Estimation and Inference in Boundary Discontinuity Designs Supplemental Appendix*

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

April 22, 2025

Abstract

This supplemental appendix presents more general theoretical results encompassing those discussed in the main paper, and their proofs.

*Cattaneo and Titiunik gratefully acknowledge financial support from the National Science Foundation (SES-2019432, DMS-2210561, and SES-2241575).

†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 **3**

SA-1.1 Notation and Definitions . . . 3

SA-1.2 Mapping between Main Paper and Supplement . . . 4

SA-2 Analysis Based on the d-variate Location Variable **5**

SA-2.1 Preliminary Lemmas . . . 6

SA-2.2 Point Estimation . . . 7

SA-2.3 Pointwise Inference . . . 8

SA-2.4 Uniform Inference . . . 8

SA-3 Analysis Based on Univariate Distance **10**

SA-3.1 Preliminary Lemmas . . . 12

SA-3.2 Pointwise Inference . . . 13

SA-3.3 Uniform Inference . . . 14

SA-4 Gaussian Strong Approximation Lemmas **16**

SA-4.1 Definitions for Function Spaces . . . 16

SA-4.2 Residual-based Empirical Process . . . 18

SA-4.3 Multiplicative-Separable Empirical Process . . . 18

SA-5 Proofs for Section SA-2 **19**

SA-5.1 Proof of Lemma SA-2.1 . . . 19

SA-5.2 Proof of Lemma SA-2.2 . . . 21

SA-5.3 Proof of Lemma SA-2.3 . . . 23

SA-5.4 Proof of Lemma SA-2.4 . . . 24

SA-5.5 Proof of Theorem SA-2.1 . . . 27

SA-5.6 Proof of Theorem SA-2.2 . . . 28

SA-5.7 Proof of Theorem SA-2.3 . . . 29

SA-5.8 Proof of Theorem SA-2.4 . . . 30

SA-5.9 Proof of Theorem SA-2.5 . . . 31

SA-5.10 Proof of Theorem SA-2.6 . . . 31

SA-5.11 Proof of Theorem SA-2.7 . . . 32

SA-5.12 Proof of Theorem SA-2.8 . . . 40

SA-5.13 Proof of Theorem SA-2.9 . . . 40

SA-6 Proofs for Section SA-3 **41**

SA-6.1 Proof of Lemma SA-3.1 . . . 41

SA-6.2 Proof of Lemma SA-3.2 . . . 41

SA-6.3 Proof of Lemma SA-3.3 . . . 42

SA-6.4 Proof of Lemma SA-3.4 . . . 44

SA-6.5 Proof of Lemma SA-3.5 . . . 44

SA-6.6 Proof of Theorem SA-3.1 . . . 46

SA-6.7 Proof of Theorem SA-3.2 . . . 46

SA-6.8 Proof of Theorem SA-3.3 47
SA-6.9 Proof of Theorem SA-3.4 47
SA-6.10 Proof of Theorem SA-3.5 48
SA-6.11 Proof of Theorem SA-3.6 49

SA-7 Proofs of Distance-Based Bias Results **50**
SA-7.1 Proof of Lemma 2 50
SA-7.2 Proof of Lemma 3 54
SA-7.3 Proof of Theorem 6 58

SA-8 Proofs for Section SA-4 **61**
SA-8.1 Proof of Lemma SA-4.1 61
SA-8.2 Proof of Lemma SA-4.2 63

SA-1 Setup

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

Assumption 1 from the main 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* $\mathcal{X} = \prod_{l=1}^d [a_l, b_l]$ *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* $\mathcal{X}$.

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

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

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

We partition $\mathcal{X}$ 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, $\mathcal{X} = \mathscr{A}_0 \cup \mathscr{A}_1$, where $\mathscr{A}_0$ and $\mathscr{A}_1$ are two disjoint regions in $\mathbb{R}^d$, and $\mathtt{cl}(\mathscr{A}_t)$ denotes the closure of $\mathscr{A}_t$, $t \in \{0, 1\}$. The observed outcome is $Y_i = \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_0)Y_i(0) + \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_1)Y_i(1)$. $\mathscr{B} = \mathtt{bd}(\mathscr{A}_0) \cap \mathtt{bd}(\mathscr{A}_1)$ 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 $\mathscr{A}_t$. The *treatment effect curve along the boundary* is

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

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 $\mathbf{u} \in \mathbb{R}^d$.

(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 $A \in \mathbb{R}^{m\times n}$, $\|A\|_p = \sup_{\|\mathbf{x}\|_p=1} \|A\mathbf{x}\|_p$, $p \in \mathbb{N} \cup \{\infty\}$. For a function f on a metric space (S, d), $\|f\|_\infty = \sup_{x\in\mathcal{X}} |f|$, $\|f\|_{\mathrm{Lip},\infty} = \sup_{\mathbf{x},\mathbf{x}'\in S} \frac{|\mathbf{x}-\mathbf{x}'|}{d(\mathbf{x},\mathbf{x}')}$. 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 sta ndard 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 $W_n(f), f \in L_2(\mathscr{X}, \mathbb{P})$ with the covariance $\mathbb{E}[W_{\mathbb{P}}(f)W_{\mathbb{P}}(g)] = \mathbb{P}(fg) - \mathbb{P}(f)\mathbb{P}(g)$, for $f, g \in L_2(\mathscr{X}, \mathbb{P})$. A class $\mathscr{F} \subseteq L_2(\mathscr{X}, \mathbb{P})$ 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 $L_2(\mathscr{X}, \mathbb{P})$ 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 $f, g \in L_2(\mathscr{X}, \mathbb{P})$.

(iv) *Geometric Measure Theory.* For a set $E \subseteq \mathscr{X}$, the *De Giorgi perimeter of E related to $\mathscr{X}$* is $\mathscr{L}(E) = \mathtt{TV}_{\{\mathbb{1}_E\},\mathscr{X}}$. 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 $\limsup_{M\to\infty} \limsup_{n\to\infty} \mathbb{P}[|\frac{a_n}{b_n}| \geq M] = 0$.

(vi) *Distributions and Statistical Distances.* For $\boldsymbol{\mu} \in \mathbb{R}^k$ and $\boldsymbol{\Sigma}$ a $k \times k$ positive definite matrix, $\mathsf{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. For $-\infty < a < b < \infty$, $\mathsf{Unif}([a, b])$ denotes the uniform distribution on $[a, b]$. $\mathsf{Bern}(p)$ denotes the Bernoulli distribution with success probability p. $\Phi(\cdot)$ denotes the standard Gaussian cumulative distribution function. For two distributions P and Q, $d_{\mathrm{KL}}(P, Q)$ denotes the KL-distance between P and Q, and $d_{\chi^2}(P, Q)$ denotes the χ^2 distance between P and Q.

SA-1.2 Mapping between Main Paper and Supplement

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

- Lemma 1 in the paper corresponds to Lemma SA-3.1 with $d = 2$.
- Lemma 2 in is proven in Section SA-7.1.
- Lemma 3 in is proven in Section SA-7.2.
- Theorem 1(i) in the paper corresponds to Theorem SA-3.1 with $d = 2$.
- Theorem 1(ii) in the paper corresponds to Theorem SA-3.3 with $d = 2$.
- Theorem 2(i) in the paper corresponds to Theorem SA-3.2 with $d = 2$.
- Theorem 2(ii) in the paper corresponds to Theorem SA-3.6 with $d = 2$.
- Theorem 3(i) in the paper corresponds to Theorem SA-2.1 with $d = 2$.
- Theorem 3(ii) in the paper corresponds to Theorem SA-2.5 with $d = 2$.
- Theorem 4 in the paper corresponds to Theorem SA-2.2 with $d = 2$.
- Theorem 5(i) in the paper corresponds to Theorem SA-2.4 with $d = 2$.
- Theorem 5(ii) in the paper corresponds to Theorem SA-2.9 with $d = 2$.
- Theorem 6 is proven in Section SA-7.3.

SA-2 Analysis Based on the d-variate Location Variable

We consider a more general setting compared to the main paper, where the parameter of interest is

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

where $\boldsymbol{\nu}$ is a multi-index with $|\boldsymbol{\nu}| \leq p$. Thus, the treatment effect curve estimator is $(\widehat{\tau}^{(\boldsymbol{\nu})}(\mathbf{x}) : \mathbf{x} \in \mathscr{B})$, where

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

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

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

with $\mathfrak{p}_p = \frac{(d+p)!}{d!p!}$, $\mathbf{R}_p(\mathbf{u}) = (1, u_1, u_2 \cdots, u_d, u_1^2, u_1u_2, u_1u_2, \cdots u_d^2, \cdots, u_1^p, u_1^{p-1}u_2, \cdots, u_2^p)^\top$ 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 d-variate kernel function and assignment boundary.

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

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

(ii) $\liminf_{h \downarrow 0} \inf_{\mathbf{x} \in \mathscr{B}} \int_{\mathscr{A}_t} K_h(\mathbf{u} - \mathbf{x}) d\mathbf{u} \gtrsim 1$.

Under the assumptions imposed, for $t \in \{0,1\}$, we have

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

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

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

In particular, $\|\mathbf{e}_{1+\boldsymbol{\nu}}^\top \mathbf{H}^{-1}\|_2 = \|\mathbf{e}_{1+\boldsymbol{\nu}}^\top \mathbf{H}^{-1}\|_\infty = h^{-|\boldsymbol{\nu}|}$.

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

$$\begin{aligned}
\boldsymbol{\Gamma}_{t,\mathbf{x}} &= \mathbb{E}\left[\mathbf{R}_p\left(\frac{\mathbf{X}_i - \mathbf{x}}{h}\right) \mathbf{R}_p\left(\frac{\mathbf{X}_i - \mathbf{x}}{h}\right)^\top K_h(\mathbf{X}_i - \mathbf{x}) \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)\right], \\
\widehat{\boldsymbol{\Sigma}}_{t,\mathbf{x}_1,\mathbf{x}_2} &= h^d \mathbb{E}_n\left[\mathbf{R}_p\left(\frac{\mathbf{X}_i - \mathbf{x}_1}{h}\right) \mathbf{R}_p\left(\frac{\mathbf{X}_i - \mathbf{x}_2}{h}\right)^\top K_h(\mathbf{X}_i - \mathbf{x}_1) K_h(\mathbf{X}_i - \mathbf{x}_2) \varepsilon_i^2 \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)\right], \\
\boldsymbol{\Sigma}_{t,\mathbf{x}_1,\mathbf{x}_2} &= h^d \mathbb{E}\left[\mathbf{R}_p\left(\frac{\mathbf{X}_i - \mathbf{x}_1}{h}\right) \mathbf{R}_p\left(\frac{\mathbf{X}_i - \mathbf{x}_2}{h}\right)^\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)\right], \\
\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}}, \quad \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}}, \quad \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 $\varepsilon_i = Y_i - \sum_{t\in\{0,1\}} \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)\widehat{\boldsymbol{\beta}}_t(\mathbf{x})^\top \mathbf{R}_p(\mathbf{X}_i - \mathbf{x})$ and $\sigma_t^2(\mathbf{x}) = \mathbb{V}[Y_i(t)|\mathbf{X}_i = \mathbf{x}]$. Denote

$$\widehat{B}_{t,\mathbf{x}}^{(\boldsymbol{\nu})} = \mathbf{e}_{1+\boldsymbol{\nu}}^\top \widehat{\boldsymbol{\Gamma}}_{t,\mathbf{x}}^{-1} \sum_{|\boldsymbol{\omega}|=p+1} \frac{\mu_t^{(\boldsymbol{\omega})}(\mathbf{x})}{\boldsymbol{\omega}!} \mathbb{E}_n\left[\mathbf{R}_p\left(\frac{\mathbf{X}_i - \mathbf{x}}{h}\right)\left(\frac{\mathbf{X}_i - \mathbf{x}}{h}\right)^{\boldsymbol{\omega}} K_h(\mathbf{X}_i - \mathbf{x})\right], \qquad \widehat{B}_{\mathbf{x}}^{(\boldsymbol{\nu})} = \widehat{B}_{1,\mathbf{x}}^{(\boldsymbol{\nu})} - \widehat{B}_{0,\mathbf{x}}^{(\boldsymbol{\nu})},$$

$$B_{t,\mathbf{x}}^{(\boldsymbol{\nu})} = \mathbf{e}_{1+\boldsymbol{\nu}}^\top \boldsymbol{\Gamma}_{t,\mathbf{x}}^{-1} \sum_{|\boldsymbol{\omega}|=p+1} \frac{\mu_t^{(\boldsymbol{\omega})}(\mathbf{x})}{\boldsymbol{\omega}!} \mathbb{E}\left[\mathbf{R}_p\left(\frac{\mathbf{X}_i - \mathbf{x}}{h}\right)\left(\frac{\mathbf{X}_i - \mathbf{x}}{h}\right)^{\boldsymbol{\omega}} K_h(\mathbf{X}_i - \mathbf{x})\right], \qquad B_{\mathbf{x}}^{(\boldsymbol{\nu})} = B_{1,\mathbf{x}}^{(\boldsymbol{\nu})} - B_{0,\mathbf{x}}^{(\boldsymbol{\nu})},$$

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

$$\widehat{V}_{t,\mathbf{x}}^{(\boldsymbol{\nu})} = \mathbf{e}_{1+\boldsymbol{\nu}}^\top \widehat{\boldsymbol{\Gamma}}_{t,\mathbf{x}}^{-1} \widehat{\boldsymbol{\Sigma}}_{t,\mathbf{x},\mathbf{x}} \widehat{\boldsymbol{\Gamma}}_{t,\mathbf{x}}^{-1} \mathbf{e}_{1+\boldsymbol{\nu}}, \qquad \widehat{V}_{\mathbf{x}}^{(\boldsymbol{\nu})} = \widehat{V}_{0,\mathbf{x}}^{(\boldsymbol{\nu})} + \widehat{V}_{1,\mathbf{x}}^{(\boldsymbol{\nu})},$$

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

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

SA-2.1 Preliminary Lemmas

In what follows, we denote $\mathbf{X} = (\mathbf{X}_1^\top, \cdots, \mathbf{X}_n^\top)$ and $\mathbf{W}_n = ((\mathbf{X}_1^\top, Y_1), \cdots, (\mathbf{X}_n^\top, Y_n))^\top$.

Lemma SA-2.1 (Gram)
Suppose Assumption SA-1(i)(ii) and Assumption SA-2 hold. If $\frac{\log(1/h)}{nh^d} = o(1)$, then

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

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

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

Lemma SA-2.2 (Bias)
Suppose Assumption SA-1(i)(ii)(iii) and Assumption SA-2 hold. If $\frac{\log(1/h)}{nh^d} = o(1)$, then

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

for $t \in \{0, 1\}$. If, in addition, $h = o(1)$, then

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

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

Lemma SA-2.3 (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

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

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

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

Lemma SA-2.4 (Covariance)
Suppose Assumptions SA–1 and SA–2 hold. If $\frac{\log(1/h)}{nh^d} = o(1)$, *then*

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

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

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

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

SA-2.2 Point Estimation

Theorem SA-2.1 (Pointwise Convergence Rate)
Suppose Assumptions SA–1 and SA–2 hold. If $nh^d \to \infty$, *then*

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

The conditional mean-squared error (MSE) is

$$\mathrm{MSE}_{\boldsymbol{\nu}}(\mathbf{x}) = \mathbb{E}\left[(\widehat{\tau}^{(\boldsymbol{\nu})}(\mathbf{x}) - \tau^{(\boldsymbol{\nu})}(\mathbf{x}))^2\middle|\mathbf{X}\right], \qquad \mathbf{x}\in\mathscr{B},$$

and, for some non-negative weighting function ω satisfying $\int_{\mathscr{B}} \omega(\mathbf{x})d\mathbf{x} < \infty$, the conditional integrated mean-squared error (IMSE) is defined to be

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

where H^{d-1} is the $(d-1)$ dimensional Hausdorff measure, also known as "area" element on $\mathscr{B}$ (Folland, 2002; Federer, 2014).

Theorem SA-2.2 (MSE Expansions)
Suppose Assumptions SA–1 and SA–2 hold. If $\frac{\log(1/h)}{nh^d} = o(1)$ *and* $h = o(1)$, *then*

$$\mathrm{MSE}_{\boldsymbol{\nu}}(\mathbf{x}) = (h^{p+1-|\boldsymbol{\nu}|}B^{(\boldsymbol{\nu})}_{\mathbf{x}})^2 + n^{-1}h^{-d-2|\boldsymbol{\nu}|}V^{(\boldsymbol{\nu})}_{\mathbf{x}} + o_{\mathbb{P}}\big(h^{2p+2-2|\boldsymbol{\nu}|} + n^{-1}h^{-d-2|\boldsymbol{\nu}|}\big), \qquad \mathbf{x}\in\mathscr{B},$$

$$\mathrm{IMSE}_{\boldsymbol{\nu}} = \int_{\mathscr{B}} \left[(h^{p+1-|\boldsymbol{\nu}}B^{(\boldsymbol{\nu})}_{\mathbf{x}})^2 + n^{-1}h^{-d-2|\boldsymbol{\nu}|}V^{(\boldsymbol{\nu})}_{\mathbf{x}}\right]\omega(\mathbf{x})dH^{d-1}(\mathbf{x}) + o_{\mathbb{P}}\big(h^{2p+2-2|\boldsymbol{\nu}|} + n^{-1}h^{-d-2|\boldsymbol{\nu}|}\big).$$

With the estimated $\widehat{B}^{(\boldsymbol{\nu})}_{\mathbf{x}}$ *and* $\widehat{V}^{(\boldsymbol{\nu})}_{\mathbf{x}}$, *suppose* $\frac{\log(1/h)}{n^{\frac{v}{2+v}}h^d} = o(1)$ *and* $h = o(1)$, *then*

$$\mathrm{MSE}_{\boldsymbol{\nu}}(\mathbf{x}) = (h^{p+1-|\boldsymbol{\nu}|}\widehat{B}^{(\boldsymbol{\nu})}_{\mathbf{x}})^2 + n^{-1}h^{-d-2|\boldsymbol{\nu}|}\widehat{V}^{(\boldsymbol{\nu})}_{\mathbf{x}} + o_{\mathbb{P}}\big(h^{2p+2-2|\boldsymbol{\nu}|} + n^{-1}h^{-d-2|\boldsymbol{\nu}|}\big), \qquad \mathbf{x}\in\mathscr{B},$$

$$\mathrm{IMSE}_{\boldsymbol{\nu}} = \int_{\mathscr{B}} \left[(h^{p+1-|\boldsymbol{\nu}}\widehat{B}^{(\boldsymbol{\nu})}_{\mathbf{x}})^2 + n^{-1}h^{-d-2|\boldsymbol{\nu}|}\widehat{V}^{(\boldsymbol{\nu})}_{\mathbf{x}}\right]\omega(\mathbf{x})dH^{d-1}(\mathbf{x}) + o_{\mathbb{P}}\big(h^{2p+2-2|\boldsymbol{\nu}|} + n^{-1}h^{-d-2|\boldsymbol{\nu}|}\big).$$

If $\widehat{B}_{\mathbf{x}}^{(\boldsymbol{\nu})} \neq 0$, the asymptotic MSE-optimal bandwidth is

$$h_{\mathrm{MSE},\boldsymbol{\nu},p}(\mathbf{x}) = \left(\frac{(d+2|\boldsymbol{\nu}|)\widehat{V}_{\mathbf{x}}^{(\boldsymbol{\nu})}}{(2p+2-2|\boldsymbol{\nu}|)n(\widehat{B}_{\mathbf{x}}^{(\boldsymbol{\nu})})^2} \right)^{\frac{1}{2p+d+2}}, \qquad \mathbf{x} \in \mathscr{B}.$$

If $\int_{\mathscr{B}} (B_{\mathbf{x}}^{(\boldsymbol{\nu})})^2 \omega(\mathbf{x}) dH^{d-1}(\mathbf{x}) \neq 0$, the asymptotic IMSE-optimal bandwidth is

$$h_{\mathrm{IMSE},\boldsymbol{\nu},p} = \left(\frac{(d+2|\boldsymbol{\nu}|) \int_{\mathscr{B}} \widehat{V}_{\mathbf{x}}^{(\boldsymbol{\nu})} \omega(\mathbf{x}) dH^{d-1}(\mathbf{x})}{(2p+2-2|\boldsymbol{\nu}|)n \int_{\mathscr{B}} (\widehat{B}_{\mathbf{x}}^{(\boldsymbol{\nu})})^2 \omega(\mathbf{x}) dH^{d-1}(\mathbf{x})} \right)^{\frac{1}{2p+d+2}}.$$

SA-2.3 Pointwise Inference

For $|\boldsymbol{\nu}| \leq p$, define the feasible t-statistics:

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

Theorem SA-2.3 (Asymptotic Normality)
Suppose Assumptions SA–1 and SA–2 hold. If $nh^d \to \infty$ and $nh^d h^{2(p+1)} \to 0$, then

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

For any $0 < \alpha < 1$, define the confidence interval:

$$\widehat{\mathrm{I}}_{\alpha}^{(\boldsymbol{\nu})}(\mathbf{x}) = \left[\widehat{\tau}^{(\boldsymbol{\nu})}(\mathbf{x}) - \mathfrak{c}_\alpha \sqrt{\widehat{\Omega}_{\mathbf{x},\mathbf{x}}^{(\boldsymbol{\nu})}}, \widehat{\tau}^{(\boldsymbol{\nu})}(\mathbf{x}) + \mathfrak{c}_\alpha \sqrt{\widehat{\Omega}_{\mathbf{x},\mathbf{x}}^{(\boldsymbol{\nu})}} \right],$$

where $\mathfrak{c}_\alpha = \inf\{c > 0 : \mathbb{P}(|\widehat{Z}| \geq c | \mathbf{W}_n) \leq \alpha\}$ with $\widehat{Z}|\mathbf{X} \sim \mathsf{Normal}(0, \widehat{\Omega}_{\mathbf{x},\mathbf{x}}^{(\boldsymbol{\nu})})$, for each $\mathbf{x} \in \mathscr{B}$.

