R}$ are classes of functions satisfying certain regularity conditions.

SA-6.1 Definitions for Function Spaces

Let $\mathscr{F}$ be a class of measurable functions from a probability space $(\mathbb{R}^q, \mathscr{B}(\mathbb{R}^q), \mathbb{P})$ to $\mathbb{R}$. We introduce several definitions that capture properties of $\mathscr{F}$.

(i) $\mathscr{F}$ is pointwise measurable if it contains a countable subset $\mathscr{G}$ such that for any $f \in \mathscr{F}$, there exists a sequence $(g_m : m \geq 1) \subseteq \mathscr{G}$ such that $\lim_{m \to \infty} g_m(\mathbf{u}) = f(\mathbf{u})$ for all $\mathbf{u} \in \mathbb{R}^q$.

(ii) Let $\operatorname{Supp}(\mathscr{F}) = \cup_{f \in \mathscr{F}} \operatorname{Supp}(f)$. A probability measure $\mathbb{Q}_{\mathscr{F}}$ on $(\mathbb{R}^q, \mathscr{B}(\mathbb{R}^q))$ is a surrogate measure for $\mathbb{P}$ with respect to $\mathscr{F}$ if

   (i) $\mathbb{Q}_{\mathscr{F}}$ agrees with $\mathbb{P}$ on $\operatorname{Supp}(\mathbb{P}) \cap \operatorname{Supp}(\mathscr{F})$.

   (ii) $\mathbb{Q}_{\mathscr{F}}(\operatorname{Supp}(\mathscr{F}) \setminus \operatorname{Supp}(\mathbb{P})) = 0$.

   Let $\mathscr{Q}_{\mathscr{F}} = \operatorname{Supp}(\mathbb{Q}_{\mathscr{F}})$.

(iii) For $q = 1$ and an interval $\mathscr{I} \subseteq \mathbb{R}$, the pointwise total variation of $\mathscr{F}$ over $\mathscr{I}$ is

$$\texttt{pTV}_{\mathscr{F}, \mathscr{I}} = \sup_{f \in \mathscr{F}} \sup_{P \geq 1} \sup_{\mathscr{P}_P \in \mathscr{I}} \sum_{i=1}^{P-1} |f(a_{i+1}) - f(a_i)|,$$

   where $\mathscr{P}_P = \{(a_1, \ldots, a_P) : a_1 \leq \cdots \leq a_P\}$ denotes the collection of all partitions of $\mathscr{I}$.

(iv) For a non-empty $\mathscr{C} \subseteq \mathbb{R}^q$, the total variation of $\mathscr{F}$ over $\mathscr{C}$ is

$$\texttt{TV}_{\mathscr{F}, \mathscr{C}} = \inf_{\mathscr{U} \in \mathscr{O}(\mathscr{C})} \sup_{f \in \mathscr{F}} \sup_{\phi \in \mathscr{D}_q(\mathscr{U})} \int_{\mathbb{R}^q} f(\mathbf{u}) \operatorname{div}(\phi)(\mathbf{u}) d\mathbf{u} / \| \|\phi\|_2 \|_\infty,$$

   where $\mathscr{O}(\mathscr{C})$ denotes the collection of all open sets that contains $\mathscr{C}$, and $\mathscr{D}_q(\mathscr{U})$ denotes the space of infinitely differentiable functions from $\mathbb{R}^q$ to $\mathbb{R}^q$ with compact support contained in $\mathscr{U}$.

(v) For a non-empty $\mathscr{C} \subseteq \mathbb{R}^q$, the local total variation constant of $\mathscr{F}$ over $\mathscr{C}$, is a positive number $\texttt{K}_{\mathscr{F}, \mathscr{C}}$ such that for any cube $\mathscr{D} \subseteq \mathbb{R}^q$ with edges of length ℓ parallel to the coordinate axises,

$$\texttt{TV}_{\mathscr{F}, \mathscr{D} \cap \mathscr{C}} \leq \texttt{K}_{\mathscr{F}, \mathscr{C}} \ell^{d-1}.$$

(vi) For a non-empty $\mathscr{C} \subseteq \mathbb{R}^q$, the envelopes of $\mathscr{F}$ over $\mathscr{C}$ are

$$\mathtt{M}_{\mathscr{F},\mathscr{C}} = \sup_{\mathbf{u}\in\mathscr{C}} M_{\mathscr{F},\mathscr{C}}(\mathbf{u}), \qquad M_{\mathscr{F},\mathscr{C}}(\mathbf{u}) = \sup_{f\in\mathscr{F}} |f(\mathbf{u})|, \qquad \mathbf{u}\in\mathscr{C}.$$

(vii) For a non-empty $\mathscr{C} \subseteq \mathbb{R}^q$, the Lipschitz constant of $\mathscr{F}$ over $\mathscr{C}$ is

$$\mathtt{L}_{\mathscr{F},\mathscr{C}} = \sup_{f\in\mathscr{F}} \sup_{\mathbf{u}_1,\mathbf{u}_2\in\mathscr{C}} \frac{|f(\mathbf{u}_1) - f(\mathbf{u}_2)|}{\|\mathbf{u}_1 - \mathbf{u}_2\|_\infty}.$$

(viii) For a non-empty $\mathscr{C} \subseteq \mathbb{R}^q$, the L_1 bound of $\mathscr{F}$ over $\mathscr{C}$ is

$$\mathtt{E}_{\mathscr{F},\mathscr{C}} = \sup_{f\in\mathscr{F}} \int_{\mathscr{C}} |f| d\mathbb{P}.$$

(ix) For a non-empty $\mathscr{C} \subseteq \mathbb{R}^q$, the uniform covering number of $\mathscr{F}$ with envelope $M_{\mathscr{F},\mathscr{C}}$ over $\mathscr{C}$ is

$$\mathtt{N}_{\mathscr{F},\mathscr{C}}(\delta, M_{\mathscr{F},\mathscr{C}}) = \sup_{\mu} N(\mathscr{F}, \|\cdot\|_{\mu,2}, \delta \|M_{\mathscr{F},\mathscr{C}}\|_{\mu,2}), \qquad \delta \in (0,\infty),$$

where the supremum is taken over all finite discrete measures on $(\mathscr{C}, \mathscr{B}(\mathscr{C}))$. We assume that $M_{\mathscr{F},\mathscr{C}}(\mathbf{u})$ is finite for every $\mathbf{u}\in\mathscr{C}$.

(x) For a non-empty $\mathscr{C} \subseteq \mathbb{R}^q$, the uniform entropy integral of $\mathscr{F}$ with envelope $M_{\mathscr{F},\mathscr{C}}$ over $\mathscr{C}$ is

$$J_{\mathscr{C}}(\delta, \mathscr{F}, M_{\mathscr{F},\mathscr{C}}) = \int_0^\delta \sqrt{1 + \log \mathtt{N}_{\mathscr{F},\mathscr{C}}(\varepsilon, M_{\mathscr{F},\mathscr{C}})} d\varepsilon,$$

where it is assumed that $M_{\mathscr{F},\mathscr{C}}(\mathbf{u})$ is finite for every $\mathbf{u}\in\mathscr{C}$.

(xi) For a non-empty $\mathscr{C} \subseteq \mathbb{R}^q$, $\mathscr{F}$ is a VC-type class with envelope $M_{\mathscr{F},\mathscr{C}}$ over $\mathscr{C}$ if (i) $M_{\mathscr{F},\mathscr{C}}$ is measurable and $M_{\mathscr{F},\mathscr{C}}(\mathbf{u})$ is finite for every $\mathbf{u}\in\mathscr{C}$, and (ii) there exist $\mathtt{c}_{\mathscr{F},\mathscr{C}} > 0$ and $\mathtt{d}_{\mathscr{F},\mathscr{C}} > 0$ such that

$$\mathtt{N}_{\mathscr{F},\mathscr{C}}(\varepsilon, M_{\mathscr{F},\mathscr{C}}) \leq \mathtt{c}_{\mathscr{F},\mathscr{C}} \varepsilon^{-\mathtt{d}_{\mathscr{F},\mathscr{C}}}, \qquad \varepsilon \in (0,1).$$

If a surrogate measure $\mathbb{Q}_{\mathscr{F}}$ for $\mathbb{P}$ with respect to $\mathscr{F}$ has been assumed, and it is clear from the context, we drop the dependence on $\mathscr{C} = \mathcal{Q}_{\mathscr{F}}$ for all quantities in the previous definitions. That is, to save notation, we set $\mathtt{TV}_{\mathscr{F}} = \mathtt{TV}_{\mathscr{F},\mathcal{Q}_{\mathscr{F}}}$, $\mathtt{K}_{\mathscr{F}} = \mathtt{K}_{\mathscr{F},\mathcal{Q}_{\mathscr{F}}}$, $\mathtt{M}_{\mathscr{F}} = \mathtt{M}_{\mathscr{F},\mathcal{Q}_{\mathscr{F}}}$, $M_{\mathscr{F}}(\mathbf{u}) = M_{\mathscr{F},\mathcal{Q}_{\mathscr{F}}}(\mathbf{u})$, $\mathtt{L}_{\mathscr{F}} = \mathtt{L}_{\mathscr{F},\mathcal{Q}_{\mathscr{F}}}$, and so on, whenever there is no confusion.

SA-6.2 Residual-based Empirical Process

The following theorem generalizes Cattaneo and Yu [2025, Theorem 2] by requiring only bounded polynomial moments for y_i conditional on $\mathbf{x}_i$.

Theorem SA-11 (Strong Approximation for Residual-based Empirical Processes)**.** *Suppose $(\mathbf{z}_i = (\mathbf{x}_i, y_i) : 1 \leq i \leq n)$ are i.i.d. random vectors taking values in $(\mathbb{R}^{d+1}, \mathscr{B}(\mathbb{R}^{d+1}))$ with common law $\mathbb{P}_Z$, where $\mathbf{x}_i$ has distribution $\mathbb{P}_X$ supported on $\mathscr{X} \subseteq \mathbb{R}^d$, y_i has distribution $\mathbb{P}_Y$ supported on $\mathscr{Y} \subseteq \mathbb{R}$, $\sup_{\mathbf{x}\in\mathscr{X}} \mathbb{E}[|y_i|^{2+v}|\mathbf{x}_i = \mathbf{x}] \leq 2$ for some $v > 0$, and the following conditions hold:*

(i) $\mathscr{G}$ *is a real-valued pointwise measurable class of functions on* $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d), \mathbb{P}_X)$.

(ii) *There exists a surrogate measure* $\mathbb{Q}_{\mathscr{G}}$ *for* $\mathbb{P}_X$ *with respect to* $\mathscr{G}$ *such that* $\mathbb{Q}_{\mathscr{G}} = \mathtt{m} \circ \phi_{\mathscr{G}}$, *where the normalizing transformation* $\phi_{\mathscr{G}} : \mathcal{Q}_{\mathscr{G}} \mapsto [0,1]^d$ *is a diffeomorphism.*

(iii) $\mathscr{G}$ *is a VC-type class with envelope* $\mathtt{M}_{\mathscr{G}}$ *over* $\mathcal{Q}_{\mathscr{G}}$ *with* $\mathtt{c}_{\mathscr{G}} \geq e$ *and* $\mathtt{d}_{\mathscr{G}} \geq 1$.

(iv) $\mathscr{R}$ *is a real-valued pointwise measurable class of functions on* $(\mathbb{R}, \mathsf{Borel}(\mathbb{R}), \mathbb{P}_Y)$.

(v) $\mathscr{R}$ *is a VC-type class with envelope* $M_{\mathscr{R},\mathscr{Y}}$ *over* $\mathscr{Y}$ *with* $\mathtt{c}_{\mathscr{R},\mathscr{Y}} \geq e$ *and* $\mathtt{d}_{\mathscr{R},\mathscr{Y}} \geq 1$, *where* $M_{\mathscr{R},\mathscr{Y}}(y) + \mathtt{pTV}_{\mathscr{R},(-|y|,|y|)} \leq \mathtt{v}(1+|y|)$ *for all* $y \in \mathscr{Y}$, *for some* $\mathtt{v} > 0$.

(vi) *There exists a constant* $\mathtt{k}$ *such that* $|\log_2 \mathtt{E}_{\mathscr{G}}| + |\log_2 \mathtt{TV}| + |\log_2 \mathtt{M}_{\mathscr{G}}| \leq \mathtt{k} \log_2 n$, *where the constant* $\mathtt{TV} = \max\{\mathtt{TV}_{\mathscr{G}}, \mathtt{TV}_{\mathscr{G} \times \mathscr{U}_{\mathscr{R}}, \mathcal{Q}_{\mathscr{G}}}\}$ *with* $\mathscr{U}_{\mathscr{R}} = \{\theta(\cdot, r, \tau) : r \in \mathscr{R}, \tau \in (0, \infty]\}$, *and* $\theta(\mathbf{x}, r, \tau) = \mathbb{E}[r(y_i)\mathbb{1}(|y_i| \leq \tau)|\mathbf{x}_i = \mathbf{x}]$ *for* $\mathbf{x} \in \mathscr{X}$.

Define the residual based empirical process

$$R_n(g, r) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(\mathbf{x}_i)(r(y_i) - \mathbb{E}[r(y_i)|\mathbf{x}_i]), \qquad g \in \mathscr{G}, r \in R.$$

Then (1) on a possibly enlarged probability space, there exists a sequence of mean-zero Gaussian processes $(Z_n^R(g, r) : g \in \mathscr{G}, r \in \mathscr{R})$ *with almost sure continuous trajectories such that:*

- $\mathbb{E}[R_n(g_1, r_1) R_n(g_2, r_2)] = \mathbb{E}[Z_n^R(g_1, r_1) Z_n^R(g_2, r_2)]$ *for all* $(g_1, r_1), (g_2, r_2) \in \mathscr{G} \times \mathscr{R}$, *and*
- $\mathbb{E}\big[\, \|R_n - Z_n^R\|_{\mathscr{G} \times \mathscr{R}} \big] \leq C \mathtt{v} \mathtt{d} \log(\mathtt{c} n) \rho_n$,

with

$$\rho_n = \sqrt{\mathtt{d} \log(\mathtt{c} n)}\, \mathtt{r}_n^{\frac{v}{v+2}} (\sqrt{\mathtt{M}_{\mathscr{G}} \mathtt{E}_{\mathscr{G}}})^{\frac{2}{v+2}} + \mathtt{M}_{\mathscr{G}} n^{-\frac{v/2}{2+v}} + \frac{\mathtt{M}_{\mathscr{G}}}{\sqrt{n}} \Big(\frac{\sqrt{\mathtt{M}_{\mathscr{G}} \mathtt{E}_{\mathscr{G}}}}{\mathtt{r}_n} \Big)^{\frac{2}{v+2}},$$

where C *is a positive universal constant,* $\mathtt{c} = \mathtt{c}_{\mathscr{G}} + \mathtt{c}_{\mathscr{R},\mathscr{Y}} + \mathtt{k}$, $\mathtt{d} = \mathtt{d}_{\mathscr{G}} \mathtt{d}_{\mathscr{R},\mathscr{Y}} \mathtt{k}$, *and*

$$\mathtt{r}_n = \min\Big\{ \frac{(\mathtt{c}_1^d \mathtt{M}_{\mathscr{G}}^{d+1} \mathtt{TV}^d \mathtt{E}_{\mathscr{G}})^{1/(2d+2)}}{n^{1/(2d+2)}}, \frac{(\mathtt{c}_1^{d/2} \mathtt{c}_2^{d/2} \mathtt{M}_{\mathscr{G}} \mathtt{TV}^{d/2} \mathtt{E}_{\mathscr{G}} \mathtt{L}^{d/2})^{1/(d+2)}}{n^{1/(d+2)}} \Big\},$$

$$\mathtt{c}_1 = d \sup_{\mathbf{x} \in \mathcal{Q}_{\mathscr{G}}} \prod_{j=1}^{d-1} \sigma_j(\nabla \phi_{\mathscr{G}}(\mathbf{x})), \qquad \mathtt{c}_2 = \sup_{\mathbf{x} \in \mathcal{Q}_{\mathscr{G}}} \frac{1}{\sigma_d(\nabla \phi_{\mathscr{G}}(\mathbf{x}))}, \qquad \mathtt{L} = \max\{\mathtt{L}_{\mathscr{G}}, \mathtt{L}_{\mathscr{G} \times \mathscr{U}_{\mathscr{R}}, \mathcal{Q}_{\mathscr{R}}}\};$$

and (2) if $\mathscr{R}$ *is a singleton, then we can replace* $\mathtt{TV}$ *and* $\mathtt{L}$ *in the previous conditions and statements by* $\mathtt{TV}_{sing} = \max\{\mathtt{TV}_{\mathscr{G}}, \mathtt{TV}_{\mathscr{G} \times \mathscr{V}_{\mathscr{R}}, \mathcal{Q}_{\mathscr{G}}}\}$, *and* $\mathtt{L}_{sing} = \max\{\mathtt{L}_{\mathscr{G}}, \mathtt{L}_{\mathscr{G} \times \mathscr{V}_{\mathscr{R}}, \mathcal{Q}_{\mathscr{R}}}\}$, *respectively, with* $\mathscr{V}_{\mathscr{R}} = \{\theta(\cdot, r) : r \in \mathscr{R}\}$, *and* $\theta(\mathbf{x}, r) = \mathbb{E}[r(y_i)|\mathbf{x}_i = \mathbf{x}]$ *for* $\mathbf{x} \in \mathscr{X}$.

Remark SA-1. *The class* $\mathscr{U}_{\mathscr{R}}$ *comprises truncated conditional means at all truncation levels. Its Lipschitz and total-variation constants can be bounded, for example, if* $f(y \mid \mathbf{x})$ *is Lipschitz in* $\mathbf{x}$ *uniformly over* $(\mathbf{x}, y)$ *in the support of* $(\mathbf{x}_i, y_i)$. *When* $\mathscr{R}$ *is a singleton, it suffices to assume regularity only for* $\mathscr{V}_{\mathscr{R}}$, *the class containing the (untruncated) conditional mean functions, which is easily justified.*

SA-7 Proofs

SA-7.1 Proof of Lemma SA-1

Assumption SA–1 (ii) implies

$$\begin{aligned}\boldsymbol{\Gamma}_{t,\mathbf{x}} &= \mathbb{E}\Big[\mathbf{r}_p\Big(\frac{\mathbf{X}_i-\mathbf{x}}{h}\Big)\mathbf{r}_p\Big(\frac{\mathbf{X}_i-\mathbf{x}}{h}\Big)^\top K_h(\mathbf{X}_i-\mathbf{x})\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_t)\Big]\\ &= \int_{\mathscr{A}_t}\mathbf{r}_p\Big(\frac{\mathbf{u}-\mathbf{x}}{h}\Big)\mathbf{r}_p\Big(\frac{\mathbf{u}-\mathbf{x}}{h}\Big)^\top K_h(\mathbf{u}-\mathbf{x})f(\mathbf{u})d\mathbf{u}\\ &= f(\mathbf{x})\int_{\mathscr{A}_t}\mathbf{r}_p\Big(\frac{\mathbf{u}-\mathbf{x}}{h}\Big)\mathbf{r}_p\Big(\frac{\mathbf{u}-\mathbf{x}}{h}\Big)^\top K_h(\mathbf{u}-\mathbf{x})d\mathbf{u}+o(1),\end{aligned}$$

where in the last line we have used $\int_{\mathscr{A}_t}(\frac{\mathbf{u}-\mathbf{x}}{h})^{\mathbf{v}}K_h(\mathbf{u}-\mathbf{x})d\mathbf{u}=O(1)$ for any multi-index $\mathbf{v}$ from standard change of variable argument.

I. Polynomial Representation of Minimum Eigenvalue

For simplicity, call

$$\mathbf{S}_{t,\mathbf{x}}=\lim_{h\to 0}\mathbf{S}_{t,\mathbf{x}}(h),\qquad \mathbf{S}_{t,\mathbf{x}}(h)=\int_{\mathscr{A}_t}\mathbf{r}_p\Big(\frac{\mathbf{u}-\mathbf{x}}{h}\Big)\mathbf{r}_p\Big(\frac{\mathbf{u}-\mathbf{x}}{h}\Big)^\top K_h(\mathbf{u}-\mathbf{x})d\mathbf{u}.$$

A change of variable gives

$$\mathbf{S}_{t,\mathbf{x}}(h)=\int\mathbf{r}_p(\mathbf{z})\mathbf{r}_p(\mathbf{z})^\top K(\mathbf{z})\mathbb{1}(\mathbf{x}+h\mathbf{z}\in\mathscr{A}_t)d\mathbf{z}.$$

Let $\mathbf{a}\in\mathbb{R}^{\mathfrak{p}_p}$, where $\mathfrak{p}_p=\frac{(d+p)!}{d!p!}$. Then the equivalent representation of minimum eigenvalue gives

$$\begin{aligned}\lambda_{\min}(\mathbf{S}_{t,\mathbf{x}}(h)) &= \min_{\|\mathbf{a}\|=1}\int(\mathbf{a}^\top\mathbf{r}_p(\mathbf{z}))^2K(\mathbf{z})\mathbb{1}(\mathbf{x}+h\mathbf{z}\in\mathscr{A}_t)d\mathbf{z}\\ &\geq \kappa\min_{\|\mathbf{a}\|=1}\int_U(\mathbf{a}^\top\mathbf{r}_p(\mathbf{z}))^2\mathbb{1}(\mathbf{x}+h\mathbf{z}\in\mathscr{A}_t)d\mathbf{z},\end{aligned}\tag{SA-1}$$

where in the last line we have used $K(\mathbf{u})\geq\kappa$ for all $u\in U$.

II. Mass Retaining Ratio in Treatment/Control Region

Denote $E_h(\mathbf{x},t)=\{\mathbf{z}\in U:\mathbf{x}+h\mathbf{z}\in\mathscr{A}_t\}$. Assumption SA–2 (ii) implies there is some upper bound $\Lambda>0$ of $K(\cdot)$. Hence for $c_0=1/2\ \liminf_{h\downarrow 0}\inf_{\mathbf{x}\in\mathscr{B}}\int_U K(\mathbf{u})\mathbb{1}(\mathbf{x}+h\mathbf{u}\in\mathscr{A}_t)d\mathbf{u}$, we have

$$\Lambda\mathfrak{m}(E_h(\mathbf{x},t))\geq\int_U K(\mathbf{u})\mathbb{1}(\mathbf{x}+h\mathbf{u}\in\mathscr{A}_t)\geq c_0$$

for small enough h, which implies

$$\mathfrak{m}(E_h(\mathbf{x},t))\geq\alpha\mathfrak{m}(U),\qquad\alpha=\frac{c_0}{\Lambda\mathfrak{m}(U)}.\tag{SA-2}$$

III. L_2 Integral of Polynomials in Full v.s. Treatment/Control Regions

Consider $S = \{f \in \mathscr{P}_{\mathfrak{p}} : \int_U f(\mathbf{u})^2 d\mathbf{u} = 1\}$, where $\mathscr{P}_{\mathfrak{p}}$ is the collection of all $\mathfrak{p}$-order polynomials. Let $(\phi_j, 1 \leq j \leq \mathfrak{p})$ be a set of orthonormal basis of $(\mathscr{P}_{\mathfrak{p}}, \|\cdot\|_{L_2})$. Then $T(\mathbf{a}) = \sum_{j=1}^{\mathfrak{p}} a_j \phi_j$ is an isometry. Since $T(S) = \{\mathbf{a} \in \mathbb{R}^{\mathfrak{p}} : \|\mathbf{a}\| = 1\}$ is compact, S is also compact in $(\mathscr{P}_{\mathfrak{p}}, \|\cdot\|_{L_2})$. Since $\mathscr{P}_{\mathfrak{p}}$ is $\mathfrak{p}$-dimensional, equivalent of norms implies that S is also compact in $(\mathscr{P}_{\mathfrak{p}}, \|\cdot\|_{L_\infty})$. Now consider

$$\Phi_q(\varepsilon) = \mathfrak{m}(\{\mathbf{u} \in U : |q(u)| < \varepsilon\}), \qquad q \in S, \varepsilon > 0,$$

and

$$\psi(q) = \sup\Big\{\varepsilon > 0 : \Phi_q(\varepsilon) \leq \frac{\alpha}{2}\mathfrak{m}(U)\Big\}.$$

Since $\int_U q^2 = 1$ and q is polynomial, $\lim_{\varepsilon \downarrow 0} \Phi_q(\varepsilon) = 0$ and $\Phi_q(\|q\|_\infty) = \mathfrak{m}(U)$. Continuity and Lipchitzness of $q \in S$ imply $\psi(q) > 0$ for all $q \in S$.