Theorem SA-2.4 (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

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

SA-2.4 Uniform Inference

Theorem SA-2.5 (Uniform Convergence Rate)
Suppose Assumptions SA–1 and SA–2 hold. If $\frac{\log(1/h)}{nh^d} = o(1)$, then

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

$\widehat{\mathrm{T}}^{(\boldsymbol{\nu})}$ is not directly a sum of i.i.d terms. For $\mathbf{x} \in \mathscr{B}$, we define the *stochastic linearization* of $\widehat{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})$ to be

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

with $u_i = Y_i - \sum_{t\in\{0,1\}} \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)\mu_t(\mathbf{X}_i)$.

Theorem SA-2.6 (Stochastic Linearization)
Suppose Assumptions SA–1 and SA–2 hold. If $\frac{\log(1/h)}{n^{\frac{v}{2+v}}h^d} = o(1)$, *then*

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

Next, we exploit a structure of $(\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x}) : \mathbf{x} \in \mathscr{B})$. Define the following function indexed by $\mathbf{x} \in \mathscr{B}$.

$$\begin{aligned} g_{\mathbf{x}}(\mathbf{u}) &= \mathbb{1}(\mathbf{u}\in\mathscr{A}_1)\mathscr{K}_1^{(\boldsymbol{\nu})}(\mathbf{u};\mathbf{x}) - \mathbb{1}(\mathbf{u}\in\mathscr{A}_0)\mathscr{K}_0^{(\boldsymbol{\nu})}(\mathbf{u};\mathbf{x}), && \mathbf{u}\in\mathscr{X},\\ \mathscr{K}_t^{(\boldsymbol{\nu})}(\mathbf{u};\mathbf{x}) &= n^{-1/2}(\Omega_{\mathbf{x},\mathbf{x}}^{(\boldsymbol{\nu})})^{-1/2}\mathbf{e}_{1+\boldsymbol{\nu}}^\top\mathbf{H}^{-1}\boldsymbol{\Gamma}_{t,\mathbf{x}}^{-1}\mathbf{R}_p\left(\frac{\mathbf{u}-\mathbf{x}}{h}\right)K_h(\mathbf{u}-\mathbf{x}), && \mathbf{u}\in\mathscr{X}, t\in\{0,1\}, \end{aligned}$$

and define the 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 $x \in \mathbb{R}$. Define the *residual-based empirical process* by

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

Then,

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

In Lemma SA-4.1, we provide a generic bound on the rate of Gaussian strong approximation for *residual-based empirical process*. This lemma generalizes Cattaneo and Yu (2025, Theorem 3) to allow for polynomial moment bound on the conditional distribution of Y_i given $\mathbf{X}_i$.

Theorem SA-2.7 (Strong Approximation of $\overline{\mathrm{T}}^{(\boldsymbol{\nu})}$)
Suppose Assumptions SA–1 and SA–2 hold. Suppose 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 $\mathscr{L}(E_{t,\mathbf{x}}) \leq Ch^{d-1}$. Suppose $\liminf_{n\to\infty} \frac{\log h}{\log n} > -\infty$ 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

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

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, $\mathbb{C}\mathsf{ov}[\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x}_1), \overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x}_2)] = \mathbb{C}\mathsf{ov}[Z^{(\boldsymbol{\nu})}(\mathbf{x}_1), Z^{(\boldsymbol{\nu})}(\mathbf{x}_2)]$ for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathscr{B}$.

For confidence bands, let $\widehat{Z}^{(\boldsymbol{\nu})}(\mathbf{x})$, $\mathbf{x} \in \mathscr{B}$, be a mean-zero Gaussian process with feasible (conditional) covariance function given by

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

Theorem SA-2.8 (Distributional Approximation for Suprema)
Suppose Assumptions SA–1 and SA–2 hold. Suppose $\liminf_{n\to\infty} \frac{\log h}{\log n} > -\infty$, $h^{p+1}\sqrt{nh^d} \to 0$ *and* $\frac{n^{\frac{v}{2+v}} h^d}{(\log n)^3} \to \infty$, *then*

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

where $\mathbf{W}_n = ((\mathbf{X}_1^\top, Y_1), \cdots, (\mathbf{X}_n^\top, Y_n))^\top$.

For any $0 < \alpha < 1$, define the confidence bands by

$$\widehat{\mathrm{I}}_\alpha^{(\boldsymbol{\nu})}(\mathbf{x}) = \left[\widehat{\tau}^{(\boldsymbol{\nu})}(\mathbf{x}) - \mathfrak{c}_\alpha \sqrt{\widehat{\Omega}^{(\boldsymbol{\nu})}_{\mathbf{x},\mathbf{x}}}, \widehat{\tau}^{(\boldsymbol{\nu})}(\mathbf{x}) + \mathfrak{c}_\alpha \sqrt{\widehat{\Omega}^{(\boldsymbol{\nu})}_{\mathbf{x},\mathbf{x}}} \right], \qquad \mathbf{x}\in\mathscr{B},$$

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

Theorem SA-2.9 (Confidence bands)
Suppose Assumptions SA–1 and SA–2 hold. Suppose $\liminf_{n\to\infty} \frac{\log h}{\log n} > -\infty$, $h^{p+1}\sqrt{nh^d} \to 0$ *and* $\frac{n^{\frac{v}{2+v}} h^d}{(\log n)^3} \to \infty$, *then*

$$\mathbb{P}\big[\mu^{(\boldsymbol{\nu})}(\mathbf{x}) \in \widehat{\mathrm{I}}_\alpha^{(\boldsymbol{\nu})}(\mathbf{x}), \forall \mathbf{x}\in\mathscr{B}\big] = 1 - \alpha - o(1).$$

SA-3 Analysis Based on Univariate Distance

The treatment effect curve estimator for $(\tau(\mathbf{x}) : \mathbf{x}\in\mathscr{B})$ is

$$\widehat{\tau}_{\mathrm{dis}}(\mathbf{x}) = \widehat{\theta}_{1,\mathbf{x}}(0) - \widehat{\theta}_{0,\mathbf{x}}(0), \qquad \mathbf{x}\in\mathscr{B},$$

where, for $t\in\{0,1\}$, $\widehat{\theta}_{t,\mathbf{x}}(0) = \mathbf{e}_1^\top \widehat{\boldsymbol{\gamma}}_t(\mathbf{x})$ with

$$\widehat{\boldsymbol{\gamma}}_t(\mathbf{x}) = \operatorname*{argmin}_{\boldsymbol{\gamma}\in\mathbb{R}^{p+1}} \mathbb{E}_n\Big[\big(Y_i - \mathbf{r}_p(D_i(\mathbf{x}))^\top \boldsymbol{\gamma} \big)^2 k_h(D_i(\mathbf{x})) \mathbb{1}_{\mathscr{I}_t}(D_i(\mathbf{x})) \Big],$$

where the univariate distance score is

$$D_i(\mathbf{x}) = \mathbb{1}(\mathbf{X}_i\in\mathscr{A}_1)\mathscr{d}(\mathbf{X}_i,\mathbf{x}) - \mathbb{1}(\mathbf{X}_i\in\mathscr{A}_0)\mathscr{d}(\mathbf{X}_i,\mathbf{x}), \qquad \mathbf{x}\in\mathscr{B},$$

$\mathbf{r}_p(u) = (1, u, \cdots, u^p)^\top$, $k_h(u) = k(u/h)/h^2$ for a univariate kernel $k(\cdot)$ and a bandwidth parameter h, and $\mathbb{1}_{\mathscr{I}_t}(D_i(\mathbf{x})) = \mathbb{1}(D_i(\mathbf{x})\in\mathscr{I}_t)$ with $\mathscr{I}_0 = (-\infty, 0)$ and $\mathscr{I}_1 = [0,\infty)$. More generally,

$$\widehat{\theta}_{t,\mathbf{x}}(D_i(\mathbf{x})) = \mathbf{r}_p(D_i(\mathbf{x}))^\top \widehat{\boldsymbol{\gamma}}_t(\mathbf{x}), \qquad t\in\{0,1\}, \quad \mathbf{x}\in\mathscr{B}.$$

We impose the following assumptions on the distance function, kernel function, and assignment boundary.

Assumption SA–3 (Regularity Conditions for Distance)
$\mathscr{d} : \mathbb{R}^d\times\mathbb{R}^d \to \mathbb{R}$ *is a metric on* $\mathbb{R}^d$ *equivalent to the Euclidean distance, that is, there exists positive constants* C_u *and* C_l *such that* $C_l\|\mathbf{x}-\mathbf{x}'\| \leq \mathscr{d}(\mathbf{x},\mathbf{x}') \leq C_u\|\mathbf{x}-\mathbf{x}'\|$ *for all* $\mathbf{x},\mathbf{x}'\in\mathscr{X}$.

Assumption SA–4 (Kernel Function)
Let $t\in\{0,1\}$.

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

(ii) $\liminf_{h \downarrow 0} \inf_{\mathbf{x} \in \mathscr{B}} \int_{\mathscr{A}_t} k_h(\mathscr{d}(\mathbf{u}, \mathbf{x})) d\mathbf{u} \gtrsim 1$.

For each $t \in \{0, 1\}$, the induced conditional expectation based on univariate distance is

$$\theta_{t,\mathbf{x}}(r) = \mathbb{E}[Y_i | D_i(\mathbf{x}) = r] = \mathbb{E}[Y_i | \mathscr{d}(\mathbf{X}_i, \mathbf{x}) = |r|, \mathbf{X}_i \in \mathscr{A}_t], \qquad r \in \mathscr{I}_t, \quad \mathbf{x} \in \mathscr{B}.$$

More rigorously, for each $t \in \{0, 1\}$, let $S_{t,\mathbf{x}}(r) = \{\mathbf{v} \in \mathscr{X} : \mathscr{d}(\mathbf{v}, \mathbf{x}) = r, \mathbf{v} \in \mathscr{A}_t\}$ for $r \geq 0$ and $\mathbf{x} \in \mathscr{B}$. Letting H_{d-1} denote the $(d-1)$-dimensional Hausdorff measure, then our definition means

$$\theta_{t,\mathbf{x}}(r) = \mathbb{E}[Y_i | \mathscr{d}(\mathbf{X}_i, \mathbf{x}) = |r|, \mathbf{X}_i \in \mathscr{A}_t] = \frac{\int_{S_{t,\mathbf{x}}(|r|)} \mu_t(\mathbf{v}) f_X(\mathbf{v}) H_{d-1}(d\mathbf{v})}{\int_{S_{t,\mathbf{x}}(|r|)} f_X(\mathbf{v}) H_{d-1}(d\mathbf{v})},$$

for $|r| > 0, \mathbf{x} \in \mathscr{B}, t \in \{0, 1\}$. For $r = 0$, $\mathbf{x} \in \mathscr{B}$, $t \in \{0, 1\}$, then

$$\theta_{t,\mathbf{x}}(0) = \lim_{r \to 0} \mathbb{E}[Y_i | \mathscr{d}(\mathbf{X}_i, \mathbf{x}) = |r|, \mathbf{X}_i \in \mathscr{A}_t] = \lim_{r \to 0} \frac{\int_{S_{t,\mathbf{x}}(|r|)} \mu_t(\mathbf{v}) f_X(\mathbf{v}) H_{d-1}(d\mathbf{v})}{\int_{S_{t,\mathbf{x}}(|r|)} f_X(\mathbf{v}) H_{d-1}(d\mathbf{v})}.$$

Under our assumptions, the above limit exists, and thus we obtain the following identification result.

Lemma SA-3.1 (Distance-Based Identification)
Suppose Assumption SA–1 (i)-(iii), and Assumption SA–3 hold. Then, $\theta_{t,\mathbf{x}}(0) = \mu_t(\mathbf{x})$, *for all* $t \in \{0, 1\}$ *and* $\mathbf{x} \in \mathscr{B}$.

For $t \in \{0, 1\}$, define the best mean square approximation

$$\theta^*_{t,\mathbf{x}}(D_i(\mathbf{x})) = \mathbf{r}_p(D_i(\mathbf{x}))^\top \boldsymbol{\gamma}^*_t(\mathbf{x}),$$

where

$$\boldsymbol{\gamma}^*_t(\mathbf{x}) = \underset{\boldsymbol{\gamma} \in \mathbb{R}^{p+1}}{\operatorname{argmin}} \mathbb{E}\Big[\big(Y_i - \mathbf{r}_p(D_i(\mathbf{x}))^\top \boldsymbol{\gamma}\big)^2 k_h(D_i(\mathbf{x})) \mathbb{1}_{\mathscr{I}_t}(D_i(\mathbf{x})) \Big].$$

The estimation error decomposes into *linear error*, *approximation error*, and *non-linear error*:

$$\begin{aligned}
\widehat{\theta}_{t,\mathbf{x}}(0) - \theta_{t,\mathbf{x}}(0) &= \mathbf{e}_1^\top \widehat{\boldsymbol{\Psi}}_{t,\mathbf{x}}^{-1} \mathbb{E}_n \left[\mathbf{r}_p \left(\frac{D_i(\mathbf{x})}{h} \right) k_h(D_i(\mathbf{x})) Y_i \right] - \theta_{t,\mathbf{x}}(0) \\
&= \mathbf{e}_1^\top \widehat{\boldsymbol{\Psi}}_{t,\mathbf{x}}^{-1} \mathbb{E}_n \left[\mathbf{r}_p \left(\frac{D_i(\mathbf{x})}{h} \right) k_h(D_i(\mathbf{x}))(Y_i - \theta^*_{t,\mathbf{x}}(D_i)) \right] + \theta^*_{t,\mathbf{x}}(0) - \theta_{t,\mathbf{x}}(0) \\
&= \underbrace{\mathbf{e}_1^\top \boldsymbol{\Psi}_{t,\mathbf{x}}^{-1} \mathbf{O}_{t,\mathbf{x}}}_{\text{linear error}} + \underbrace{\theta^*_{t,\mathbf{x}}(0) - \theta_{t,\mathbf{x}}(0)}_{\text{approximation error}} + \underbrace{\mathbf{e}_1^\top (\widehat{\boldsymbol{\Psi}}_{t,\mathbf{x}}^{-1} - \boldsymbol{\Psi}_{t,\mathbf{x}}^{-1}) \mathbf{O}_{t,\mathbf{x}}}_{\text{non-linear error}}, \qquad \text{(SA-3.1)}
\end{aligned}$$

for all $t \in \{0,1\}$ and $\mathbf{x} \in \mathscr{B}$, where

$$\widehat{\mathbf{\Psi}}_{t,\mathbf{x}} = \mathbb{E}_n \left[\mathbf{r}_p \left(\frac{D_i(\mathbf{x})}{h} \right) \mathbf{r}_p \left(\frac{D_i(\mathbf{x})}{h} \right)^\top k_h(D_i(\mathbf{x})) \mathbb{1}_{\mathscr{I}_t}(D_i(\mathbf{x})) \right],$$

$$\mathbf{\Psi}_{t,\mathbf{x}} = \mathbb{E} \left[\mathbf{r}_p \left(\frac{D_i(\mathbf{x})}{h} \right) \mathbf{r}_p \left(\frac{D_i(\mathbf{x})}{h} \right)^\top k_h(D_i(\mathbf{x})) \mathbb{1}_{\mathscr{I}_t}(D_i(\mathbf{x})) \right],$$

$$\mathbf{O}_{t,\mathbf{x}} = \mathbb{E}_n \left[\mathbf{r}_p \left(\frac{D_i(\mathbf{x})}{h} \right) k_h(D_i(\mathbf{x}))(Y_i - \theta^*_{t,\mathbf{x}}(D_i(\mathbf{x}))) \mathbb{1}_{\mathscr{I}_t}(D_i(\mathbf{x})) \right],$$

and the misspecification bias is

$$\mathfrak{B}_{n,t}(\mathbf{x}) = \theta^*_{t,\mathbf{x}}(0) - \theta_{t,\mathbf{x}}(0). \tag{SA-3.2}$$

In the main text, $\mathfrak{B}_n(\mathbf{x}) = \mathfrak{B}_{n,1}(\mathbf{x}) - \mathfrak{B}_{n,0}(\mathbf{x})$. Define the following for variance analysis: For $t \in \{0,1\}$, $\mathbf{x}_1, \mathbf{x}_2 \in \mathscr{B}$,

$$\begin{aligned}
\widehat{\mathbf{\Upsilon}}_{t,\mathbf{x}_1,\mathbf{x}_2} &= h^d \mathbb{E}_n \Bigg[\mathbf{r}_p \left(\frac{D_i(\mathbf{x}_1)}{h} \right) \mathbf{r}_p \left(\frac{D_i(\mathbf{x}_2)}{h} \right)^\top k_h \left(D_i(\mathbf{x}_1) \right) k_h \left(D_i(\mathbf{x}_2) \right) \left(Y_i - \widehat{\theta}_{t,\mathbf{x}_1}(D_i(\mathbf{x}_1)) \right) \\
&\qquad (Y_i - \widehat{\theta}_{t,\mathbf{x}_2}(D_i(\mathbf{x}_2))) \mathbb{1}_{\mathscr{I}_t}(D_i(\mathbf{x}_1)) \Bigg], \\
\mathbf{\Upsilon}_{t,\mathbf{x}_1,\mathbf{x}_2} &= h^d \mathbb{E} \Bigg[\mathbf{r}_p \left(\frac{D_i(\mathbf{x}_1)}{h} \right) \mathbf{r}_p \left(\frac{D_i(\mathbf{x}_2)}{h} \right)^\top k_h(D_i(\mathbf{x}_1)) k_h(D_i(\mathbf{x}_2))(Y_i - \theta^*_{t,\mathbf{x}_1}(D_i(\mathbf{x}_1))) \\
&\qquad (Y_i - \theta^*_{t,\mathbf{x}_2}(D_i(\mathbf{x}_2))) \mathbb{1}_{\mathscr{I}_t}(D_i(\mathbf{x}_1)) \Bigg], \\
\widehat{\Xi}_{t,\mathbf{x}_1,\mathbf{x}_2} &= \frac{1}{nh^d} \mathbf{e}_1^\top \widehat{\mathbf{\Psi}}^{-1}_{t,\mathbf{x}_1} \widehat{\mathbf{\Upsilon}}_{t,\mathbf{x}_1,\mathbf{x}_2} \widehat{\mathbf{\Psi}}^{-1}_{t,\mathbf{x}_2} \mathbf{e}_1, \qquad \widehat{\Xi}_{\mathbf{x}_1,\mathbf{x}_2} = \widehat{\Xi}_{0,\mathbf{x}_1,\mathbf{x}_2} + \widehat{\Xi}_{1,\mathbf{x}_1,\mathbf{x}_2}, \\
\Xi_{t,\mathbf{x}_1,\mathbf{x}_2} &= \frac{1}{nh^d} \mathbf{e}_1^\top \mathbf{\Psi}^{-1}_{t,\mathbf{x}_1} \mathbf{\Upsilon}_{t,\mathbf{x}_1,\mathbf{x}_2} \mathbf{\Psi}^{-1}_{t,\mathbf{x}_2} \mathbf{e}_1, \qquad \Xi_{\mathbf{x}_1,\mathbf{x}_2} = \Xi_{0,\mathbf{x}_1,\mathbf{x}_2} + \Xi_{1,\mathbf{x}_1,\mathbf{x}_2}.
\end{aligned}$$

SA-3.1 Preliminary Lemmas

Lemma SA-3.2 (Gram)

Suppose Assumption SA-1 (i)(ii), Assumption SA-3 and Assumption SA-4 hold. If $\frac{nh^d}{\log(1/h)} \to \infty$, then

$$\begin{aligned}
\sup_{x \in \mathscr{B}} \|\widehat{\mathbf{\Psi}}_{t,\mathbf{x}} - \mathbf{\Psi}_{t,\mathbf{x}}\| &\lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}}, \\
1 \lesssim_{\mathbb{P}} \inf_{\mathbf{x} \in \mathscr{B}} \|\widehat{\mathbf{\Psi}}_{t,\mathbf{x}}\| &\lesssim \sup_{\mathbf{x} \in \mathscr{B}} \|\widehat{\mathbf{\Psi}}_{t,\mathbf{x}}\| \lesssim_{\mathbb{P}} 1, \\
\sup_{\mathbf{x} \in \mathscr{B}} \|\widehat{\mathbf{\Psi}}^{-1}_{t,\mathbf{x}} - \mathbf{\Psi}^{-1}_{t,\mathbf{x}}\| &\lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}},
\end{aligned}$$

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

Lemma SA-3.3 (Stochastic Linear Approximation)

Suppose Assumption SA-1 (i)(ii)(iii)(v), Assumption SA-3 and Assumption SA-4 hold. If $\frac{nh^d}{\log(1/h)} \to \infty$, *then*

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

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

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

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

Lemma SA-3.4 (Approximation Error: Minimal Guarantee)
Suppose Assumption SA–1 (i)(ii)(iii), Assumption SA–3 and Assumption SA–4 hold. Then,

$$\sup_{\mathbf{x}\in\mathscr{B}} |\mathfrak{B}_n(\mathbf{x})| \lesssim h.$$

Lemma SA-3.5 (Covariance)
Suppose Assumptions SA–1, SA–3 and SA–4 hold. If $\frac{nh^d}{\log(1/h)} \to \infty$*, then*

$$\max_{t\in\{0,1\}} \sup_{\mathbf{x}_1,\mathbf{x}_2\in\mathscr{B}} \|\widehat{\mathbf{\Upsilon}}_{t,\mathbf{x}_1,\mathbf{x}_2} - \mathbf{\Upsilon}_{t,\mathbf{x}_1,\mathbf{x}_2}\| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{v}{2+v}} h^d},$$

$$\max_{t\in\{0,1\}} \sup_{\mathbf{x}_1,\mathbf{x}_2\in\mathscr{B}} nh^d |\widehat{\Xi}_{t,\mathbf{x}_1,\mathbf{x}_2} - \Xi_{t,\mathbf{x}_1,\mathbf{x}_2}| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{v}{2+v}} h^d}.$$

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

$$\inf_{\mathbf{x}\in\mathscr{B}} \lambda_{\min}(\widehat{\mathbf{\Upsilon}}_{t,\mathbf{x},\mathbf{x}}) \gtrsim_{\mathbb{P}} 1,$$

$$\inf_{\mathbf{x}\in\mathscr{B}} \widehat{\Xi}_{t,\mathbf{x},\mathbf{x}} \gtrsim_{\mathbb{P}} (nh^d)^{-1},$$

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

Since we consider a covariance estimator based on the best linear approximation, instead of the population conditional mean functions, no bias condition appears in the estimates above.