Next, we want to show ψ is lower-semicontinous function on $(\mathscr{P}_{\mathfrak{p}}, \|\cdot\|_{L_\infty})$. Suppose $q_n \to q$ uniformly on U. For every $\varepsilon_0 \in (0, \psi(q))$, there exists $\eta > 0$ such that $\Phi_q(\varepsilon_0) \leq \frac{\alpha}{2}\mathfrak{m}(U) - \eta$. Continuity of polynomials and the fact that level sets of polynomials have zero Lebesgue measure imply $\mathbb{1}_{\{|q_n| < \varepsilon_0\}}(\cdot) \to \mathbb{1}_{\{|q| < \varepsilon_0\}}(\cdot)$ almost surely. By Dominated Convergence Theorem, $\Phi_{q_n}(\varepsilon_0) \to \Phi_q(\varepsilon_0)$. Hence for large enough n, $\Phi_{q_n}(\varepsilon_0) \leq \frac{\alpha}{2}\mathfrak{m}(U)$, which implies $\varepsilon_0 \leq \psi(q_n)$. This implies $\liminf_{n\to\infty} \psi(q_n) \geq \varepsilon_0$. Since ε_0 is arbitrary in $(0, \psi(q))$, we have $\liminf_{n\to\infty} \psi(q_n) \geq \psi(q)$.

Compactness of S and lower-semicontinuity of ψ implies ψ attains its minimum on S. Since $\psi(q) > 0$ for all $q \in S$, we know $\varepsilon_* = \inf_{q \in S} \psi(q) > 0$. Then for every $q \in S$,

$$\begin{aligned}
\int_{E_h(\mathbf{x},t)} q^2 &\geq \varepsilon_*^2\, \mathfrak{m}\Big(E_h(\mathbf{x},t) \setminus \{|q| \leq \varepsilon_*\}\Big) \\
&\geq \varepsilon_*^2\, \Big(\mathfrak{m}(E_h(\mathbf{x},t)) - \mathfrak{m}(\{|q| \leq \varepsilon_*\})\Big) \\
&\geq \varepsilon_*^2\, \frac{\alpha}{2}\mathfrak{m}(U).
\end{aligned}$$

Scaling q from S gives

$$\int_{E_h(\mathbf{x},t)} q^2 \geq \varepsilon_*^2\, \frac{\alpha}{2} \int_U q^2, \qquad q \in \mathscr{P}_{\mathfrak{p}}. \tag{SA-3}$$

IV. Lower Bound of Minimum Eigenvalue

Equations (SA-1), (SA-2) and (SA-3) together give for small enough h,

$$\begin{aligned}
\inf_{\mathbf{x} \in \mathscr{B}} \lambda_{\min}(\mathbf{S}_{t,\mathbf{x}}(h)) &\geq \kappa \inf_{\mathbf{x} \in \mathscr{B}} \min_{\|\mathbf{a}\|=1} \int_{E_h(\mathbf{x},t)} (\mathbf{a}^\top \mathbf{r}_p(\mathbf{z}))^2 d\mathbf{z}, \\
&\geq \kappa \varepsilon_*^2\, \frac{\alpha}{2} \min_{\|\mathbf{a}\|=1} \int_U (\mathbf{a}^\top \mathbf{r}_p(\mathbf{z}))^2 d\mathbf{z} \\
&\geq \kappa \varepsilon_*^2\, \frac{\alpha}{2} \lambda_{\min}\Big(\int_U \mathbf{r}_p(\mathbf{z}) \mathbf{r}_p(\mathbf{z})^\top d\mathbf{z}\Big),
\end{aligned}$$

which implies $\liminf_{h\to 0} \inf_{\mathbf{x} \in \mathscr{B}} \lambda_{\min}(\mathbf{S}_{t,\mathbf{x}}(h)) > 0$.

SA-7.2 Proof of Lemma SA-2

Since $\widehat{\mathbf{\Gamma}}_{t,\mathbf{x}}$ is a finite dimensional matrix, it suffices to show the stated rate of convergence for each entry. Let $\mathbf{v}$ be a multi-index such that $|\mathbf{v}| \leq 2p$. Define

$$g_n(\xi, \mathbf{x}) = \left(\frac{\xi - \mathbf{x}}{h}\right)^{\mathbf{v}} \frac{1}{h^d} K\left(\frac{\xi - \mathbf{x}}{h}\right) \mathbb{1}(\xi \in \mathscr{A}_t), \qquad \xi \in \mathscr{X}, \mathbf{x} \in \mathscr{B}.$$

and $\mathscr{F} = \{g_n(\cdot, \mathbf{x}) : \mathbf{x} \in \mathscr{B}\}$. We will show $\mathscr{F}$ is a VC-type of class. In order to do this, we study the following quantities.

Constant Envelope Function. We assume K is continuous and has compact support, or $K = \mathbb{1}(\cdot \in [-1,1]^d)$. Hence there exists a constant C_1 such that for all $l \in \mathscr{F}$, for all $\mathbf{x} \in \mathscr{B}$, $|l(\mathbf{x})| \leq C_1 h^{-d} = F$.

Diameter of $\mathscr{F}$ in L_2. $\sup_{l \in \mathscr{F}} \|l\|_{\mathbb{P},2} = \sup_{\mathbf{x} \in \mathscr{B}} (\int_{\frac{\mathscr{A}_t - \mathbf{x}}{h}} \frac{1}{h^d} \mathbf{y}^{2\mathbf{v}} K(\mathbf{y})^2 f_X(\mathbf{x} + h\mathbf{y}) d\mathbf{y})^{1/2} \leq C_2 h^{-d/2}$ for some constant C_2. We can take C_1 large enough so that $\sigma = C_2 h^{-d/2} \leq F = C_1 h^{-d}$.

Ratio. For some constant C_3, $\delta = \frac{\sigma}{F} = C_3 \sqrt{h^d}$.

Covering Numbers. Case 1: K is Lipschitz. Let $\mathbf{x}, \mathbf{x}' \in \mathscr{B}$. Then, for a generic evaluation points $\mathbf{x} = (x_1, \dots, x_d)^\top$ and $\mathbf{x}' = (x_1', \dots, x_d')^\top$,

$$\begin{aligned}
\sup_{\xi \in \mathscr{X}} |g_n(\xi, \mathbf{x}) - g_n(\xi, \mathbf{x}')| &\leq \left| \left(\frac{\xi_1 - x_1}{h}\right)^{v_1} \cdots \left(\frac{\xi - x_d}{h}\right)^{v_d} - \left(\frac{\xi_1 - x_1'}{h}\right)^{v_1} \cdots \left(\frac{\xi - x_d'}{h}\right)^{v_d} \right| K_h(\xi - \mathbf{x}) \\
&\quad + \left(\frac{\xi_1 - x_1'}{h}\right)^{v_1} \cdots \left(\frac{\xi - x_d'}{h}\right)^{v_d} \left| K_h(\xi - \mathbf{x}) - K_h(\xi - \mathbf{x}') \right| \\
&\lesssim h^{-d-1} \|\mathbf{x} - \mathbf{x}'\|_\infty,
\end{aligned}$$

since we have assumed that K has compact support and is Lipschitz continuous. Hence, for any $\varepsilon \in (0,1]$ and for any finitely supported measure Q and metric $\|\cdot\|_{Q,2}$ based on $L_2(Q)$,

$$N(\mathscr{F}, \|\cdot\|_{Q,2}, \varepsilon \|F\|_{Q,2}) \leq N(\mathscr{X}, \|\cdot\|_\infty, \varepsilon \|F\|_{Q,2} h^{d+1}) \overset{(i)}{\lesssim} \left(\frac{\operatorname{diam}(\mathscr{X})}{\varepsilon \|F\|_{Q,2} h^{d+1}}\right)^d \lesssim \left(\frac{\operatorname{diam}(\mathscr{X})}{\varepsilon h}\right)^d,$$

where in (i) we used the fact that $\varepsilon \|F\|_{Q,2} h^{d+1} \lesssim \varepsilon h \lesssim 1$. Hence, $\mathscr{F}$ forms a VC-type class, and taking $A_1 = \operatorname{diam}(\mathscr{X})/h$ and $A_2 = d$, $\sup_Q N(\mathscr{F}, \|\cdot\|_{Q,2}, \varepsilon \|F\|_{Q,2}) \lesssim (A_1/\varepsilon)^{A_2}$, $\varepsilon \in (0,1]$, and where the supremum is over all finite discrete measure.

Case 2: $K = \mathbb{1}(\cdot \in [-1,1]^d)$. Consider

$$m_n(\xi, \mathbf{x}) = \left(\frac{\xi - \mathbf{x}}{h}\right)^{\mathbf{v}} \frac{1}{h^d} \mathbb{1}(\xi \in \mathscr{A}_t), \qquad \xi, \mathbf{x} \in \mathscr{X},$$

$\mathscr{M} = \{m_n(\cdot, \mathbf{x}) : \mathbf{x} \in \mathscr{B}\}$ and the constant envelope function $M = C_4 h^{-|\mathbf{v}|-d}$, for some constant C_4 only depending on diameter of $\mathscr{X}$. The same argument as before shows that for any discrete measure Q, we have

$$N(\mathscr{M}, \|\cdot\|_{Q,2}, \varepsilon \|M\|_{Q,2}) \leq N(\mathscr{X}, \|\cdot\|_\infty, \varepsilon \|M\|_{Q,2} h^{d+|\mathbf{v}|+1}) \lesssim \left(\frac{\operatorname{diam}(\mathscr{X})}{\varepsilon \|M\|_{Q,2} h^{d+|\mathbf{v}|+1}}\right)^d \lesssim \left(\frac{\operatorname{diam}(\mathscr{X})}{\varepsilon h}\right)^d.$$

The class $\mathscr{G} = \{\mathbb{1}(\cdot - \mathbf{x} \in [-1,1]^d) : \mathbf{x} \in \mathscr{B}\}$ has VC dimension no greater than $2d$ [van der Vaart and Wellner, 1996, Example 2.6.1], and by van der Vaart and Wellner [1996, Theorem 2.6.4], for any discrete measure Q, $N(\mathscr{G}, \|\cdot\|_{Q,2}, \varepsilon) \leq 2d(4e)^{2d}\varepsilon^{-4d}$, $0 < \varepsilon \leq 1$. It then follows that for any discrete measure Q,

$$N(\mathscr{F}, \|\cdot\|_{Q,2}, \varepsilon \|H\|_{Q,2}) \lesssim N(\mathscr{H}, \|\cdot\|_{Q,2}, \varepsilon/2 \|H\|_{Q,2}) + N(\mathscr{G}, \|\cdot\|_{Q,2}, \varepsilon/2) \lesssim 2^d h^{-d}\varepsilon^{-d} + 2d(32e)^d\varepsilon^{-4d}.$$

Hence, taking $A_1 = (2^d h^{-d} + 2d(32e)^d)h^{-|\mathbf{v}|}$ and $A_2 = 4d$, $\sup_Q N(\mathscr{F}, \|\cdot\|_{Q,2}, \varepsilon \|F\|_{Q,2}) \lesssim (A_1/\varepsilon)^{A_2}$, $\varepsilon \in (0, 1]$, where the supremum is over all finite discrete measure.

Maximal Inequality. By Corollary 5.1 in Chernozhukov et al. [2014b] for the empirical process on class $\mathscr{F}$,

$$\begin{aligned}\mathbb{E}\Big[\sup_{\mathbf{x}\in\mathscr{B}} \big|\mathbb{E}_n[g_n(\mathbf{X}_i, \mathbf{x})] - \mathbb{E}[g_n(\mathbf{X}_i, \mathbf{x})]\big|\Big] &\lesssim \frac{\sigma}{\sqrt{n}}\sqrt{A_2 \log(A_1/\delta)} + \frac{\|F\|_{\mathbb{P},2} A_2 \log(A_1/\delta)}{n} \\ &\lesssim \sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{nh^d},\end{aligned}$$

where $A_1, A_2, \sigma, F, \delta$ are all given previously. Assuming $\frac{\log(h^{-1})}{nh^d} \to 0$ as $n \to \infty$, we conclude that $\sup_{\mathbf{x}\in\mathscr{B}} \|\widehat{\mathbf{\Gamma}}_{t,\mathbf{x}} - \mathbf{\Gamma}_{t,\mathbf{x}}\| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}}$. Hence, $1 \lesssim_{\mathbb{P}} \inf_{\mathbf{x}\in\mathscr{B}} \|\widehat{\mathbf{\Gamma}}_{t,\mathbf{x}}\| \lesssim_{\mathbb{P}} \sup_{\mathbf{x}\in\mathscr{B}} \|\widehat{\mathbf{\Gamma}}_{t,\mathbf{x}}\| \lesssim_{\mathbb{P}} 1$. By Weyl's Theorem, $\sup_{\mathbf{x}\in\mathscr{B}} |\lambda_{\min}(\widehat{\mathbf{\Gamma}}_{t,\mathbf{x}}) - \lambda_{\min}(\mathbf{\Gamma}_{t,\mathbf{x}})| \leq \sup_{\mathbf{x}\in\mathscr{B}} \|\widehat{\mathbf{\Gamma}}_{t,\mathbf{x}} - \mathbf{\Gamma}_{t,\mathbf{x}}\| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}}$. Assuming that $\lambda_{\min}(\mathbf{\Gamma}_{t,\mathbf{x}}) \gtrsim 1$ (which we will verify in the last part of the proof), then we can lower the minimum eigenvalue by $\inf_{\mathbf{x}\in\mathscr{B}} \lambda_{\min}(\widehat{\mathbf{\Gamma}}_{t,\mathbf{x}}) \geq \inf_{\mathbf{x}\in\mathscr{B}} \lambda_{\min}(\mathbf{\Gamma}_{t,\mathbf{x}}) - \sup_{\mathbf{x}\in\mathscr{B}} |\lambda_{\min}(\widehat{\mathbf{\Gamma}}_{t,\mathbf{x}}) - \lambda_{\min}(\mathbf{\Gamma}_{t,\mathbf{x}})| \gtrsim_{\mathbb{P}} 1$. It follows that $\sup_{\mathbf{x}\in\mathscr{B}} \|\widehat{\mathbf{\Gamma}}_{t,\mathbf{x}}^{-1}\| \lesssim_{\mathbb{P}} 1$ and hence $\sup_{\mathbf{x}\in\mathscr{B}} \|\widehat{\mathbf{\Gamma}}_{t,\mathbf{x}}^{-1} - \mathbf{\Gamma}_{t,\mathbf{x}}^{-1}\| \leq \sup_{\mathbf{x}\in\mathscr{B}} \|\mathbf{\Gamma}_{t,\mathbf{x}}^{-1}\| \|\mathbf{\Gamma}_{t,\mathbf{x}} - \widehat{\mathbf{\Gamma}}_{t,\mathbf{x}}\| \|\widehat{\mathbf{\Gamma}}_{t,\mathbf{x}}^{-1}\| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}}$.

SA-7.3 Proof of Lemma SA-3

The proof is similar to the proof of Lemma SA-2. Let $\mathbf{v}$ be a multi-index such that $0 \leq |\mathbf{v}| \leq p$. Let

$$g_n(\xi, \mathbf{x}) = \Big(\frac{\xi - \mathbf{x}}{h}\Big)^{\mathbf{v}} K_h(\xi - \mathbf{x})\mathbb{1}(\xi \in \mathscr{A}_t), \qquad \xi, \mathbf{x} \in \mathscr{X}.$$

Define the class of functions $\mathscr{F} = \{(\xi, u) \in \mathscr{X} \times \mathbb{R} \mapsto g_n(\xi, \mathbf{x}) : \mathbf{x} \in \mathscr{B}\}$.

Envelope Function. Since K is continuous on its compact support, there exists a constant $C_1 > 0$ such that $|g_n(\xi, \mathbf{x})u| \leq C_1 h^{-d}|u|$, for $\xi, \mathbf{x} \in \mathscr{X}$ and $u \in \mathbb{R}$. We define the envelope function $F(\xi, u) = C_1 h^{-d}|u|$, for $\xi \in \mathscr{X}$ and $u \in \mathbb{R}$. Moreover, by Assumption SA–1(v), let $M = \max_{1\leq i\leq n} F(\mathbf{X}_i, u_i)$, then

$$\mathbb{E}[M^2]^{1/2} \lesssim h^{-d}\mathbb{E}\big[\max_{1\leq i\leq n} |u_i|^2\big]^{1/2} \lesssim h^{-d}\mathbb{E}\big[\max_{1\leq i\leq n} |u_i|^{2+v}\big]^{1/(2+v)} \lesssim n^{1/(2+v)}h^{-d}.$$

Diameter of $\mathscr{F}$ in L_2. Recall we denote $u_i = Y_i - \mathbb{E}[Y_i|\mathbf{X}_i]$, then

$$\sup_{l\in\mathscr{F}} \mathbb{E}[l(\mathbf{X}_i, u_i)^2]^{1/2} \leq \sup_{\xi\in\mathscr{X}} \mathbb{E}[u_i^2|\mathbf{X}_i = \xi]^{1/2} \sup_{\xi\in\mathscr{X}} \mathbb{E}[g_n(\mathbf{X}_i, \xi)^2]^{1/2} \leq C_3 h^{-d/2} = \sigma.$$

Ratio. We set $\delta = \frac{\sigma}{\|F\|_{\mathbb{P},2}} \lesssim h^{d/2}$.

Covering Numbers. Case 1: K is Lipschitz. Let $\mathbb{Q}$ be a finite distribution on $(\mathscr{X} \times \mathbb{R}, \mathscr{B}(\mathscr{X}) \otimes \mathsf{Borel}(\mathbb{R}))$. Let $\mathbf{x}, \mathbf{x}' \in \mathscr{X}$. In the proof of Lemma SA-2, we showed that $\sup_{\xi\in\mathscr{X}} \sup_{\mathbf{x},\mathbf{x}'\in\mathscr{X}} \frac{|g_n(\xi,\mathbf{x}) - g_n(\xi,\mathbf{x}')|}{\|\mathbf{x}-\mathbf{x}'\|_\infty} \lesssim h^{-d-1}$. Hence,

$$\|g_n(\mathbf{X}_i, \mathbf{x})u_i - g_n(\mathbf{X}_i, \mathbf{x}')u_i\|_{\mathbb{Q},2} \leq \|g_n(\cdot, \mathbf{x}) - g_n(\cdot, \mathbf{x}')\|_\infty \|u_i\|_{\mathbb{Q},2} \lesssim h^{-1}\|F\|_{\mathbb{Q},2} \|\mathbf{x} - \mathbf{x}'\|_\infty.$$

It follows that $\sup_Q N(\mathscr{F}, \|\cdot\|_{Q,2}, \epsilon\|F\|_{Q,2}) \lesssim (\frac{\operatorname{diam}(\mathcal{X})}{\epsilon h}))^d$, where sup is over all finite probability distributions on $(\mathcal{X} \times \mathbb{R}, \mathscr{B}(\mathcal{X}) \otimes \mathsf{Borel}(\mathbb{R}))$. Letting $A_1 = \frac{\operatorname{diam}(\mathcal{X})}{h}$ and $A_2 = d$, we conclude that

$$\sup_Q N(\mathscr{F}, \|\cdot\|_{Q,2}, \epsilon \|F\|_{Q,2}) \lesssim (A_1/\epsilon)^{A_2}, \qquad \epsilon \in (0,1].$$

Case 2: K is the uniform kernel. Let

$$m_n(\xi, \mathbf{x}) = \left(\frac{\xi - \mathbf{x}}{h}\right)^{\mathbf{v}} \frac{1}{h^d} \mathbb{1}(\xi \in \mathscr{A}_t), \qquad \xi, \mathbf{x} \in \mathcal{X},$$

with $\mathscr{M} = \{(\xi, u) \in \mathcal{X} \times \mathbb{R} \to m_n(\xi, \mathbf{x})u : \mathbf{x} \in \mathscr{B}\}$ and envelop function $M(\mathbf{x}, u) = C_1 h^{-d-|\mathbf{v}|}|u|$, for a positive constant C_1 depending only on K. By similar arguments as Case 1 and the proof of Lemma SA-2, it follows that $\sup_Q N(\mathscr{M}, \|\cdot\|_{Q,2}, \varepsilon\|M\|_{Q,2}) \lesssim \left(\frac{\operatorname{diam}(\mathcal{X})}{\varepsilon h}\right)^d$, where the supremum is taken over all finite discrete measures. Taking $\mathscr{G} = \{\mathbb{1}(\cdot - \mathbf{x} \in [-1,1]^d) : \mathbf{x} \in \mathscr{B}\}$, the proof of Lemma SA-2 shows that

$$\sup_Q N(\mathscr{G}, \|\cdot\|_{Q,2}, \varepsilon) \leq 2d(4e)^{2d} \varepsilon^{-4d}, \qquad \varepsilon \in [0,1],$$

where the supremum is taken over all finite discrete measures. Taking $A_1 = (2^d h^{-d} + 2d(32e)^d) h^{-|\mathbf{v}|}$ and $A_2 = 4d$, we have

$$\sup_Q N(\mathscr{F}, \|\cdot\|_{Q,2}, \varepsilon \|F\|_{Q,2}) \lesssim (A_1/\varepsilon)^{A_2}, \qquad \varepsilon \in (0,1],$$

the supremum is over all finite discrete measure.