SA-3.2 Pointwise Inference

Theorem SA-3.1 (Convergence Rate)
Suppose Assumptions SA–1, SA–3 and SA–4 hold. If $nh^d \to \infty$*, then*

$$|\widehat{\tau}_{\mathrm{dis}}(\mathbf{x}) - \tau(\mathbf{x})| \lesssim_{\mathbb{P}} \frac{1}{\sqrt{nh^d}} + \frac{1}{n^{\frac{1+v}{2+v}} h^d} + |\mathfrak{B}_n(\mathbf{x})|,$$

for all $\mathbf{x} \in \mathscr{B}$.

Define the feasible t-statistics by

$$\widehat{\mathrm{T}}_{\mathrm{dis}}(\mathbf{x}) = \frac{\widehat{\tau}_{\mathrm{dis}}(\mathbf{x}) - \tau(\mathbf{x})}{\sqrt{\widehat{\Xi}_{\mathbf{x},\mathbf{x}}}}, \qquad \mathbf{x} \in \mathscr{B}.$$

Theorem SA-3.2 (Asymptotic Normality)
Suppose Assumptions SA–1, SA–3 and SA–4 hold. If $n^{\frac{v}{2+v}} h^d \to \infty$ *and* $\sqrt{nh^d} \sup_{\mathbf{x}\in\mathscr{B}} |\mathfrak{B}_n(\mathbf{x})| \to 0$, *then*

$$\sup_{u\in\mathbb{R}} \left| \mathbb{P}\left(\widehat{\mathrm{T}}_{\mathrm{dis}}(\mathbf{x}) \leq u\right) - \Phi(u) \right| = o(1), \qquad \forall \mathbf{x} \in \mathscr{B}.$$

For any $0 < \alpha < 1$, *take* $\mathfrak{c}_\alpha = \inf\{c > 0 : \mathbb{P}(|Z| \geq c) \leq \alpha\}$ *where* $Z \sim N(0,1)$, *and define* $\widehat{\mathrm{I}}_{\mathrm{dis}}(\mathbf{x},\alpha) = \left(\widehat{\tau}_{\mathrm{dis}}(\mathbf{x}) - \mathfrak{c}_\alpha\sqrt{\widehat{\Xi}_{\mathbf{x},\mathbf{x}}}, \widehat{\tau}_{\mathrm{dis}}(\mathbf{x}) + \mathfrak{c}_\alpha\sqrt{\widehat{\Xi}_{\mathbf{x},\mathbf{x}}}\right)$. *Then,*

$$\mathbb{P}\left(\tau(\mathbf{x}) \in \widehat{\mathrm{I}}_{\mathrm{dis}}(\mathbf{x},\alpha)\right) \to 1 - \alpha, \qquad \mathbf{x} \in \mathscr{B}.$$

SA-3.3 Uniform Inference

Theorem SA-3.3 (Uniform Convergence Rate)
Suppose Assumptions SA–1, SA–3 and SA–4 hold. If $\frac{nh^d}{\log(1/h)} \to \infty$, *then*

$$\sup_{\mathbf{x}\in\mathscr{B}} |\widehat{\tau}_{\mathrm{dis}}(\mathbf{x}) - \tau(\mathbf{x})| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{1+v}{2+v}} h^d} + \sup_{\mathbf{x}\in\mathscr{B}} |\mathfrak{B}_n(\mathbf{x})|.$$

Define $\overline{\mathrm{T}}_{\mathrm{dis}}(\mathbf{x})$ to be the stochastic linearization of $\widehat{\mathrm{T}}_{\mathrm{dis}}(\mathbf{x})$, that is, we define

$$\overline{\mathrm{T}}_{\mathrm{dis}}(\mathbf{x}) = \Xi_{\mathbf{x},\mathbf{x}}^{-1/2}(\mathbf{e}_1^\top \mathbf{\Psi}_{1,\mathbf{x}}^{-1}\mathbf{O}_{1,\mathbf{x}} - \mathbf{e}_1^\top \mathbf{\Psi}_{0,\mathbf{x}}^{-1}\mathbf{O}_{0,\mathbf{x}}), \qquad \mathbf{x} \in \mathscr{B}.$$

Theorem SA-3.4 (Stochastic Linearization)
Suppose Assumptions SA–1, SA–3 and SA–4 hold. Suppose $\frac{nh^d}{\log(1/h)} \to \infty$. *Then,*

$$\sup_{\mathbf{x}\in\mathscr{B}} \left|\widehat{\mathrm{T}}_{\mathrm{dis}}(\mathbf{x}) - \overline{\mathrm{T}}_{\mathrm{dis}}(\mathbf{x})\right| \lesssim_{\mathbb{P}} \sqrt{\log(1/h)}\left(\sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{v}{2+v}} h^d}\right) + \sqrt{nh^d} \sup_{\mathbf{x}\in\mathscr{B}} |\mathfrak{B}_n(\mathbf{x})|.$$

To establish a Gaussian strong approximation for $\overline{\mathrm{T}}_{\mathrm{dis}}(\mathbf{x})$, consider the class of functions $\mathscr{G} = \{g_{\mathbf{x}} : \mathbf{x} \in \mathscr{B}\}$ and $\mathscr{H} = \{h_{\mathbf{x}} : \mathbf{x} \in \mathscr{B}\}$, where

$$\begin{aligned}
g_{\mathbf{x}}(\mathbf{u}) &= \mathbb{1}_{\mathscr{A}_1}(\mathbf{u})\mathfrak{K}_1(\mathbf{u};\mathbf{x}) - \mathbb{1}_{\mathscr{A}_0}(\mathbf{u})\mathfrak{K}_0(\mathbf{u};\mathbf{x}), && \mathbf{u} \in \mathscr{X}, \\
\mathfrak{K}_t(\mathbf{u};\mathbf{x}) &= \frac{1}{\sqrt{n\Xi_{\mathbf{x},\mathbf{x}}}} \mathbf{e}_1^\top \mathbf{\Psi}_{t,\mathbf{x}}^{-1} \mathbf{r}_p\left(\frac{\mathscr{d}(\mathbf{u},\mathbf{x})}{h}\right) k_h(\mathscr{d}(\mathbf{u},\mathbf{x})), && \mathbf{u} \in \mathscr{X}, \mathbf{x} \in \mathscr{B}, t \in \{0,1\}, \\
h_{\mathbf{x}}(\mathbf{u}) &= -\mathbb{1}_{\mathscr{A}_1}(\mathbf{u})\mathfrak{K}_1(\mathbf{u};\mathbf{x})\theta_{1,\mathbf{x}}^*(\mathscr{d}(\mathbf{u},\mathbf{x})) + \mathbb{1}_{\mathscr{A}_0}(\mathbf{u})\mathfrak{K}_0(\mathbf{u};\mathbf{x})\theta_{0,\mathbf{x}}^*(\mathscr{d}(\mathbf{u},\mathbf{x})), && \mathbf{u} \in \mathscr{X}, \mathbf{x} \in \mathscr{B},
\end{aligned} \tag{SA-3.3}$$

and $\mathscr{R}$ is the singleton of identity function $\mathrm{Id} : \mathbb{R} \mapsto \mathbb{R}$, $\mathrm{Id}(x) = x$. For classes of functions $\mathscr{G}$, $\mathscr{H}$ from $\mathbb{R}^d$ to

$\mathbb{R}$ and $\mathscr{R}$ from $\mathbb{R}$ to $\mathbb{R}$. Then, for $\mathbf{x} \in \mathscr{B}$, $\overline{\mathrm{T}}_{\mathrm{dis}}(\mathbf{x})$ can be represented by

$$\overline{\mathrm{T}}_{\mathrm{dis}}(\mathbf{x}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big[g_{\mathbf{x}}(\mathbf{X}_i) \operatorname{Id}(y_i) + h_{\mathbf{x}}(\mathbf{X}_i) - \mathbb{E}[g_{\mathbf{x}}(\mathbf{X}_i) \operatorname{Id}(y_i) + h_{\mathbf{x}}(\mathbf{X}_i)] \Big].$$

Define the multiplicative separable empirical processes by

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

Then, $\overline{\mathrm{T}}_{\mathrm{dis}}(, \mathbf{x})$ has the representation

$$\overline{\mathrm{T}}_{\mathrm{dis}}(\mathbf{x}) = M_n(g_{\mathbf{x}}, \mathrm{Id}) + M_n(h_{\mathbf{x}}, 1), \qquad \mathbf{x} \in \mathscr{B}.$$

In Lemma SA-4.2, we give upper bounds for Gaussian strong approximation of *additive empirical process* of the form $(M_n(g,r) + M_n(h,s) : g \in \mathscr{G}, r \in \mathscr{R}, h \in \mathscr{H}, s \in \mathscr{S})$. Since upper bounds for empirical processes of the form $(M_n(g,r) : g \in \mathscr{G}, r \in \mathscr{R})$ has already been studied in (Cattaneo and Yu, 2025, Theorem SA.1), Lemma SA-4.2 is given as its simple extension, considering the worse case between $\mathscr{G}$ and $\mathscr{H}$, and between $\mathscr{R}$ and $\mathscr{S}$. Applying Lemma SA-4.2, we get the following theorem on Gaussian strong approximation of $(\overline{\mathrm{T}}_{\mathrm{dis}}(\mathbf{x}) : \mathbf{x} \in \mathscr{B})$.

Theorem SA-3.5 (Strong Approximation of t-statistics)
Suppose Assumption SA–1, SA–3 and SA–4 hold. Suppose 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 $E_{t,\mathbf{x}} = \{\mathbf{y} \in \mathscr{A}_t : (\mathbf{y} - \mathbf{x})/h \in \operatorname{Supp}(K)\}$ satisfies $\mathscr{L}(E_{t,\mathbf{x}}) \le C h^{d-1}$. Suppose $\liminf_{n\to\infty} \frac{\log h}{\log n} > -\infty$ and $nh^d \to \infty$ as $n \to \infty$. Then, on a possibly enlarged probability space there exists a mean-zero Gaussian process z indexed by $\mathscr{B}$ with almost surely continuous sample path such that

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

where $\lesssim$ is up to a universal constant. Moreover, z has the same covariance structure as $\overline{\mathrm{T}}_{\mathrm{dis}}$, that is,

$$\mathbb{C}\mathsf{ov}\left[\overline{\mathrm{T}}_{\mathrm{dis}}(\mathbf{x}_1), \overline{\mathrm{T}}_{\mathrm{dis}}(\mathbf{x}_2) \right] = \mathbb{C}\mathsf{ov}\left[z(\mathbf{x}_1), z(\mathbf{x}_2) \right], \qquad \forall \mathbf{x}, \mathbf{y} \in \mathscr{B}.$$

Theorem SA-3.6 (Confidence Bands)
Suppose Assumption SA–1, SA–3 and SA–4 hold. Suppose $\liminf_{n\to\infty} \frac{\log h}{\log n} > -\infty$, $\frac{n^{\frac{v}{2+v}} h^d}{(\log n)^3} \to \infty$, and $\sqrt{nh^d} \sup_{\mathbf{x}\in\mathscr{B}} \sum_{t\in\{0,1\}} |\mathfrak{B}_{n,t}(\mathbf{x})| \to 0$. Suppose $\widehat{z}$ is a mean-zero Gaussian process indexed by $\mathscr{B}$ s.t.

$$\mathbb{C}\mathsf{ov}\left[\widehat{z}(\mathbf{x}_1), \widehat{z}(\mathbf{x}_2) \right] = \frac{\widehat{\Xi}_{\mathbf{x}_1,\mathbf{x}_2}}{\sqrt{\widehat{\Xi}_{\mathbf{x}_1,\mathbf{x}_1} \widehat{\Xi}_{\mathbf{x}_2,\mathbf{x}_2}}}, \qquad \mathbf{x}_1, \mathbf{x}_2 \in \mathscr{B}.$$

Let $\mathscr{U}_n$ be the σ-algebra generated by $((Y_i, (D_i(\mathbf{x}) : \mathbf{x} \in \mathscr{B})) : 1 \le i \le n)$. Then

$$\sup_{\mathbf{u}\in\mathbb{R}} \left| \mathbb{P}\left(\sup_{\mathbf{x}\in\mathscr{B}} \left| \widehat{\mathrm{T}}_{\mathrm{dis}}(\mathbf{x}) \right| \le u \right) - \mathbb{P}\left(\sup_{\mathbf{x}\in\mathscr{B}} |\widehat{z}(\mathbf{x})| \le u \middle| \mathscr{U}_n \right) \right| = o_{\mathbb{P}}(1).$$

For any $0 < \alpha < 1$, *if we define* $\mathbf{c}_\alpha = \inf\{c > 0 : \mathbb{P}(\sup_{\mathbf{x}\in\mathscr{B}} |\widehat{z}(\mathbf{x})| \geq c|\mathscr{U}_n) \leq \alpha\}$ *and define* $\widehat{\mathrm{I}}_\alpha(\mathbf{x}) = \left(\widehat{\tau}_{\mathrm{dis}}(\mathbf{x}) - \mathbf{c}_\alpha\sqrt{\widehat{\Xi}_{\mathbf{x},\mathbf{x}}}, \widehat{\tau}_{\mathrm{dis}}(\mathbf{x}) + \mathbf{c}_\alpha\sqrt{\widehat{\Xi}_{\mathbf{x},\mathbf{x}}}\right)$ *for all* $\mathbf{x} \in \mathscr{B}$, *then*

$$\mathbb{P}\left(\tau(\mathbf{x}) \in \widehat{\mathrm{I}}_\alpha(\mathbf{x}), \forall \mathbf{x} \in \mathscr{B}\right) = 1 - \alpha - o(1).$$

SA-4 Gaussian Strong Approximation Lemmas

We present two Gaussian strong approximation lemmas that are the key technical tools behind Theorem SA-2.7 and Theorem SA-3.5, building on and generalizing 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. In addition, consider the *multiplicative-separable empirical process* given by

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

SA-4.1 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 $\mathrm{Supp}(\mathscr{F}) = \cup_{f\in\mathscr{F}} \mathrm{Supp}(f)$. 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 $\mathrm{Supp}(\mathbb{P}) \cap \mathrm{Supp}(\mathscr{F})$.

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

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

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

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

   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

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

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 $\mathtt{K}_{\mathscr{F},\mathscr{C}}$ such that for any cube $\mathscr{D} \subseteq \mathbb{R}^q$ with edges of length ℓ parallel to the coordinate axises,

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

(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 $\mathscr{C} = \mathscr{Q}_{\mathscr{F}}$ for all quantities in the previous definitions. That is, to save notation, we set $\mathtt{TV}_{\mathscr{F}} = \mathtt{TV}_{\mathscr{F},\mathscr{Q}_{\mathscr{F}}}$, $\mathtt{K}_{\mathscr{F}} = \mathtt{K}_{\mathscr{F},\mathscr{Q}_{\mathscr{F}}}$, $\mathtt{M}_{\mathscr{F}} = \mathtt{M}_{\mathscr{F},\mathscr{Q}_{\mathscr{F}}}$, $M_{\mathscr{F}}(\mathbf{u}) = M_{\mathscr{F},\mathscr{Q}_{\mathscr{F}}}(\mathbf{u})$, $\mathtt{L}_{\mathscr{F}} = \mathtt{L}_{\mathscr{F},\mathscr{Q}_{\mathscr{F}}}$, and so on, whenever there is no confusion.

SA-4.2 Residual-based Empirical Process

The following Lemma SA-4.1 generalizes Cattaneo and Yu (2025, Theorem 2) by allowing y_i to have bounded moments conditional on $\mathbf{x}_i$.

Lemma SA-4.1 (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 $\mathcal{X} \subseteq \mathbb{R}^d$, y_i has distribution $\mathbb{P}_Y$ supported on $\mathcal{Y} \subseteq \mathbb{R}$, $\sup_{\mathbf{x}\in\mathcal{X}} \mathbb{E}[|y_i|^{2+v}|\mathbf{x}_i = \mathbf{x}] \leq 2$ 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 $\mathbb{Q}_\mathscr{G} = \mathfrak{m} \circ \phi_\mathscr{G}$, where the normalizing transformation $\phi_\mathscr{G} : \mathbb{Q}_\mathscr{G} \mapsto [0, 1]^d$ is a diffeomorphism.*

(iii) *$\mathscr{G}$ is a VC-type class with envelope $\mathtt{M}_\mathscr{G}$ over $\mathbb{Q}_\mathscr{G}$ 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 $(\mathbb{R}, \mathscr{B}(\mathbb{R}), \mathbb{P}_Y)$.*

(v) *$\mathscr{R}$ is a VC-type class with envelope $M_{\mathscr{R},\mathcal{Y}}$ over $\mathcal{Y}$ with $\mathtt{c}_{\mathscr{R},\mathcal{Y}} \geq e$ and $\mathtt{d}_{\mathscr{R},\mathcal{Y}} \geq 1$, where $M_{\mathscr{R},\mathcal{Y}}(y) + \mathtt{pTV}_{\mathscr{R},(-|y|,|y|)} \leq \mathtt{v}(1 + |y|)$ for all $y \in \mathcal{Y}$, 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 $\mathtt{TV} = \max\{\mathtt{TV}_\mathscr{G}, \mathtt{TV}_{\mathscr{G}\times\mathscr{V}_\mathscr{R},\mathbb{Q}_\mathscr{G}}\}$ with $\mathscr{V}_\mathscr{R} = \{\theta(\cdot, r) : r \in \mathscr{R}\}$, and $\theta(\mathbf{x}, r) = \mathbb{E}[r(y_i)|\mathbf{x}_i = \mathbf{x}]$.*

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

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

$$\mathtt{r}_n = \min\Big\{\frac{(\mathtt{c}_1^d \mathtt{M}_\mathscr{G}^{d+1} \mathtt{TV}^d \mathtt{E}_\mathscr{G})^{1/(2d+2)}}{n^{1/(2d+2)}}, \frac{(\mathtt{c}_1^{d/2} \mathtt{c}_2^{d/2} \mathtt{M}_\mathscr{G} \mathtt{TV}^{d/2} \mathtt{E}_\mathscr{G} \mathtt{L}^{d/2})^{1/(d+2)}}{n^{1/(d+2)}}\Big\},$$
$$\mathtt{c}_1 = d \sup_{\mathbf{x}\in\mathcal{Q}_\mathscr{G}} \prod_{j=1}^{d-1} \sigma_j(\nabla\phi_\mathscr{G}(\mathbf{x})), \qquad \mathtt{c}_2 = \sup_{\mathbf{x}\in\mathcal{Q}_\mathscr{G}} \frac{1}{\sigma_d(\nabla\phi_\mathscr{G}(\mathbf{x}))}.$$

SA-4.3 Multiplicative-Separable Empirical Process

The following Lemma SA-4.2 generalizes Cattaneo and Yu (2025, Theorem SA.1) by allowing y_i to have bounded moments conditional on $\mathbf{x}_i$.

Lemma SA-4.2 (Strong Approximation for $(M_n(g,r)+M_n(h,s): g\in\mathscr{G}, r\in\mathscr{R}, h\in\mathscr{H}, s\in\mathscr{S})$)
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 $\mathcal{X}\subseteq\mathbb{R}^d$, y_i has distribution $\mathbb{P}_Y$ supported on $\mathscr{Y}\subseteq\mathbb{R}$, $\sup_{\mathbf{x}\in\mathcal{X}}\mathbb{E}[|y_i|^{2+v}|\mathbf{x}_i=\mathbf{x}]\leq 2$ for some $v>0$, and the following conditions hold.

(i) *$\mathscr{G}$ and $\mathscr{H}$ are real-valued pointwise measurable classes of functions on $(\mathbb{R}^d,\mathscr{B}(\mathbb{R}^d),\mathbb{P}_X)$.*

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

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

(iv) *$\mathscr{R}$ and $\mathcal{S}$ are real-valued pointwise measurable classes of functions on $(\mathbb{R},\mathscr{B}(\mathbb{R}),\mathbb{P}_Y)$.*

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

(vi) *There exists a constant $\mathtt{k}$ such that $|\log_2\mathtt{E}|+|\log_2\mathtt{TV}|+|\log_2\mathtt{M}|\leq\mathtt{k}\log_2(n)$, where $\mathtt{E}=\max\{\mathtt{E}_{\mathscr{G},\mathcal{Q}_{\mathscr{G}\cup\mathscr{H}}},\mathtt{E}_{\mathscr{H},\mathcal{Q}_{\mathscr{G}\cup\mathscr{H}}}\}$, $\mathtt{TV}=\max\{\mathtt{TV}_{\mathscr{G},\mathcal{Q}_{\mathscr{G}\cup\mathscr{H}}},\mathtt{TV}_{\mathscr{H},\mathcal{Q}_{\mathscr{G}\cup\mathscr{H}}}\}$ and $\mathtt{M}=\max\{\mathtt{M}_{\mathscr{G},\mathcal{Q}_{\mathscr{G}\cup\mathscr{H}}},\mathtt{M}_{\mathscr{H},\mathcal{Q}_{\mathscr{G}\cup\mathscr{H}}}\}$.*

Consider the empirical process

$$A_n(g,h,r,s)=M_n(g,r)+M_n(h,s),\qquad g\in\mathscr{G},r\in\mathscr{R},h\in\mathscr{H},s\in\mathcal{S}.$$

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

- *$\mathbb{E}[A_n(g_1,h_1,r_1,s_1)A_n(g_2,h_2,r_2,s_2)]=\mathbb{E}[Z_n^A(g_1,h_1,r_1,s_1)Z_n^A(g_2,h_2,r_2,s_2)]$ holds for all (g_1,h_1,r_1,s_1), $(g_2,h_2,r_2,s_2)\in\mathscr{G}\times\mathscr{H}\times\mathscr{R}\times\mathcal{S}$, and*

- *$\mathbb{E}\big[\|A_n-Z_n^A\|_{\mathscr{G}\times\mathscr{H}\times\mathscr{R}\times\mathcal{S}}\big]\leq C\mathtt{v}((\mathtt{d}\log(\mathtt{c}n))^{\frac{3}{2}}\mathtt{r}_n^{\frac{v}{v+2}}(\sqrt{\mathtt{ME}})^{\frac{2}{v+2}}+\mathtt{d}\log(\mathtt{c}n)\mathtt{M}n^{-\frac{v/2}{2+v}}+\mathtt{d}\log(\mathtt{c}n)\mathtt{M}n^{-\frac{1}{2}}\Big(\frac{\sqrt{\mathtt{ME}}}{\mathtt{r}_n}\Big)^{\frac{2}{v+2}})$,*

where C is a universal constant, $\mathtt{c}=\mathtt{c}_{\mathscr{G},\mathcal{Q}_{\mathscr{G}\cup\mathscr{H}}}+\mathtt{c}_{\mathscr{H},\mathcal{Q}_{\mathscr{G}\cup\mathscr{H}}}+\mathtt{c}_{\mathscr{R},\mathscr{Y}}+\mathtt{c}_{\mathcal{S},\mathscr{Y}}+\mathtt{k}$, $\mathtt{d}=\mathtt{d}_{\mathscr{G},\mathcal{Q}_{\mathscr{G}\cup\mathscr{H}}}\mathtt{d}_{\mathscr{H},\mathcal{Q}_{\mathscr{G}\cup\mathscr{H}}}\mathtt{d}_{\mathscr{R},\mathscr{Y}}\mathtt{d}_{\mathcal{S},\mathscr{Y}}\mathtt{k}$,

$$\mathtt{r}_n=\min\Big\{\frac{(\mathtt{c}_1^d\mathtt{M}^{d+1}\mathtt{TV}^d\mathtt{E})^{1/(2d+2)}}{n^{1/(2d+2)}},\frac{(\mathtt{c}_1^{\frac{d}{2}}\mathtt{c}_2^{\frac{d}{2}}\mathtt{M}\mathtt{TV}^{\frac{d}{2}}\mathtt{EL}^{\frac{d}{2}})^{1/(d+2)}}{n^{1/(d+2)}}\Big\},$$
$$\mathtt{c}_1=d\sup_{\mathbf{x}\in\mathcal{Q}_{\mathscr{G}\cup\mathscr{H}}}\prod_{j=1}^{d-1}\sigma_j(\nabla\phi_{\mathscr{G}\cup\mathscr{H}}(\mathbf{x})),\qquad\mathtt{c}_2=\sup_{\mathbf{x}\in\mathcal{Q}_{\mathscr{G}\cup\mathscr{H}}}\frac{1}{\sigma_d(\nabla\phi_{\mathscr{G}\cup\mathscr{H}}(\mathbf{x}))}.$$

SA-5 Proofs for Section SA-2

SA-5.1 Proof of Lemma SA-2.1

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

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

Define $\mathscr{F} = \{g_n(\cdot, \mathbf{x}) : \mathbf{x} \in \mathscr{B}\}$. 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: When K is Lipschitz.* Let $\mathbf{x}, \mathbf{x}' \in \mathscr{B}$. Then,

$$\begin{aligned}
\sup_{\xi\in\mathscr{X}} |g_n(\xi, \mathbf{x}) - g_n(\xi, \mathbf{x}')| &\leq \left| \left(\frac{\xi_1 - \mathbf{x}_1}{h}\right)^{v_1} \cdots \left(\frac{\xi - \mathbf{x}_d}{h}\right)^{v_d} - \left(\frac{\xi_1 - \mathbf{x}_1'}{h}\right)^{v_1} \cdots \left(\frac{\xi - \mathbf{x}_d'}{h}\right)^{v_d} \right| K_h(\xi - \mathbf{x}) \\
&\quad + \left(\frac{\xi_1 - \mathbf{x}_1'}{h}\right)^{v_1} \cdots \left(\frac{\xi - \mathbf{x}_d'}{h}\right)^{v_d} |K_h(\xi - \mathbf{x}) - K_h(\xi - \mathbf{x}')| \\
&\lesssim h_n^{-d-1} \|\mathbf{x} - \mathbf{x}'\|_\infty,
\end{aligned}$$

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 $L_2(Q)$,

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

where in (1) we used the fact that $\varepsilon\|F\|_{\mathbb{Q},2} h^{d+1} \lesssim \varepsilon h \lesssim 1$. Hence $\mathscr{F}$ forms a VC-type class, and taking $A_1 = \operatorname{diam}(\mathscr{X})/h$ and $A_2 = d$,

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

where the supremum is over all finite discrete measure.

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

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

$\mathscr{H} = \{h_n(\cdot, \mathbf{x}) : \mathbf{x} \in \mathscr{B}\}$ and the constant envelope function $H = C_4 h^{-|\mathbf{v}|-d}$, 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

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

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, we have

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

It then follows that for any discrete measure Q,

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

Hence taking $A_1 = (2^d h^{-d} + 2d(32e)^d)h^{-|\mathbf{v}|}$ and $A_2 = 4d$,

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

the supremum is over all finite discrete measure.

Maximal Inequality: Using Corollary 5.1 in Chernozhukov et al. (2014b) for the empirical process on class $\mathscr{F}$,

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

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 $\sup_{\mathbf{x}\in\mathscr{B}}\|\widehat{\boldsymbol{\Gamma}}_{t,\mathbf{x}} - \boldsymbol{\Gamma}_{t,\mathbf{x}}\| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}}$. Hence, $1 \lesssim_{\mathbb{P}} \inf_{\mathbf{x}\in\mathscr{B}}\|\widehat{\boldsymbol{\Gamma}}_{t,\mathbf{x}}\| \lesssim_{\mathbb{P}} \sup_{\mathbf{x}\in\mathscr{B}}\|\widehat{\boldsymbol{\Gamma}}_{t,\mathbf{x}}\| \lesssim_{\mathbb{P}} 1$. By Weyl's Theorem, $\sup_{\mathbf{x}\in\mathscr{B}} |\lambda_{\min}(\widehat{\boldsymbol{\Gamma}}_{t,\mathbf{x}}) - \lambda_{\min}(\boldsymbol{\Gamma}_{t,\mathbf{x}})| \leq \sup_{\mathbf{x}\in\mathscr{B}}\|\widehat{\boldsymbol{\Gamma}}_{t,\mathbf{x}} - \boldsymbol{\Gamma}_{t,\mathbf{x}}\| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}}$. Therefore, we can lower bound the minimum eigenvalue by

$$\inf_{\mathbf{x}\in\mathscr{B}} \lambda_{\min}(\widehat{\boldsymbol{\Gamma}}_{t,\mathbf{x}}) \geq \inf_{\mathbf{x}\in\mathscr{B}} \lambda_{\min}(\boldsymbol{\Gamma}_{t,\mathbf{x}}) - \sup_{\mathbf{x}\in\mathscr{B}} |\lambda_{\min}(\widehat{\boldsymbol{\Gamma}}_{t,\mathbf{x}}) - \lambda_{\min}(\boldsymbol{\Gamma}_{t,\mathbf{x}})| \gtrsim_{\mathbb{P}} 1.$$

It follows that $\sup_{\mathbf{x}\in\mathscr{B}}\|\widehat{\boldsymbol{\Gamma}}_{t,\mathbf{x}}^{-1}\| \lesssim_{\mathbb{P}} 1$ and hence

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

This completes the proof. ∎

SA-5.2 Proof of Lemma SA-2.2

We introduce the following notation for an approximation error and an empirical average:

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

Since we have assumed 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

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

where $\sup_{\mathbf{x}\in\mathscr{B}}\max_{t\in\{0,1\}}\max_{1\leq i\leq n}\|\boldsymbol{\alpha}_{\mathbf{x},\mathbf{X}_i,t}\| \lesssim 1$. Assume $\frac{\log(1/h)}{nh^d} = o(1)$, the same argument as the proof of Lemma SA-2.1 shows

$$\frac{1}{n}\sum_{i=1}^n \Big\|\mathbf{R}_p\left(\frac{\mathbf{X}_i-\mathbf{x}}{h}\right)K_h(\mathbf{X}_i-\mathbf{x})\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_t)\mathbf{R}_p\left(\frac{\mathbf{X}_i-\mathbf{x}}{h}\right)^\top\Big\|^2 \lesssim_{\mathbb{P}} 1.$$

It then follows from Lemma SA-2.1 that

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

Now assume further that $h=o(1)$. Since $\gamma_{\mathbf{v}}(\xi;\mathbf{x}) = \frac{|\mathbf{v}|}{\mathbf{v}!}\int_0^1(1-t)^{|\mathbf{v}|-1}\partial_{\mathbf{v}}\mu_t(\mathbf{x}+t(\xi-\mathbf{x}))dt$, then for all $\mathbf{x}\in\mathscr{B}$, $\xi\in\mathscr{X}$,

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

By Assumption SA–1(iii), $\partial^{\mathbf{v}}\mu_t$ is uniformly continuous on the compact set $\mathscr{X}$. This implies that when $h=o(1)$, $M_n=o(1)$. Denote

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

then

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

where in the last equality, we have used the same as in the proof of Lemma SA-2.1 maximal inequality to bound the deviation of the term on the left hand side from its expectation. 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-2.1 and maximal inequality as in the proof of Lemma SA-2.1, we can show

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

Since $\max_{t\in\{0,1\}} \sup_{\mathbf{x}\in\mathscr{B}} |B^{(\boldsymbol{\nu})}_{t,\mathbf{x}}| \lesssim 1$, the inequality above implies

$$\max_{t\in\{0,1\}} \sup_{\mathbf{x}\in\mathscr{B}} |\widehat{B}^{(\boldsymbol{\nu})}_{t,\mathbf{x}}| \lesssim_{\mathbb{P}} 1.$$

This completes the proof. ■

SA-5.3 Proof of Lemma SA-2.3

The proof will be similar to the proof of Lemma SA-2.1. Let $\mathbf{v}$ be a multi-index such that $0 \le |\mathbf{v}| \le \mathfrak{p}$. Denote

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

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

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

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

Diameter of $\mathscr{F}$ in L_2: By (v) in Assumption SA–1, recall we denote $u_i = Y_i - \mathbb{E}[Y_i|\mathbf{X}_i]$,

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

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

Covering Numbers: *Case 1: K is Lipschitz.* Let $\mathbb{Q}$ be a finite distribution on $(\mathscr{X}\times\mathbb{R},\mathscr{B}(\mathscr{X})\otimes\mathscr{B}(\mathbb{R}))$. Let $\mathbf{x},\mathbf{x}'\in\mathscr{X}$. In the proof of Lemma SA-2.1, we have shown $\sup_{\xi\in\mathscr{X}}\sup_{\mathbf{x},\mathbf{x}'\in\mathscr{X}}\frac{|g_n(\xi,\mathbf{x})-g_n(\xi,\mathbf{x}')|}{\|\mathbf{x}-\mathbf{x}'\|_\infty} \lesssim h^{-d-1}$. Hence

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

It follows that $\sup_{\mathbb{Q}} N(\mathscr{F},\|\cdot\|_{\mathbb{Q},2},\epsilon\|F\|_{\mathbb{Q},2}) \lesssim (\frac{\mathrm{diam}(\mathscr{X})}{\epsilon h}))^d$, where sup is over all finite probability distribution on $(\mathscr{X}\times\mathbb{R},\mathscr{B}(\mathscr{X})\otimes\mathscr{B}(\mathbb{R}))$. Denote $A_1 = \frac{\mathrm{diam}(\mathscr{X})}{h}$, $A_2 = d$. We have

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

Case 2: K is the uniform kernel. Consider

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

and $\mathscr{H} = \{(\xi,u)\in\mathscr{X}\times\mathbb{R}\to h_n(\xi,\mathbf{x})u : \mathbf{x}\in\mathscr{B}\}$. By similar arguments as *Case 1* and the proof of Lemma SA-2.1, we can show

$$\sup_Q N(\mathscr{H},\|\cdot\|_{Q,2},\varepsilon\|H\|_{\mathbb{Q},2}) \lesssim \left(\frac{\mathrm{diam}(\mathscr{X})}{\varepsilon h}\right)^d,$$

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-2.1 shows

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

where the supremum is taken over all finite discrete measures. 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),

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

Since $\mathbf{Q}_{t,\mathbf{x}}$ is finite-dimensional, entrywise-convergence implies convergence in norm in the same rate. Hence $\sup_{\mathbf{x}\in\mathscr{X}}\|\mathbf{Q}_{t,\mathbf{x}}\| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{1+v}{2+v}}h^d}$. By Lemma SA-2.1,

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

This completes the proof. ■

SA-5.4 Proof of Lemma SA-2.4

Recall we denote $\varepsilon_i = Y_i - \sum_{t\in\{0,1\}}\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)\widehat{\boldsymbol{\beta}}_t(\mathbf{x})^\top\mathbf{R}_p(\mathbf{X}_i - \mathbf{x})$, and $u_i = Y_i - \sum_{t\in\{0,1\}}\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)\mu_t(\mathbf{X}_i)$. Denote $\eta_i = \sum_{t\in\{0,1\}}\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)(\mu_t(\mathbf{X}_i) - \widehat{\boldsymbol{\beta}}_t(\mathbf{x})^\top\mathbf{R}_p(\mathbf{X}_i - \mathbf{x}))$. Then, for all $\mathbf{x},\mathbf{y} \in \mathscr{B}$, the difference betwen estimated and true sandwich matrix can be decomposed by

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

where

$$\mathbf{M}_{1,\mathbf{x},\mathbf{y}} = \mathbb{E}_n \left[\mathbf{R}_p\left(\frac{\mathbf{X}_i-\mathbf{x}}{h}\right)\mathbf{R}_p\left(\frac{\mathbf{X}_i-\mathbf{y}}{h}\right)^\top \frac{1}{h^d}K\left(\frac{\mathbf{X}_i-\mathbf{x}}{h}\right)K\left(\frac{\mathbf{X}_i-\mathbf{y}}{h}\right)\eta_i^2\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_t)\right],$$

$$\mathbf{M}_{2,\mathbf{x},\mathbf{y}} = 2\mathbb{E}_n \left[\mathbf{R}_p\left(\frac{\mathbf{X}_i-\mathbf{x}}{h}\right)\mathbf{R}_p\left(\frac{\mathbf{X}_i-\mathbf{y}}{h}\right)^\top \frac{1}{h^d}K\left(\frac{\mathbf{X}_i-\mathbf{x}}{h}\right)K\left(\frac{\mathbf{X}_i-\mathbf{y}}{h}\right)\eta_i u_i\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_t)\right],$$

$$\mathbf{M}_{3,\mathbf{x},\mathbf{y}} = \mathbb{E}_n \left[\mathbf{R}_p\left(\frac{\mathbf{X}_i-\mathbf{x}}{h}\right)\mathbf{R}_p\left(\frac{\mathbf{X}_i-\mathbf{y}}{h}\right)^\top \frac{1}{h^d}K\left(\frac{\mathbf{X}_i-\mathbf{x}}{h}\right)K\left(\frac{\mathbf{X}_i-\mathbf{y}}{h}\right)(u_i^2-\sigma_t(\mathbf{X}_i)^2)\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_t)\right],$$

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

Let $\mathbf{u},\mathbf{v}$ be multi-indices. Denote $g_n(\mathbf{X}_i;\mathbf{x},\mathbf{y}) = \frac{1}{h^d}(\frac{\mathbf{X}_i-\mathbf{x}}{h})^{\mathbf{u}}(\frac{\mathbf{X}_i-\mathbf{y}}{h})^{\mathbf{v}}K(\frac{\mathbf{X}_i-\mathbf{x}}{h})K(\frac{\mathbf{X}_i-\mathbf{y}}{h})\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_t)$. For notational simplicity, denote in what follows

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

First, we present a bound on $\max_{1\le i\le n}|\eta_i|\mathbb{1}((\mathbf{X}_i-\mathbf{x})/h\in\mathrm{Supp}(K))$. By Lemma SA-2.2 and Lemma SA-2.3, and multi-index $\boldsymbol{\nu}$ such that $|\boldsymbol{\nu}|\le p$,

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

Since K is compactly supported, we have

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

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

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

It follows that

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

Term $\mathbf{M}_{1,\mathbf{x},\mathbf{y}}$. From the proof for Lemma SA-2.1, $\sup_{\mathbf{x},\mathbf{y}\in\mathcal{X}}|\mathbb{E}_n[g_n(\mathbf{X}_i;\mathbf{x},\mathbf{y})]-\mathbb{E}[g_n(\mathbf{X}_i;\mathbf{x},\mathbf{y})]| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}}$. Moreover, $\sup_{\mathbf{x},\mathbf{y}\in\mathcal{X}}|\mathbb{E}[g_n(\mathbf{X}_i;\mathbf{x},\mathbf{y})]| \lesssim_{\mathbb{P}} 1$. Hence $\sup_{\mathbf{x},\mathbf{y}\in\mathcal{X}}|\mathbb{E}_n[g_n(\mathbf{X}_i;\mathbf{x},\mathbf{y})]| \lesssim_{\mathbb{P}} 1$. So

$$\begin{aligned}\sup_{\mathbf{x},\mathbf{y}\in\mathscr{B}}\left|\mathbb{E}_n\left[g_n(\mathbf{X}_i;\mathbf{x},\mathbf{y})\eta_i^2\right]\right| &\le \max_{1\le i\le n}|\eta_i|\mathbb{1}((\mathbf{X}_i-\mathbf{x})/h\in\mathrm{Supp}(K))\cdot\sup_{\mathbf{x},\mathbf{y}\in\mathcal{X}}|\mathbb{E}_n\left[g_n(\mathbf{X}_i;\mathbf{x},\mathbf{y})\right]|\\ &\lesssim_{\mathbb{P}} \left(h^{p+1}+\alpha_n\right)^2,\end{aligned}$$

where we have use Theorem SA-2.5, which does not depend on this lemma, for $\sup_{\mathbf{x}\in\mathscr{B}}|\widehat{\mu}_t(\mathbf{x})-\mu_t(\mathbf{x})| \lesssim_{\mathbb{P}}$

$h^{p+1}+\alpha_n$. Finite dimensionality of $\mathbf{M}_{1,\mathbf{x},\mathbf{y}}$ then implies

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

Term $\mathbf{M}_{2,\mathbf{x},\mathbf{y}}$. From the proof of Lemma SA-2.3, $\sup_{\mathbf{x},\mathbf{y}\in\mathscr{X}}|\mathbb{E}_n[g_n(\mathbf{X}_i;\mathbf{x},\mathbf{y})u_i]-\mathbb{E}[g_n(\mathbf{X}_i;\mathbf{x},\mathbf{y})u_i]| \lesssim_{\mathbb{P}} \alpha_n$. Moreover, $\sup_{\mathbf{x},\mathbf{y}\in\mathscr{X}}|\mathbb{E}[g_n(\mathbf{X}_i;\mathbf{x},\mathbf{y})u_i]| \lesssim_{\mathbb{P}} 1$. Hence $\sup_{\mathbf{x},\mathbf{y}\in\mathscr{X}}|\mathbb{E}_n[g_n(\mathbf{X}_i;\mathbf{x},\mathbf{y})u_i]| \lesssim_{\mathbb{P}} 1$.

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

implying

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

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

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

and consider $\mathscr{L}=\{l_n(\cdot,\cdot;\mathbf{x},\mathbf{y}):\mathbf{x},\mathbf{y}\in\mathscr{X}\}$. Define $L:\mathscr{X}\times\mathbb{R}\to\mathbb{R}$ by $L(\xi,\varepsilon)=\frac{c}{h^d}|\varepsilon^2-\sigma_t(\xi)^2|$ where $c=\sup_{\mathbf{x},\mathbf{y}\in\mathscr{B}}\left|\left(\frac{\xi-\mathbf{x}}{h}\right)^{\mathbf{u}}\left(\frac{\xi-\mathbf{y}}{h}\right)^{\mathbf{v}}K\left(\frac{\xi-\mathbf{x}}{h}\right)K\left(\frac{\xi-\mathbf{y}}{h}\right)\right|$. By similar argument as in the proof for Lemma SA-2.3, we can show $\mathscr{L}$ is a VC-type class such that $\mathbb{E}[l_n(\mathbf{X}_i,u_i;\mathbf{x},\mathbf{y})]=0, \forall\mathbf{x},\mathbf{y}\in\mathscr{X}$ and

$$\begin{aligned}
\sup_{\mathbf{x},\mathbf{y}\in\mathscr{X}}\mathbb{E}\left[l_n(\mathbf{X}_i,\varepsilon;\mathbf{x},\mathbf{y})^2\right]^{\frac{1}{2}} &\lesssim \sup_{\mathbf{x},\mathbf{y}\in\mathscr{B}}\mathbb{E}\left[g_n(\mathbf{X}_i,u_i;\mathbf{x},\mathbf{y})^2\right]^{\frac{1}{2}}\sup_{\xi\in\mathscr{X}}\mathbb{V}[u_i^2|\mathbf{X}_i=\xi] \lesssim h^{-d/2},\\
\mathbb{E}\left[\max_{1\leq i\leq n}L(\mathbf{X}_i,u_i)^2\right]^{\frac{1}{2}} &\lesssim h^{-d}\mathbb{E}\left[\max_{1\leq i\leq n}u_i^4\right]^{1/2} \lesssim h^{-d}\mathbb{E}\left[\max_{1\leq i\leq n}u_i^{2+v}\right]^{\frac{2}{2+v}} \lesssim h^{-d}n^{\frac{2}{2+v}}.
\end{aligned}$$

Apply Corollary 5.1 in Chernozhukov et al. (2014b), we get

$$\sup_{\mathbf{x},\mathbf{y}\in\mathscr{B}}|\mathbb{E}_n[l_n(\mathbf{X}_i,u_i;\mathbf{x},\mathbf{y})]| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}}+\frac{\log(1/h)}{n^{\frac{v}{2+v}}h^d}, \qquad \sup_{\mathbf{x},\mathbf{y}\in\mathscr{B}}\|\mathbf{M}_{3,\mathbf{x},\mathbf{y}}\| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}}+\frac{\log(1/h)}{n^{\frac{v}{2+v}}h^d}.$$

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 suitable C and

$$\begin{aligned}
&\sup_{\mathbf{x},\mathbf{y}\in\mathscr{B}}\sup_{\xi\in\mathscr{X}}\left|g_n(\xi;\mathbf{x},\mathbf{y})\sigma^2(\xi)\right| \lesssim h^{-d},\\
&\sup_{\mathbf{x},\mathbf{y}\in\mathscr{B}} E\left[g_n(\mathbf{X}_i;\mathbf{x},\mathbf{y})^2\sigma_t(\mathbf{X}_i)^2\right]^{\frac{1}{2}} \lesssim h^{-d/2}.
\end{aligned}$$