Maximal Inequality. By Corollary 5.1 in Chernozhukov et al. [2014b],

$$\begin{aligned}
\mathbb{E}\Big[\sup_{\mathbf{x} \in \mathcal{X}} \big|\mathbb{E}_n[g_n(\mathbf{X}_i, \mathbf{x}) u_i]\big|\Big] &\lesssim \frac{\sigma}{\sqrt{n}} \sqrt{A_2 \log(A_1/\delta)} + \frac{\|M\|_{\mathbb{P},2} A_2 \log(A_1/\delta)}{n} \\
&\lesssim \sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{1+v}{2+v}} h^d}.
\end{aligned}$$

Since $\mathbf{Q}_{t,\mathbf{x}}$ is finite-dimensional, entry-wise convergence implies convergence in norm with the same rate. Hence, $\sup_{\mathbf{x} \in \mathcal{X}} \|\mathbf{Q}_{t,\mathbf{x}}\| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{1+v}{2+v}} h^d}$. By Lemma SA-2,

$$\begin{aligned}
\sup_{\mathbf{x} \in \mathcal{X}} \big|\widehat{\mu}_t^{(\boldsymbol{\nu})}(\mathbf{x}) - \mathbb{E}[\widehat{\mu}_t^{(\boldsymbol{\nu})}(\mathbf{x})|\mathbf{X}] - \mathbf{e}_{1+\boldsymbol{\nu}}^\top \mathbf{H}^{-1} \boldsymbol{\Gamma}_{t,\mathbf{x}}^{-1} \mathbf{Q}_{t,\mathbf{x}}\big| &= \sup_{\mathbf{x} \in \mathcal{X}} \big|\mathbf{e}_{1+\boldsymbol{\nu}}^\top \mathbf{H}^{-1} (\widehat{\boldsymbol{\Gamma}}_{t,\mathbf{x}}^{-1} - \boldsymbol{\Gamma}_{t,\mathbf{x}}^{-1}) \mathbf{Q}_{t,\mathbf{x}}\big| \\
&\lesssim_{\mathbb{P}} h^{-|\boldsymbol{\nu}|} \sqrt{\frac{\log(1/h)}{nh^d}} \Big(\sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{1+v}{2+v}} h^d}\Big),
\end{aligned}$$

and

$$\sup_{\mathbf{x} \in \mathcal{X}} \big|\widehat{\mu}_t^{(\boldsymbol{\nu})}(\mathbf{x}) - \mathbb{E}[\widehat{\mu}_t^{(\boldsymbol{\nu})}(\mathbf{x})|\mathbf{X}]\big| \lesssim_{\mathbb{P}} h^{-|\boldsymbol{\nu}|} \Big(\sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{1+v}{2+v}} h^d}\Big),$$

which completes the proof. □

SA-7.4 Proof of Lemma SA-4

Let $\eta_i(\mathbf{x}) = \sum_{t\in\{0,1\}} \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)(\mu_t(\mathbf{X}_i) - \widehat{\boldsymbol{\beta}}_t(\mathbf{x})^\top \mathbf{R}_p(\mathbf{X}_i - \mathbf{x}))$. Then, for all $\mathbf{x}, \mathbf{y} \in \mathscr{B}$, the difference between the estimated and true variance matrices is

$$\widehat{\boldsymbol{\Sigma}}_{t,\mathbf{x},\mathbf{y}} - \boldsymbol{\Sigma}_{t,\mathbf{x},\mathbf{y}} = \mathbf{M}_{1,\mathbf{x},\mathbf{y}} + \mathbf{M}_{2,\mathbf{x},\mathbf{y}} + \mathbf{M}_{3,\mathbf{x},\mathbf{y}} + \mathbf{M}_{4,\mathbf{x},\mathbf{y}}$$

where

$$\begin{aligned}
\mathbf{M}_{1,\mathbf{x},\mathbf{y}} &= \mathbb{E}_n\Big[\mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{x}}{h}\Big)\mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{y}}{h}\Big)^\top \frac{1}{h^d} K\Big(\frac{\mathbf{X}_i - \mathbf{x}}{h}\Big) K\Big(\frac{\mathbf{X}_i - \mathbf{y}}{h}\Big)\eta_i(\mathbf{x})\eta_i(\mathbf{y})\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)\Big],\\
\mathbf{M}_{2,\mathbf{x},\mathbf{y}} &= \mathbb{E}_n\Big[\mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{x}}{h}\Big)\mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{y}}{h}\Big)^\top \frac{1}{h^d} K\Big(\frac{\mathbf{X}_i - \mathbf{x}}{h}\Big) K\Big(\frac{\mathbf{X}_i - \mathbf{y}}{h}\Big)(\eta_i(\mathbf{x}) + \eta_i(\mathbf{y}))u_i\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)\Big],\\
\mathbf{M}_{3,\mathbf{x},\mathbf{y}} &= \mathbb{E}_n\Big[\mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{x}}{h}\Big)\mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{y}}{h}\Big)^\top \frac{1}{h^d} K\Big(\frac{\mathbf{X}_i - \mathbf{x}}{h}\Big) K\Big(\frac{\mathbf{X}_i - \mathbf{y}}{h}\Big)(u_i^2 - \sigma_t(\mathbf{X}_i)^2)\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)\Big],\\
\mathbf{M}_{4,\mathbf{x},\mathbf{y}} &= \mathbb{E}_n\Big[\mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{x}}{h}\Big)\mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{y}}{h}\Big)^\top \frac{1}{h^d} K\Big(\frac{\mathbf{X}_i - \mathbf{x}}{h}\Big) K\Big(\frac{\mathbf{X}_i - \mathbf{y}}{h}\Big)\sigma_t(\mathbf{X}_i)^2\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)\Big]\\
&\quad - \mathbb{E}\Big[\mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{x}}{h}\Big)\mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{y}}{h}\Big)^\top \frac{1}{h^d} K\Big(\frac{\mathbf{X}_i - \mathbf{x}}{h}\Big) K\Big(\frac{\mathbf{X}_i - \mathbf{y}}{h}\Big)\sigma_t(\mathbf{X}_i)^2\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)\Big].
\end{aligned}$$

For $\mathbf{u}$ and $\mathbf{v}$ multi-indices, let $g_n(\mathbf{X}_i; \mathbf{x}, \mathbf{y}) = \frac{1}{h^d}(\frac{\mathbf{X}_i - \mathbf{x}}{h})^{\mathbf{u}}(\frac{\mathbf{X}_i - \mathbf{y}}{h})^{\mathbf{v}} K(\frac{\mathbf{X}_i - \mathbf{x}}{h})K(\frac{\mathbf{X}_i - \mathbf{y}}{h})\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)$. Set

$$\mathfrak{R}_n = \sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{1+v}{2+v}}h^d}.$$

First, we present a bound on $\max_{1\le i\le n} |\eta_i(\mathbf{x})|\mathbb{1}((\mathbf{X}_i - \mathbf{x})/h \in \mathrm{Supp}(K))$. By Lemma SA-5 and Lemma SA-3, and multi-index $\boldsymbol{\nu}$ such that $|\boldsymbol{\nu}| \le p$,

$$\sup_{\mathbf{x}\in\mathscr{B}} |\mathbf{e}_{1+\boldsymbol{\nu}}^\top \widehat{\mu}_t(\mathbf{x}) - \mathbf{e}_{1+\boldsymbol{\nu}}^\top \mu_t(\mathbf{x})| \lesssim_{\mathbb{P}} h^{-|\boldsymbol{\nu}|}(h^{p+1} + \mathfrak{R}_n).$$

Since K is compactly supported, we have

$$\max_{1\le i\le n}\Big|\sum_{t\in\{0,1\}} \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)(\widehat{\boldsymbol{\beta}}_t(\mathbf{x}) - \boldsymbol{\beta}_t(\mathbf{x}))^\top \mathbf{R}_p(\mathbf{X}_i - \mathbf{x})\mathbb{1}((\mathbf{X}_i - \mathbf{x})/h \in \mathrm{Supp}(K))\Big| \lesssim_{\mathbb{P}} h^{p+1} + \mathfrak{R}_n.$$

Since μ_t is $p+1$ times continuously differentiable,

$$\max_{1\le i\le n}\Big|\sum_{t\in\{0,1\}} \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)(\mu_t(\mathbf{X}_i) - \boldsymbol{\beta}_t(\mathbf{x})^\top \mathbf{R}_p(\mathbf{X}_i - \mathbf{x}))\mathbb{1}((\mathbf{X}_i - \mathbf{x})/h \in \mathrm{Supp}(K))\Big| \lesssim h^{p+1}.$$

It follows that

$$\sup_{\mathbf{x}\in\mathscr{B}} \max_{1\le i\le n} |\eta_i(\mathbf{x})|\mathbb{1}((\mathbf{X}_i - \mathbf{x})/h \in \mathrm{Supp}(K)) \lesssim_{\mathbb{P}} h^{p+1} + \mathfrak{R}_n.$$

Term $\mathbf{M}_{1,\mathbf{x},\mathbf{y}}$. From the proof for Lemma SA-2, $\sup_{\mathbf{x},\mathbf{y}\in\mathscr{X}} \big|\mathbb{E}_n[g_n(\mathbf{X}_i; \mathbf{x}, \mathbf{y})] - \mathbb{E}[g_n(\mathbf{X}_i; \mathbf{x}, \mathbf{y})]\big| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}}$. Moreover, $\sup_{\mathbf{x},\mathbf{y}\in\mathscr{X}} \big|\mathbb{E}[g_n(\mathbf{X}_i; \mathbf{x}, \mathbf{y})]\big| \lesssim_{\mathbb{P}} 1$. Hence $\sup_{\mathbf{x},\mathbf{y}\in\mathscr{X}} \big|\mathbb{E}_n[g_n(\mathbf{X}_i; \mathbf{x}, \mathbf{y})]\big| \lesssim_{\mathbb{P}} 1$. Thus,

$$\sup_{\mathbf{x},\mathbf{y}\in\mathscr{B}} \big|\mathbb{E}_n[g_n(\mathbf{X}_i; \mathbf{x}, \mathbf{y})\eta_i(\mathbf{x})\eta_i(\mathbf{y})]\big| \le \sup_{\mathbf{x}\in\mathscr{X}} \max_{1\le i\le n} |\eta_i(\mathbf{x})|\mathbb{1}((\mathbf{X}_i - \mathbf{x})/h \in \mathrm{Supp}(K)) \cdot \sup_{\mathbf{x},\mathbf{y}\in\mathscr{X}} \big|\mathbb{E}_n[g_n(\mathbf{X}_i; \mathbf{x}, \mathbf{y})]\big|$$

$$\lesssim_{\mathbb{P}} (h^{p+1} + \mathfrak{R}_n)^2,$$

where we have used Theorem SA-1, which does not depend on this lemma, for $\sup_{\mathbf{x}\in\mathscr{B}} \left|\widehat{\mu}_t(\mathbf{x}) - \mu_t(\mathbf{x})\right| \lesssim_{\mathbb{P}} h^{p+1} + \mathfrak{R}_n$. Finite dimensionality of $\mathbf{M}_{1,\mathbf{x},\mathbf{y}}$ then implies

$$\sup_{\mathbf{x},\mathbf{y}\in\mathscr{B}} \|\mathbf{M}_{1,\mathbf{x},\mathbf{y}}\| \lesssim_{\mathbb{P}} (h^{p+1} + \mathfrak{R}_n)^2.$$

Term $\mathbf{M}_{2,\mathbf{x},\mathbf{y}}$. From the proof of Lemma SA-3, $\sup_{\mathbf{x},\mathbf{y}\in\mathscr{X}} \left|\mathbb{E}_n[g_n(\mathbf{X}_i;\mathbf{x},\mathbf{y})u_i] - \mathbb{E}[g_n(\mathbf{X}_i;\mathbf{x},\mathbf{y})u_i]\right| \lesssim_{\mathbb{P}} \mathfrak{R}_n$. Moreover, $\sup_{\mathbf{x},\mathbf{y}\in\mathscr{X}} \left|\mathbb{E}[g_n(\mathbf{X}_i;\mathbf{x},\mathbf{y})u_i]big\| \lesssim_{\mathbb{P}} 1\right.$. Hence, $\sup_{\mathbf{x},\mathbf{y}\in\mathscr{X}} \left|\mathbb{E}_n[g_n(\mathbf{X}_i;\mathbf{x},\mathbf{y})u_i]\right| \lesssim_{\mathbb{P}} 1$. Thus,

$$\sup_{\mathbf{x},\mathbf{y}\in\mathscr{B}} \left|\mathbb{E}_n[g_n(\mathbf{X}_i;\mathbf{x},\mathbf{y})(\eta_i(\mathbf{x}) + \eta_i(\mathbf{y}))u_i]\right| \leq \sup_{\mathbf{x},\mathbf{y}\in\mathscr{B}} \left|\widehat{\mu}_t(\mathbf{x}) - \mu_t(\mathbf{x})\right| \sup_{\mathbf{x},\mathbf{y}\in\mathscr{B}} \mathbb{E}_n[|g_n(\mathbf{X}_i;\mathbf{x},\mathbf{y})u_i|] \lesssim_{\mathbb{P}} h^{p+1} + \mathfrak{R}_n,$$

which implies that

$$\sup_{\mathbf{x},\mathbf{y}\in\mathscr{B}} \|\mathbf{M}_{2,\mathbf{x},\mathbf{y}}\| \lesssim_{\mathbb{P}} h^{p+1} + \mathfrak{R}_n.$$

Term $\mathbf{M}_{3,\mathbf{x},\mathbf{y}}$. Define $l_n(\cdot,\cdot;\mathbf{x},\mathbf{y}) : \mathscr{X} \times \mathbb{R} \to \mathbb{R}$ as

$$l_n(\xi,\varepsilon;\mathbf{x},\mathbf{y}) = \frac{1}{h^d}\Big(\frac{\xi - \mathbf{x}}{h}\Big)^{\mathbf{u}}\Big(\frac{\xi - \mathbf{y}}{h}\Big)^{\mathbf{v}} K\Big(\frac{\xi - \mathbf{x}}{h}\Big) K\Big(\frac{\xi - \mathbf{y}}{h}\Big)\mathbb{1}(\xi \in \mathscr{A}_t)(\varepsilon^2 - \sigma_t^2(\xi)),$$

and consider the function class $\mathscr{L} = \{l_n(\cdot,\cdot;\mathbf{x},\mathbf{y}) : \mathbf{x},\mathbf{y} \in \mathscr{X}\}$. Let $L : \mathscr{X} \times \mathbb{R} \to \mathbb{R}$ be $L(\xi,\varepsilon) = \frac{c}{h^d}|\varepsilon^2 - \sigma_t^2(\xi)|$ with $c = \sup_{\mathbf{x},\mathbf{y}\in\mathscr{B}} \left|\big(\frac{\xi-\mathbf{x}}{h}\big)^{\mathbf{u}}\big(\frac{\xi-\mathbf{y}}{h}\big)^{\mathbf{v}} K\big(\frac{\xi-\mathbf{x}}{h}\big) K\big(\frac{\xi-\mathbf{y}}{h}\big)\right|$. By similar argument as in the proof for Lemma SA-3, we can show $\mathscr{L}$ is a VC-type class such that $\mathbb{E}[l_n(\mathbf{X}_i,u_i;\mathbf{x},\mathbf{y})] = 0$, for all $\mathbf{x},\mathbf{y} \in \mathscr{X}$,

$$\sup_{\mathbf{x},\mathbf{y}\in\mathscr{X}} \mathbb{E}[l_n(\mathbf{X}_i,\varepsilon;\mathbf{x},\mathbf{y})^2]^{\frac{1}{2}} \lesssim \sup_{\mathbf{x},\mathbf{y}\in\mathscr{B}} \mathbb{E}[g_n(\mathbf{X}_i,u_i;\mathbf{x},\mathbf{y})^2]^{\frac{1}{2}} \sup_{\xi\in\mathscr{X}} \mathbb{V}[u_i^2|\mathbf{X}_i = \xi] \lesssim h^{-d/2}$$

and

$$\mathbb{E}\big[\max_{1\leq i\leq n} L(\mathbf{X}_i,u_i)^2\big]^{\frac{1}{2}} \lesssim h^{-d}\mathbb{E}\big[\max_{1\leq i\leq n} u_i^4\big]^{1/2} \lesssim h^{-d}\mathbb{E}\big[\max_{1\leq i\leq n} u_i^{2+v}\big]^{\frac{2}{2+v}} \lesssim h^{-d} n^{\frac{2}{2+v}}.$$

Applying Corollary 5.1 in Chernozhukov et al. [2014b], we obtain

$$\sup_{\mathbf{x},\mathbf{y}\in\mathscr{B}} \left|\mathbb{E}_n[l_n(\mathbf{X}_i,u_i;\mathbf{x},\mathbf{y})]\right| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{v}{2+v}} h^d}$$

and

$$\sup_{\mathbf{x},\mathbf{y}\in\mathscr{B}} \|\mathbf{M}_{3,\mathbf{x},\mathbf{y}}\| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{v}{2+v}} h^d}.$$

Term $\mathbf{M}_{4,\mathbf{x},\mathbf{y}}$. Notice that $\{g_n(\cdot;\mathbf{x},\mathbf{y})\sigma_t^2(\cdot) : \mathbf{x},\mathbf{y} \in \mathscr{B}\}$ is a VC-type of class with constant envelope function Ch^{-d} for some positive constant C, where $\sup_{\mathbf{x},\mathbf{y}\in\mathscr{B}} \sup_{\xi\in\mathscr{X}} |g_n(\xi;\mathbf{x},\mathbf{y})\sigma^2(\xi)| \lesssim h^{-d}$ and

$\sup_{\mathbf{x},\mathbf{y}\in\mathscr{B}} \mathbb{E}[g_n(\mathbf{X}_i;\mathbf{x},\mathbf{y})^2\sigma_t(\mathbf{X}_i)^2]^{\frac{1}{2}} \lesssim h^{-d/2}$. Then, similar to the proof of $\mathbf{M}_{1,\mathbf{x},\mathbf{y}}$, we conclude that

$$\sup_{\mathbf{x},\mathbf{y}\in\mathscr{B}} \big|\mathbb{E}_n[g_n(\mathbf{X}_i;\mathbf{x},\mathbf{y})] - \mathbb{E}[g_n(\mathbf{X}_i;\mathbf{x},\mathbf{y})]\big| \lesssim \sqrt{\frac{\log(1/h)}{nh^d}}$$

and

$$\sup_{\mathbf{x},\mathbf{y}\in\mathscr{B}} \|\mathbf{M}_{4,\mathbf{x},\mathbf{y}}\| \lesssim \sqrt{\frac{\log(1/h)}{nh^d}}.$$

Final result. Combining the the upper bounds of the four terms,

$$\sup_{\mathbf{x},\mathbf{y}\in\mathscr{B}} \big\|\widehat{\mathbf{\Sigma}}_{1,\mathbf{x},\mathbf{y}} - \mathbf{\Sigma}_{1,\mathbf{x},\mathbf{y}}\big\| \lesssim_{\mathbb{P}} h^{p+1} + \sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{v}{2+v}}h^d},$$

which implies $\sup_{\mathbf{x},\mathbf{y}\in\mathscr{B}} \|\widehat{\mathbf{\Sigma}}_{1,\mathbf{x},\mathbf{y}}\| \lesssim_{\mathbb{P}} 1$. It follows that

$$\begin{aligned}
\sup_{\mathbf{x},\mathbf{y}\in\mathscr{B}} |\widehat{\Omega}^{(\boldsymbol{\nu})}_{1,\mathbf{x},\mathbf{y}} - \Omega^{(\boldsymbol{\nu})}_{1,\mathbf{x},\mathbf{y}}| &\leq \frac{1}{nh^{d+2|\boldsymbol{\nu}|}}\Big(\sup_{\mathbf{x},\mathbf{y}\in\mathscr{B}} \|\widehat{\mathbf{\Gamma}}^{-1}_{1,\mathbf{x}} - \mathbf{\Gamma}^{-1}_{1,\mathbf{x}}\|\|\widehat{\mathbf{\Sigma}}_{1,\mathbf{x},\mathbf{y}}\|\|\widehat{\mathbf{\Gamma}}^{-1}_{1,\mathbf{y}}\| \\
&\qquad + \sup_{\mathbf{x},\mathbf{y}\in\mathscr{B}} \|\mathbf{\Gamma}^{-1}_{1,\mathbf{x}}\|\|\widehat{\mathbf{\Sigma}}_{1,\mathbf{x},\mathbf{y}} - \mathbf{\Sigma}_{1,\mathbf{x},\mathbf{y}}\|\|\widehat{\mathbf{\Gamma}}^{-1}_{1,\mathbf{y}}\| \\
&\qquad + \sup_{\mathbf{x},\mathbf{y}\in\mathscr{B}} \|\mathbf{\Gamma}^{-1}_{1,\mathbf{x}}\|\|\mathbf{\Sigma}_{1,\mathbf{x},\mathbf{y}}\|\|\widehat{\mathbf{\Gamma}}^{-1}_{1,\mathbf{y}} - \mathbf{\Gamma}^{-1}_{1,\mathbf{y}}\|\Big) \\
&\leq \frac{1}{nh^{d+2|\boldsymbol{\nu}|}}\Big(h^{p+1} + \sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{v}{2+v}}h^d}\Big).
\end{aligned}$$

By Assumption SA–1(iv) and Assumption SA–2(ii), $\inf_{\mathbf{x}\in\mathscr{B}} \Omega^{(\boldsymbol{\nu})}_{\mathbf{x},\mathbf{x}} \gtrsim_{\mathbb{P}} (nh^{d+2|\boldsymbol{\nu}|})^{-1}$. Therefore, $\inf_{\mathbf{x}\in\mathscr{B}} \widehat{\Omega}^{(\boldsymbol{\nu})}_{\mathbf{x},\mathbf{x}} \gtrsim (nh^{d+2|\boldsymbol{\nu}|})^{-1}$. Furthermore,

$$\sup_{\mathbf{x}\in\mathscr{B}} \Big|\sqrt{\widehat{\Omega}^{(\boldsymbol{\nu})}_{\mathbf{x},\mathbf{x}}} - \sqrt{\Omega^{(\boldsymbol{\nu})}_{\mathbf{x},\mathbf{x}}}\Big| \lesssim_{\mathbb{P}} \sup_{\mathbf{x}\in\mathscr{B}} \sqrt{nh^{d+2|\boldsymbol{\nu}|}}\Big|\widehat{\Omega}^{(\boldsymbol{\nu})}_{\mathbf{x},\mathbf{x}} - \Omega^{(\boldsymbol{\nu})}_{\mathbf{x},\mathbf{x}}\Big| \lesssim_{\mathbb{P}} \frac{1}{\sqrt{nh^{d+2|\boldsymbol{\nu}|}}}\Big(h^{p+1} + \sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{v}{2+v}}h^d}\Big)$$

and

$$\sup_{\mathbf{x}\in\mathscr{B}} \left|\frac{h^{-|\boldsymbol{\nu}|}}{\sqrt{\widehat{\Omega}^{(\boldsymbol{\nu})}_{\mathbf{x},\mathbf{x}}}} - \frac{h^{-|\boldsymbol{\nu}|}}{\sqrt{\Omega^{(\boldsymbol{\nu})}_{\mathbf{x},\mathbf{x}}}}\right| = h^{-|\boldsymbol{\nu}|} \sup_{\mathbf{x}\in\mathscr{B}} \left|\frac{\sqrt{\widehat{\Omega}^{(\boldsymbol{\nu})}_{\mathbf{x},\mathbf{x}}} - \sqrt{\Omega^{(\boldsymbol{\nu})}_{\mathbf{x},\mathbf{x}}}}{\sqrt{\widehat{\Omega}^{(\boldsymbol{\nu})}_{\mathbf{x},\mathbf{x}}\Omega^{(\boldsymbol{\nu})}_{\mathbf{x},\mathbf{x}}}}\right| \lesssim_{\mathbb{P}} \sqrt{nh^d}\Big(h^{p+1} + \sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{v}{2+v}}h^d}\Big),$$

which completes the proof. □

SA-7.5 Proof of Lemma SA-5

Define

$$\boldsymbol{\chi}_{t,\mathbf{x}} = \mathbb{E}_n\Big[\mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{x}}{h}\Big)K_h(\mathbf{X}_i - \mathbf{x})\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)\mathfrak{r}_t(\mathbf{X}_i;\mathbf{x})\Big], \qquad \mathfrak{r}_t(\xi;\mathbf{x}) = \mu_t(\xi) - \sum_{0\leq|\boldsymbol{\omega}|\leq p} \frac{\mu_t^{(\boldsymbol{\omega})}(\mathbf{x})}{\boldsymbol{\omega}!}(\xi - \mathbf{x})^{\boldsymbol{\omega}}.$$