Then, similar to the proof of $\mathbf{M}_{1,\mathbf{x},\mathbf{y}}$ we can get

$$\sup_{\mathbf{x},\mathbf{y}\in\mathscr{B}}|\mathbb{E}_n[g_n(\mathbf{X}_i;\mathbf{x},\mathbf{y})]-\mathbb{E}[g_n(\mathbf{X}_i;\mathbf{x},\mathbf{y})]| \lesssim \sqrt{\frac{\log(1/h)}{nh^d}}, \qquad \sup_{\mathbf{x},\mathbf{y}\in\mathscr{B}}\|\mathbf{M}_{4,\mathbf{x},\mathbf{y}}\| \lesssim \sqrt{\frac{\log(1/h)}{nh^d}}.$$

Putting Together. Combining the the upper bounds of the four terms, we get

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

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

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

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

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

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

This completes the proof. ■

SA-5.5 Proof of Theorem SA-2.1

We can use the same argument as in the proof for Lemma SA-2.1, SA-2.2 and SA-2.3, with $\mathscr{B}=\{\mathbf{x}\}$, to get that under the conditions specified we have

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

and

$$|\widehat{\mu}^{(\boldsymbol{\nu})}_t(\mathbf{x})-\mathbb{E}\big[\widehat{\mu}^{(\boldsymbol{\nu})}_t(\mathbf{x})\big|\mathbf{X}\big]|\lesssim_{\mathbb{P}} h^{-|\boldsymbol{\nu}|}\left(\frac{1}{\sqrt{nh^d}}+\frac{1}{n^{\frac{1+v}{2+v}}h^d}\right).$$

In particular, when applying concentration inequalities, we always apply the singleton class of functions that correspond to the point of evaluation $\mathbf{x}$. Putting together, we get the claimed result. ■

For the proof of Theorem SA-2.2, we define the following matrices: For $\mathbf{x}, \mathbf{y} \in \mathscr{B}$,

$$\overline{\mathbf{\Sigma}}_{t,\mathbf{x},\mathbf{y}} = \mathbb{E}_n\left[\mathbf{R}_p\left(\frac{\mathbf{X}_i - \mathbf{x}}{h}\right)\mathbf{R}_p\left(\frac{\mathbf{X}_i - \mathbf{y}}{h}\right)^\top h^d K_h\left(\mathbf{X}_i - \mathbf{x}_1\right) K_h\left(\mathbf{X}_i - \mathbf{y}\right)\sigma_t^2(\mathbf{X}_i)\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)\right],$$
$$\overline{\Omega}^{(\boldsymbol{\nu})}_{t,\mathbf{x},\mathbf{y}} = \frac{1}{nh^{d+2|\boldsymbol{\nu}|}}\mathbf{e}_{1+\boldsymbol{\nu}}^\top \mathbf{\Gamma}_{t,\mathbf{x}}^{-1}\overline{\mathbf{\Sigma}}_{t,\mathbf{x},\mathbf{y}}\mathbf{\Gamma}_{t,\mathbf{y}}^{-1}\mathbf{e}_{1+\boldsymbol{\nu}}, \qquad \overline{\Omega}^{(\boldsymbol{\nu})}_{\mathbf{x},\mathbf{y}} = \overline{\Omega}^{(\boldsymbol{\nu})}_{0,\mathbf{x},\mathbf{y}} + \overline{\Omega}^{(\boldsymbol{\nu})}_{1,\mathbf{x},\mathbf{y}},$$
$$\overline{V}^{(\boldsymbol{\nu})}_{t,\mathbf{x}} = \mathbf{e}_{1+\boldsymbol{\nu}}^\top \mathbf{\Gamma}_{t,\mathbf{x}}^{-1}\overline{\mathbf{\Sigma}}_{t,\mathbf{x},\mathbf{x}}\mathbf{\Gamma}_{t,\mathbf{y}}^{-1}\mathbf{e}_{1+\boldsymbol{\nu}}, \qquad \overline{V}^{(\boldsymbol{\nu})}_{\mathbf{x}} = \overline{V}^{(\boldsymbol{\nu})}_{0,\mathbf{x}} + \overline{V}^{(\boldsymbol{\nu})}_{1,\mathbf{x}}.$$

The following lemma is used for the convergence of $\overline{\Omega}^{(\boldsymbol{\nu})}_{t,\mathbf{x},\mathbf{y}}$.

Lemma SA-5.1 (Conditional Variance)
Suppose Assumption SA–1 (i), (ii), (iv) and Assumption SA–2 hold. If $\frac{\log(1/h)}{nh^d} = o(1)$, *then*

$$\sup_{\mathbf{x}\in\mathscr{B}}\|\mathbf{\Sigma}_{t,\mathbf{x},\mathbf{x}} - \overline{\mathbf{\Sigma}}_{t,\mathbf{x},\mathbf{x}}\| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}}, \quad t \in \{0,1\}, \qquad \textit{and} \ \sup_{\mathbf{x}\in\mathscr{B}}\left|V^{(\boldsymbol{\nu})}_{\mathbf{x}} - \overline{V}^{(\boldsymbol{\nu})}_{\mathbf{x}}\right| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}}.$$

Proof of Lemma SA-5.1. The proof will be similar to the proof of Lemma SA-2.1. Let $\mathbf{u}$, $\mathbf{v}$ be multi-indices such that $|\mathbf{u}| \leq p$ and $|\mathbf{v}| \leq p$. Fix $t \in \{0,1\}$. For $\xi \in \mathscr{X}$ and $\mathbf{x} \in \mathscr{B}$, define

$$g_n(\xi, \mathbf{x}) = \left(\frac{\xi - \mathbf{x}}{h}\right)^{\mathbf{u}+\mathbf{v}}\frac{1}{h^d}K^2\left(\frac{\xi - \mathbf{x}}{h}\right)\sigma_t^2(\xi)\mathbb{1}(\xi \in \mathscr{A}_t).$$

Consider the class of functions $\mathscr{F} = \{g_n(\cdot, \mathbf{x}) : \mathbf{x} \in \mathscr{B}\}$. Then, by the same maximal inequality argument as in the proof of Lemma SA-2.1,

$$\mathbb{E}\left[\sup_{\mathbf{x}\in\mathscr{B}}\left|\mathbb{E}_n\left[g_n(\mathbf{X}_i, \mathbf{x})\right] - \mathbb{E}[g_n(\mathbf{X}_i, \mathbf{x})]\right|\right] \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}}.$$

Since $\mathbf{\Sigma}_{t,\mathbf{x},\mathbf{x}}$ is finite dimensional, $\sup_{\mathbf{x}\in\mathscr{B}}\|\mathbf{\Sigma}_{t,\mathbf{x},\mathbf{x}} - \overline{\mathbf{\Sigma}}_{t,\mathbf{x},\mathbf{x}}\| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}}$ and hence $\sup_{\mathbf{x}\in\mathscr{B}}\|V_{\mathbf{x}} - \overline{V}_{\mathbf{x}}\| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}}$. ■

SA-5.6 Proof of Theorem SA-2.2

For conditional bias, by Lemma SA-2.2,

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

Since we know $\sup_{\mathbf{x}\in\mathscr{B}}|\widehat{B}^{(\boldsymbol{\nu})}_{t,\mathbf{x}} - B^{(\boldsymbol{\nu})}_{t,\mathbf{x}}| \lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}}$ from Lemma SA-2.2,

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

For conditional variance, by Lemma SA-5.1,

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

Since $(nh^{d+2|\boldsymbol{\nu}|})^{-1}\sup_{\mathbf{x}\in\mathscr{B}}|V_{\mathbf{x}}^{(\boldsymbol{\nu})}-\widehat{V}_{\mathbf{x}}^{(\boldsymbol{\nu})}|=\sup_{\mathbf{x}\in\mathscr{B}}|\Omega_{\mathbf{x},\mathbf{x}}^{(\boldsymbol{\nu})}-\widehat{\Omega}_{\mathbf{x},\mathbf{x}}^{(\boldsymbol{\nu})}|=o_{\mathbb{P}}((nh^{d+2|\boldsymbol{\nu}|})^{-1})$ from Lemma SA-2.4,

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

Putting together we get the two MSE results. And

$$\begin{aligned}
&\left|\mathrm{IMSE}_{\boldsymbol{\nu}}-\int_{\mathscr{B}}\left[(h^{p+1-|\boldsymbol{\nu}|}B_{\mathbf{x}}^{(\boldsymbol{\nu})})^2+(nh^{d+2|\boldsymbol{\nu}|})^{-1}V_{\mathbf{x}}^{(\boldsymbol{\nu})}\right]\omega(\mathbf{x})dH^{d-1}(\mathbf{x})\right|\\
&\leq\int_{\mathscr{B}}\omega(\mathbf{x})dH^{d-1}(\mathbf{x})\cdot\sup_{\mathbf{x}\in\mathscr{B}}\left|\mathrm{MSE}_{\boldsymbol{\nu}}(\mathbf{x})-(h^{p+1-|\boldsymbol{\nu}|}B_{\mathbf{x}}^{(\boldsymbol{\nu})})^2-(nh^{d+2|\boldsymbol{\nu}|})^{-1}V_{\mathbf{x}}^{(\boldsymbol{\nu})}\right|\\
&=o_{\mathbb{P}}\left(h^{2p+2-2|\boldsymbol{\nu}|}+(nh^{d+2|\boldsymbol{\nu}|})^{-1}\right).
\end{aligned}$$

Similarly, we can get

$$\left|\mathrm{IMSE}_{\boldsymbol{\nu}}-\int_{\mathscr{B}}\left[(h^{p+1-|\boldsymbol{\nu}|}\widehat{B}_{\mathbf{x}}^{(\boldsymbol{\nu})})^2+(nh^{d+2|\boldsymbol{\nu}|})^{-1}\widehat{V}_{\mathbf{x}}^{(\boldsymbol{\nu})}\right]\omega(\mathbf{x})dH^{d-1}(\mathbf{x})\right|=o_{\mathbb{P}}\left(h^{2p+2-2|\boldsymbol{\nu}|}+(nh^{d+2|\boldsymbol{\nu}|})^{-1}\right).$$

■

SA-5.7 Proof of Theorem SA-2.3

Consider $\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})=(\Omega_{\mathbf{x},\mathbf{x}}^{(\boldsymbol{\nu})})^{-1/2}\mathbf{e}_{1+\boldsymbol{\nu}}^{\top}\mathbf{H}^{-1}\boldsymbol{\Gamma}_{t,\mathbf{x}}^{-1}\mathbf{Q}_{t,\mathbf{x}}$ and $u_i=Y_i-\sum_{t\in\{0,1\}}\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_t)\mu_t(\mathbf{X}_i)$. Define

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

Then, $\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})=\sum_{i=1}^n Z_i$ and $\mathbb{E}\left[Z_i\right]=0$ and $\mathbb{V}[Z_i]=n^{-1}$. By Berry-Essen Theorem,

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

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

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

where in the second line we used Assumption SA–1(v), in the third line we used

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

in the fourth line we used the definition of $\Omega_{\mathbf{x},\mathbf{x}}^{(\boldsymbol{\nu})}$. Hence the Berry-Esseen inequality gives

$$\mathtt{s}_n=\sup_{u\in\mathbb{R}}\left|\mathbb{P}\left(\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})\leq u\right)-\Phi(u)\right|=o(1). \tag{SA-5.1}$$

Although Lemma SA-2.1 to Lemma SA-2.4 provides convergence results uniformly in $\mathbf{x}$, for pointwise result with fix $\mathbf{x}\in\mathscr{B}$, we can replace the class of functions in the proof by one *singleton* corresponding to $\mathbf{x}$, and get: If $h^{p+1}\sqrt{nh^d}\to 0$ and $n^{\frac{v}{2+v}}h^d\to 0$, then

$$\left|\widehat{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})-\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})\right|\lesssim_{\mathbb{P}}\mathtt{r}_n, \tag{SA-5.2}$$

where $\mathtt{r}_n=h^{p+1}\sqrt{nh^d}+1/\sqrt{nh^d}+1/(n^{\frac{v}{2+v}}h^d)$. Take Z to be a standard Gaussian random variable and using anti-concentration arguments, for any $t\in\mathbb{R}$,

$$\begin{aligned}
\mathbb{P}(\widehat{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})\leq t)&=\mathbb{P}(\widehat{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})\leq t,|\widehat{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})-\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})|\leq\mathtt{r}_n)+\mathbb{P}(\widehat{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})\leq t,|\widehat{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})-\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})|\geq\mathtt{r}_n)\\
&\leq\mathbb{P}(\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})\leq t+\mathtt{r}_n)+\mathbb{P}(|\widehat{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})-\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})|\geq\mathtt{r}_n)\\
&\leq\mathbb{P}(Z\leq t+\mathtt{r}_n)+\mathbb{P}(|\widehat{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})-\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})|\geq\mathtt{r}_n)+\mathtt{s}_n\\
&=\Phi(t)+\sup_{t\in\mathbb{R}}|\mathbb{P}(t\leq Z\leq t+\mathtt{r}_n)|+\mathbb{P}(|\widehat{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})-\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})|\geq\mathtt{r}_n)+\mathtt{s}_n,
\end{aligned}$$

where in the third line we used Equation (SA-5.1), in the fourth line we used Equation (SA-5.2) and $\mathbb{P}(t\leq Z\leq t+\mathtt{r}_n)=o(1)$. Similarly, for any $t\in\mathbb{R}$,

$$\begin{aligned}
\mathbb{P}(\widehat{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})\leq t)&=\mathbb{P}(\widehat{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})\leq t,|\widehat{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})-\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})|\leq\mathtt{r}_n)+\mathbb{P}(\widehat{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})\leq t,|\widehat{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})-\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})|\geq\mathtt{r}_n)\\
&\geq\mathbb{P}(\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})\leq t-\mathtt{r}_n)-\mathbb{P}(|\widehat{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})-\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})|\geq\mathtt{r}_n)\\
&\geq\mathbb{P}(Z\leq t-\mathtt{r}_n)-\mathbb{P}(|\widehat{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})-\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})|\geq\mathtt{r}_n)-\mathtt{s}_n\\
&=\Phi(t)-\sup_{t\in\mathbb{R}}|\mathbb{P}(t-\mathtt{r}_n\leq Z\leq t)|-\mathbb{P}(|\widehat{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})-\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})|\geq\mathtt{r}_n)-\mathtt{s}_n.
\end{aligned}$$

It follows that

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}(\widehat{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})\leq t)-\Phi(t)\right|\leq\sup_{t\in\mathbb{R}}|\mathbb{P}(t-\mathtt{r}_n\leq Z\leq t+\mathtt{r}_n)|+\mathbb{P}(|\widehat{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})-\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})|\geq\mathtt{r}_n)+\mathtt{s}_n=o(1).$$

■

SA-5.8 Proof of Theorem SA-2.4

The proof follows directly from Theorem SA-2.3. ■

SA-5.9 Proof of Theorem SA-2.5

The result follows from Lemma SA-2.2 and Lemma SA-2.3. ■

SA-5.10 Proof of Theorem SA-2.6

The feasible t-statistic can be decomposed as follows:

$$\widehat{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x}) = \frac{\widehat{\tau}^{(\boldsymbol{\nu})}(\mathbf{x}) - \tau^{(\boldsymbol{\nu})}(\mathbf{x})}{\sqrt{\widehat{\Omega}^{(\boldsymbol{\nu})}_{\mathbf{x},\mathbf{x}}}} = \overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x}) + G_1^{(\boldsymbol{\nu})}(\mathbf{x}) + G_2^{(\boldsymbol{\nu})}(\mathbf{x}), \qquad \mathbf{x} \in \mathscr{B},$$

where

$$\begin{aligned}
G_1^{(\boldsymbol{\nu})}(\mathbf{x}) &= \Big(\mathbb{E}\big[\widehat{\tau}^{(\boldsymbol{\nu})}(\mathbf{x})\big|\mathbf{X}\big] - \tau^{(\boldsymbol{\nu})}(\mathbf{x})\Big)\,(\widehat{\Omega}^{(\boldsymbol{\nu})}_{\mathbf{x},\mathbf{x}})^{-1/2},\\
G_2^{(\boldsymbol{\nu})}(\mathbf{x}) &= \mathbf{e}_{1+\boldsymbol{\nu}}^\top \mathbf{H}^{-1}\Bigg[\Big(\widehat{\boldsymbol{\Gamma}}_{1,\mathbf{x}}^{-1}\mathbf{Q}_{1,\mathbf{x}} - \widehat{\boldsymbol{\Gamma}}_{0,\mathbf{x}}^{-1}\mathbf{Q}_{0,\mathbf{x}}\Big)\,(\widehat{\Omega}^{(\boldsymbol{\nu})}_{\mathbf{x},\mathbf{x}})^{-\frac{1}{2}} - \big(\boldsymbol{\Gamma}_{1,\mathbf{x}}^{-1}\mathbf{Q}_{1,\mathbf{x}} - \boldsymbol{\Gamma}_{0,\mathbf{x}}^{-1}\mathbf{Q}_{0,\mathbf{x}}\big)\,(\Omega^{(\boldsymbol{\nu})}_{\mathbf{x},\mathbf{x}})^{-\frac{1}{2}}\Bigg].
\end{aligned}$$

By Lemma SA-2.2 and Lemma SA-2.4,

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

By Lemma SA-2.1, Lemma SA-2.3 and Lemma SA-2.4, for $t \in \{0,1\}$ we have

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

By Lemma SA-2.1, Lemma SA-2.3 and Lemma SA-2.4, for $t \in \{0,1\}$ we have

$$\begin{aligned}
&\sup_{\mathbf{x}\in\mathscr{B}}\left|e_{1+\boldsymbol{\nu}}^\top \mathbf{H}^{-1}\boldsymbol{\Gamma}_{t,\mathbf{x}}^{-1}\mathbf{Q}_{t,\mathbf{x}}\left[(\widehat{\Omega}^{(\boldsymbol{\nu})}_{\mathbf{x},\mathbf{x}})^{-1/2} - (\Omega^{(\boldsymbol{\nu})}_{\mathbf{x},\mathbf{x}})^{-1/2}\right]\right|\\
&\lesssim_{\mathbb{P}} h^{-|\boldsymbol{\nu}|}\cdot\Bigg(\sqrt{\frac{\log n}{nh^d}} + \frac{\log n}{n^{\frac{1+v}{2+v}}h^d}\Bigg)\cdot\sqrt{nh^{d+2\boldsymbol{\nu}}}\Bigg(\sqrt{\frac{\log n}{nh^d}} + \frac{\log n}{n^{\frac{v}{2+v}}h^d} + h^{p+1}\Bigg)\\
&\lesssim \frac{\log n}{\sqrt{nh^d}} + \frac{(\log n)^{3/2}}{n^{\frac{v}{2+v}}h^d}.
\end{aligned}$$

Combining the previous two displays, we get

$$\sup_{\mathbf{x}\in\mathscr{B}}|G_2^{(\boldsymbol{\nu})}(\mathbf{x})| \lesssim_{\mathbb{P}} \sqrt{\log n}\Bigg(\sqrt{\frac{\log n}{nh^d}} + \frac{\log n}{n^{\frac{v}{2+v}}h^d}\Bigg).$$

It follows from the decomposition of $\widehat{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x}) - \overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})$ that

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

■

The following lemma is used in the proof of Theorem SA-3.5.

Lemma SA-5.2 (VC Class to VC2 Class)

Assume $\mathscr{F}$ is a VC class on a measure space $(\mathcal{X}, \mathscr{B})$ in the sense that there exists an envelope function F and positive constants $c(\mathscr{F}), d(\mathscr{F})$ such that for all $0 < \varepsilon < 1$,

$$\sup_{\mathbb{Q} \in \mathscr{A}(\mathcal{X})} N(\mathscr{F}, \|\cdot\|_{\mathbb{Q},1}, \varepsilon \|F\|_{\mathbb{Q},1}) \leq c(\mathscr{F}) \varepsilon^{-d(\mathscr{F})}.$$

Then, $\mathscr{F}$ is also VC2 in the sense that for all $0 < \varepsilon < 1$,

$$\sup_{\mathbb{Q} \in \mathscr{A}(\mathcal{X})} N(\mathscr{F}, \|\cdot\|_{\mathbb{Q},2}, \varepsilon \|F\|_{\mathbb{Q},2}) \leq c(\mathscr{F}) (\varepsilon^2/2)^{-d(\mathscr{F})}.$$

Proof of Lemma SA-5.2. Let $\mathbb{Q}$ be a finite discrete probability measure. Let $f, g \in \mathscr{F}$. Then

$$\int |f - g|^2 d\mathbb{Q} \leq 2 \int |f - g||F| d\mathbb{Q}.$$

Suppose $\mathbb{Q}$ is supported on $\{c_1, \ldots, c_p\}$. Define another probability measure $\tilde{\mathbb{Q}}(c_k) = F(c_k)\mathbb{Q}(c_k)/\|F\|_{\mathbb{Q},1}$. Then,

$$\begin{aligned}
\int |f - g|^2 d\mathbb{Q} &\leq 2\|F\|_{\mathbb{Q},1} \int |f - g| d\tilde{\mathbb{Q}} \\
&\leq 2\|F\|_{\mathbb{Q},1} \|f - g\|_{\tilde{\mathbb{Q}},1}.
\end{aligned}$$

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 $\|h - h_1\|_{\tilde{\mathbb{Q}},1} \leq \varepsilon^2/2\|F\|_{\tilde{\mathbb{Q}},1}$, and hence

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

Hence $\sup_{\mathbb{Q} \in \mathscr{A}(\mathcal{X})} N(\mathscr{F}, \|\cdot\|_{\mathbb{Q},2}, \varepsilon\|F\|_{\mathbb{Q},2}) \leq c(\mathscr{F})(\varepsilon^2/2)^{-d(\mathscr{F})}$. ■

SA-5.11 Proof of Theorem SA-2.7

First, we consider the class of functions $\mathscr{F}_t = \{\mathscr{K}_t^{(\boldsymbol{\nu})}(\cdot; \mathbf{x}) : \mathbf{x} \in \mathscr{B}\}$, $t \in \{0, 1\}$. W.l.o.g., we can 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 Lemma SA-4.1). This implies the constants $\mathtt{c}_1$ and $\mathtt{c}_2$ from Lemma SA-4.1 are all 1.

I. Properties of $\mathscr{F}_t$

Envelope Function: By Lemma SA-2.1 and Lemma SA-2.4 and the fact that $\mathrm{Supp}(K)$ is compact,

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

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 of $\mathscr{G}$.

L_1 **Bound:**

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

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

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

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

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

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

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

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

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

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,

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

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

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

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 $\mathbf{u}\in\mathscr{X}$, $\mathbf{x}\in\mathscr{B}$. Consider $\tilde{\mathscr{F}}_t = \{\tilde{\mathscr{K}}^{(\boldsymbol{\nu})}(\cdot;\mathbf{x}) : \mathbf{x}\in\mathscr{B}\}$, $t\in\{0,1\}$. Then, the argument above implies

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

Consider $\mathscr{L} = \{\mathbb{1}((\cdot-\mathbf{x})/h\in[-1,1]^d) : \mathbf{x}\in\mathscr{B}\}$. 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: Suppose K is Lipschitz.* We will use Cattaneo et al. (2024, Lemma 7). To make the notation consistent, define

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

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

(i) boundedness	$\sup_{\mathbf{z}} \sup_{\mathbf{z}'} \lvert f_{\mathbf{z}}(\mathbf{z}')\rvert \leq \mathbf{c}$,
(ii) compact support	$\mathrm{supp}(f_{\mathbf{z}}(\cdot)) \subseteq [-\mathbf{c}, \mathbf{c}]^d, \forall \mathbf{z} \in \mathscr{X}$,
(iii) Lipschitz continuity	$\sup_{\mathbf{z}} \lvert f_{\mathbf{z}}(\mathbf{z}') - f_{\mathbf{z}}(\mathbf{z}'')\rvert \leq \mathbf{c}\lvert\mathbf{z}' - \mathbf{z}''\rvert$
	$\sup_{\mathbf{z}} \lvert f_{\mathbf{z}'}(\mathbf{z}) - f_{\mathbf{z}''}(\mathbf{z})\rvert \leq \mathbf{c}h^{-1}\lvert\mathbf{z}' - \mathbf{z}''\rvert$.

Then, by Cattaneo et al. (2024, Lemma 7), there exists a constant $\mathbf{c}'$ only depending on $\mathbf{c}$ and d that for any $0 \leq \varepsilon \leq 1$,

$$\sup_{Q \in \mathscr{A}(\mathscr{X})} N\left(\mathscr{H}, \|\cdot\|_{Q,1}, (2c+1)^{d+1}\varepsilon\right) \leq \mathbf{c}'\varepsilon^{-d-1} + 1,$$

where $\mathscr{A}(\mathscr{X})$ denotes the collections of all finite discrete measures on $\mathscr{X} = [0,1]^d$. It then follows from Lemma SA-5.2 that with the constant envelope function $\mathtt{M}_{\mathscr{F}_t} = h^{-d/2}$, for any $0 \leq \varepsilon \leq 1$,

$$\sup_{Q \in \mathscr{A}(\mathscr{X})} N\left(\mathscr{F}_t, \|\cdot\|_{Q,2}, (2c+1)^{d+1}\varepsilon \mathtt{M}_{\mathscr{F}_t}\right) \leq \mathbf{c}'\varepsilon^{-d-1} + 1.$$

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

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

where $\tilde{\mathscr{F}}_t = h^{-d/2}$. 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),

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

Putting together, we have

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

where C_1, C_2 are constants only depending on d.

II. Properties of $\mathscr{G}$

Recall for each $\mathbf{x} \in \mathscr{B}$,

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

and $\mathscr{G} = \{g_{\mathbf{x}} : \mathbf{x} \in \mathscr{B}\}$. Hence

$$\mathtt{M}_{\mathscr{G}} \lesssim h^{-d/2}, \qquad \mathtt{E}_{\mathscr{G}} \lesssim h^{d/2}, \qquad \sup_{Q} N(\mathscr{G}, \|\cdot\|_{Q,2}, \varepsilon(2c+1)^{d+1}\mathtt{M}_{\mathscr{G}}) \leq 2\mathbf{c}'\varepsilon^{-d-1} + 2.$$

Total Variation: Observe that $\mathbb{1}_{\mathscr{A}_1}(\mathbf{u})\mathscr{K}_t^{(\boldsymbol{\nu})}(\mathbf{u};\mathbf{x}) \neq 0$ implies $E_{t,\mathbf{x}} = \mathbf{u} \in \{\mathbf{y} \in \mathscr{A}_t : (\mathbf{y}-\mathbf{x})/h \in \mathrm{Supp}(K)\}$, and

$$\mathbb{1}(\mathbf{u} \in \mathscr{A}_t)\mathscr{K}_t^{(\boldsymbol{\nu})}(\mathbf{u};\mathbf{x}) = \mathbb{1}(\mathbf{u} \in E_{t,\mathbf{x}})\mathscr{K}_t^{(\boldsymbol{\nu})}(\mathbf{u};\mathbf{x}), \qquad \forall \mathbf{u} \in \mathscr{X}.$$

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 $\mathtt{TV}_{\{gh\}} \leq \mathtt{M}_{\{g\}}\mathtt{TV}_{\{h\}} + \mathtt{M}_{\{h\}}\mathtt{TV}_{\{g\}}$, we have

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

Then, by Lemma SA-4.1, on a possibly enlarged probability space, there exists a mean-zero Gaussian process $Z^{(\boldsymbol{\nu})}$ with the same covariance structure such that

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

■

To build up the proof for confidence bands, we need the following lemmas.

Lemma SA-5.3 (Distance Between Infeasible Gaussian and Bahadur Representation) *Suppose the conditions of Theorem SA-2.7 hold. Then, for any multi-index $|\boldsymbol{\nu}| \leq p$, we have*

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

Proof of Lemma SA-5.3. Denote $R_n = (\log n)^{\frac{3}{2}}\left(\frac{1}{nh^d}\right)^{\frac{1}{d+2}\cdot\frac{v}{v+2}} + \log(n)\sqrt{\frac{1}{n^{\frac{v}{2+v}}h^d}}$. Let α_n to be determined. For any $u > 0$,

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\mathbf{x}\in\mathscr{B}}\left|\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})\right| \leq u\right) \\
&\leq \mathbb{P}\left(\sup_{\mathbf{x}\in\mathscr{B}}\left|Z^{(\boldsymbol{\nu})}(\mathbf{x})\right| \leq \sup_{\mathbf{x}\in\mathscr{B}}\left|\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x}) - Z^{(\boldsymbol{\nu})}(\mathbf{x})\right| + u\right) \\
&\leq \mathbb{P}\left(\sup_{\mathbf{x}\in\mathscr{B}}\left|Z^{(\boldsymbol{\nu})}(\mathbf{x})\right| \leq u + \alpha_n\right) + \mathbb{P}\left(\sup_{\mathbf{x}\in\mathscr{B}}\left|Z^{(\boldsymbol{\nu})}(\mathbf{x}) - \overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})\right| > \alpha_n\right) \\
&\leq \mathbb{P}\left(\sup_{\mathbf{x}\in\mathscr{B}}\left|Z^{(\boldsymbol{\nu})}(\mathbf{x})\right| \leq u\right) + 4\alpha_n\left(\mathbb{E}\left[\sup_{\mathbf{x}\in\mathscr{B}}\left|Z^{(\boldsymbol{\nu})}(\mathbf{x})\right|\right] + 1\right) + \mathbb{P}\left(\sup_{\mathbf{x}\in\mathscr{B}}\left|Z^{(\boldsymbol{\nu})}(\mathbf{x}) - \overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})\right| > \alpha_n\right) \\
&\leq \mathbb{P}\left(\sup_{\mathbf{x}\in\mathscr{B}}\left|Z^{(\boldsymbol{\nu})}(\mathbf{x})\right| \leq u\right) + 4\alpha_n\left(\mathbb{E}\left[\sup_{\mathbf{x}\in\mathscr{B}}\left|Z^{(\boldsymbol{\nu})}(\mathbf{x})\right|\right] + 1\right) + \frac{CR_n}{\alpha_n},
\end{aligned}$$

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-2.7. Similarly, for any

$u > 0$, we have the lower bound

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\mathbf{x}\in\mathscr{B}}\left|\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})\right| \leq u\right)\\
&\geq \mathbb{P}\left(\sup_{\mathbf{x}\in\mathscr{B}}\left|Z^{(\boldsymbol{\nu})}(\mathbf{x})\right| \leq u - \sup_{\mathbf{x}\in\mathscr{B}}\left|\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x}) - Z^{(\boldsymbol{\nu})}(\mathbf{x})\right|\right)\\
&\geq \mathbb{P}\left(\sup_{\mathbf{x}\in\mathscr{B}}\left|Z^{(\boldsymbol{\nu})}(\mathbf{x})\right| \leq u - \alpha_n\right) - \mathbb{P}\left(\sup_{\mathbf{x}\in\mathscr{B}}\left|Z^{(\boldsymbol{\nu})}(\mathbf{x}) - \overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})\right| > \alpha_n\right)\\
&\geq \mathbb{P}\left(\sup_{\mathbf{x}\in\mathscr{B}}\left|Z^{(\boldsymbol{\nu})}(\mathbf{x})\right| \leq u\right) - 4\alpha_n\left(\mathbb{E}\left[\sup_{\mathbf{x}\in\mathscr{B}}\left|Z^{(\boldsymbol{\nu})}(\mathbf{x})\right|\right] + 1\right) - \mathbb{P}\left(\sup_{\mathbf{x}\in\mathscr{B}}\left|Z^{(\boldsymbol{\nu})}(\mathbf{x}) - \overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})\right| > \alpha_n\right)\\
&\geq \mathbb{P}\left(\sup_{\mathbf{x}\in\mathscr{B}}\left|Z^{(\boldsymbol{\nu})}(\mathbf{x})\right| \leq u\right) - 4\alpha_n\left(\mathbb{E}\left[\sup_{\mathbf{x}\in\mathscr{B}}\left|Z^{(\boldsymbol{\nu})}(\mathbf{x})\right|\right] + 1\right) - \frac{CR_n}{\alpha_n}.
\end{aligned}$$

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

$$\begin{aligned}
d\left(Z^{(\boldsymbol{\nu})}(\mathbf{x}), Z^{(\boldsymbol{\nu})}(\mathbf{y})\right) &= \mathbb{E}\left[\left(Z^{(\boldsymbol{\nu})}(\mathbf{x}) - Z^{(\boldsymbol{\nu})}(\mathbf{y})\right)^2\right]^{\frac{1}{2}} = \mathbb{E}\left[\left(\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x}) - G_0^{(\boldsymbol{\nu})}(\mathbf{y})\right)^2\right]^{\frac{1}{2}}\\
&= \mathbb{E}\left[\left(\mathscr{K}(\mathbf{X}_i,\mathbf{x}) - \mathscr{K}(\mathbf{X}_i,\mathbf{y})\right)^2\sigma(\mathbf{X}_i)^2\right]^{\frac{1}{2}} \leq C' l_{n,2}\|\mathbf{x}-\mathbf{y}\|_\infty,\\
\sup_{\mathbf{x}\in\mathscr{B}} d(Z^{(\boldsymbol{\nu})}(\mathbf{x}), Z^{(\boldsymbol{\nu})}(\mathbf{x})) &= \sup_{\mathbf{x}\in\mathscr{B}} \mathbb{E}\left[\mathscr{K}(\mathbf{X}_i,\mathbf{x})^2\sigma^2(\mathbf{X}_i)\right] \lesssim 1.
\end{aligned}$$

where C' is a constant and $l_{n,2} \asymp h_n^{-1}$. Then, by Corollary 2.2.8 in van der Vaart and Wellner (1996), we have

$$\mathbb{E}\left[\sup_{\mathbf{x}\in\mathscr{B}}\left|Z^{(\boldsymbol{\nu})}(\mathbf{x})\right|\right] \leq \mathbb{E}\left[|Z_n(\mathbf{x}_0)|\right] + \int_0^{2\sup_{\mathbf{x}\in\mathscr{B}} d(Z^{(\boldsymbol{\nu})}(\mathbf{x}), Z^{(\boldsymbol{\nu})}(\mathbf{x}))} \sqrt{d\log\left(\frac{C'' l_{n,2}}{\varepsilon}\right)} \lesssim 1.$$

Hence by choosing $\alpha_n^* \asymp \sqrt{R_n}$, we have

$$\sup_{u\in\mathbb{R}}\left|\mathbb{P}\left(\sup_{\mathbf{x}\in\mathscr{B}}\left|\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})\right| \leq u\right) - \mathbb{P}\left(\sup_{\mathbf{x}\in\mathscr{B}}\left|Z^{(\boldsymbol{\nu})}(\mathbf{x})\right| \leq u\right)\right| \lesssim \sqrt{R_n}.$$

■

Lemma SA-5.4 (Distance Between Bahadur Representation and t-statistics)
Suppose the conditions in Theorem SA-2.7 hold. Then

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

For notational simplicity, define r_n and α_n to be sequences such that

$$\begin{aligned}
r_n &= \left[(\log n)^{\frac{3}{2}}\left(\frac{1}{nh^d}\right)^{\frac{1}{d+2}\cdot\frac{v}{v+2}} + \sqrt{\frac{(\log n)^2}{n^{\frac{v}{v+2}}h^d}}\right]^{1/2},\\
\alpha_n &\ll \sqrt{\log(1/h)}\left(\sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{v}{2+v}}h^d}\right) + h_n^{p+1}\sqrt{nh_n^d}.
\end{aligned}$$

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

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\mathbf{x}\in\mathscr{B}}\left|\widehat{\mathrm{T}}(\mathbf{x})\right|\leq u\right)\\
&\leq \mathbb{P}\left(\sup_{\mathbf{x}\in\mathscr{B}}\left|\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})\right|\leq u+\alpha_n\right)+\mathbb{P}\left(\sup_{\mathbf{x}\in\mathscr{B}}\left|\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})-\widehat{\mathrm{T}}(\mathbf{x})\right|\geq\alpha_n\right)\\
&\leq \mathbb{P}\left(\sup_{\mathbf{x}\in\mathscr{B}}\left|Z^{(\boldsymbol{\nu})}(\mathbf{x})\right|\leq u+\alpha_n\right)+r_n+o(1)\\
&\leq \mathbb{P}\left(\sup_{\mathbf{x}\in\mathscr{B}}\left|Z^{(\boldsymbol{\nu})}(\mathbf{x})\right|\leq u\right)+4\alpha_n\left(\mathbb{E}\left[\sup_{\mathbf{x}\in\mathscr{B}}\left|Z^{(\boldsymbol{\nu})}(\mathbf{x})\right|\right]+1\right)+r_n+o(1)\\
&\leq \mathbb{P}\left(\sup_{\mathbf{x}\in\mathscr{B}}\left|\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})\right|\leq u\right)+4\alpha_n\left(\mathbb{E}\left[\sup_{\mathbf{x}\in\mathscr{B}}\left|Z^{(\boldsymbol{\nu})}(\mathbf{x})\right|\right]+1\right)+2r_n+o(1),
\end{aligned}$$

where in the third line we have used Lemma SA-5.3 and $\sup_{\mathbf{x}\in\mathscr{B}}\left|\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})-\widehat{\mathrm{T}}(\mathbf{x})\right|=o_{\mathbb{P}}(\alpha_n)$, in the fourth line we use the (Chernozhukov et al., 2014a, Theorem 2.1), and in the last line we have used Lemma SA-5.3 again. Similarly,

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\mathbf{x}\in\mathscr{B}}\left|\widehat{\mathrm{T}}(\mathbf{x})\right|\leq u\right)\\
&\geq \mathbb{P}\left(\sup_{\mathbf{x}\in\mathscr{B}}\left|\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})\right|\leq u-\alpha_n\right)-\mathbb{P}\left(\sup_{\mathbf{x}\in\mathscr{B}}\left|\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})-\widehat{\mathrm{T}}(\mathbf{x})\right|\geq\alpha_n\right)\\
&\geq \mathbb{P}\left(\sup_{\mathbf{x}\in\mathscr{B}}\left|Z^{(\boldsymbol{\nu})}(\mathbf{x})\right|\leq u-\alpha_n\right)-r_n+o(1)\\
&\geq \mathbb{P}\left(\sup_{\mathbf{x}\in\mathscr{B}}\left|Z^{(\boldsymbol{\nu})}(\mathbf{x})\right|\leq u\right)-4\alpha_n\left(\mathbb{E}\left[\sup_{\mathbf{x}\in\mathscr{B}}\left|Z^{(\boldsymbol{\nu})}(\mathbf{x})\right|\right]+1\right)-r_n+o(1)\\
&\geq \mathbb{P}\left(\sup_{\mathbf{x}\in\mathscr{B}}\left|\overline{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})\right|\leq u\right)-4\alpha_n\left(\mathbb{E}\left[\sup_{\mathbf{x}\in\mathscr{B}}\left|Z^{(\boldsymbol{\nu})}(\mathbf{x})\right|\right]+1\right)-2r_n+o(1).
\end{aligned}$$

From the proof of Lemma SA-5.3, $\mathbb{E}\left[\sup_{\mathbf{x}\in\mathscr{B}}\left|Z^{(\boldsymbol{\nu})}(\mathbf{x})\right|\right]\lesssim 1$. Hence under the rate restrictions in this lemma,

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

■

Lemma SA-5.5 (Distance Between Feasible Gaussian and Infeasible Gaussian)
Suppose the conditions for Theorem SA-2.7 hold. Then, for any multi-index $|\boldsymbol{\nu}|\leq p$,

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

Proof of Lemma SA-5.5. First, using Lemma SA-2.4, we provide an upper bound between covariance functions of the feasible Gaussian process and the infeasible Gaussian process.

$$\begin{aligned}
&\sup_{\mathbf{x},\mathbf{y}\in\mathscr{X}}\left|\boldsymbol{\Pi}_{\mathbf{x},\mathbf{y}}-\widehat{\boldsymbol{\Pi}}_{\mathbf{x},\mathbf{y}}\right| \\
&=\sup_{\mathbf{x},\mathbf{y}\in\mathscr{X}}\left|\Omega_{\mathbf{x},\mathbf{y}}/\sqrt{\Omega_{\mathbf{x},\mathbf{x}}\Omega_{\mathbf{y}}}-\widehat{\Omega}_{\mathbf{x}_1,\mathbf{x}_2}/\sqrt{\widehat{\Omega}_{\mathbf{x},\mathbf{x}}\widehat{\Omega}_{\mathbf{y}}}\right| \\
&=\sup_{\mathbf{x},\mathbf{y}\in\mathscr{X}}\left|\left(\Omega_{\mathbf{x},\mathbf{y}}-\widehat{\Omega}_{\mathbf{x},\mathbf{y}}\right)/\sqrt{\Omega_{\mathbf{x},\mathbf{x}}\Omega_{\mathbf{y}}}+\frac{\widehat{\Omega}_{\mathbf{x},\mathbf{y}}}{\sqrt{\widehat{\Omega}_{\mathbf{x},\mathbf{x}}\widehat{\Omega}_{\mathbf{y}}}}\left(\sqrt{\frac{\widehat{\Omega}_{\mathbf{x},\mathbf{x}}\widehat{\Omega}_{\mathbf{y}}}{\Omega_{\mathbf{x},\mathbf{x}}\Omega_{\mathbf{y}}}}-1\right)\right|
\end{aligned}$$

From Lemma SA-2.4 and the fact that $|\sqrt{x}-\sqrt{y}|\leq(x\wedge y)^{-1/2}|x-y|/2$ for $x,y>0$,

$$\begin{aligned}
&\sup_{\mathbf{x},\mathbf{y}\in\mathscr{X}}\frac{\left|\left(\widehat{\Omega}_{\mathbf{x},\mathbf{x}}\widehat{\Omega}_{\mathbf{y}}\right)^{1/2}-\left(\Omega_{\mathbf{x},\mathbf{x}}\Omega_{\mathbf{y}}\right)^{1/2}\right|}{\left(\Omega_{\mathbf{x},\mathbf{x}}\Omega_{\mathbf{y}}\right)^{1/2}}\lesssim\frac{\sup_{\mathbf{x},\mathbf{y}\in\mathscr{X}}\left|\widehat{\Omega}_{\mathbf{x},\mathbf{x}}\widehat{\Omega}_{\mathbf{y}}-\Omega_{\mathbf{x},\mathbf{x}}\Omega_{\mathbf{y}}\right|}{\inf_{\mathbf{x},\mathbf{y}}\widehat{\Omega}_{\mathbf{x},\mathbf{x}}\widehat{\Omega}_{\mathbf{y}}\wedge\inf_{\mathbf{x},\mathbf{y}}\Omega_{\mathbf{x},\mathbf{x}}\Omega_{\mathbf{y}}}\lesssim_{\mathbb{P}}h^{p+1}+\sqrt{\frac{\log n}{nh^d}}+\frac{\log n}{n^{\frac{v}{2+v}}h^d}, \\
&\sup_{\mathbf{x},\mathbf{y}\in\mathscr{X}}\frac{\left|\Omega_{\mathbf{x}_1,\mathbf{x}_2}-\widehat{\Omega}_{\mathbf{x}_1,\mathbf{x}_2}\right|}{\sqrt{\Omega_{\mathbf{x},\mathbf{x}}\Omega_{\mathbf{y}}}}\lesssim_{\mathbb{P}}h^{p+1}+\sqrt{\frac{\log n}{nh^d}}+\frac{\log n}{n^{\frac{v}{2+v}}h^d}.
\end{aligned}$$

For simplicity, denote $a_n=\sqrt{\frac{\log n}{nh^d}}+\frac{\log n}{n^{\frac{v}{2+v}}h^d}$. Then, it follows that

$$\sup_{\mathbf{x},\mathbf{y}\in\mathscr{X}}\left|\boldsymbol{\Pi}_{\mathbf{x},\mathbf{y}}-\widehat{\boldsymbol{\Pi}}_{\mathbf{x},\mathbf{y}}\right|\lesssim_{\mathbb{P}}h^{p+1}+a_n.$$

Then, we bound the Kolmogorov-Smirnov distance between the maximum of Z_n and $\widehat{Z}^{(\boldsymbol{\nu})}$ on a δ_n-net of $\mathscr{X}$, denoted by $\mathscr{X}_{\delta_n}$, i.e. 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 $M:=\operatorname{Card}(\mathscr{X}_{\delta_n})\lesssim\delta_n^{-d}$. 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 the Gaussian Comparison Inequality Theorem 2.1 from Chernozhuokov et al. (2022),

$$\sup_{\mathbf{y}\in\mathbb{R}^M}\left|\mathbb{P}\left(\mathbf{Z}_n^{\delta_n}\leq\mathbf{y}\right)-\mathbb{P}\left(\widehat{\mathbf{Z}}_n^{\delta_n}\leq\mathbf{y}\middle|\mathbf{X}\right)\right|\lesssim\log M\left(\sup_{\mathbf{x},\mathbf{y}\in\mathscr{X}}\left|\boldsymbol{\Pi}_{\mathbf{x},\mathbf{y}}-\widehat{\boldsymbol{\Pi}}_{\mathbf{x},\mathbf{y}}\right|\right)^{\frac{1}{2}}\lesssim_{\mathbb{P}}\log M\left(a_n+h_n^{p+1}\right)^{\frac{1}{2}}.$$