Since μ_t is $(p+1)$-times continuously differentiable, there exists $\boldsymbol{\alpha}_{\mathbf{x},\mathbf{X}_i,t} \in \mathbb{R}^{p+1}$ such that

$$\begin{aligned}
\|\boldsymbol{\chi}_{t,\mathbf{x}}\|^2 &= \Big\| \frac{1}{n} \sum_{i=1}^n \mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{x}}{h}\Big) K_h(\mathbf{X}_i - \mathbf{x}) \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t) \mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{x}}{h}\Big)^\top (\mathbf{0}^\top, \boldsymbol{\alpha}_{\mathbf{x},\mathbf{X}_i,t}^\top)^\top \Big\|^2 h^{2(p+1)} \\
&\leq \Big(\mathbb{E}_n\Big[\Big\|\mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{x}}{h}\Big) K_h(\mathbf{X}_i - \mathbf{x}) \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t) \mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{x}}{h}\Big)^\top \Big\|^2\Big]\Big) \Big(\mathbb{E}_n\big[\|\boldsymbol{\alpha}_{\mathbf{x},\mathbf{X}_i,t}\|^2\big]\Big) h^{2(p+1)},
\end{aligned}$$

where $\sup_{\mathbf{x}\in\mathscr{B}} \max_{t\in\{0,1\}} \max_{1\leq i\leq n} \|\boldsymbol{\alpha}_{\mathbf{x},\mathbf{X}_i,t}\| \lesssim 1$. Since $\frac{\log(1/h)}{nh^d} = o(1)$, the same argument as the proof of Lemma SA-2 shows

$$\mathbb{E}_n\Big[\Big\|\mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{x}}{h}\Big) K_h(\mathbf{X}_i - \mathbf{x}) \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t) \mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{x}}{h}\Big)^\top \Big\|^2\Big] \lesssim_{\mathbb{P}} 1.$$

It then follows from Lemma SA-2 that

$$\sup_{\mathbf{x}\in\mathscr{B}} \big|\mathbb{E}[\widehat{\mu}_t^{(\boldsymbol{\nu})}(\mathbf{x})|\mathbf{X}] - \mu_t^{(\boldsymbol{\nu})}(\mathbf{x})\big| = \sup_{\mathbf{x}\in\mathscr{B}} \big|\mathbf{e}_{1+\boldsymbol{\nu}}^\top \mathbf{H}^{-1} \widehat{\boldsymbol{\Gamma}}_{t,\mathbf{x}}^{-1} \boldsymbol{\chi}_{t,\mathbf{x}}\big| \lesssim_{\mathbb{P}} h^{p+1-|\boldsymbol{\nu}|}.$$

Now also assume that $h = o(1)$. Then, for all $\mathbf{x} \in \mathscr{B}$ and $\xi \in \mathscr{X}$,

$$\mathbb{1}\left(K_h(\xi - \mathbf{x}) \neq 0\right) \left|\gamma_{\mathbf{v}}(\xi; \mathbf{x}) - \frac{|\mathbf{v}|}{\mathbf{v}!} \partial^{\mathbf{v}} \mu_t(\mathbf{x})\right| \leq \frac{|\mathbf{v}|}{\mathbf{v}!} \sup_{\|\mathbf{u}-\mathbf{u}'\|\leq h} |\partial^{\mathbf{v}} \mu_t(\mathbf{u}) - \partial^{\mathbf{v}} \mu_t(\mathbf{u}')| = M_n,$$

where $\gamma_{\mathbf{v}}(\xi; \mathbf{x}) = \frac{|\mathbf{v}|}{\mathbf{v}!} \int_0^1 (1-t)^{|\mathbf{v}|-1} \partial_{\mathbf{v}} \mu_t(\mathbf{x} + t(\xi - \mathbf{x})) dt$. By Assumption SA–1(iii), $\partial^{\mathbf{v}} \mu_t$ is uniformly continuous on the compact set $\mathscr{X}$. This implies that when $h = o(1)$, $M_n = o(1)$. Letting

$$\widetilde{\boldsymbol{\chi}}_{t,\mathbf{x}} = \mathbb{E}_n\Big[\mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{x}}{h}\Big) K_h(\mathbf{X}_i - \mathbf{x}) \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t) (\sum_{|\mathbf{v}|=p+1} \frac{|\mathbf{v}|}{\mathbf{v}!} \partial^{\mathbf{v}} \mu_t(\mathbf{x}) (\mathbf{X}_i - \mathbf{x})^{\mathbf{v}})\Big],$$

we conclude that

$$\sup_{\mathbf{x}\in\mathscr{B}} \|\boldsymbol{\chi}_{t,\mathbf{x}} - \widetilde{\boldsymbol{\chi}}_{t,\mathbf{x}}\| \lesssim M_n \sup_{\mathbf{x}\in\mathscr{B}} \Big\|\mathbb{E}_n\Big[\mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{x}}{h}\Big) K_h(\mathbf{X}_i - \mathbf{x}) \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t) \Big(\sum_{|\mathbf{v}|=p+1} \frac{|\mathbf{v}|}{\mathbf{v}!} |\mathbf{X}_i - \mathbf{x}|^{\mathbf{v}}\Big)\Big]\Big\| = o_{\mathbb{P}}(h^{p+1}),$$

where the last equality employs the same arguments as in the proof of Lemma SA-2. Hence,

$$\sup_{\mathbf{x}\in\mathscr{B}} \left|\mathbb{E}[\widehat{\mu}_t^{(\boldsymbol{\nu})}(\mathbf{x})|\mathbf{X}] - \mu_t^{(\boldsymbol{\nu})}(\mathbf{x}) - h^{p+1-|\boldsymbol{\nu}|} \widehat{B}_{t,\mathbf{x}}^{(\boldsymbol{\nu})}\right| = \sup_{\mathbf{x}\in\mathscr{B}} \left|\mathbf{e}_{1+\boldsymbol{\nu}}^\top \mathbf{H}^{-1} \widehat{\boldsymbol{\Gamma}}_{t,\mathbf{x}}^{-1} \boldsymbol{\chi}_{t,\mathbf{x}} - \mathbf{e}_{1+\boldsymbol{\nu}}^\top \mathbf{H}^{-1} \widehat{\boldsymbol{\Gamma}}_{t,\mathbf{x}}^{-1} \widetilde{\boldsymbol{\chi}}_{t,\mathbf{x}}\right| = o_{\mathbb{P}}(h^{p+1-|\boldsymbol{\nu}|}).$$

Using Lemma SA-2 and the maximal inequality as in the proof of Lemma SA-2, we conclude that

$$\max_{t\in\{0,1\}} \sup_{\mathbf{x}\in\mathscr{B}} |\widehat{B}_{t,\mathbf{x}}^{(\boldsymbol{\nu})} - B_{t,\mathbf{x}}^{(\boldsymbol{\nu})}| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}}.$$

Since $\max_{t\in\{0,1\}} \sup_{\mathbf{x}\in\mathscr{B}} |B_{t,\mathbf{x}}^{(\boldsymbol{\nu})}| \lesssim 1$, it follows that $\max_{t\in\{0,1\}} \sup_{\mathbf{x}\in\mathscr{B}} |\widehat{B}_{t,\mathbf{x}}^{(\boldsymbol{\nu})}| \lesssim_{\mathbb{P}} 1$. □

SA-7.6 Proof of Theorem SA-1

The results follow from Lemma SA-5 and Lemma SA-3. □

SA-7.7 Proof of Theorem SA-2

For the conditional bias, by Lemma SA-5,

$$\begin{aligned}
&\sup_{\mathbf{x}\in\mathscr{B}} \big|\mathbb{E}\big[\widehat{\tau}^{(\boldsymbol{\nu})}(\mathbf{x}) - \tau^{(\boldsymbol{\nu})}(\mathbf{x})\big|\mathbf{X}\big]^2 - (h^{p+1-|\boldsymbol{\nu}|}B_{\mathbf{x}}^{(\boldsymbol{\nu})})^2\big| \\
&\leq \sup_{\mathbf{x}\in\mathscr{B}} \big|\mathbb{E}\big[\widehat{\tau}^{(\boldsymbol{\nu})}(\mathbf{x}) - \tau^{(\boldsymbol{\nu})}(\mathbf{x})\big|\mathbf{X}\big] - h^{p+1-|\boldsymbol{\nu}|}B_{\mathbf{x}}^{(\boldsymbol{\nu})}\big| \cdot \sup_{\mathbf{x}\in\mathscr{B}} \big|\mathbb{E}\big[\widehat{\tau}^{(\boldsymbol{\nu})}(\mathbf{x}) - \tau^{(\boldsymbol{\nu})}(\mathbf{x})\big|\mathbf{X}\big] + h^{p+1-|\boldsymbol{\nu}|}B_{\mathbf{x}}^{(\boldsymbol{\nu})}\big| \\
&= o_{\mathbb{P}}(h^{p+1-|\boldsymbol{\nu}|}).
\end{aligned}$$

Since $\sup_{\mathbf{x}\in\mathscr{B}} |B_{t,\mathbf{x}}^{(\boldsymbol{\nu})} - B_{t,\mathbf{x}}^{(\boldsymbol{\nu})}| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}}$ from Lemma SA-5,

$$\sup_{\mathbf{x}\in\mathscr{B}} \big|\mathbb{E}\big[\widehat{\tau}^{(\boldsymbol{\nu})}(\mathbf{x}) - \tau^{(\boldsymbol{\nu})}(\mathbf{x})\big|\mathbf{X}\big]^2 - (h^{p+1-|\boldsymbol{\nu}|}B_{\mathbf{x}}^{(\boldsymbol{\nu})})^2\big| = o_{\mathbb{P}}(h^{p+1-|\boldsymbol{\nu}|}).$$

For the conditional variance, by Lemma SA-4,

$$\sup_{\mathbf{x}\in\mathscr{B}} \big|\mathbb{V}\big[\widehat{\tau}^{(\boldsymbol{\nu})}(\mathbf{x})\big|\mathbf{X}\big] - (nh^{d+2|\boldsymbol{\nu}|})^{-1}V_{\mathbf{x}}^{(\boldsymbol{\nu})}\big| = o_{\mathbb{P}}((nh^{d+2|\boldsymbol{\nu}|})^{-1}).$$

The pointwise MSE expansion follows directly. For the IMSE expansion, notice that

$$\begin{aligned}
&\Big|\mathrm{IMSE}_{\boldsymbol{\nu}} - \int_{\mathscr{B}} \big[(h^{p+1-|\boldsymbol{\nu}|}B_{\mathbf{x}}^{(\boldsymbol{\nu})})^2 + (nh^{d+2|\boldsymbol{\nu}|})^{-1}V_{\mathbf{x}}^{(\boldsymbol{\nu})}\big]w(\mathbf{x})d\mathfrak{H}^{d-1}(\mathbf{x})\Big| \\
&\leq \int_{\mathscr{B}} |w(\mathbf{x})|d\mathfrak{H}^{d-1}(\mathbf{x}) \cdot \sup_{\mathbf{x}\in\mathscr{B}} \big|\mathrm{MSE}_{\boldsymbol{\nu}}(\mathbf{x}) - (h^{p+1-|\boldsymbol{\nu}|}B_{\mathbf{x}}^{(\boldsymbol{\nu})})^2 - (nh^{d+2|\boldsymbol{\nu}|})^{-1}V_{\mathbf{x}}^{(\boldsymbol{\nu})}\big| \\
&= o_{\mathbb{P}}\big(h^{2p+2-2|\boldsymbol{\nu}|} + (nh^{d+2|\boldsymbol{\nu}|})^{-1}\big),
\end{aligned}$$

which completes the proof. □

SA-7.8 Proof of Theorem SA-3

We have $\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x}) = \sum_{i=1}^n Z_i$ with

$$Z_i = \sum_{t\in\{0,1\}} n^{-1}(\Omega_{\mathbf{x},\mathbf{x}}^{(\boldsymbol{\nu})})^{-1/2}\mathbf{e}_{1+\boldsymbol{\nu}}^\top \mathbf{H}^{-1}\boldsymbol{\Gamma}_{t,\mathbf{x}}^{-1}\mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{x}}{h}\Big)K_h(\mathbf{X}_i - \mathbf{x})\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)u_i,$$

where $\mathbb{E}[Z_i] = 0$ and $\mathbb{V}[Z_i] = n^{-1}$. By the Berry-Essen Theorem,

$$\sup_{u\in\mathbb{R}} \Big|\mathbb{P}\big(\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x}) \leq u\big) - \Phi(u)\Big| \lesssim B_n^{-1}\sum_{i=1}^n \mathbb{E}[|Z_i|^3],$$

where $B_n = \sum_{i=1}^n \mathbb{V}[Z_i] = 1$. Moreover,

$$\begin{aligned}
\sum_{i=1}^n \mathbb{E}[|Z_i|^3] &= n^{-3}(\Omega_{\mathbf{x},\mathbf{x}}^{(\boldsymbol{\nu})})^{-3/2}\sum_{i=1}^n \mathbb{E}\Big[\Big|\sum_{t\in\{0,1\}} \mathbf{e}_{1+\boldsymbol{\nu}}^\top \mathbf{H}^{-1}\boldsymbol{\Gamma}_{t,\mathbf{x}}^{-1}\mathbf{r}_p\left(\frac{\mathbf{X}_i - \mathbf{x}}{h}\right)K_h\left(\mathbf{X}_i - \mathbf{x}\right)\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)u_i\Big|^3\Big] \\
&\lesssim n^{-3}(\Omega_{\mathbf{x},\mathbf{x}}^{(\boldsymbol{\nu})})^{-3/2}\sum_{i=1}^n \mathbb{E}\Big[\Big|\sum_{t\in\{0,1\}} \mathbf{e}_{1+\boldsymbol{\nu}}^\top \mathbf{H}^{-1}\boldsymbol{\Gamma}_{t,\mathbf{x}}^{-1}\mathbf{r}_p\left(\frac{\mathbf{X}_i - \mathbf{x}}{h}\right)K_h\left(\mathbf{X}_i - \mathbf{x}\right)\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)\Big|^3\Big]
\end{aligned}$$

$$
\begin{aligned}
&\lesssim n^{-2}h^{-|\boldsymbol{\nu}|-d}(\Omega_{\mathbf{x},\mathbf{x}}^{(\boldsymbol{\nu})})^{-3/2}\mathbb{E}\Big[\Big|\sum_{t\in\{0,1\}}\mathbf{e}_{1+\boldsymbol{\nu}}^{\top}\mathbf{H}^{-1}\boldsymbol{\Gamma}_{t,\mathbf{x}}^{-1}\mathbf{r}_p\left(\frac{\mathbf{X}_i-\mathbf{x}}{h}\right)K_h\left(\mathbf{X}_i-\mathbf{x}\right)\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_t)\Big|^2\Big]\\
&\lesssim n^{-1}h^{-|\boldsymbol{\nu}|-d}(\Omega_{\mathbf{x},\mathbf{x}}^{(\boldsymbol{\nu})})^{-1/2}\\
&\lesssim (nh^d)^{-1/2},
\end{aligned}
$$

where the second line uses Assumption SA–1(v), the third line uses

$$
\Big|\sum_{t\in\{0,1\}}\mathbf{e}_{1+\boldsymbol{\nu}}^{\top}\mathbf{H}^{-1}\boldsymbol{\Gamma}_{t,\mathbf{x}}^{-1}\mathbf{r}_p\left(\frac{\mathbf{X}_i-\mathbf{x}}{h}\right)K_h\left(\mathbf{X}_i-\mathbf{x}\right)\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_t)\Big|\lesssim h^{-|\boldsymbol{\nu}|-d}
$$

and the fourth line uses the definition of $\Omega_{\mathbf{x},\mathbf{x}}^{(\boldsymbol{\nu})}$.

Finally, although Lemma SA-2 through Lemma SA-4 provide convergence results uniformly in $\mathbf{x}$, for pointwise results with fix $\mathbf{x}\in\mathscr{B}$, we can replace the class of functions in those proofs by one containing a *singleton* (corresponding to the evaluation point $\mathbf{x}$). Thus, we obtain the following result:

$$
\left|\widehat{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})-\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})\right|\lesssim_{\mathbb{P}} h^{p+1}\sqrt{nh^d}+1/\sqrt{nh^d}+1/(n^{\frac{v}{2+v}}h^d), \tag{SA-4}
$$

provided that $h^{p+1}\sqrt{nh^d}\to 0$ and $n^{\frac{v}{2+v}}h^d\to 0$.

The final results follow by weak convergence to a Gaussian distribution, and properties of the distribution function. □

SA-7.9 Proof of Theorem SA-4

For all $\mathbf{x}\in\mathscr{B}$, we have $\widehat{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})=\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})+G_1^{(\boldsymbol{\nu})}(\mathbf{x})+G_2^{(\boldsymbol{\nu})}(\mathbf{x})$, where

$$
G_1^{(\boldsymbol{\nu})}(\mathbf{x})=\Big(\mathbb{E}\big[\widehat{\tau}^{(\boldsymbol{\nu})}(\mathbf{x})\big|\mathbf{X}\big]-\tau^{(\boldsymbol{\nu})}(\mathbf{x})\Big)(\widehat{\Omega}_{\mathbf{x},\mathbf{x}}^{(\boldsymbol{\nu})})^{-1/2},
$$

and

$$
G_2^{(\boldsymbol{\nu})}(\mathbf{x})=\mathbf{e}_{1+\boldsymbol{\nu}}^{\top}\mathbf{H}^{-1}\Big[\big(\widehat{\boldsymbol{\Gamma}}_{1,\mathbf{x}}^{-1}\mathbf{Q}_{1,\mathbf{x}}-\widehat{\boldsymbol{\Gamma}}_{0,\mathbf{x}}^{-1}\mathbf{Q}_{0,\mathbf{x}}\big)(\widehat{\Omega}_{\mathbf{x},\mathbf{x}}^{(\boldsymbol{\nu})})^{-\frac{1}{2}}-\big(\boldsymbol{\Gamma}_{1,\mathbf{x}}^{-1}\mathbf{Q}_{1,\mathbf{x}}-\boldsymbol{\Gamma}_{0,\mathbf{x}}^{-1}\mathbf{Q}_{0,\mathbf{x}}\big)(\Omega_{\mathbf{x},\mathbf{x}}^{(\boldsymbol{\nu})})^{-\frac{1}{2}}\Big].
$$

By Lemma SA-5 and Lemma SA-4,

$$
\sup_{\mathbf{x}\in\mathscr{B}}\big|G_1^{(\boldsymbol{\nu})}(\mathbf{x})\big|\lesssim_{\mathbb{P}} h^{p+1-|\boldsymbol{\nu}|}(nh^{d+2|\boldsymbol{\nu}|})^{1/2}\lesssim h^{p+1}\sqrt{nh^d}.
$$

By Lemma SA-2, Lemma SA-3 and Lemma SA-4,

$$
\sup_{\mathbf{x}\in\mathscr{B}}\Big|e_{1+\boldsymbol{\nu}}^{\top}\mathbf{H}^{-1}\big[\widehat{\boldsymbol{\Gamma}}_{t,\mathbf{x}}^{-1}-\boldsymbol{\Gamma}_{t,\mathbf{x}}^{-1}\big]\mathbf{Q}_{t,\mathbf{x}}(\widehat{\Omega}_{\mathbf{x},\mathbf{x}}^{(\boldsymbol{\nu})})^{-1/2}\Big|\lesssim_{\mathbb{P}}\sqrt{\log(1/h)}\bigg(\sqrt{\frac{\log(1/h)}{nh^d}}+\frac{\log(1/h)}{n^{\frac{1+v}{2+v}}h^d}\bigg)
$$

and

$$
\begin{aligned}
&\sup_{\mathbf{x}\in\mathscr{B}}\Big|e_{1+\boldsymbol{\nu}}^{\top}\mathbf{H}^{-1}\boldsymbol{\Gamma}_{t,\mathbf{x}}^{-1}\mathbf{Q}_{t,\mathbf{x}}\Big[(\widehat{\Omega}_{\mathbf{x},\mathbf{x}}^{(\boldsymbol{\nu})})^{-1/2}-(\Omega_{\mathbf{x},\mathbf{x}}^{(\boldsymbol{\nu})})^{-1/2}\Big]\Big|\\
&\lesssim_{\mathbb{P}} h^{-|\boldsymbol{\nu}|}\cdot\bigg(\sqrt{\frac{\log(1/h)}{nh^d}}+\frac{\log(1/h)}{n^{\frac{1+v}{2+v}}h^d}\bigg)\cdot\sqrt{nh^{d+2\boldsymbol{\nu}}}\bigg(\sqrt{\frac{\log(1/h)}{nh^d}}+\frac{\log(1/h)}{n^{\frac{v}{2+v}}h^d}+h^{p+1}\bigg)
\end{aligned}
$$

$$\lesssim \frac{\log(1/h)}{\sqrt{nh^d}} + \frac{(\log(1/h))^{3/2}}{n^{\frac{v}{2+v}} h^d}.$$

The result now follows from combining the bounds above. $\square$

SA-7.10 Proof of Theorem SA-5

We verify the high-level conditions of Theorem SA-11. We will employ the following technical lemma.

Lemma SA-8 (VC Class to VC2 Class)**.** *Assume $\mathscr{F}$ is a VC class on a measure space $(\mathcal{X}, \mathscr{B})$: there exists an envelope function F and positive constants $c(\mathscr{F}), d(\mathscr{F})$ such that for all $\varepsilon \in (0,1)$,*

$$\sup_Q N(\mathscr{F}, \|\cdot\|_{Q,1}, \varepsilon \|F\|_{Q,1}) \leq c(\mathscr{F})\varepsilon^{-d(\mathscr{F})},$$

where the supremum is taken over all finite discrete measures. Then, $\mathscr{F}$ is also VC2 class: for all $\varepsilon \in (0,1)$,

$$\sup_Q N(\mathscr{F}, \|\cdot\|_{Q,2}, \varepsilon \|F\|_{Q,2}) \leq c(\mathscr{F})(\varepsilon^2/2)^{-d(\mathscr{F})},$$

where the supremum is taken over all finite discrete measures.