Consequently,

$$\begin{aligned}
\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\|\mathbf{Z}_n^{\delta_n}\|_\infty\leq x\right)-\mathbb{P}\left(\|\widehat{\mathbf{Z}}_n^{\delta_n}\|_\infty\leq x\middle|\mathbf{X}\right)\right|&\leq\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(-x\mathbf{1}\leq\mathbf{Z}_n^{\delta_n}\leq x\mathbf{1}\right)-\mathbb{P}\left(-x\mathbf{1}\leq\widehat{\mathbf{Z}}_n^{\delta_n}\leq x\mathbf{1}\middle|\mathbf{X}\right)\right| \\
&\lesssim_{\mathbb{P}}\log M\left(a_n+h_n^{p+1}\right)^{\frac{1}{2}}=R_M.
\end{aligned}$$

Then, we bound the Kolmogorov-Smirnov distance on the whole $\mathscr{X}$ with the help of some $\alpha_n>0$ to be determined. For simplicity, denote

$$\begin{aligned}
\Phi_{\delta_n}(\alpha_n)&=\mathbb{P}\left(\sup_{\|\mathbf{x}-\mathbf{y}\|_\infty\leq\delta_n}\left|Z^{(\boldsymbol{\nu})}(\mathbf{x})-Z^{(\boldsymbol{\nu})}(\mathbf{y})\right|\geq\alpha_n\right), \\
\widehat{\Phi}_{\delta_n}(\alpha_n)&=\mathbb{P}\left(\sup_{\|\mathbf{x}-\mathbf{y}\|_\infty\leq\delta_n}\left|\widehat{Z}^{(\boldsymbol{\nu})}(\mathbf{x})-\widehat{Z}^{(\boldsymbol{\nu})}(\mathbf{y})\right|\geq\alpha_n\middle|\mathbf{X}\right),
\end{aligned}$$

then for all $t > 0$,

$$
\begin{aligned}
&\mathbb{P}\left(\sup_{\mathbf{x}\in\mathscr{B}}\left|Z^{(\boldsymbol{\nu})}(\mathbf{x})\right|\leq t\right)\\
&\leq \mathbb{P}\left(\sup_{\mathbf{x}\in\mathscr{B}_{\delta_n}}\left|Z^{(\boldsymbol{\nu})}(\mathbf{x})\right|\leq t+\alpha_n\right)+\Phi_{\delta_n}\left(\alpha_n\right)\\
&\leq \mathbb{P}\left(\sup_{\mathbf{x}\in\mathscr{B}_{\delta_n}}\left|\widehat{Z}^{(\boldsymbol{\nu})}(\mathbf{x})\right|\leq t+\alpha_n\middle|\mathbf{X}\right)+\Phi_{\delta_n}\left(\alpha_n\right)+R_M\\
&\leq \mathbb{P}\left(\sup_{\mathbf{x}\in\mathscr{B}}\left|\widehat{Z}^{(\boldsymbol{\nu})}(\mathbf{x})\right|\leq t+\alpha_n\middle|\mathbf{X}\right)+\Phi_{\delta_n}\left(\alpha_n\right)+\widehat{\Phi}_{\delta_n}\left(\alpha_n\right)+R_M\\
&\leq \mathbb{P}\left(\sup_{\mathbf{x}\in\mathscr{B}}\left|\widehat{Z}^{(\boldsymbol{\nu})}(\mathbf{x})\right|\leq t\middle|\mathbf{X}\right)+4\alpha_n\left(\mathbb{E}\left[\sup_{\mathbf{x}\in\mathscr{B}}\left|\widehat{Z}^{(\boldsymbol{\nu})}(\mathbf{x})\right|\middle|\mathbf{X}\right]+1\right)+\Phi_{\delta_n}\left(\alpha_n\right)+\widehat{\Phi}_{\delta_n}\left(\alpha_n\right)+R_M.
\end{aligned}
$$

Similary, we get for all $t > 0$,

$$
\begin{aligned}
&\mathbb{P}\left(\sup_{\mathbf{x}\in\mathscr{B}}\left|Z^{(\boldsymbol{\nu})}(\mathbf{x})\right|\leq t\right)\\
&\geq \mathbb{P}\left(\sup_{\mathbf{x}\in\mathscr{B}}\left|\widehat{Z}^{(\boldsymbol{\nu})}(\mathbf{x})\right|\leq t\middle|\mathbf{X}\right)-4\alpha_n\left(\mathbb{E}\left[\sup_{\mathbf{x}\in\mathscr{B}}\left|\widehat{Z}^{(\boldsymbol{\nu})}(\mathbf{x})\right|\middle|\mathbf{X}\right]+1\right)-\Phi_{\delta_n}\left(\alpha_n\right)-\widehat{\Phi}_{\delta_n}\left(\alpha_n\right)-R_M.
\end{aligned}
$$

Heuristically, 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 R_M term will dominates the terms $\Phi_{\delta_n}\left(\alpha_n\right)$ and $\widehat{\Phi}_{\delta_n}\left(\alpha_n\right)$. Precisely, for any δ,

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

where in the last line we have used the scale of covariance matrices from Lemma SA-2.4, the scale of Gram matrices from Lemma SA-2.1, and the almost sure bound on the Lipschitz constant from the proof of Theorem SA-2.7 and $C > 0$ is a constant. Similarly, for any $\delta > 0$,

$$
\begin{aligned}
&\sup_{\|\mathbf{x}-\mathbf{y}\|_\infty\leq\delta}\mathbb{E}\left[\left(Z^{(\boldsymbol{\nu})}(\mathbf{x})-Z^{(\boldsymbol{\nu})}(\mathbf{y})\right)^2\right]\\
&=\sup_{\|\mathbf{x}-\mathbf{y}\|_\infty\leq\delta}\mathbb{E}\left[\left(\mathscr{K}(\mathbf{X}_i,\mathbf{x})-\mathscr{K}(\mathbf{X}_i,\mathbf{y})\right)^2\varepsilon_i^2\right]\leq C'h_n^{-2}\delta^2,
\end{aligned}
$$

Then, by Corollary 2.2.5 from van der Vaart and Wellner (1996),

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

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

Also using the fact that $\mathbb{E}\left[\sup_{\mathbf{x}\in\mathscr{B}}\left|\widehat{Z}^{(\boldsymbol{\nu})}(\mathbf{x})\right|\middle|\mathbf{X}\right] \lesssim 1$, by choosing $\alpha_n^* \asymp \left(\sqrt{\log n} h_n^{-d/2-1}\delta_n\right)^{\frac{1}{2}}$ and $\delta_n \asymp n^{-s}$ for some large constant $s > 0$, we have

$$\begin{aligned}
&4\alpha_n\left(\mathbb{E}\left[\sup_{\mathbf{x}\in\mathscr{B}}\left|\widehat{Z}^{(\boldsymbol{\nu})}(\mathbf{x})\right|\middle|\mathbf{X}\right] + 1\right) + \Phi_{\delta_n}(\alpha_n) + \widehat{\Phi}_{\delta_n}(\alpha_n) + R_M \\
&\lesssim_{\mathbb{P}} \left(\sqrt{\log n} h_n^{-d/2-1}\delta_n\right)^{\frac{1}{2}} + d\log\left(\delta_n^{-1}\right)\left(a_n + h_n^{p+1}\right)^{\frac{1}{2}} \\
&\lesssim_{\mathbb{P}} d\log n\left(a_n + h_n^{p+1}\right)^{\frac{1}{2}}.
\end{aligned}$$

Putting together, we have

$$\sup_{\mathbf{u}\in\mathbb{R}}\left|\mathbb{P}\left(\sup_{\mathbf{x}\in\mathscr{B}}\left|Z^{(\boldsymbol{\nu})}(\mathbf{x})\right| \leq u\right) - \mathbb{P}\left(\sup_{\mathbf{x}\in\mathscr{B}}\left|\widehat{Z}^{(\boldsymbol{\nu})}(\mathbf{x})\right| \leq u\middle|\mathbf{X}\right)\right| \lesssim \log n\left(a_n + h_n^{p+1}\right)^{\frac{1}{2}}.$$

■

SA-5.12 Proof of Theorem SA-2.8

The result follows from Lemma SA-5.3, Lemma SA-5.4 and Lemma SA-5.5. ■

SA-5.13 Proof of Theorem SA-2.9

Lemma SA-2.8 and Dominated Convergence Theorem give

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

Then, by definition of $\widehat{B}_\alpha^{(\boldsymbol{\nu})}(\mathbf{x})$,

$$\begin{aligned}
\mathbb{P}\big[\mu^{(\boldsymbol{\nu})}(\mathbf{x}) \in \widehat{B}_\alpha^{(\boldsymbol{\nu})}(\mathbf{x}), \forall \mathbf{x}\in\mathscr{B}\big] &= \mathbb{P}\left[\sup_{\mathbf{x}\in\mathscr{B}}|\widehat{\mathrm{T}}^{(\boldsymbol{\nu})}(\mathbf{x})| \leq \mathfrak{c}_\alpha\right] \\
&= \mathbb{P}\left[\sup_{\mathbf{x}\in\mathscr{B}}\left|\widehat{Z}^{(\boldsymbol{\nu})}(\mathbf{x})\right| \leq \mathfrak{c}_\alpha\right] + o(1) \\
&= \mathbb{E}\left[\mathbb{P}\left[\sup_{\mathbf{x}\in\mathscr{B}}\left|\widehat{Z}^{(\boldsymbol{\nu})}(\mathbf{x})\right| \leq \mathfrak{c}_\alpha\middle|\mathbf{W}_n\right]\right] + o(1) \\
&= 1 - \alpha + o(1).
\end{aligned}$$

■

SA-6 Proofs for Section SA-3

SA-6.1 Proof of Lemma SA-3.1

By Assumption SA–1(iii) and Assumption SA–3, for any $r \neq 0$, for any $\mathbf{x} \in \mathscr{B}$ and $\mathbf{y} \in S_{t,\mathbf{x}}(r)$,

$$|\mu_t(\mathbf{y}) - \mu_t(\mathbf{x})| \lesssim |r|.$$

Hence for any $r \neq 0$, for any $\mathbf{x} \in \mathscr{B}$, $t \in \{0,1\}$,

$$|\theta_{t,\mathbf{x}}(r) - \mu_t(\mathbf{x})| \leq \frac{\int_{S_{t,\mathbf{x}}(|r|)} |\mu_t(\mathbf{y}) - \mu_t(\mathbf{x})| f_X(\mathbf{y}) H_{d-1}(d\mathbf{y})}{\int_{S_{t,\mathbf{x}}(|r|)} f_X(\mathbf{y}) H_{d-1}(d\mathbf{y})} \lesssim r.$$

implying

$$|\theta_{t,\mathbf{x}}(0) - \mu_t(\mathbf{x})| \leq \lim_{r \to 0} |\theta_{t,\mathbf{x}}(r) - \mu_t(\mathbf{x})| = 0.$$

■

SA-6.2 Proof of Lemma SA-3.2

The proof will be similar to the proof of Lemma SA-2.1. Let $0 \leq v \leq p$. Instead of g_n, we study the function k_n defined by

$$k_n(\xi, \mathbf{x}) = \left(\frac{d(\xi, \mathbf{x})}{h}\right)^v \frac{1}{h^d} K\left(\frac{d(\xi, \mathbf{x})}{h}\right), \xi, \mathbf{x} \in \mathcal{X}.$$

Define $\mathscr{H} = \{k_n(\cdot, \mathbf{x}) \mathbb{1}(\cdot \in \mathscr{A}_t) : \mathbf{x} \in \mathcal{X}\}$. We will show $\mathscr{H}$ is a VC-type of class.

Constant Envelope Function We assume K is continuous and has compact support. Hence there exists a constant C_1 such that $\sup_{\mathbf{x} \in \mathcal{X}} \|k_n(\cdot, \mathbf{x})\|_\infty \leq C_1 h^{-d} = H$.

Diameter in L_2 For each $\mathbf{x} \in \mathcal{X}$, $k_n(\cdot, \mathbf{x})$ is supported in $\{\xi : \mathscr{d}(\xi, \mathbf{x}) \leq h\}$. By Assumption SA–1(ii) and Assumption SA–3(i), $\sup_{\mathbf{x} \in \mathcal{X}} \mathbb{P}\left(\mathscr{d}(\mathbf{X}_i, \mathbf{x}) \leq h\right) \lesssim h^d$. It follows that $\sup_{\mathbf{x} \in \mathcal{X}} \|k_n(\cdot, \mathbf{x})\|_{\mathbb{P},2} \leq C_2 h^{-d/2}$ 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 $\delta = \frac{\sigma}{F} = C_3 \sqrt{h^d}$, for some constant C_3.

Covering Numbers *Case 1: k is Lipschitz.* Let $\mathbf{x}, \mathbf{x}' \in \mathcal{X}$. By Assumption SA–3 and Assumption SA–2,

$$\begin{aligned}
&\sup_{\xi \in \mathcal{X}} |k_n(\xi, \mathbf{x}) - k_n(\xi, \mathbf{x}')| \\
&\leq \sup_{\xi \in \mathcal{X}} \left[\left(\frac{\mathscr{d}(\xi, \mathbf{x})}{h}\right)^v - \left(\frac{\mathscr{d}(\xi, \mathbf{x}')}{h}\right)^v\right] k_h\left(\mathscr{d}(\xi, \mathbf{x})\right) + \left(\frac{\mathscr{d}(\xi, \mathbf{x}')}{h}\right)^v \left[k_h\left(\mathscr{d}(\xi, \mathbf{x})\right) - k_h\left(\mathscr{d}(\xi, \mathbf{x}')\right)\right] \\
&\lesssim h_n^{-d-1} \|\mathbf{x} - \mathbf{x}'\|_\infty,
\end{aligned}$$

By Lipschitz continuity property of $\mathscr{F}$, for any $\varepsilon \in (0,1]$ and for any finitely supported measure Q and metric $\|\cdot\|_{Q,2}$ based on $L_2(Q)$,

$$N(\{k_n(\cdot, \mathbf{x}) : \mathbf{x} \in \mathcal{X}\}, \|\cdot\|_{Q,2}, \varepsilon \|H\|_{\mathbb{Q},2}) \leq N(\mathcal{X}, \|\cdot\|_\infty, \varepsilon \|H\|_{\mathbb{Q},2} h^{d+1}) \overset{(1)}{\lesssim} \left(\frac{\operatorname{diam}(\mathcal{X})}{\varepsilon \|H\|_{\mathbb{Q},2} h^{d+1}}\right)^d \lesssim \left(\frac{\operatorname{diam}(\mathcal{X})}{\varepsilon h}\right)^d,$$

where in (1) we used the fact that $\varepsilon\|H\|_{\mathbb{Q},2}h^{d+1} \lesssim \varepsilon h \lesssim 1$. Hence $\mathscr{H}$ forms a VC-type class in that $\sup_Q N(\mathscr{H}, \|\cdot\|_{Q,2}, \varepsilon\|H\|_{\mathbb{Q},2}) \lesssim (C_1/\epsilon)^{C_2}$ for all $\epsilon \in (0,1]$ with $C_1 = \frac{\operatorname{diam}(\mathcal{X})}{h}$ and $C_2 = d$. Moreover, for any discrete measure Q, and for any $\mathbf{x}, \mathbf{x}' \in \mathcal{X}$, $\|k_n(\cdot,\mathbf{x})\mathbb{1}(\cdot \in A_t) - k_n(\cdot,\mathbf{x}')\mathbb{1}(\cdot \in A_t)\|_{Q,2} \leq \|k_n(\cdot,\mathbf{x}) - k_n(\cdot,\mathbf{x}')\|_{Q,2}$. Hence

$$\sup_{Q \in \mathscr{A}(\mathcal{X})} N(\mathscr{H}, \|\cdot\|_{Q,2}, \varepsilon\|H\|_{\mathbb{Q},2}) \leq N(\{k_n(\cdot,\mathbf{x}) : \mathbf{x} \in \mathcal{X}\}, \|\cdot\|_{Q,2}, \varepsilon\|H\|_{\mathbb{Q},2}) \leq (C_1/\varepsilon)^{C_2}, \qquad \varepsilon \in (0,1],$$

where $\mathscr{A}(\mathcal{X})$ denotes the collection of all finite discrete measures on $\mathcal{X}$.

Case 2: $k = \mathbb{1}(\cdot \in [-1,1])$. The same argument as in the proof of Lemma SA-4.1 and the fact that $\mathscr{L} = \{\mathbb{1}((\cdot - \mathbf{x})/h \in [-1,1]^d) : \mathbf{x} \in \mathscr{B}\}$ has VC dimension no greater than $2d$ implies again we have,

$$\sup_{Q \in \mathscr{A}(\mathcal{X})} N(\mathscr{H}, \|\cdot\|_{Q,2}, \varepsilon\|H\|_{\mathbb{Q},2}) \leq N(\{k_n(\cdot,\mathbf{x}) : \mathbf{x} \in \mathcal{X}\}, \|\cdot\|_{Q,2}, \varepsilon\|H\|_{\mathbb{Q},2}) \leq (C_1/\varepsilon)^{C_2}, \varepsilon \in (0,1].$$

Hence, by Chernozhukov et al. (2014b, Corollary 5.1)

$$\begin{aligned}
\mathbb{E}\left[\sup_{l \in \mathscr{H}} |\mathbb{E}_n[l(\mathbf{X}_i)] - \mathbb{E}[l(\mathbf{X}_i)]|\right] &\lesssim \frac{\sigma}{\sqrt{n}}\sqrt{C_2 \log(C_1/\delta)} + \frac{\|M\|_{\mathbb{P},2} C_2 \log(C_1/\delta)}{n} \\
&\lesssim \frac{1}{\sqrt{nh^d}}\sqrt{d \log\left(\frac{\operatorname{diam}(\mathcal{X})}{h^{1+d/2}}\right)} + \frac{1}{nh^d} d \log\left(\frac{\operatorname{diam}(\mathcal{X})}{h^{1+d/2}}\right) \lesssim \sqrt{\frac{\log n}{nh^d}}.
\end{aligned}$$

We conclude that $\sup_{\mathbf{x} \in \mathcal{X}} \|\widehat{\boldsymbol{\Psi}}_{t,\mathbf{x}} - \boldsymbol{\Psi}_{t,\mathbf{x}}\| \lesssim_{\mathbb{P}} \sqrt{\frac{\log n}{nh^d}}$. By Weyl's Theorem, $\sup_{\mathbf{x} \in \mathcal{X}} |\lambda_{\min}(\widehat{\boldsymbol{\Psi}}_{t,\mathbf{x}}) - \lambda_{\min}(\boldsymbol{\Psi}_{t,\mathbf{x}})| \leq \sup_{\mathbf{x} \in \mathcal{X}} \|\widehat{\boldsymbol{\Psi}}_{t,\mathbf{x}} - \boldsymbol{\Psi}_{t,\mathbf{x}}\| \lesssim_{\mathbb{P}} \sqrt{\frac{\log n}{nh^d}}$. Therefore we can lower bound the minimum eigenvalue by $\inf_{\mathbf{x} \in \mathcal{X}} \lambda_{\min}(\widehat{\boldsymbol{\Psi}}_{t,\mathbf{x}}) \geq \inf_{\mathbf{x} \in \mathcal{X}} \lambda_{\min}(\boldsymbol{\Psi}_{t,\mathbf{x}}) - \sup_{\mathbf{x} \in \mathcal{X}} |\lambda_{\min}(\widehat{\boldsymbol{\Psi}}_{t,\mathbf{x}}) - \lambda_{\min}(\boldsymbol{\Psi}_{t,\mathbf{x}})| \gtrsim_{\mathbb{P}} 1$. It follows that $\sup_{\mathbf{x} \in \mathcal{X}} \|\widehat{\boldsymbol{\Psi}}_{t,\mathbf{x}}^{-1}\| \lesssim_{\mathbb{P}} 1$ and hence

$$\sup_{\mathbf{x} \in \mathcal{X}} \|\widehat{\boldsymbol{\Psi}}_{t,\mathbf{x}}^{-1} - \boldsymbol{\Psi}_{t,\mathbf{x}}^{-1}\| \leq \sup_{\mathbf{x} \in \mathcal{X}} \|\boldsymbol{\Psi}_{t,\mathbf{x}}^{-1}\| \|\boldsymbol{\Psi}_{t,\mathbf{x}} - \widehat{\boldsymbol{\Psi}}_{t,\mathbf{x}}\| \|\widehat{\boldsymbol{\Psi}}_{t,\mathbf{x}}^{-1}\| \lesssim_{\mathbb{P}} \sqrt{\frac{\log n}{nh^d}}.$$

■

SA-6.3 Proof of Lemma SA-3.3

Consider the class $\mathscr{F} = \{(\mathbf{z}, u) \mapsto \mathbf{e}_\nu^\top g_{\mathbf{x}}(\mathbf{z})(u - h_{\mathbf{x}}(\mathbf{z})) : \mathbf{x} \in \mathscr{B}\}$, $0 \leq \boldsymbol{\nu} \leq p$, where for $\mathbf{z} \in \mathcal{X}$,

$$g_{\mathbf{x}}(\mathbf{z}) = \mathbf{r}_p\left(\frac{\mathscr{d}(\mathbf{z},\mathbf{x})}{h}\right) k_h(\mathscr{d}(\mathbf{z},\mathbf{x})), \qquad h_{\mathbf{x}}(\mathbf{z}) = \boldsymbol{\gamma}_t^*(\mathbf{x})^\top \mathbf{r}_p\left(\mathscr{d}(\mathbf{z},\mathbf{x})\right).$$