Proof of Lemma SA-8. Let Q be a finite discrete probability measure. Let $f, g \in \mathscr{F}$. Then, $\int |f-g|^2 dQ \leq 2\int |f-g||F|dQ$. Define another probability measure $\tilde{Q}(c_k) = F(c_k)Q(c_k)/\|F\|_{Q,1}$ on the support of Q, denoted by $\{c_1, \dots, c_k, \dots\}$. Then,

$$\int |f-g|^2 dQ \leq 2\|F\|_{Q,1} \int |f-g| d\tilde{Q} \leq 2\|F\|_{Q,1} \|f-g\|_{\tilde{Q},1}.$$

Hence, if we take an $\varepsilon^2/2$-net in $(\mathscr{F}, \|\cdot\|_{\tilde{Q},1})$ with cardinality no greater than $c(\mathscr{F})\varepsilon^{-d(\mathscr{F})}$, then for any $f \in \mathscr{F}$, there exists a $g \in \mathscr{F}$ such that $\|f-g\|_{\tilde{Q},1} \leq \varepsilon^2/2 \|F\|_{\tilde{Q},1}$, and hence

$$\|f-g\|_{Q,2}^2 \leq 2\varepsilon^2/2 \|F\|_{Q,1} \|F\|_{\tilde{Q},1} \leq \varepsilon^2 \|F\|_{Q,2}^2,$$

which gives the result. $\square$

Without loss of generality, we assume $\mathcal{X} = [0,1]^d$, and $\mathbb{Q}_{\mathscr{F}_t} = \mathbb{P}_X$ is a valid surrogate measure for $\mathbb{P}_X$ with respect to $\mathscr{F}_t$, and $\phi_{\mathscr{F}_t} = \mathrm{Id}$ is a valid normalizing transformation (as in). This implies the constants $\mathtt{c}_1$ and $\mathtt{c}_2$ from Theorem SA-11 are all 1.

Consider first the class of functions $\mathscr{F}_t = \{\mathscr{K}_t^{(\boldsymbol{\nu})}(\cdot; \mathbf{x}) : \mathbf{x} \in \mathscr{B}\}$, for $t \in \{0,1\}$.

Envelope Function. By Lemma SA-2 and Lemma SA-4 and the fact that $\mathrm{Supp}(K)$ is compact,

$$\sup_{\mathbf{x}\in\mathscr{B}} \sup_{\xi\in\mathcal{X}} \left|\mathscr{K}_t^{(\boldsymbol{\nu})}(\xi; \mathbf{x})\right| \lesssim \frac{1}{\sqrt{n h^{d+\boldsymbol{\nu}}}} \sup_{\mathbf{x}\in\mathscr{B}} \left(\|\boldsymbol{\Gamma}_{1,\mathbf{x}}^{-1}\| + \|\boldsymbol{\Gamma}_{0,\mathbf{x}}^{-1}\|\right) \sup_{\mathbf{x}\in\mathscr{B}} \left|(\Omega_{\mathbf{x},\mathbf{x}}^{(\boldsymbol{\nu})})^{-\frac{1}{2}}\right| \lesssim h^{-d/2}.$$

Hence, there exists a constant $C_1 > 0$ such that $\mathtt{M}_{\mathscr{F}_t} = C_1 h^{-d/2}$ is a constant envelope function.

L_1 Bound. We have $\mathtt{E}_{\mathscr{F}_t} = \sup_{\mathbf{x}\in\mathscr{B}} \mathbb{E}[|\mathscr{K}_t^{(\boldsymbol{\nu})}(\mathbf{X}_i; \mathbf{x})|] \lesssim h^{d/2}$.

Uniform Variation. Case 1: K is Lipschitz. By Assumption SA–1(iv) and Assumption SA–2,

$$\mathtt{L}_{\mathscr{F}_t} = \sup_{\mathbf{x}\in\mathscr{B}} \sup_{\xi,\xi'\in\mathcal{X}} \frac{|\mathscr{K}_t^{(\boldsymbol{\nu})}(\xi;\mathbf{x}) - \mathscr{K}_t^{(\boldsymbol{\nu})}(\xi';\mathbf{x})|}{\|\xi-\xi'\|_\infty} \lesssim h^{-d/2-1}.$$

Each entry of $\boldsymbol{\Gamma}_{t,\mathbf{x}}$ and $\boldsymbol{\Sigma}_{t,\mathbf{x}}$ are of the form $\int (\frac{\xi-\mathbf{x}}{h})^{\mathbf{u}+\mathbf{v}} K_h(\xi-\mathbf{x})\mathbb{1}(\xi\in\mathscr{A}_t)f(\xi)d\xi$ and $\int (\frac{\xi-\mathbf{x}}{h})^{\mathbf{u}+\mathbf{v}} K_h(\xi-\mathbf{x})\sigma_t(\xi)^2\mathbb{1}(\xi\in\mathscr{A}_t)d\xi$ for some multi-index $\mathbf{u}$ and $\mathbf{v}$, respectively. Hence, by Assumption SA–2, each entry of $\boldsymbol{\Gamma}_{t,\mathbf{x}}$ and $\boldsymbol{\Sigma}_{t,\mathbf{x}}$ are h^{-1}-Lipschitz in $\mathbf{x}$. It follows that there exists a constant C_2 such that for all $\mathbf{x},\mathbf{x}'\in\mathscr{B}$,

$$\left\|\boldsymbol{\Gamma}_{t,\mathbf{x}}^{-1} - \boldsymbol{\Gamma}_{t,\mathbf{x}'}^{-1}\right\| \le \|\boldsymbol{\Gamma}_{t,\mathbf{x}}^{-1}\|\|\boldsymbol{\Gamma}_{t,\mathbf{x}} - \boldsymbol{\Gamma}_{t,\mathbf{x}}\|\|\boldsymbol{\Gamma}_{t,\mathbf{x}'}^{-1}\| \le C_2 h^{-1}\left\|\mathbf{x}-\mathbf{x}'\right\|.$$

Also, by definition of $\Omega_{t,\mathbf{x}}$ and Assumption SA–2(iv), there exists C_3 such that for all $\mathbf{x},\mathbf{x}'\in\mathcal{X}$,

$$\left|\Omega_{t,\mathbf{x}}^{(\boldsymbol{\nu})} - \Omega_{t,\mathbf{x}'}^{(\boldsymbol{\nu})}\right| \le C_3(nh^{d+2|\boldsymbol{\nu}|+1})^{-1}\|\mathbf{x}-\mathbf{x}'\|_\infty,$$

and

$$\left|(\Omega_{t,\mathbf{x}}^{(\boldsymbol{\nu})})^{-1/2} - (\Omega_{t,\mathbf{x}'}^{(\boldsymbol{\nu})})^{-1/2}\right| \le \frac{1}{2}\inf_{\mathbf{z}\in\mathcal{X}}(\Omega_{t,\mathbf{z}}^{(\boldsymbol{\nu})})^{-3/2}\left|\Omega_{t,\mathbf{x}}^{(\boldsymbol{\nu})} - \Omega_{t,\mathbf{x}'}^{(\boldsymbol{\nu})}\right| \le \frac{1}{2}C_3 h^{-1}(nh^{d+2|\boldsymbol{\nu}|})^{1/2}\|\mathbf{x}-\mathbf{x}'\|_\infty.$$

It then follows that we have a uniform Lipschitz property with respect to the point of evaluation:

$$\mathtt{l}_{\mathscr{F}_t} = \sup_{\xi\in\mathcal{X}} \sup_{\mathbf{x},\mathbf{x}'\in\mathscr{B}} \frac{\left|\mathscr{K}_t^{(\boldsymbol{\nu})}(\xi;\mathbf{x}) - \mathscr{K}_t^{(\boldsymbol{\nu})}(\xi;\mathbf{x}')\right|}{\|\mathbf{x}-\mathbf{x}'\|_\infty} \lesssim h^{-d/2-1}.$$

Let $\mathbf{x}\in\mathscr{B}$. Then, $\mathscr{K}_t^{(\boldsymbol{\nu})}(\cdot;\mathbf{x})$ is supported on $\mathbf{x}+\mathbf{c}[-h,h]^d$. Then,

$$\mathtt{TV}_{\mathscr{F}_t} \lesssim \mathfrak{m}\big(\mathbf{c}[-h,h]^d\big)\mathtt{L}_{\mathscr{F}_t} \lesssim h^{d/2-1}$$

Case 2: $K=\mathbb{1}(\cdot\in[-1,1]^d)$. Consider

$$\tilde{\mathscr{K}}_t^{(\boldsymbol{\nu})}(\mathbf{u};\mathbf{x}) = n^{-1/2}(\Omega_{\mathbf{x},\mathbf{x}}^{(\boldsymbol{\nu})})^{-1/2}\mathbf{e}_{1+\boldsymbol{\nu}}^\top\mathbf{H}^{-1}\boldsymbol{\Gamma}_{t,\mathbf{x}}^{-1}\mathbf{r}_p\Big(\frac{\mathbf{u}-\mathbf{x}}{h}\Big)h^{-d}, \qquad \mathbf{u}\in\mathcal{X}, \quad t\in\{0,1\}.$$

Then, $\mathscr{K}^{(\boldsymbol{\nu})}(\mathbf{u};\mathbf{x}) = \tilde{\mathscr{K}}^{(\boldsymbol{\nu})}(\mathbf{u};\mathbf{x})\mathbb{1}(\mathbf{u}-\mathbf{x}\in[-1,1]^d)$ for all $\mathbf{u}\in\mathcal{X}$ and $\mathbf{x}\in\mathscr{B}$, and we set $\tilde{\mathscr{F}}_t=\{\tilde{\mathscr{K}}^{(\boldsymbol{\nu})}(\cdot;\mathbf{x}):\mathbf{x}\in\mathscr{B}\}$, $t\in\{0,1\}$. Then, the argument above implies that $\mathtt{TV}_{\tilde{\mathscr{F}}_t}\lesssim\mathfrak{m}\left(\mathbf{c}[-h,h]^d\right)\mathtt{L}_{\mathscr{F}_t}\lesssim h^{d/2-1}$. Next, set $\mathscr{L}=\{\mathbb{1}((\cdot-\mathbf{x})/h\in[-1,1]^d):\mathbf{x}\in\mathscr{B}\}$. Then, using a product rule, we have

$$\mathtt{TV}_{\mathscr{F}_t} \le \mathtt{TV}_{\tilde{\mathscr{F}}_t}\mathtt{M}_{\mathscr{L}} + \mathtt{M}_{\tilde{\mathscr{F}}_t}\mathtt{TV}_{\mathscr{L}} \lesssim h^{d/2-1}\cdot 1 + h^{-d/2}h^{d-1} \lesssim h^{d/2-1}.$$

VC-type Class. Case 1: K is Lipschitz. We apply Cattaneo et al. [2024, Lemma 7]. To make the notation consistent, define

$$f_{\mathbf{x}}(\cdot) = \frac{1}{\sqrt{n\Omega_{\mathbf{x},\mathbf{x}}^{(\boldsymbol{\nu})}}}\mathbf{e}_{1+\boldsymbol{\nu}}^\top\mathbf{H}^{-1}\boldsymbol{\Gamma}_t^{-1}\mathbf{r}_p(\cdot)\,K(\cdot)\,, \qquad \mathbf{x}\in\mathscr{B},$$

and $\mathscr{H}=\{g_{\mathbf{x}}\left(\frac{\cdot-\mathbf{x}}{h}\right):\mathbf{x}\in\mathscr{B}\}$. Notice that $f_{\mathbf{x}}(\frac{\cdot-\mathbf{x}}{h}) = h^d\frac{1}{\sqrt{n\Omega_{\mathbf{x},\mathbf{x}}^{(\boldsymbol{\nu})}}}\mathbf{e}_{1+\boldsymbol{\nu}}^\top\mathbf{H}^{-1}\boldsymbol{\Gamma}^{-1}\mathbf{r}_p(\frac{\cdot-\mathbf{x}}{h})K_h(\cdot-\mathbf{x})$. Then, the following conditions in Cattaneo et al. [2024, Lemma 7] hold (for $\mathbf{z},\mathbf{z}',\mathbf{z}''\in\mathcal{X}$):

(i) boundedness: $\sup_{\mathbf{z}} \sup_{\mathbf{z}'} |f_{\mathbf{z}}(\mathbf{z}')| \leq \mathbf{c}$,

(ii) compact support: $\operatorname{supp}(f_{\mathbf{z}}(\cdot)) \subseteq [-\mathbf{c}, \mathbf{c}]^d$,

(iii) Lipschitz continuity: $\sup_{\mathbf{z}} |f_{\mathbf{z}}(\mathbf{z}') - f_{\mathbf{z}}(\mathbf{z}'')| \leq \mathbf{c}|\mathbf{z}' - \mathbf{z}''|$ and $\sup_{\mathbf{z}} |f_{\mathbf{z}'}(\mathbf{z}) - f_{\mathbf{z}''}(\mathbf{z})| \leq \mathbf{c}h^{-1}|\mathbf{z}' - \mathbf{z}''|$,

and therefore there exists a constant $\mathbf{c}'$ only depending on $\mathbf{c}$ and d that for any $0 \leq \varepsilon \leq 1$,

$$\sup_Q N\big(\mathscr{H}, \|\cdot\|_{Q,1}, (2c+1)^{d+1}\varepsilon\big) \leq \mathbf{c}'\varepsilon^{-d-2} + 1,$$

where the supremum is taken over all finite discrete measures on $\mathcal{X} = [0,1]^d$. It then follows from Lemma SA-8 that with the constant envelope function $\mathtt{M}_{\mathscr{F}_t} = h^{-d/2}$, for any $0 \leq \varepsilon \leq 1$,

$$\sup_Q N\big(\mathscr{F}_t, \|\cdot\|_{Q,2}, (2c+1)^{d+1}\varepsilon \mathtt{M}_{\mathscr{F}_t}\big) \leq \mathbf{c}' 2^{2d+4}\varepsilon^{-2d-4} + 1,$$

where the supremum is taken over all finite discrete measures.

Case 2: Suppose $K = \mathbb{1}(\cdot \in [-1,1]^d)$. Recall $\tilde{\mathscr{F}}_t$ and $\mathscr{L}$ defined in the analysis of *uniform variation*. The same argument as before shows

$$\sup_Q N\big(\tilde{\mathscr{F}}_t, \|\cdot\|_{Q,2}, (2c+1)^{d+1}\varepsilon \mathtt{M}_{\tilde{\mathscr{F}}_t}\big) \leq \mathbf{c}' 2^{2d+4}\varepsilon^{-2d-4} + 1, \qquad \varepsilon \in (0,1],$$

where the supremum is taken over all finite discrete measures, and $\tilde{\mathscr{F}}_t = h^{-d/2}$. By van der Vaart and Wellner [1996, Example 2.6.1], the class $\mathscr{L} = \{\mathbb{1}((\cdot - \mathbf{x})/h \in [-1,1]^d) : \mathbf{x} \in \mathscr{B}\}$ has VC dimension no greater than $2d$, and by van der Vaart and Wellner [1996, Theorem 2.6.4],

$$\sup_Q N(\mathscr{L}, \|\cdot\|_{Q,2}, \varepsilon) \leq 2d(4e)^{2d}\varepsilon^{-4d}, \qquad 0 < \varepsilon \leq 1,$$

where the supremum is taken over all finite discrete measures on $\mathcal{X} = [0,1]^d$. Putting together, we have

$$\sup_Q N(\mathscr{F}_t, \|\cdot\|_{Q,2}, \varepsilon C_1 \mathtt{M}_{\tilde{\mathscr{F}}_t}) \leq C_2 \varepsilon^{-4d},$$

where C_1, C_2 are constants only depending on d, and the supremum is taken over all finite discrete measures on $\mathcal{X} = [0,1]^d$.

Consider next the class of functions $\mathscr{G} = \{g_{\mathbf{x}} : \mathbf{x} \in \mathscr{B}\}$, where $g_{\mathbf{x}}(\mathbf{u}) = \mathbb{1}(\mathbf{u} \in \mathscr{A}_1)\mathscr{K}_1^{(\boldsymbol{\nu})}(\mathbf{u}; \mathbf{x}) - \mathbb{1}(\mathbf{u} \in \mathscr{A}_0)\mathscr{K}_0^{(\boldsymbol{\nu})}(\mathbf{u}; \mathbf{x})$. We have immediately that $\mathtt{M}_{\mathscr{G}} \lesssim h^{-d/2}$, $\mathtt{E}_{\mathscr{G}} \lesssim h^{d/2}$, and

$$\sup_Q N(\mathscr{G}, \|\cdot\|_{Q,2}, \varepsilon(2c+1)^{d+1}\mathtt{M}_{\mathscr{G}}) \leq 2\mathbf{c}'\varepsilon^{-4d-4} + 2,$$

where the supremum is taken over all finite discrete measures.

Total Variation. Observe that $\mathbb{1}(\mathbf{u} \in \mathscr{A}_t)\mathscr{K}_t^{(\boldsymbol{\nu})}(\mathbf{u}; \mathbf{x}) \neq 0$ implies $E_{t,\mathbf{x}} = \mathbf{u} \in \{\mathbf{y} \in \mathscr{A}_t : (\mathbf{y} - \mathbf{x})/h \in \operatorname{Supp}(K)\}$, and $\mathbb{1}(\mathbf{u} \in \mathscr{A}_t)\mathscr{K}_t^{(\boldsymbol{\nu})}(\mathbf{u}; \mathbf{x}) = \mathbb{1}(\mathbf{u} \in E_{t,\mathbf{x}})\mathscr{K}_t^{(\boldsymbol{\nu})}(\mathbf{u}; \mathbf{x})$, for all $\mathbf{u} \in \mathcal{X}$. By the assumption that the De Giorgi perimeter of $E_{t,\mathbf{x}}$ satisfies $\mathscr{L}(E_{t,\mathbf{x}}) \leq Ch^{d-1}$ and using $\mathtt{TV}_{\{gf\}} \leq \mathtt{M}_{\{g\}}\mathtt{TV}_{\{f\}} + \mathtt{M}_{\{f\}}\mathtt{TV}_{\{g\}}$ for any two functions g and f, we have

$$\mathtt{TV}_{\mathscr{G}} = \sup_{\mathbf{x}\in\mathscr{B}} \mathtt{TV}_{\{g_{\mathbf{x}}\}} \leq \sup_{\mathbf{x}\in\mathscr{B}} \sum_{t\in\{0,1\}} \mathtt{TV}_{\{\mathbb{1}_{\mathscr{A}_t}\mathscr{K}_t^{(\nu)}(\cdot;\mathbf{x})\}} \leq \sup_{\mathbf{x}\in\mathscr{B}} \sum_{t\in\{0,1\}} \mathtt{TV}_{\{\mathscr{K}_t^{(\nu)}(\cdot;\mathbf{x})\}} + \mathtt{M}_{\mathscr{F}_t}\mathtt{TV}_{\{\mathbb{1}_{E_{t,\mathbf{x}}}\}} \lesssim h^{d/2-1}.$$

We completed the verification of all the high-level sufficient conditions of Theorem SA-11, which immediately give the result. □

SA-7.11 Proof of Theorem SA-6

The proof is divided in three technical lemmas.

Lemma SA-9 (KS Distance Between $\overline{\mathrm{T}}^{(\boldsymbol{\nu})}$ and $Z^{(\boldsymbol{\nu})}$)**.** *Suppose the conditions of Theorem SA-5 hold. Then, for any multi-index $|\boldsymbol{\nu}| \leq p$,*

$$\sup_{u\in\mathbb{R}} \Big|\mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}} \big|\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})\big| \leq u\Big) - \mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}} \big|Z^{(\boldsymbol{\nu})}(\mathbf{x})\big| \leq u\Big)\Big| \lesssim \left((\log n)^{\frac{3}{2}} \Big(\frac{1}{nh^d}\Big)^{\frac{1}{d+2}\cdot\frac{v}{v+2}} + \log(n)\sqrt{\frac{1}{n^{\frac{v}{v+2}} h^d}}\right)^{1/2}.$$

Proof of Lemma SA-9. Let $\mathfrak{R}_n = (\log n)^{\frac{3}{2}}(\frac{1}{nh^d})^{\frac{1}{d+2}\cdot\frac{v}{v+2}} + \log(n)\sqrt{\frac{1}{n^{\frac{v}{2+v}} h^d}}$, and a_n positive sequence to be determined below. For any $u > 0$,

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}} \big|\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})\big| \leq u\Big) \\
&\leq \mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}} \big|Z^{(\boldsymbol{\nu})}(\mathbf{x})\big| \leq \sup_{\mathbf{x}\in\mathscr{B}} \big|\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x}) - Z^{(\boldsymbol{\nu})}(\mathbf{x})\big| + u\Big) \\
&\leq \mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}} \big|Z^{(\boldsymbol{\nu})}(\mathbf{x})\big| \leq u + a_n\Big) + \mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}} \big|Z^{(\boldsymbol{\nu})}(\mathbf{x}) - \overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})\big| > a_n\Big) \\
&\leq \mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}} \big|Z^{(\boldsymbol{\nu})}(\mathbf{x})\big| \leq u\Big) + 4a_n\Big(\mathbb{E}\Big[\sup_{\mathbf{x}\in\mathscr{B}} \big|Z^{(\boldsymbol{\nu})}(\mathbf{x})\big|\Big] + 1\Big) + \mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}} \big|Z^{(\boldsymbol{\nu})}(\mathbf{x}) - \overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})\big| > a_n\Big) \\
&\leq \mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}} \big|Z^{(\boldsymbol{\nu})}(\mathbf{x})\big| \leq u\Big) + 4a_n\Big(\mathbb{E}\Big[\sup_{\mathbf{x}\in\mathscr{B}} \big|Z^{(\boldsymbol{\nu})}(\mathbf{x})\big|\Big] + 1\Big) + \frac{C\mathfrak{R}_n}{a_n},
\end{aligned}$$

where in the fourth line we have used the Gaussian Anti-concentration Inequality in [Chernozhukov et al., 2014a, Theorem 2.1], and in the last line we have used the tail bound in Theorem SA-5. Similarly, for any $u > 0$, we have the lower bound

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}} \big|\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})\big| \leq u\Big) \\
&\geq \mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}} \big|Z^{(\boldsymbol{\nu})}(\mathbf{x})\big| \leq u - \sup_{\mathbf{x}\in\mathscr{B}} \big|\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x}) - Z^{(\boldsymbol{\nu})}(\mathbf{x})\big|\Big) \\
&\geq \mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}} \big|Z^{(\boldsymbol{\nu})}(\mathbf{x})\big| \leq u - a_n\Big) - \mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}} \big|Z^{(\boldsymbol{\nu})}(\mathbf{x}) - \overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})\big| > a_n\Big) \\
&\geq \mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}} \big|Z^{(\boldsymbol{\nu})}(\mathbf{x})\big| \leq u\Big) - 4a_n\Big(\mathbb{E}\Big[\sup_{\mathbf{x}\in\mathscr{B}} \big|Z^{(\boldsymbol{\nu})}(\mathbf{x})\big|\Big] + 1\Big) - \mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}} \big|Z^{(\boldsymbol{\nu})}(\mathbf{x}) - \overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})\big| > a_n\Big) \\
&\geq \mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}} \big|Z^{(\boldsymbol{\nu})}(\mathbf{x})\big| \leq u\Big) - 4a_n\Big(\mathbb{E}\Big[\sup_{\mathbf{x}\in\mathscr{B}} \big|Z^{(\boldsymbol{\nu})}(\mathbf{x})\big|\Big] + 1\Big) - \frac{C\mathfrak{R}_n}{a_n}.
\end{aligned}$$

Notice that $Z^{(\boldsymbol{\nu})}(\mathbf{x}), \mathbf{x} \in \mathscr{B}$ is a mean-zero Gaussian process satisfying

$$\mathbb{E}\big[\big(Z^{(\boldsymbol{\nu})}(\mathbf{x}) - Z^{(\boldsymbol{\nu})}(\mathbf{y})\big)^2\big]^{\frac{1}{2}} = \mathbb{E}\big[(\mathscr{K}(\mathbf{X}_i,\mathbf{x}) - \mathscr{K}(\mathbf{X}_i,\mathbf{y}))^2\sigma(\mathbf{X}_i)^2\big]^{\frac{1}{2}} \leq C' l_{n,2}\|\mathbf{x} - \mathbf{y}\|_\infty,$$

where C' is a constant and $l_{n,2} \asymp h^{-1}$, and hence

$$\sup_{\mathbf{x}\in\mathscr{B}} \mathbb{E}\big[\big(Z^{(\boldsymbol{\nu})}(\mathbf{x}) - Z^{(\boldsymbol{\nu})}(\mathbf{y})\big)^2\big]^{\frac{1}{2}} = \sup_{\mathbf{x}\in\mathscr{B}} \mathbb{E}\big[\mathscr{K}(\mathbf{X}_i,\mathbf{x})^2\sigma^2(\mathbf{X}_i)\big] \lesssim 1.$$

Then, by Corollary 2.2.8 in van der Vaart and Wellner [1996], we have $\mathbb{E}\big[\sup_{\mathbf{x}\in\mathscr{B}}\big|Z^{(\boldsymbol{\nu})}(\mathbf{x})\big|\big] \lesssim 1$. Choosing $a_n \asymp \sqrt{\mathfrak{R}_n}$, the result now follows. □

Lemma SA-10 (KS Distance Between $\widehat{\mathrm{T}}^{(\boldsymbol{\nu})}$ and $\overline{\mathrm{T}}^{(\boldsymbol{\nu})}$)**.** *Suppose the conditions in Theorem SA-5 hold. Then, for any multi-index $|\boldsymbol{\nu}| \leq p$,*

$$\sup_{u\in\mathbb{R}} \Big|\mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}} \big|\widehat{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})\big| \leq u\Big) - \mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}} \big|\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})\big| \leq u\Big)\Big| = o(1).$$

Proof of Lemma SA-10. Let $\mathfrak{R}_n = (\log n)^{\frac{3}{2}}(\frac{1}{nh^d})^{\frac{1}{d+2}\cdot\frac{v}{v+2}} + \log(n)\sqrt{\frac{1}{n^{\frac{v}{2+v}}h^d}}$ and

$$a_n = o\Big(\sqrt{\log(1/h)}\Big(\sqrt{n^{-1}h^{-d}\log(1/h)} + \frac{\log(1/h)}{n^{\frac{v}{2+v}}h^d}\Big) + h^{p+1}\sqrt{nh^d}\Big).$$

Then, $\sup_{\mathbf{x}\in\mathscr{B}}\big|\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x}) - \widehat{\mathrm{T}}(\mathbf{x})\big| = o_{\mathbb{P}}(a_n)$. Hence, for any $u > 0$,

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}}\big|\widehat{\mathrm{T}}(\mathbf{x})\big| \leq u\Big)\\
&\leq \mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}}\big|\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})\big| \leq u + a_n\Big) + \mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}}\big|\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x}) - \widehat{\mathrm{T}}(\mathbf{x})\big| \geq a_n\Big)\\
&\leq \mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}}\big|Z^{(\boldsymbol{\nu})}(\mathbf{x})\big| \leq u + a_n\Big) + \sqrt{\mathfrak{R}_n} + o(1)\\
&\leq \mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}}\big|Z^{(\boldsymbol{\nu})}(\mathbf{x})\big| \leq u\Big) + 4a_n\Big(\mathbb{E}\Big[\sup_{\mathbf{x}\in\mathscr{B}}\big|Z^{(\boldsymbol{\nu})}(\mathbf{x})\big|\Big] + 1\Big) + \sqrt{\mathfrak{R}_n} + o(1)\\
&\leq \mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}}\big|\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})\big| \leq u\Big) + 4a_n\Big(\mathbb{E}\Big[\sup_{\mathbf{x}\in\mathscr{B}}\big|Z^{(\boldsymbol{\nu})}(\mathbf{x})\big|\Big] + 1\Big) + 2\sqrt{\mathfrak{R}_n} + o(1),
\end{aligned}$$

where the third line uses Lemma SA-9 and $\sup_{\mathbf{x}\in\mathscr{B}}\big|\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x}) - \widehat{\mathrm{T}}(\mathbf{x})\big| = o_{\mathbb{P}}(a_n)$, the fourth line uses [Chernozhukov et al., 2014a, Theorem 2.1], and the last line uses Lemma SA-9 again. Similarly,

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}}\big|\widehat{\mathrm{T}}(\mathbf{x})\big| \leq u\Big)\\
&\geq \mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}}\big|\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})\big| \leq u - a_n\Big) - \mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}}\big|\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x}) - \widehat{\mathrm{T}}(\mathbf{x})\big| \geq a_n\Big)\\
&\geq \mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}}\big|Z^{(\boldsymbol{\nu})}(\mathbf{x})\big| \leq u - a_n) - \sqrt{\mathfrak{R}_n} + o(1)\\
&\geq \mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}}\big|Z^{(\boldsymbol{\nu})}(\mathbf{x})\big| \leq u\Big) - 4a_n\Big(\mathbb{E}\Big[\sup_{\mathbf{x}\in\mathscr{B}}\big|Z^{(\boldsymbol{\nu})}(\mathbf{x})\big|\Big] + 1\Big) - \sqrt{\mathfrak{R}_n} + o(1)\\
&\geq \mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}}\big|\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})\big| \leq u\Big) - 4a_n\Big(\mathbb{E}\Big[\sup_{\mathbf{x}\in\mathscr{B}}\big|Z^{(\boldsymbol{\nu})}(\mathbf{x})\big|\Big] + 1\Big) - 2\sqrt{\mathfrak{R}_n} + o(1).
\end{aligned}$$

From the proof of Lemma SA-9, $\mathbb{E}\left[\sup_{\mathbf{x}\in\mathscr{B}}\big|Z^{(\boldsymbol{\nu})}(\mathbf{x})\big|\right] \lesssim 1$. Hence, the result follows. □

Lemma SA-11 (KS Distance Between $Z^{(\boldsymbol{\nu})}$ and $\widehat{Z}^{(\boldsymbol{\nu})}$)**.** *Suppose the conditions for Theorem SA-5 hold.*

Then, for any multi-index $|\boldsymbol{\nu}| \leq p$,

$$\sup_{\mathbf{u}\in\mathbb{R}} \Big|\mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}} \big|Z^{(\boldsymbol{\nu})}(\mathbf{x})\big| \leq u\Big) - \mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}} \big|\widehat{Z}^{(\boldsymbol{\nu})}(\mathbf{x})\big| \leq u\big|\mathbf{W}\Big)\Big| \lesssim_{\mathbb{P}} \log(n)\Big(\sqrt{\frac{\log n}{nh^d}} + \frac{\log n}{n^{\frac{v}{2+v}} h^d} + h^{p+1}\Big)^{1/2}.$$

Proof of Lemma SA-11. First, using Lemma SA-4, we provide an upper bound between covariance functions of the feasible Gaussian process and the infeasible Gaussian process. Letting $\boldsymbol{\Pi}_{\mathbf{x},\mathbf{y}} = \Omega_{\mathbf{x},\mathbf{y}}/\sqrt{\Omega_{\mathbf{x},\mathbf{x}}\Omega_{\mathbf{y}}}$ and $\widehat{\boldsymbol{\Pi}}_{\mathbf{x},\mathbf{y}} = \widehat{\Omega}_{\mathbf{x},\mathbf{y}}/\sqrt{\widehat{\Omega}_{\mathbf{x},\mathbf{x}}\widehat{\Omega}_{\mathbf{y}}}$,

$$\sup_{\mathbf{x},\mathbf{y}\in\mathscr{X}} \big|\boldsymbol{\Pi}_{\mathbf{x},\mathbf{y}} - \widehat{\boldsymbol{\Pi}}_{\mathbf{x},\mathbf{y}}\big| = \sup_{\mathbf{x},\mathbf{y}\in\mathscr{X}} \Big|\frac{\Omega_{\mathbf{x},\mathbf{y}} - \widehat{\Omega}_{\mathbf{x},\mathbf{y}}}{\sqrt{\Omega_{\mathbf{x},\mathbf{x}}\Omega_{\mathbf{y}}}} + \frac{\widehat{\Omega}_{\mathbf{x},\mathbf{y}}}{\sqrt{\widehat{\Omega}_{\mathbf{x},\mathbf{x}}\widehat{\Omega}_{\mathbf{y}}}}\Big(\sqrt{\frac{\widehat{\Omega}_{\mathbf{x},\mathbf{x}}\widehat{\Omega}_{\mathbf{y}}}{\Omega_{\mathbf{x},\mathbf{x}}\Omega_{\mathbf{y}}}} - 1\Big)\Big|$$

From Lemma SA-4 and the fact that $|\sqrt{x} - \sqrt{y}| \leq (x \wedge y)^{-1/2}|x - y|/2$ for $x, y > 0$,

$$\sup_{\mathbf{x},\mathbf{y}\in\mathscr{X}} \frac{\big|\big(\widehat{\Omega}_{\mathbf{x},\mathbf{x}}\widehat{\Omega}_{\mathbf{y}}\big)^{1/2} - \big(\Omega_{\mathbf{x},\mathbf{x}}\Omega_{\mathbf{y}}\big)^{1/2}\big|}{\big(\Omega_{\mathbf{x},\mathbf{x}}\Omega_{\mathbf{y}}\big)^{1/2}} \lesssim \frac{\sup_{\mathbf{x},\mathbf{y}\in\mathscr{X}} \big|\widehat{\Omega}_{\mathbf{x},\mathbf{x}}\widehat{\Omega}_{\mathbf{y}} - \Omega_{\mathbf{x},\mathbf{x}}\Omega_{\mathbf{y}}\big|}{\inf_{\mathbf{x},\mathbf{y}} \widehat{\Omega}_{\mathbf{x},\mathbf{x}}\widehat{\Omega}_{\mathbf{y}} \wedge \inf_{\mathbf{x},\mathbf{y}} \Omega_{\mathbf{x},\mathbf{x}}\Omega_{\mathbf{y}}} \lesssim_{\mathbb{P}} h^{p+1} + \sqrt{\frac{\log n}{nh^d}} + \frac{\log n}{n^{\frac{v}{2+v}} h^d}$$

and

$$\sup_{\mathbf{x},\mathbf{y}\in\mathscr{X}} \frac{\big|\Omega_{\mathbf{x},\mathbf{y}} - \widehat{\Omega}_{\mathbf{x},\mathbf{y}}\big|}{\sqrt{\Omega_{\mathbf{x},\mathbf{x}}\Omega_{\mathbf{y}}}} \lesssim_{\mathbb{P}} h^{p+1} + \sqrt{\frac{\log n}{nh^d}} + \frac{\log n}{n^{\frac{v}{2+v}} h^d}.$$

Therefore, letting $\mathfrak{R}_n = \sqrt{\frac{\log n}{nh^d}} + \frac{\log n}{n^{\frac{v}{2+v}} h^d}$, it follows that $\sup_{\mathbf{x},\mathbf{y}\in\mathscr{X}} |\boldsymbol{\Pi}_{\mathbf{x},\mathbf{y}} - \widehat{\boldsymbol{\Pi}}_{\mathbf{x},\mathbf{y}}| \lesssim_{\mathbb{P}} h^{p+1} + \mathfrak{R}_n$. Then, we bound the KS distance between the maximum of Z_n and $\widehat{Z}^{(\boldsymbol{\nu})}$ on a δ_n-net of $\mathscr{X}$, denoted by $\mathscr{X}_{\delta_n}$: for all $\mathbf{x} \in \mathscr{B}$, there exists $\mathbf{z} \in \mathscr{X}_{\delta_n}$ such that $\|\mathbf{x} - \mathbf{z}\|_\infty \leq \delta_n$. Since $\mathscr{X}$ is compact, we can assume $M := \mathrm{Card}\,(\mathscr{X}_{\delta_n}) \lesssim \delta_n^{-d}$. Denote $\mathbf{Z}_n^{\delta_n}$ and $\widehat{\mathbf{Z}}_n^{\delta_n}$ to the process Z_n and $\widehat{Z}^{(\boldsymbol{\nu})}$ restricted on $\mathscr{X}_{\delta_n}$, respectively. Then, by [Chernozhuokov et al., 2022, Theorem 2.1],

$$\sup_{\mathbf{y}\in\mathbb{R}^M} \big|\mathbb{P}(\mathbf{Z}_n^{\delta_n} \leq \mathbf{y}) - \mathbb{P}(\widehat{\mathbf{Z}}_n^{\delta_n} \leq \mathbf{y}|\mathbf{W})\big| \lesssim \log(M) \sup_{\mathbf{x},\mathbf{y}\in\mathscr{X}} \big|\boldsymbol{\Pi}_{\mathbf{x},\mathbf{y}} - \widehat{\boldsymbol{\Pi}}_{\mathbf{x},\mathbf{y}}\big|^{\frac{1}{2}} \lesssim_{\mathbb{P}} \log(M)(\mathfrak{R}_n + h^{p+1})^{\frac{1}{2}},$$

and hence

$$\begin{aligned}\sup_{x\in\mathbb{R}} \big|\mathbb{P}(\|\mathbf{Z}_n^{\delta_n}\|_\infty \leq x) - \mathbb{P}(\|\widehat{\mathbf{Z}}_n^{\delta_n}\|_\infty \leq x|\mathbf{W})\big| &\leq \sup_{x\in\mathbb{R}} \big|\mathbb{P}(-x\mathbf{1} \leq \mathbf{Z}_n^{\delta_n} \leq x\mathbf{1}) - \mathbb{P}(-x\mathbf{1} \leq \widehat{\mathbf{Z}}_n^{\delta_n} \leq x\mathbf{1}|\mathbf{W})\big| \\ &\lesssim_{\mathbb{P}} \log(M)(\mathfrak{R}_n + h^{p+1})^{\frac{1}{2}} =: \mathfrak{R}_M.\end{aligned}$$

Finally, we bound the KS distance on the whole $\mathscr{X}$ with the help of a sequence $a_n > 0$ to be determined. Let

$$\Psi_{\delta_n}(a_n) = \mathbb{P}\Big(\sup_{\|\mathbf{x}-\mathbf{y}\|_\infty \leq \delta_n} \big|Z^{(\boldsymbol{\nu})}(\mathbf{x}) - Z^{(\boldsymbol{\nu})}(\mathbf{y})\big| \geq a_n\Big)$$

and

$$\widehat{\Psi}_{\delta_n}(a_n) = \mathbb{P}\Big(\sup_{\|\mathbf{x}-\mathbf{y}\|_\infty \leq \delta_n} \big|\widehat{Z}^{(\boldsymbol{\nu})}(\mathbf{x}) - \widehat{Z}^{(\boldsymbol{\nu})}(\mathbf{y})\big| \geq a_n\Big|\mathbf{W}\Big).$$

Then, for all $t > 0$,

$$
\begin{aligned}
&\mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}}\big|Z^{(\boldsymbol{\nu})}(\mathbf{x})\big|\leq t\Big)\\
&\leq \mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}_{\delta_n}}\big|Z^{(\boldsymbol{\nu})}(\mathbf{x})\big|\leq t+a_n\Big)+\Psi_{\delta_n}(a_n)\\
&\leq \mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}_{\delta_n}}\big|\widehat{Z}^{(\boldsymbol{\nu})}(\mathbf{x})\big|\leq t+a_n\Big|\mathbf{W}\Big)+\Psi_{\delta_n}(a_n)+\mathfrak{R}_M\\
&\leq \mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}}\big|\widehat{Z}^{(\boldsymbol{\nu})}(\mathbf{x})\big|\leq t+a_n\Big|\mathbf{W}\Big)+\Psi_{\delta_n}(a_n)+\widehat{\Psi}_{\delta_n}(a_n)+\mathfrak{R}_M\\
&\leq \mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}}\big|\widehat{Z}^{(\boldsymbol{\nu})}(\mathbf{x})\big|\leq t\Big|\mathbf{W}\Big)+4a_n\Big(\mathbb{E}\Big[\sup_{\mathbf{x}\in\mathscr{B}}\big|\widehat{Z}^{(\boldsymbol{\nu})}(\mathbf{x})\big|\Big|\mathbf{W}\Big]+1\Big)+\Psi_{\delta_n}(a_n)+\widehat{\Psi}_{\delta_n}(a_n)+\mathfrak{R}_M.
\end{aligned}
$$

Similarly, for all $t > 0$,

$$
\begin{aligned}
\mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}}\Big|Z^{(\boldsymbol{\nu})}(\mathbf{x})\Big|\leq t\Big)\geq \mathbb{P}\Big(\sup_{\mathbf{x}\in\mathscr{B}}\big|\widehat{Z}^{(\boldsymbol{\nu})}(\mathbf{x})\big|\leq t\Big|\mathbf{W}\Big)&-4a_n\Big(\mathbb{E}\Big[\sup_{\mathbf{x}\in\mathscr{B}}\big|\widehat{Z}^{(\boldsymbol{\nu})}(\mathbf{x})\big|\Big|\mathbf{W}\Big]+1\Big)\\
&-\Psi_{\delta_n}(a_n)-\widehat{\Psi}_{\delta_n}(a_n)-\mathfrak{R}_M.
\end{aligned}
$$

Since $\mathfrak{R}_M$ depends on δ_n through $\log M \asymp \log(\delta_n^{-d})$, by choosing $\delta_n = n^{-s}$ for large enough s, the term $\mathfrak{R}_M$ will dominate the terms $\Psi_{\delta_n}(a_n)$ and $\widehat{\Psi}_{\delta_n}(a_n)$. More precisely, for any δ,

$$
\begin{aligned}
&\sup_{\|\mathbf{x}-\mathbf{y}\|_\infty\leq\delta}\mathbb{E}\Big[\big(\widehat{Z}^{(\boldsymbol{\nu})}(\mathbf{x})-\widehat{Z}^{(\boldsymbol{\nu})}(\mathbf{y})\big)^2\Big|\mathbf{W}\Big]\\
&=\sup_{\|\mathbf{x}-\mathbf{y}\|_\infty\leq\delta}\big(\widehat{\Omega}_{\mathbf{x},\mathbf{x}}\widehat{\Omega}_{\mathbf{y}}\big)^{-\frac{1}{2}}\Big(\frac{1}{nh^d}\Big)^2\sum_{i=1}^n\widehat{\varepsilon}_i^2\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_1)\\
&\qquad\cdot\Big(\mathbf{e}_1^T\widehat{\boldsymbol{\Gamma}}_{1,\mathbf{x}}^{-1}\mathbf{r}_p\Big(\frac{\mathbf{X}_i-\mathbf{x}}{h}\Big)K\Big(\frac{\mathbf{X}_i-\mathbf{x}}{h}\Big)-\mathbf{e}_1^T\widehat{\boldsymbol{\Gamma}}_{1,\mathbf{x}}^{-1}\mathbf{r}_p\Big(\frac{\mathbf{X}_i-\mathbf{y}}{h}\Big)K\Big(\frac{\mathbf{X}_i-\mathbf{y}}{h}\Big)\Big)^2\\
&\quad+\sup_{\|\mathbf{x}-\mathbf{y}\|_\infty\leq\delta}\big(\widehat{\Omega}_{\mathbf{x},\mathbf{x}}\widehat{\Omega}_{\mathbf{y}}\big)^{-\frac{1}{2}}\Big(\frac{1}{nh^d}\Big)^2\sum_{i=1}^n\widehat{\varepsilon}_i{}^2\mathbb{1}\left(\mathbf{X}_i\in\mathscr{A}_0\right)\\
&\qquad\cdot\Big(\mathbf{e}_1^T\widehat{\boldsymbol{\Gamma}}_{0,\mathbf{x}}^{-1}\mathbf{r}_p\Big(\frac{\mathbf{X}_i-\mathbf{x}}{h}\Big)K\Big(\frac{\mathbf{X}_i-\mathbf{x}}{h}\Big)-\mathbf{e}_1^T\widehat{\boldsymbol{\Gamma}}_{0,\mathbf{x}}^{-1}\mathbf{r}_p\Big(\frac{\mathbf{X}_i-\mathbf{y}}{h}\Big)K\Big(\frac{\mathbf{X}_i-\mathbf{y}}{h}\Big)\Big)^2\\
&\lesssim_{\mathbb{P}} h^{-d-2}\delta^2,
\end{aligned}
$$

where the last line uses Lemma SA-4, Lemma SA-2, and the almost sure bound on the Lipschitz constant from the proof of Theorem SA-5, for some constant $C > 0$. Similarly, for any $\delta > 0$,

$$
\sup_{\|\mathbf{x}-\mathbf{y}\|_\infty\leq\delta}\mathbb{E}\Big[\big(Z^{(\boldsymbol{\nu})}(\mathbf{x})-Z^{(\boldsymbol{\nu})}(\mathbf{y})\big)^2\Big]=\sup_{\|\mathbf{x}-\mathbf{y}\|_\infty\leq\delta}\mathbb{E}\Big[\big(\mathscr{K}(\mathbf{X}_i,\mathbf{x})-\mathscr{K}(\mathbf{X}_i,\mathbf{y})\big)^2\varepsilon_i^2\Big]\leq C'h^{-2}\delta^2,
$$

Then, by [van der Vaart and Wellner, 1996, Corollary 2.2.5],

$$
\mathbb{E}\Big[\sup_{\|\mathbf{x}-\mathbf{y}\|_\infty\leq\delta_n}\big|\widehat{Z}^{(\boldsymbol{\nu})}(\mathbf{x})-\widehat{Z}^{(\boldsymbol{\nu})}(\mathbf{y})\big|\Big|\mathbf{W}\Big]\lesssim_{\mathbb{P}}\int_0^{Ch^{-d/2-1}\delta_n}\sqrt{d\log\left(\frac{1}{\varepsilon h^{d/2+1}}\right)}d\varepsilon\lesssim\sqrt{\log n}h^{-d/2-1}\delta_n
$$

and

$$
\mathbb{E}\Big[\sup_{\|\mathbf{x}-\mathbf{y}\|_\infty\leq\delta_n}\big|Z^{(\boldsymbol{\nu})}(\mathbf{x})-Z^{(\boldsymbol{\nu})}(\mathbf{y})\big|\Big]\lesssim\int_0^{Ch^{-1}\delta_n}\sqrt{d\log\left(\frac{1}{\varepsilon h}\right)}d\varepsilon\lesssim\sqrt{\log n}h^{-1}\delta_n.
$$

In addition, using the fact that $\mathbb{E}\big[\sup_{\mathbf{x}\in\mathscr{B}}\big|\widehat{Z}^{(\boldsymbol{\nu})}(\mathbf{x})\big|\big|\mathbf{W}\big] \lesssim 1$, and choosing $a_n \asymp (\sqrt{\log n}h^{-d/2-1}\delta_n)^{1/2}$ and $\delta_n \asymp n^{-s}$ for some large constant $s > 0$, we conclude that

$$\begin{aligned}
&4a_n\Big(\mathbb{E}\Big[\sup_{\mathbf{x}\in\mathscr{B}}\big|\widehat{Z}^{(\boldsymbol{\nu})}(\mathbf{x})\big|\Big|\mathbf{W}\Big]+1\Big)+\Psi_{\delta_n}(a_n)+\widehat{\Psi}_{\delta_n}(a_n)+\mathfrak{R}_M\\
&\lesssim_{\mathbb{P}} \big(\sqrt{\log n}h^{-d/2-1}\delta_n\big)^{1/2}+d\log(\delta_n^{-1})(a_n+h^{p+1})^{1/2}\lesssim d\log(n)(a_n+h^{p+1})^{1/2},
\end{aligned}$$

and putting all the intermediate results together, the lemma follows. □

The proof of Theorem SA-6 now follows directly from Lemma SA-9, Lemma SA-10 and Lemma SA-11. Furthermore, by definition of $\widehat{\mathrm{I}}_{\alpha}^{(\boldsymbol{\nu})}(\mathbf{x})$,

$$\begin{aligned}
\mathbb{P}\Big[\mu^{(\boldsymbol{\nu})}(\mathbf{x})\in\widehat{\mathrm{I}}_{\alpha}^{(\boldsymbol{\nu})}(\mathbf{x}),\ \text{for all } \mathbf{x}\in\mathscr{B}\Big] &= \mathbb{P}\Big[\sup_{\mathbf{x}\in\mathscr{B}}\big|\widehat{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})\big|\le \mathscr{q}_{\alpha}\Big]\\
&= \mathbb{P}\Big[\sup_{\mathbf{x}\in\mathscr{B}}\big|\widehat{Z}^{(\boldsymbol{\nu})}(\mathbf{x})\big|\le \mathscr{q}_{\alpha}\Big]+o(1)\\
&= \mathbb{E}\Big[\mathbb{P}\Big[\sup_{\mathbf{x}\in\mathscr{B}}\big|\widehat{Z}^{(\boldsymbol{\nu})}(\mathbf{x})\big|\le \mathscr{q}_{\alpha}\Big|\mathbf{W}\Big]\Big]+o(1)\\
&= 1-\alpha+o(1),
\end{aligned}$$

which completes the proof of the theorem. □

SA-7.12 Proof of Lemma SA-6

Follows from Lemma SA-5 and the assumption that $\int_{\mathscr{B}}|w(\mathbf{x})|d\mathfrak{H}^{d-1}(\mathbf{x})<\infty$. □

SA-7.13 Proof of Lemma SA-7

Since $\mathbb{V}[\widehat{\tau}_{\mathtt{WBATE}}|\mathbf{X}] = \mathbb{V}[\int_{\mathscr{B}}\widehat{\mu}_0(\mathbf{b})w(\mathbf{b})d\mathfrak{H}^{d-1}(\mathbf{b})|\mathbf{X}]+\mathbb{V}[\int_{\mathscr{B}}\widehat{\mu}_1(\mathbf{b})w(\mathbf{b})d\mathfrak{H}^{d-1}(\mathbf{b})|\mathbf{X}]$, it is enough to consider only one treatment assignment group $t\in\{0,1\}$. In addition,

$$\mathbb{V}\Big[\int_{\mathscr{B}}\widehat{\mu}_t(\mathbf{b})w(\mathbf{b})d\mathfrak{H}^{d-1}(\mathbf{b})\Big|\mathbf{X}\Big]=\int_{\mathscr{B}}\int_{\mathscr{B}}\mathbb{C}\mathrm{ov}\big[\widehat{\mu}_t(\mathbf{b}_1),\widehat{\mu}_t(\mathbf{b}_2)\big|\mathbf{X}\big]w(\mathbf{b}_1)w(\mathbf{b}_2)d\mathfrak{H}^{d-1}(\mathbf{b}_1)d\mathfrak{H}^{d-1}(\mathbf{b}_2)$$

and

$$\Omega_{t,\mathtt{WBATE}}=\int_{\mathscr{B}}\int_{\mathscr{B}}\Omega_{t,\mathbf{b}_1,\mathbf{b}_2}w(\mathbf{b}_1)w(\mathbf{b}_2)d\mathfrak{H}^{d-1}(\mathbf{b}_1)d\mathfrak{H}^{d-1}(\mathbf{b}_2).$$

Proceeding as in the proof of Lemma SA-4, we have

$$\sup_{\mathbf{b}_1,\mathbf{b}_2\in\mathscr{B}}\big|\mathbb{C}\mathrm{ov}[\widehat{\mu}_t(\mathbf{b}_1),\widehat{\mu}_t(\mathbf{b}_2)|\mathbf{X}]-\Omega_{t,\mathbf{b}_1,\mathbf{b}_2}\big|\lesssim_{\mathbb{P}}\frac{\log(1/h)^{1/2}}{(nh^d)^{3/2}}.$$

Since K is supported on a compact set, let $R\in(0,\infty)$ denote the diameter of the support, and define the "effective domain" $\mathscr{E}(h)=\{(\mathbf{x},\mathbf{y})\in\mathscr{B}\times\mathscr{B}:\|\mathbf{x}-\mathbf{y}\|\le hR\}$. Since $\mathscr{B}$ is $(d-1)$ dimensional, we have

$\nu_d(\mathscr{E}(h)) \lesssim h^{d-1}$, where ν_d is the product measure $\mathfrak{H}^{d-1} \times \mathfrak{H}^{d-1}$. Therefore,

$$\begin{aligned}
&\Big|\mathbb{V}\Big[\int_{\mathscr{B}} \widehat{\mu}_t(\mathbf{b})w(\mathbf{b})d\mathfrak{H}^{d-1}(\mathbf{b})\Big|\mathbf{X}\Big] - \Omega_t\Big| \\
&= \Big|\int_{\mathscr{B}}\int_{\mathscr{B}} \big(\mathbb{C}\mathrm{ov}[\widehat{\mu}_t(\mathbf{b}_1), \widehat{\mu}_t(\mathbf{b}_2)|\mathbf{X}] - \Omega_{t,\mathbf{b}_1,\mathbf{b}_2}\big)w(\mathbf{b}_1)w(\mathbf{b}_2)d\mathfrak{H}^{d-1}(\mathbf{b}_1)d\mathfrak{H}^{d-1}(\mathbf{b}_2)\Big| \\
&\lesssim \sup_{\mathbf{b}_1,\mathbf{b}_2\in\mathscr{B}} \big|\mathbb{C}\mathrm{ov}[\widehat{\mu}_t(\mathbf{b}_1), \widehat{\mu}_t(\mathbf{b}_2)|\mathbf{X}] - \Omega_{t,\mathbf{b}_1,\mathbf{b}_2}\big| \int_{\mathscr{B}}\int_{\mathscr{B}} \mathbb{1}((\mathbf{b}_1,\mathbf{b}_2)\in\mathscr{E}(h))w(\mathbf{b}_1)w(\mathbf{b}_2)d\mathfrak{H}^{d-1}(\mathbf{b}_1)d\mathfrak{H}^{d-1}(\mathbf{b}_2) \\
&\lesssim_{\mathbb{P}} h^{d-1}\frac{\log(1/h)^{1/2}}{(nh^d)^{3/2}} = o_{\mathbb{P}}((nh)^{-1}),
\end{aligned}$$

because $\frac{\log(1/h)}{nh^d} = o(1)$. This proves the first claim. Next,

$$\begin{aligned}
\Omega_{t,\mathtt{WBATE}} &= \int_{\mathscr{B}}\int_{\mathscr{B}} \Omega_{t,\mathbf{b}_1,\mathbf{b}_2} w(\mathbf{b}_1)w(\mathbf{b}_2)d\mathfrak{H}^{d-1}(\mathbf{b}_1)d\mathfrak{H}^{d-1}(\mathbf{b}_2) \\
&\leq \sup_{\mathbf{b}_1,\mathbf{b}_2\in\mathscr{B}} \big|\Omega_{t,\mathbf{b}_1,\mathbf{b}_2}\big| \int_{\mathscr{B}}\int_{\mathscr{B}} \mathbb{1}((\mathbf{b}_1,\mathbf{b}_2)\in\mathscr{E}(h))w(\mathbf{b}_1)w(\mathbf{b}_2)d\mathfrak{H}^{d-1}(\mathbf{b}_1)d\mathfrak{H}^{d-1}(\mathbf{b}_2) \\
&\lesssim (nh^d)^{-1}\nu_d(\mathscr{E}(h)) \lesssim (nh)^{-1},
\end{aligned}$$

which verifies the upper bound. For the lower bound, let $\mathbf{b}_1 \in \mathscr{B}$ and $\mathbf{b}_2 = \mathbf{b}_1 + h\boldsymbol{\delta}$ for some vector $\boldsymbol{\delta}$ such that $\sup_{\mathbf{x}\in\mathcal{X}} K_h(\mathbf{x}-\mathbf{b}_1)K_h(\mathbf{x}-\mathbf{b}_2) > 0$. For multi-indexes $\mathbf{u}$ and $\mathbf{v}$, and using change of variables, a typical element of $\boldsymbol{\Sigma}_{t,\mathbf{b}_1,\mathbf{b}_2}$ is

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\frac{\mathbf{X}_i - \mathbf{b}_1}{h}\Big)^{\mathbf{u}}\Big(\frac{\mathbf{X}_i - \mathbf{b}_1 - \boldsymbol{\delta}h}{h}\Big)^{\mathbf{v}}\frac{1}{h^d}K\Big(\frac{\mathbf{X}_i - \mathbf{b}_1}{h}\Big)K\Big(\frac{\mathbf{X}_i - \mathbf{b}_1 - \boldsymbol{\delta}h}{h}\Big)\sigma_t^2(\mathbf{X}_i)\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_t)\Big] \\
&= \int_{\mathbf{b}_1 + h\mathscr{A}_t} \mathbf{s}^{\mathbf{u}}(\mathbf{s}-\boldsymbol{\delta})^{\mathbf{v}}K(\mathbf{s})K(\mathbf{s}+\boldsymbol{\delta})\sigma_t^2(\mathbf{b}_1 + h\mathbf{s})f(\mathbf{s})d\mathbf{s} \gtrsim 1,
\end{aligned}$$

which implies that $|\Omega_{t,\mathbf{b}_1,\mathbf{b}_2}| \gtrsim (nh^d)^{-1}$ for $(\mathbf{b}_1,\mathbf{b}_2)$ on a set $\mathscr{E}'(h)$ such that $\nu_d(\mathscr{E}'(h)) \gtrsim h^{d-1}$. This verifies lower bound in the second claim.

The third and final claim of the lemma follows from Lemma SA-4 and the same analysis as above.

SA-7.14 Proof of Theorem SA-7

Follows from Lemma SA-6 and Lemma SA-7. □

SA-7.15 Proof of Theorem SA-8

Since $\widehat{\tau}_{\mathtt{WBATE}} - \tau_{\mathtt{WBATE}} = (\widehat{\mu}_{1,\mathtt{WBATE}} - \mu_{1,\mathtt{WBATE}}) - (\widehat{\mu}_{0,\mathtt{WBATE}} - \mu_{0,\mathtt{WBATE}})$, it is enough to start with only one treatment assignment group $t \in \{0,1\}$. Furthermore,

$$\begin{aligned}
\widehat{\mu}_{t,\mathtt{WBATE}} - \mu_{t,\mathtt{WBATE}} &= \int_{\mathscr{B}} (\widehat{\mu}_1(\mathbf{b}) - \mu_1(\mathbf{b}))w(\mathbf{b})d\mathfrak{H}^{d-1}(\mathbf{b}) \\
&= \int_{\mathscr{B}} \mathbf{e}_1^\top \boldsymbol{\Gamma}_{t,\mathbf{b}}^{-1}\mathbf{Q}_{t,\mathbf{b}}w(\mathbf{b})d\mathfrak{H}^{d-1}(\mathbf{b}) + \int_{\mathscr{B}} \mathbf{e}_1^\top(\widehat{\boldsymbol{\Gamma}}_{t,\mathbf{b}}^{-1} - \boldsymbol{\Gamma}_{t,\mathbf{b}}^{-1})\mathbf{Q}_{t,\mathbf{b}}w(\mathbf{b})d\mathfrak{H}^{d-1}(\mathbf{b}) + O_{\mathbb{P}}(h^{p+1})
\end{aligned}$$

using Lemma SA-5 to bound the approximation error.

For the second integral, let

$$\overline{\boldsymbol{\Sigma}}_{t,\mathbf{x}_1,\mathbf{x}_2} = \mathbb{E}_n\Big[\mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{x}_1}{h}\Big)\mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{x}_2}{h}\Big)^\top h^d K_h(\mathbf{X}_i - \mathbf{x}_1)K_h(\mathbf{X}_i - \mathbf{x}_2)\sigma_t^2(\mathbf{X}_i)\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)\Big],$$

and since $\overline{\boldsymbol{\Sigma}}_{t,\mathbf{b}_1,\mathbf{b}_2} = 0$ if $\mathbf{b}_1$ and $\mathbf{b}_2$ are farther away form each other than the diameter of $\mathrm{Supp}(K)$,

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\int_{\mathscr{B}} \mathbf{e}_1^\top(\widehat{\boldsymbol{\Gamma}}_{t,\mathbf{b}}^{-1} - \boldsymbol{\Gamma}_{t,\mathbf{b}}^{-1})\mathbf{Q}_{t,\mathbf{b}} w(\mathbf{b}) d\mathfrak{H}^{d-1}(\mathbf{b})\Big)^2 \Big| \mathbf{X}\Big] \\
&= \int_{\mathscr{B}}\int_{\mathscr{B}} \mathbf{e}_1^\top(\widehat{\boldsymbol{\Gamma}}_{t,\mathbf{b}_1}^{-1} - \boldsymbol{\Gamma}_{t,\mathbf{b}_1}^{-1})(nh^d)^{-1}\overline{\boldsymbol{\Sigma}}_{t,\mathbf{b}_1,\mathbf{b}_2}(\widehat{\boldsymbol{\Gamma}}_{t,\mathbf{b}_2}^{-1} - \boldsymbol{\Gamma}_{t,\mathbf{b}_2}^{-1})\mathbf{e}_1 w(\mathbf{b}_1)w(\mathbf{b}_2)d\mathfrak{H}^{d-1}(\mathbf{b}_1)d\mathfrak{H}^{d-1}(\mathbf{b}_2), \\
&\leq \sup_{\mathbf{b}\in\mathscr{B}} \|\widehat{\boldsymbol{\Gamma}}_{t,\mathbf{b}}^{-1} - \boldsymbol{\Gamma}_{t,\mathbf{b}}^{-1}\|^2 \sup_{\mathbf{b}_1,\mathbf{b}_2\in\mathscr{B}} \|\overline{\boldsymbol{\Sigma}}_{t,\mathbf{b}_1,\mathbf{b}_2}\| \sup_{\mathbf{b}\in\mathscr{B}} |w(\mathbf{b})|(nh^d)^{-1}\mathfrak{m}(\mathscr{E}(h)) \\
&\lesssim_{\mathbb{P}} (nh)^{-1},
\end{aligned}$$

and hence $\int_{\mathscr{B}} \mathbf{e}_1^\top(\widehat{\boldsymbol{\Gamma}}_{t,\mathbf{b}}^{-1} - \boldsymbol{\Gamma}_{t,\mathbf{b}}^{-1})\mathbf{Q}_{t,\mathbf{b}} w(\mathbf{b})d\mathfrak{H}^{d-1}(\mathbf{b}) = o_{\mathbb{P}}((nh)^{-1})$.

Next, using Lemma SA-7 and the previous results,

$$\widehat{\mathrm{T}}_{\mathtt{WBATE}} - \overline{\mathrm{T}}_{\mathtt{WBATE}} = \big(\widehat{\Omega}_{\mathtt{WBATE}}^{-1/2} - \Omega_{\mathtt{WBATE}}^{-1/2}\big)\int_{\mathscr{B}} \mathbf{e}_1^\top \boldsymbol{\Gamma}_{t,\mathbf{b}}^{-1}\mathbf{Q}_{t,\mathbf{b}} d\mathfrak{H}^{d-1}(\mathbf{b}) + o_{\mathbb{P}}(1) = o_{\mathbb{P}}(1),$$

where

$$\overline{\mathrm{T}}_{\mathtt{WBATE}} = \Omega_{\mathtt{WBATE}}^{-1/2}\int_{\mathscr{B}} \mathbf{e}_1^\top \boldsymbol{\Gamma}_{t,\mathbf{b}}^{-1}\mathbf{Q}_{t,\mathbf{b}} d\mathfrak{H}^{d-1}(\mathbf{b}).$$

Finally, we apply the Berry-Esseen lemma to the statistic $\overline{\mathrm{T}}_w = \sum_{i=1}^n Z_i$, where

$$Z_i = n^{-1}\Omega_{\mathtt{WBATE}}^{-1/2}\int_{\mathscr{B}} \mathbf{e}_1^\top \boldsymbol{\Gamma}_{t,\mathbf{b}}^{-1}\mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{b}}{h}\Big)K_h(\mathbf{X}_i - \mathbf{b})\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)u_i w(\mathbf{b})d\mathfrak{H}^{d-1}(\mathbf{b}),$$

which satisfies $\mathbb{E}[Z_i] = 0$. The definition of $\Omega_{\mathtt{WBATE}}$ implies that $\sum_{i=1}^n \mathbb{V}[Z_i] = \Omega_{\mathtt{WBATE}}^{-1/2}\Omega_{\mathtt{WBATE}}\Omega_{\mathtt{WBATE}}^{-1/2} = 1$. Hence, it remains to bound

$$\sum_{i=1}^n \mathbb{E}[|Z_i|^3] = n^{-3}\Omega_{\mathtt{WBATE}}^{-3/2}\sum_{i=1}^n \mathbb{E}\Big[\Big|\int_{\mathscr{B}} \mathbf{e}_1^\top \boldsymbol{\Gamma}_{t,\mathbf{b}}^{-1}\mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{b}}{h}\Big)K_h(\mathbf{X}_i - \mathbf{b})\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)u_i w(\mathbf{b})d\mathfrak{H}^{d-1}(\mathbf{b})\Big|^3\Big].$$

Let R denote the diameter of the (compact) support of K, and define $\mathscr{E}(h) = \{(\mathbf{b}_1,\mathbf{b}_2,\mathbf{b}_3) \in \mathscr{B}^3 : \|\mathbf{b}_i - \mathbf{b}_j\| \leq R, j = 1,2,3\}$. Since $\mathscr{B}$ is $d-1$ dimensional, $\mathfrak{m}(\mathscr{E}(h)) \lesssim h^{2d-2}$. Then,

$$\begin{aligned}
&\mathbb{E}\Big[\Big|\int_{\mathscr{B}} \mathbf{e}_1^\top \boldsymbol{\Gamma}_{t,\mathbf{b}}^{-1}\mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{b}}{h}\Big)K_h(\mathbf{X}_i - \mathbf{b})\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)u_i w(\mathbf{b})d\mathfrak{H}^{d-1}(\mathbf{b})\Big|^3\Big] \\
&\leq \mathbb{E}\Big[\int_{\mathbf{b}_1\in\mathscr{B}}\int_{\mathbf{b}_2\in\mathscr{B}}\int_{\mathbf{b}_3\in\mathscr{B}} |G(\mathbf{b}_1,\mathbf{b}_2,\mathbf{b}_3)|w(\mathbf{b}_1)w(\mathbf{b}_2)w(\mathbf{b}_3)d\mathfrak{H}^{d-1}(\mathbf{b}_1)d\mathfrak{H}^{d-1}(\mathbf{b}_2)d\mathfrak{H}^{d-1}(\mathbf{b}_3)\Big] \\
&\lesssim \mathfrak{m}(\mathscr{E}(h)) \sup_{\mathbf{b}_1,\mathbf{b}_2,\mathbf{b}_3\in\mathscr{B}} \mathbb{E}[|G(\mathbf{b}_1,\mathbf{b}_2,\mathbf{b}_3)|],
\end{aligned}$$

where $G(\mathbf{b}_1,\mathbf{b}_2,\mathbf{b}_3) = g(\mathbf{X}_i,u_i,\mathbf{b}_1)g(\mathbf{X}_i,u_i,\mathbf{b}_2)g(\mathbf{X}_i,u_i,\mathbf{b}_3)$ with

$$g(\mathbf{X}_i,u_i,\mathbf{b}) = \mathbf{e}_1^\top \boldsymbol{\Gamma}_{t,\mathbf{b}}^{-1}\mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{b}}{h}\Big)K_h(\mathbf{X}_i - \mathbf{b})\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)u_i.$$

Proceeding as in the proof of Lemma SA-2 and Lemma SA-7, it can be shown that

$$\sup_{\mathbf{b}_1,\mathbf{b}_2,\mathbf{b}_3\in\mathscr{B}} \mathbb{E}[|G(\mathbf{b}_1,\mathbf{b}_2,\mathbf{b}_3)|] \lesssim h^{-2d}$$

provided that $\frac{\log(1/n)}{nh^d} = o(1)$. Therefore, together with the rate of $\Omega_{\texttt{WBATE}}$ from Lemma SA-7, we have $\sum_{i=1}^n \mathbb{E}[|Z_i^3|] \lesssim (nh)^{-1/2}$, and the result follows. □

SA-7.16 Proof of Theorem SA-9

Follows by Theorem SA-1 after noting that $|\widehat{\tau}_{\texttt{LBATE}} - \tau_{\texttt{LBATE}}| \leq \sup_{\mathbf{x}\in\mathscr{B}} |\widehat{\tau}(\mathbf{x}) - \tau(\mathbf{x})|$. □

SA-7.17 Proof of Theorem SA-10

Consider the event $E = \Big\{ \sup_{\mathbf{b}\in\mathscr{B}} \frac{|\widehat{\tau}(\mathbf{b})-\tau(\mathbf{b})|}{\widehat{\Omega}^{1/2}_{\mathbf{b},\mathbf{b}}} \leq \mathscr{q}_\alpha \Big\}$. Theorem SA-6 implies that $\mathbb{P}(E) = 1-\alpha+o(1)$. On the event E, we also have

$$\widehat{\tau}(\mathbf{b}) - \mathscr{q}_\alpha \widehat{\Omega}^{1/2}_{\mathbf{b},\mathbf{b}} \leq \tau(\mathbf{b}) \leq \widehat{\tau}(\mathbf{b}) + \mathscr{q}_\alpha \widehat{\Omega}^{1/2}_{\mathbf{b},\mathbf{b}}, \qquad \forall\, \mathbf{b} \in \mathscr{B},$$

which implies

$$\sup_{\mathbf{b}\in\mathscr{B}} \widehat{\tau}(\mathbf{b}) - \mathscr{q}_\alpha \widehat{\Omega}^{1/2}_{\mathbf{b},\mathbf{b}} \leq \sup_{\mathbf{b}\in\mathscr{B}} \tau(\mathbf{b}) \leq \sup_{\mathbf{b}\in\mathscr{B}} \widehat{\tau}(\mathbf{b}) + \mathscr{q}_\alpha \widehat{\Omega}^{1/2}_{\mathbf{b},\mathbf{b}}.$$

The stated result then follows. □

SA-7.18 Proof of Theorem SA-11

We will use a truncation argument. Let $\kappa_n > 0$ be the level of truncation. For each $r \in \mathscr{R}$, define

$$\tilde{r}(y) = r(y)\mathbb{1}(|y| \leq \kappa_n), \qquad y \in \mathbb{R},$$

and define the class $\tilde{\mathscr{R}} = \{\tilde{r} : r \in \mathscr{R}\}$. For an overview of our argument, suppose Z_n^R is some mean-zero Gaussian process indexed by $\mathscr{G}\times\mathscr{R} \cup \mathscr{G}\times\tilde{\mathscr{R}}$, whose existence will be shown below, then we can decompose by:

$$R_n(g,r) - Z_n^R(g,r) = \big[R_n(g,\tilde{r}) - Z_n^R(g,\tilde{r})\big] + \big[R_n(g,r) - R_n(g,\tilde{r})\big] + \big[Z_n^R(g,r) - Z_n^R(g,\tilde{r})\big].$$

Part 1: Strong approximation for truncated residual empirical process.

Observe that $\mathtt{M}_{\tilde{\mathscr{R}},\mathscr{Y}} \lesssim \kappa_n$ and $\mathtt{pTV}_{\tilde{\mathscr{R}},\mathscr{Y}} \lesssim \kappa_n$, and $\tilde{\mathscr{R}}$ is a VC-type class with envelope $M_{\tilde{\mathscr{R}},\mathscr{Y}} = M_{\mathscr{R},\mathscr{Y}}\mathbb{1}(|\cdot| \leq \kappa_n)$ over $\mathscr{Y}$ with constants $\mathtt{c}_{\mathscr{R},\mathscr{Y}}$ and $\mathtt{d}_{\mathscr{R},\mathscr{Y}}$. Then, Cattaneo and Yu [2025, Theorem 2] with $\mathtt{v} = \kappa_n$ and $\alpha = 0$ for the class of functions $\mathscr{G}$ and $\tilde{\mathscr{R}}$ implies on a possibly enlarged probability space, there exists a sequence of mean-zero Gaussian processes $(Z_n^R(g,r) : (g,r) \in \mathscr{G}\times\tilde{\mathscr{R}})$ with almost sure continuous trajectories on

$(\mathscr{G}\times\tilde{\mathscr{R}},\rho_{\mathbb{P}})$ such that $\mathbb{E}[R_n(g_1,r_1)R_n(g_2,r_2)]=\mathbb{E}[Z_n^R(g_1,r_1)Z_n^R(g_2,r_2)]$ for all $(g_1,r_1),(g_2,r_2)\in\mathscr{G}\times\tilde{\mathscr{R}}$, and

$$\begin{aligned}
&\mathbb{E}[\|R_n(g,\tilde{r})-Z_n^R(g,\tilde{r})\|_{\mathscr{G}\times\mathscr{R}}]\\
&\leq C_1\mathtt{v}\kappa_n\Bigg(\sqrt{d}\min\Big\{\frac{(\mathtt{c}_1^d\mathtt{M}_{\mathscr{G}}^{d+1}\mathtt{TV}^d\mathtt{E}_{\mathscr{G}})^{\frac{1}{2d+2}}}{n^{1/(2d+2)}},\frac{(\mathtt{c}_1^{\frac{d}{2}}\mathtt{c}_2^{\frac{d}{2}}\mathtt{M}_{\mathscr{G}}\mathtt{TV}^{\frac{d}{2}}\mathtt{E}_{\mathscr{G}}\mathtt{L}^{\frac{d}{2}})^{\frac{1}{d+2}}}{n^{1/(d+2)}}\Big\}((\mathtt{d}+\mathtt{k})\log(\mathtt{c}n))^{3/2}+\frac{(\mathtt{d}+\mathtt{k})\log(\mathtt{c}n)}{\sqrt{n}}\mathtt{M}_{\mathscr{G}}\Bigg)\\
&=C_1\mathtt{v}\kappa_n\Bigg(\sqrt{d}\mathtt{r}_n((\mathtt{d}+\mathtt{k})\log(\mathtt{c}n))^{\frac{3}{2}}+\frac{(\mathtt{d}+\mathtt{k})\log(\mathtt{c}n)}{\sqrt{n}}\mathtt{M}_{\mathscr{G}}\Bigg),
\end{aligned}$$

where C_1 is some positive universal constant. Notice that we use $\mathtt{TV}=\max\{\mathtt{TV}_{\mathscr{G}},\mathtt{TV}_{\mathscr{G}\times\mathscr{U}_{\mathscr{R}},\mathscr{Q}_{\mathscr{G}}}\}$ as an upper bound for $\max\{\mathtt{TV}_{\mathscr{G}},\mathtt{TV}_{\mathscr{G}\times\mathscr{V}_{\tilde{\mathscr{R}}},\mathscr{Q}_{\mathscr{G}}}\}$, and similarly $\mathtt{L}$ as an upper bound for $\max\{\mathtt{L}_{\mathscr{G}},\mathtt{L}_{\mathscr{G}\times\mathscr{V}_{\tilde{\mathscr{R}}},\mathscr{Q}_{\mathscr{G}}}\}$.

In the special case that $\mathscr{R}=\{r_*\}$ is a singleton, take $\tilde{y}_i=r_*(y_i)\mathbb{1}(|y_i|\leq\kappa_n)/(\mathtt{v}\kappa_n)$, then we have $\mathbb{E}[\exp(|\tilde{y}_i|)]\leq 2$. Also $\tilde{y}_i$ is supported on $\tilde{\mathscr{Y}}=[-1,1]$. Moreover,

$$\frac{1}{\mathtt{v}\kappa_n}R_n(g,\tilde{r}_*)=\frac{1}{n}\sum_{i=1}^n g(\mathbf{x}_i)(\tilde{y}_i-\mathbb{E}[\tilde{y}_i]),\qquad g\in\mathscr{G}.$$

In particular, the right hand side can be viewed as a residual empirical process based on sample $(\mathbf{x}_i,\tilde{y}_i),1\leq i\leq n$, indexed by $\mathscr{G}\times\{\mathrm{Id}\}$, where $\mathrm{Id}:\mathbb{R}\to\mathbb{R}$ is the identity function. Then we can apply Cattaneo and Yu [2025, Theorem 2] with $\mathtt{v}=1$ and $\alpha=0$ on the latter empirical process to get the upper bound with $\mathtt{TV}$ and $\mathtt{L}$ replaced by $\mathtt{TV}_{\mathrm{sing}}$ and $\mathtt{L}_{\mathrm{sing}}$.

Part 2: Truncation error for the empirical process — $\|R_n(g,r)-R_n(g,\widetilde{r})\|_{\mathscr{G}\times\mathscr{R}}$

Consider the class of differences due to truncation, that is, $\Delta\mathscr{R}=\{r-\tilde{r}:r\in\mathscr{R}\}$. Our assumptions imply $\mathscr{G}\times\Delta\mathscr{R}$ is VC-type in the sense that for all $0<\varepsilon<1$,

$$\sup_{\mathscr{Q}}N(\mathscr{G}\times\Delta\mathscr{R},\|\cdot\|_{\mathscr{Q},2},\varepsilon\|\mathtt{M}_{\mathscr{G}}(M_{\mathscr{R},\mathscr{Y}}-M_{\tilde{\mathscr{R}},\mathscr{Y}})\|_{\mathscr{Q},2})\leq\mathtt{c}_{\mathscr{G}}\mathtt{c}_{\mathscr{R},\mathscr{Y}}(\varepsilon^2/4)^{-\mathtt{d}_{\mathscr{G}}-\mathtt{d}_{\mathscr{R},\mathscr{Y}}},$$

where sup is over all finite discrete measure on $\mathbb{R}^{d+1}$, and $M_{\tilde{\mathscr{R}},\mathscr{Y}}(y)=M_{\mathscr{R},\mathscr{Y}}(y)\mathbb{1}(|y|\leq\kappa_n)$. We can check that $\mathtt{M}_{\mathscr{G}}(M_{\mathscr{R},\mathscr{Y}}-M_{\tilde{\mathscr{R}},\mathscr{Y}})$ is an envelope function for $\mathscr{G}\times\Delta\mathscr{R}$, since all functions in $\Delta\mathscr{R}$ are evaluated to zero on $[-\kappa_n,\kappa_n]$. Denote $\mathbf{X}=(\mathbf{x}_i)_{1\leq i\leq n}$,

$$\begin{aligned}
\mathbb{E}\Big[\max_{1\leq i\leq n}\mathtt{M}_{\mathscr{G}}^2(M_{\mathscr{R},\mathscr{Y}}(y_i)-M_{\tilde{\mathscr{R}},\mathscr{Y}}(y_i))^2\Big|\mathbf{X}\Big]^{\frac{1}{2}}&\lesssim\mathtt{M}_{\mathscr{G}}\mathbb{E}\Big[\big(\max_{1\leq i\leq n}M_{\mathscr{R},\mathscr{Y}}(y_i)\big)^2\Big|\mathbf{X}\Big]^{\frac{1}{2}}\lesssim\mathtt{M}_{\mathscr{G}}n^{\frac{1}{2+v}},\\
\sup_{(g,r)\in\mathscr{G}\times\mathscr{R}}\mathbb{E}[g(\mathbf{x}_i)^2r(y_i)^2\mathbb{1}(|y_i|\geq\kappa_n^{1/\alpha})]^{\frac{1}{2}}&\lesssim\sup_{(g,r)\in\mathscr{G}\times\mathscr{R}}\mathbb{E}\left[g(\mathbf{x}_i)^2\mathbb{E}[r(y_i)^{2+v}|\mathbf{x}_i]^{\frac{2}{2+v}}\mathbb{P}(|y_i|\geq\kappa_n|\mathbf{x}_i)^{\frac{v}{2+v}}\right]\\
&\lesssim\sqrt{\mathtt{M}_{\mathscr{G}}\mathtt{E}_{\mathscr{G}}\kappa_n}.
\end{aligned}$$

By Jensen's inequality, we also have

$$\begin{aligned}
&\mathbb{E}\Big[\max_{1\leq i\leq n}\mathtt{M}_{\mathscr{G}}^2(\mathbb{E}[M_{\mathscr{R},\mathscr{Y}}(y_i)-M_{\tilde{\mathscr{R}},\mathscr{Y}}(y_i)|\mathbf{x}_i])^2\Big|\mathbf{X}\Big]^{\frac{1}{2}}\lesssim\mathtt{M}_{\mathscr{G}}n^{\frac{1}{2+v}},\\
&\sup_{(g,r)\in\mathscr{G}\times\mathscr{R}}\mathbb{E}[g(\mathbf{x}_i)^2\mathbb{E}[r(y_i)-\tilde{r}(y_i)|\mathbf{x}_i]^2]^{\frac{1}{2}}\lesssim\sqrt{\mathtt{M}_{\mathscr{G}}\mathtt{E}_{\mathscr{G}}\kappa_n^{-v}},\\
&\mathbb{E}[\mathtt{M}_{\mathscr{G}}^2(M_{\mathscr{R},\mathscr{Y}}(y_i)-M_{\tilde{\mathscr{R}},\mathscr{Y}}(y_i))^2]^{1/2}\lesssim\mathtt{M}_{\mathscr{G}}\kappa_n^{-v/2}.
\end{aligned}$$

Denote $A = (\mathtt{c}_{\mathscr{G}}\mathtt{c}_{\mathscr{R}})^{\frac{1}{2\mathtt{d}_{\mathscr{G}}+2\mathtt{d}_{\mathscr{R}}}}/4$ and $D = 2\mathtt{d}_{\mathscr{G}} + 2\mathtt{d}_{\mathscr{R}}$, Chernozhukov et al. [2014b, Corollary 5.1] gives

$$\begin{aligned}\mathbb{E}\left[\|R_n(g,r) - R_n(g\widetilde{r})\|_{\mathscr{G}\times\mathscr{R}}\right] &\lesssim \mathbb{E}\left[\sup_{g\in\mathscr{G}}\sup_{h\in\Delta\mathscr{R}} \frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(\mathbf{x}_i)(h(y_i) - \mathbb{E}[h(y_i)|\mathbf{x}_i])\right] \\ &\lesssim \sqrt{D\mathtt{M}_{\mathscr{G}}\mathtt{E}_{\mathscr{G}}\kappa_n^{-v}\log(A\sqrt{\mathtt{M}_{\mathscr{G}}/\mathtt{E}_{\mathscr{G}}})} + \frac{D\mathtt{M}_{\mathscr{G}}n^{\frac{1}{2+V}}}{\sqrt{n}}\log(A\sqrt{\mathtt{M}_{\mathscr{G}}/\mathtt{E}_{\mathscr{G}}}) \\ &\lesssim \sqrt{D\log(A\sqrt{\mathtt{M}_{\mathscr{G}}/\mathtt{E}_{\mathscr{G}}})}\sqrt{\mathtt{M}_{\mathscr{G}}\mathtt{E}_{\mathscr{G}}}\kappa_n^{-v/2} + \frac{D\log(A\sqrt{\mathtt{M}_{\mathscr{G}}/\mathtt{E}_{\mathscr{G}}})\mathtt{M}_{\mathscr{G}}}{\sqrt{n^{\frac{v}{2+v}}}}.\end{aligned}$$

Part 3: Truncation error for the Gaussian process — $\|Z_n^R(g,r) - Z_n^R(g,\tilde{r})\|_{\mathscr{G}\times\mathscr{R}}$

Our assumptions imply $\mathscr{G}\times\tilde{\mathscr{R}}\cup\mathscr{G}\times\mathscr{R}$ is VC-type w.r.p envelope function $2\mathtt{M}_{\mathscr{G}}\mathtt{M}_{\mathscr{R},\mathscr{Y}}$ in the sense that for all $0<\varepsilon<1$,

$$\sup_{Q} N(\mathscr{G}\times\mathscr{R}\cup\mathscr{G}\times\tilde{\mathscr{R}}, \|\cdot\|_{Q,2}, 2\varepsilon\|\mathtt{M}_{\mathscr{G}}\mathtt{M}_{\mathscr{R},\mathscr{Y}}\|_{Q,2}) \leq \mathtt{c}_{\mathscr{G}}\mathtt{c}_{\mathscr{R}}(\varepsilon^2/4)^{-\mathtt{d}_{\mathscr{G}}-\mathtt{d}_{\mathscr{R}}},$$

where sup is over all finite discrete measure on $\mathbb{R}^{d+1}$. Hence $\mathscr{G}\times\tilde{\mathscr{R}}\cup\mathscr{G}\times\mathscr{R}$ is pre-Gaussian, and on some probability space, there exists a mean-zero Gaussian process $\bar{Z}_n^R$ indexed by $\mathscr{F} = \mathscr{G}\times\tilde{\mathscr{R}}\cup\mathscr{G}\times\mathscr{R}$ with the same covariance structure as R_n, and has almost sure continuous path w.r.p the metric ρ, given by

$$\rho((g_1,r_1),(g_2,r_2)) = \mathbb{E}[(Z_n^R(g_1,r_1) - Z_n^R(g_2,r_2))^2]^{\frac{1}{2}} = \mathbb{E}[(R_n(g_1,r_1) - R_n(g_2,r_2))^2]^{\frac{1}{2}}, (g_1,r_1),(g_2,r_2)\in\mathscr{F}.$$

Recall the definition of $\mathscr{G}\times\Delta\mathscr{R}$ in Part 2. Then, we have shown previously that

$$\sigma \equiv \sup_{f\in\mathscr{G}\times\Delta\mathscr{R}} \rho(f,f) \leq \sqrt{\mathtt{M}_{\mathscr{G}}\mathtt{E}_{\mathscr{G}}\kappa_n^{-v}},$$

Our assumptions imply for all $0<\varepsilon<1$,

$$N(\mathscr{G}\times\mathscr{R}\cup\mathscr{G}\times\tilde{\mathscr{R}}, \rho, \rho(2\varepsilon\mathtt{M}_{\mathscr{G}}M_{\mathscr{R},\mathscr{Y}}, 2\varepsilon\|\mathtt{M}_{\mathscr{G}}M_{\mathscr{R},\mathscr{Y}})^{1/2}) \leq \mathtt{c}_{\mathscr{G}}\mathtt{c}_{\mathscr{R}}(\varepsilon^2/4)^{-\mathtt{d}_{\mathscr{G}}-\mathtt{d}_{\mathscr{R}}}$$

Denote $A = (\mathtt{c}_{\mathscr{G}}\mathtt{c}_{\mathscr{R}})^{\frac{1}{2\mathtt{d}_{\mathscr{G}}+2\mathtt{d}_{\mathscr{R}}}}/4$ and $D = 2\mathtt{d}_{\mathscr{G}} + 2\mathtt{d}_{\mathscr{R}}$. Then, by van der Vaart and Wellner [1996, Corollary 2.2.8], choose any $(g_0,r_0)\in\mathscr{G}\times\mathscr{R}$, we have

$$\begin{aligned}\mathbb{E}\left[\|\bar{Z}_n^R(g,r) - \bar{Z}_n^R(g,\tilde{r})\|_{\mathscr{G}\times\mathscr{R}}\right] &\lesssim \mathbb{E}\left[|\bar{Z}_n^R(g_0,r_0) - \bar{Z}_n^R(g_0,\tilde{r}_0)|\right] + \int_0^{\sigma}\sqrt{\log\left(\mathtt{c}_{\mathscr{G}}\mathtt{c}_{\mathscr{R}}(\frac{\mathtt{M}_{\mathscr{G}}}{\varepsilon})^{\mathtt{d}_{\mathscr{G}}+\mathtt{d}_{\mathscr{R}}}\right)}d\varepsilon \\ &\leq \sqrt{D\log(A\sqrt{\mathtt{M}_{\mathscr{G}}/\mathtt{E}_{\mathscr{G}}})}\sqrt{\mathtt{M}_{\mathscr{G}}\mathtt{E}_{\mathscr{G}}}\kappa_n^{-v/2} \\ &\lesssim \sqrt{(\mathtt{d}_{\mathscr{G}}+\mathtt{d}_{\mathscr{R},\mathscr{Y}})\log(\mathtt{c}_{\mathscr{G}}\mathtt{c}_{\mathscr{R},\mathscr{Y}}\mathtt{k}n)}\sqrt{\mathtt{M}_{\mathscr{G}}\mathtt{E}_{\mathscr{G}}}\kappa_n^{-v/2}.\end{aligned}$$

Since $(\bar{Z}_n^R(g,r) : g\in\mathscr{G}, r\in\mathscr{R})$ has the same distribution as $(Z_n^R(g,r) : g\in\mathscr{G}, r\in\mathscr{R})$, we know from Vorob'ev–Berkes–Philipp theorem [Dudley, 2014, Theorem 1.31] that $\bar{Z}_n^R$ can be constructed on the same probability space as $(\mathbf{x}_i,y_i)_{1\leq i\leq n}$ and Z_n^R, such that $\bar{Z}_n^R$ and Z_n^R coincide on $\mathscr{G}\times\mathscr{R}$. By an abuse of notation, call $\bar{Z}_n^R$ now Z_n^R, the outputted Gaussian process.

Part 4: Putting Together

If follows from the definition of $\tilde{\mathscr{R}}$ and the previous three parts that if we choose κ_n such that

$$\mathtt{r}_n\kappa_n \asymp \sqrt{\mathtt{M}_{\mathscr{G}}\mathtt{E}_{\mathscr{G}}}\kappa_n^{-v/2},$$

then the approximation error can be bounded by

$$\begin{aligned}\mathbb{E}\big[\|R_n - Z_n^R\|_{\mathscr{G}\times\mathscr{R}}\big] \lesssim (\mathtt{d}\log(\mathtt{c}n))^{3/2}\mathtt{r}_n^{\frac{v}{v+2}}(\sqrt{\mathtt{M}_{\mathscr{G}}\mathtt{E}_{\mathscr{G}}})^{\frac{2}{v+2}} + \mathtt{d}\log(\mathtt{c}n)\mathtt{M}_{\mathscr{G}}n^{-\frac{v/2}{2+v}}\\ + \mathtt{d}\log(\mathtt{c}n)\mathtt{M}_{\mathscr{G}}n^{-1/2}\Big(\frac{\sqrt{\mathtt{M}_{\mathscr{G}}\mathtt{E}_{\mathscr{G}}}}{\mathtt{r}_n}\Big)^{\frac{2}{v+2}},\end{aligned}$$

where $\mathtt{d} = \mathtt{d}_{\mathscr{G}} + \mathtt{d}_{\mathscr{R},\mathscr{Y}} + \mathtt{k}$, and $\mathtt{c} = \mathtt{c}_{\mathscr{G}}\mathtt{c}_{\mathscr{R},\mathscr{Y}}\mathtt{k}$. □

References

Matias D. Cattaneo and Ruiqi (Rae) Yu. Strong approximations for empirical processes indexed by lipschitz functions. *Annals of Statistics*, 53(3):1203–1229, 2025.

Matias D. Cattaneo, Rajita Chandak, Michael Jansson, and Xinwei Ma. Boundary adaptive local polynomial conditional density estimators. *Bernoulli*, 30(4):3193–3223, 2024.

Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Anti-concentration and honest, adaptive confidence bands. *Annals of Statistics*, 42(5):1787–1818, 2014a.

Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Gaussian approximation of suprema of empirical processes. *Annals of Statistics*, 42(4):1564–1597, 2014b.

Victor Chernozhuokov, Denis Chetverikov, Kengo Kato, and Yuta Koike. Improved central limit theorem and bootstrap approximations in high dimensions. *Annals of Statistics*, 50(5):2562–2586, 2022.

Richard M Dudley. *Uniform central limit theorems*, volume 142. Cambridge university press, 2014.

Herbert Federer. *Geometric measure theory.* Springer, 2014.

G.B. Folland. *Advanced Calculus.* Featured Titles for Advanced Calculus Series. Prentice Hall, 2002.

Evarist Giné and Richard Nickl. *Mathematical Foundations of Infinite-dimensional Statistical Models.* Cambridge University Press, New York, 2016.

Leon Simon et al. *Lectures on geometric measure theory.* Centre for Mathematical Analysis, Australian National University Canberra, 1984.

Aad W. van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes.* Springer, 1996.