By definition of $\boldsymbol{\gamma}_t^*(\mathbf{x})$,

$$\boldsymbol{\gamma}_t^*(\mathbf{x}) = \mathbf{H}^{-1}\boldsymbol{\Psi}_{t,\mathbf{x}}^{-1}\mathbf{S}_{t,\mathbf{x}}, \qquad \mathbf{S}_{t,\mathbf{x}} = \mathbb{E}\left[\mathbf{r}_p\left(\frac{D_i(\mathbf{x})}{h}\right) k_h(D_i(\mathbf{x})) Y_i \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)\right]. \tag{SA-6.1}$$

Assumption SA–1 implies $\mathbf{S}_{t,\mathbf{x}}$ is continuous in $\mathbf{x}$, hence $\sup_{\mathbf{x} \in \mathcal{X}} \|\mathbf{S}_{t,\mathbf{x}}\| \lesssim 1$. And by Assumption SA–2(ii), $\inf_{\mathbf{x} \in \mathcal{X}} \lambda_{\min}(\boldsymbol{\Psi}_{t,\mathbf{x}}) \gtrsim 1$. Hence

$$\sup_{\mathbf{x} \in \mathscr{B}} \|\boldsymbol{\Psi}_{t,\mathbf{x}}^{-1}\mathbf{S}_{t,\mathbf{x}}\| \lesssim 1. \tag{SA-6.2}$$

Now, consider properties of $\mathscr{F}$. Definition of $\boldsymbol{\gamma}_t^*(\mathbf{x})$ implies $\mathbb{E}[f(\mathbf{X}_i, Y_i)] = 0$ for all $f \in \mathscr{F}$. Since K is compactly supported, there exists $C_1, C_2 > 0$ such that $F(\mathbf{z}, u) = C_1 h^{-d}(|u| + C_2)$ is an envelope function for $\mathscr{F}$. Denote $M = \max_{1 \leq i \leq n} F(\mathbf{X}_i, Y_i)$, then

$$\begin{aligned}\mathbb{E}[M^2]^{1/2} &\lesssim h^{-d}\mathbb{E}\left[\max_{1\leq i\leq n}|Y_i|^2 + 1\right]^{1/2} \lesssim h^{-d}\mathbb{E}\left[\max_{1\leq i\leq n}|Y_i|^{2+v}\right]^{1/(2+v)} \\ &\lesssim h^{-d}\left[\sum_{i=1}^n \mathbb{E}[|\varepsilon_i + \sum_{t\in\{0,1\}} \mathbb{1}(\mathbf{x} \in \mathscr{A}_t)\mu_t(\mathbf{x})|^{2+v}]\right]^{1/(2+v)} \lesssim h^{-d} n^{1/(2+v)},\end{aligned}$$

where we have used $\mathbf{X}$ is compact and μ_t is continuous, hence $\sup_{\mathbf{x}\in\mathscr{X}} |\sum_{t\in\{0,1\}} \mathbb{1}(\mathbf{x} \in \mathscr{A}_t)\mu_t(\mathbf{x})| \lesssim 1$. Denote $\sigma = \sup_{f\in\mathscr{F}} \mathbb{E}[f(\mathbf{X}_i, Y_i)^2]^{1/2}$. Then,

$$\sigma^2 \lesssim \sup_{\mathbf{x}\in\mathscr{B}} \mathbb{E}[\|\mathbf{e}_\nu^\top g_\mathbf{x}\|_\infty^2 (|Y_i| + \|\mathbf{e}_\nu^\top h_\mathbf{x}\|_\infty)^2 \mathbb{1}(k_h(D_i(\mathbf{x})) \neq 0)] \lesssim h^{-d}.$$

To check for the covering number of $\mathscr{F}$, notice that compare to the proof of Lemma SA-2.1, we have one more term $\mathbf{e}_{\boldsymbol{\nu}}^\top g_\mathbf{x} h_\mathbf{x} = \mathbf{r}_p\left(\frac{\mathscr{d}(\mathbf{z},\mathbf{x})}{h}\right) k_h(\mathscr{d}(\mathbf{z},\mathbf{x}))\boldsymbol{\gamma}_t^*(\mathbf{x})^\top \mathbf{r}_p(\mathscr{d}(\mathbf{z},\mathbf{x}))$. All terms except for $\boldsymbol{\gamma}_t^*(\mathbf{x})$ can be handled as in the proof of Lemma SA-2.1. Recall Equation (SA-6.1), and consider $l_{t,\mathbf{x}} = \mathbf{e}_\mathbf{v}^\top[\mathbf{R}(\mathscr{d}(\cdot,\mathbf{x})/h)k_h(\mathscr{d}(\cdot,\mathbf{x}))\mu_t \mathbb{1}(\cdot \in \mathscr{A}_t)$ and $\mathscr{L}_t = \{l_{t,\mathbf{x}} : \mathbf{x} \in \mathscr{B}\}$, $\mathbf{v}$ is a any multi-index. Then, for any $\mathbf{x}_1, \mathbf{x}_2 \in \mathscr{B}$,

$$|\mathbf{S}_{t,\mathbf{x}_1} - \mathbf{S}_{t,\mathbf{x}_2}| \leq \|l_{t,\mathbf{x}_1} - l_{t,\mathbf{x}_2}\|_{\mathbb{P}_X,2},$$

and hence

$$N(\{\mathbf{e}_\mathbf{v}^\top \mathbf{S}_{t,\mathbf{x}} : \mathbf{x} \in \mathscr{B}\}, |\cdot|, \varepsilon h^{-d}) \leq N(\mathscr{L}_t, \|\cdot\|_{\mathbb{P}_X,2}, \varepsilon h^{-d}) \leq \sup_Q N(\mathscr{L}_t, \|\cdot\|_{Q,2}, \varepsilon h^{-d}),$$

Same argument as paragraph **Covering Numbers** in the proof of Lemma SA-3.2 then shows

$$\begin{aligned}\sup_Q N(\{g_\mathbf{x} : \mathbf{x} \in \mathscr{B}\}, \|\cdot\|_{Q,2}, \varepsilon C_1 h^{-d}) &\leq \left(\frac{\operatorname{diam}(\mathscr{X})}{h\varepsilon}\right)^d, \qquad 0 < \varepsilon \leq 1, \\ \sup_Q N(\{g_\mathbf{x} h_\mathbf{x} : \mathbf{x} \in \mathscr{B}\}, \|\cdot\|_{Q,2}, \varepsilon C_1 h^{-d}) &\leq \left(\frac{\operatorname{diam}(\mathscr{X})}{h\varepsilon}\right)^d, \qquad 0 < \varepsilon \leq 1,\end{aligned}$$

where sup is taken over all discrete measures on $\mathscr{X}$. Product $\{g_\mathbf{x} : \mathbf{x} \in \mathscr{B}\}$ with the singleton of identity function $\{u \mapsto u, u \in \mathbb{R}\}$, and adding $\{g_\mathbf{x} h_\mathbf{x} : \mathbf{x} \in \mathscr{B}\}$,

$$\sup_Q N(\mathscr{F}, \|\cdot\|_{Q,2}, \varepsilon\|F\|_{Q,2}) \leq 2\left(\frac{2\operatorname{diam}(\mathscr{X})}{h\varepsilon}\right)^d, \qquad 0 < \varepsilon \leq 1,$$

where sup is taken over all discrete measures on $\mathscr{X} \times \mathbb{R}$. Denote $\mathtt{C}_1 = d$, $\mathtt{C}_2 = \frac{2(2\operatorname{diam}(\mathscr{X}))^d}{h^d}$. Hence, by

Chernozhukov et al. (2014b, Corollary 5.1)

$$\begin{aligned}
\mathbb{E}\Big[\sup_{\mathbf{x}\in\mathscr{B}}|\mathbf{e}_{\boldsymbol{\nu}}^{\top}\mathbf{O}_{t,\mathbf{x}}|\Big] &= \mathbb{E}\Big[\sup_{f\in\mathscr{F}}|\mathbb{E}_n\left[f(\mathbf{X}_i,Y_i)\right]-\mathbb{E}[f(\mathbf{X}_i,Y_i)]|\Big] \\
&\lesssim \frac{\sigma}{\sqrt{n}}\sqrt{\mathtt{C}_2\log(\mathtt{C}_1\|M\|_{\mathbb{P},2}/\sigma)}+\frac{\|M\|_{\mathbb{P},2}\mathtt{C}_2\log(\mathtt{C}_1\|M\|_{\mathbb{P},2}/\sigma)}{n} \\
&\lesssim \frac{1}{\sqrt{nh^d}}\sqrt{d\log\left(\frac{\operatorname{diam}(\mathscr{X})}{h^{1+d/2}}\right)}+\frac{1}{n^{\frac{1+v}{2+v}}h^d}d\log\left(\frac{\operatorname{diam}(\mathscr{X})}{h^{1+d/2}}\right) \\
&\lesssim \sqrt{\frac{\log(1/h)}{nh^d}}+\frac{\log(1/h)}{n^{\frac{1+v}{2+v}}h^d}.
\end{aligned}$$

The rest follows from finite dimensionality of $\mathbf{O}_{t,\mathbf{x}}$, and Lemma SA-3.2. ■

SA-6.4 Proof of Lemma SA-3.4

By Lemma SA-3.1 and Equation (SA-6.1), we have

$$\begin{aligned}
\sup_{\mathbf{x}\in\mathscr{B}}|\mathfrak{B}_{n,t}(\mathbf{x})| &= \sup_{\mathbf{x}\in\mathscr{B}}\left|\mathbf{e}_1^{\top}\boldsymbol{\Psi}_{t,\mathbf{x}}^{-1}\mathbf{S}_{t,\mathbf{x}}-\mu_t(\mathbf{x})\right| \\
&= \sup_{\mathbf{x}\in\mathscr{B}}\left|\mathbf{e}_1^{\top}\boldsymbol{\Psi}_{t,\mathbf{x}}^{-1}\mathbb{E}\Big[\mathbf{r}_p\left(\frac{D_i(\mathbf{x})}{h}\right)k_h(D_i(\mathbf{x}))\mathbf{R}_p(D_i(\mathbf{x}))^{\top}(\mu_t(\mathbf{X}_i)-\mu_t(\mathbf{x}),0,\cdots,0))\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_t)\Big]\right| \\
&\lesssim \sup_{\mathbf{x}\in\mathscr{B}}\sup_{\mathbf{z}\in\mathscr{X}}|\mu_t(\mathbf{x})-\mu_t(\mathbf{z})|\mathbb{1}(k_h(\mathscr{d}(\mathbf{z},\mathbf{x}))>0) \\
&\lesssim h.
\end{aligned}$$

■

SA-6.5 Proof of Lemma SA-3.5

Denote $\eta_{i,t,\mathbf{x}}=Y_i-\theta^*_{t,\mathbf{x}}(D_i(\mathbf{x}))$ and $\xi_{i,t,\mathbf{x}}=\theta^*_{t,\mathbf{x}}(D_i(\mathbf{x}))-\widehat{\theta}_{t,\mathbf{x}}(D_i(\mathbf{x}))$. Then

$$\widehat{\boldsymbol{\Upsilon}}_{t,\mathbf{x},\mathbf{y}}=\mathbb{E}_n\left[\mathbf{r}_p\left(\frac{D_i(\mathbf{x})}{h}\right)\mathbf{r}_p\left(\frac{D_i(\mathbf{y})}{h}\right)^{\top}h^dk_h\left(D_i(\mathbf{x})\right)k_h\left(D_i(\mathbf{y})\right)(\eta_{i,t,\mathbf{x}}+\xi_{i,t,\mathbf{x}})^2\mathbb{1}_{\mathscr{I}_t}(D_i(\mathbf{x}))\right],$$

and we decompose the error into

$$\begin{aligned}
&\widehat{\boldsymbol{\Upsilon}}_{t,\mathbf{x},\mathbf{y}}-\boldsymbol{\Upsilon}_{t,\mathbf{x},\mathbf{y}}=\Delta_{1,t,\mathbf{x},\mathbf{y}}+\Delta_{2,t,\mathbf{x},\mathbf{y}}+\Delta_{3,t,\mathbf{x},\mathbf{y}}, \\
&\Delta_{1,t,\mathbf{x},\mathbf{y}}=\mathbb{E}_n\left[\mathbf{r}_p\left(\frac{D_i(\mathbf{x})}{h}\right)\mathbf{r}_p\left(\frac{D_i(\mathbf{y})}{h}\right)^{\top}h^dk_h\left(D_i(\mathbf{x})\right)k_h\left(D_i(\mathbf{y})\right)\xi^2_{i,t,\mathbf{x}}\mathbb{1}_{\mathscr{I}_t}(D_i(\mathbf{x}))\right], \\
&\Delta_{2,t,\mathbf{x},\mathbf{y}}=2\mathbb{E}_n\left[\mathbf{r}_p\left(\frac{D_i(\mathbf{x})}{h}\right)\mathbf{r}_p\left(\frac{D_i(\mathbf{y})}{h}\right)^{\top}h^dk_h\left(D_i(\mathbf{x})\right)k_h\left(D_i(\mathbf{y})\right)\eta_{i,t,\mathbf{x}}\xi_{i,t,\mathbf{x}}\mathbb{1}_{\mathscr{I}_t}(D_i(\mathbf{x}))\right], \\
&\Delta_{3,t,\mathbf{x},\mathbf{y}}=\mathbb{E}_n\left[\mathbf{r}_p\left(\frac{D_i(\mathbf{x})}{h}\right)\mathbf{r}_p\left(\frac{D_i(\mathbf{y})}{h}\right)^{\top}h^dk_h\left(D_i(\mathbf{x})\right)k_h\left(D_i(\mathbf{y})\right)\eta^2_{i,t,\mathbf{x}}\mathbb{1}_{\mathscr{I}_t}(D_i(\mathbf{x}))\right] \\
&\qquad-\mathbb{E}\left[\mathbf{r}_p\left(\frac{D_i(\mathbf{x})}{h}\right)\mathbf{r}_p\left(\frac{D_i(\mathbf{y})}{h}\right)^{\top}h^dk_h\left(D_i(\mathbf{x})\right)k_h\left(D_i(\mathbf{y})\right)\eta^2_{i,t,\mathbf{x}}\mathbb{1}_{\mathscr{I}_t}(D_i(\mathbf{x}))\right].
\end{aligned}$$

By Assumption SA–2, $k_h(D_i(\mathbf{x})) \neq 0$ implies $\|\mathbf{r}_p(D_i(\mathbf{x})/h)\|_2 \lesssim 1$. Hence by Lemma SA-3.2 and SA-3.3,

$$\begin{aligned}
&\max_{t\in\{0,1\}} \max_{1\leq i\leq n} \sup_{\mathbf{x}\in\mathscr{B}} |\xi_{i,t,\mathbf{x}}| \\
&= \max_{t\in\{0,1\}} \max_{1\leq i\leq n} \sup_{\mathbf{x}\in\mathscr{B}} |\mathbf{r}_p(D_i(\mathbf{x}))^\top(\widehat{\boldsymbol{\gamma}}_{t,\mathbf{x}} - \boldsymbol{\gamma}^*_{t,\mathbf{x}})|\mathbb{1}(k_h(D_i(\mathbf{x})) \geq 0) \\
&= \max_{t\in\{0,1\}} \max_{1\leq i\leq n} \sup_{\mathbf{x}\in\mathscr{B}} |\mathbf{r}_p(D_i(\mathbf{x}))^\top \mathbf{H}^{-1}(\widehat{\boldsymbol{\Psi}}_{t,\mathbf{x}}^{-1}\mathbf{O}_{t,\mathbf{x}} + (\widehat{\boldsymbol{\Psi}}_{t,\mathbf{x}}^{-1} - \boldsymbol{\Psi}_{t,\mathbf{x}}^{-1})\mathbf{U}_{t,\mathbf{x}})|\mathbb{1}(k_h(D_i(\mathbf{x})) \geq 0) \\
&\leq \max_{t\in\{0,1\}} \sup_{\mathbf{x}\in\mathscr{B}} \|\widehat{\boldsymbol{\Psi}}_{t,\mathbf{x}}^{-1}\mathbf{O}_{t,\mathbf{x}}\|_2 + \max_{t\in\{0,1\}} \sup_{\mathbf{x}\in\mathscr{B}} \|(\widehat{\boldsymbol{\Psi}}_{t,\mathbf{x}}^{-1} - \boldsymbol{\Psi}_{t,\mathbf{x}}^{-1})\mathbf{U}_{t,\mathbf{x}}\|_2 \\
&\lesssim \sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{1+v}{2+v}}h^d},
\end{aligned}$$

where

$$\mathbf{U}_{t,\mathbf{x}} = \mathbb{E}_n\left[\mathbf{r}_p\left(\frac{D_i(\mathbf{x})}{h}\right) k_h(D_i(\mathbf{x}))\theta^*_{t,\mathbf{x}}(\mathbf{X}_i)\mathbb{1}_{\mathscr{I}_t}(D_i(\mathbf{x}))\right].$$

Assuming $\frac{\log(1/h)}{n^{\frac{1+v}{2+v}}h^d} \to \infty$, similar maximal inequality as in the proof of Lemma SA-3.2 shows

$$\begin{aligned}
\sup_{\mathbf{x},\mathbf{y}\in\mathscr{X}} \|\Delta_{1,t,\mathbf{x},\mathbf{y}}\| &\lesssim_{\mathbb{P}} \max_{t\in\{0,1\}} \max_{1\leq i\leq n} \sup_{\mathbf{x}\in\mathscr{B}} |\xi_{i,t,\mathbf{x}}|^2 \lesssim \left(\sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{1+v}{2+v}}h^d}\right)^2, \\
\sup_{\mathbf{x},\mathbf{y}\in\mathscr{X}} \|\Delta_{2,t,\mathbf{x},\mathbf{y}}\| &\lesssim_{\mathbb{P}} \max_{t\in\{0,1\}} \max_{1\leq i\leq n} \sup_{\mathbf{x}\in\mathscr{B}} |\xi_{i,t,\mathbf{x}}| \lesssim \sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{1+v}{2+v}}h^d}.
\end{aligned} \tag{SA-6.3}$$

Consider the (μ,ν) entry of $\Delta_{3,t,\mathbf{x},\mathbf{y}}$. Consider the class

$$\mathscr{F} = \left\{(\mathbf{z},u) \mapsto \left(\frac{\mathscr{d}(\mathbf{z},\mathbf{x})}{h}\right)^{\mu+\nu} h^d k_h(\mathscr{d}(\mathbf{z},\mathbf{x}))k_h(\mathscr{d}(\mathbf{z},\mathbf{y}))(u - \mathbf{r}_p(\mathscr{d}(\mathbf{z},\mathbf{x}))^\top\boldsymbol{\gamma}^*_{t,\mathbf{x}})^2 : \mathbf{x},\mathbf{y}\in\mathscr{X}\right\}.$$

By Assumption SA–2 and SA–1(v), we have $\sup_{f\in\mathscr{F}} \mathbb{E}[f(\mathbf{X}_i,Y_i)^2]^{1/2} \lesssim h^{-d/2}$. Moreover, Assumption SA–2 and Equation (SA-6.2) imply there exists $C_1, C_2 > 0$ such that $F(\mathbf{z},u) = C_1 h^{-d}(u^2 + C_2)$ is an envelope function for $\mathscr{F}$, with

$$\mathbb{E}\big[\max_{1\leq i\leq n} F(\mathbf{X}_i,Y_i)^2\big]^{\frac{1}{2}} \lesssim C_1 h^{-d}(\mathbb{E}\big[\max_{1\leq i\leq n} Y_i^4\big]^{\frac{1}{2}} + C_2) \lesssim C_1 h^{-d}(\mathbb{E}\big[\max_{1\leq i\leq n} Y_i^{2+v}\big]^{\frac{2}{2+v}} + C_2) \lesssim h^{-d} n^{\frac{2}{2+v}}.$$

Apply Chernozhukov et al. (2014b, Corollary 5.1) similarly as in Lemma SA-3.3 gives

$$\mathbb{E}\left[\sup_{f\in\mathscr{F}} |\mathbb{E}_n[f(\mathbf{X}_i,Y_i)] - \mathbb{E}[f(\mathbf{X}_i,Y_i)]|\right] \lesssim \sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{v}{2+v}}h^d}.$$

Finite dimensionality of $\Delta_{3,t,\mathbf{x},\mathbf{y}}$ then implies

$$\mathbb{E}\left[\sup_{\mathbf{x},\mathbf{y}\in\mathscr{X}} \|\Delta_{3,t,\mathbf{x},\mathbf{y}}\|\right] \lesssim \sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{v}{2+v}}h^d}. \tag{SA-6.4}$$

Putting together Equations (SA-6.3), (SA-6.4) and Lemma SA-3.2 gives the result. ■

SA-6.6 Proof of Theorem SA-3.1

All analysis in Lemma SA-3.2 and Lemma SA-3.3 can be done when the index set is the singleton $\{\mathbf{x}\}$ instead of $\mathscr{B}$, replacing (Chernozhukov et al., 2014b, Corollary 5.1) by Bernstein inequality, and gives for any $\mathbf{x} \in \mathscr{B}$,

$$\left|\mathbf{e}_1^\top \mathbf{\Psi}_{t,\mathbf{x}}^{-1} \mathbf{O}_{t,\mathbf{x}}\right| \lesssim_{\mathbb{P}} \sqrt{\frac{1}{nh^d}} + \frac{1}{n^{\frac{1+v}{2+v}} h^d},$$
$$\left|\mathbf{e}_1^\top (\widehat{\mathbf{\Psi}}_{t,\mathbf{x}}^{-1} - \mathbf{\Psi}_{t,\mathbf{x}}^{-1}) \mathbf{O}_{t,\mathbf{x}}\right| \lesssim_{\mathbb{P}} \sqrt{\frac{1}{nh^d}} \left(\sqrt{\frac{1}{nh^d}} + \frac{1}{n^{\frac{1+v}{2+v}} h^d}\right).$$

The decomposition Equation (SA-3.1) then gives the result. ■

SA-6.7 Proof of Theorem SA-3.2

Define $\overline{\mathrm{T}}_{\mathrm{dis}}(\mathbf{x}) = \Xi_{\mathbf{x},\mathbf{x}}^{-1/2} \mathbf{e}_1^\top \mathbf{\Psi}_{t,\mathbf{x}}^{-1} \mathbf{O}_{t,\mathbf{x}}$. Notice that if we define

$$Z_{n,i} = \frac{1}{n} \Xi_{\mathbf{x},\mathbf{x}}^{-1/2} \mathbf{e}_1^\top \mathbf{\Psi}_{t,\mathbf{x}}^{-1} \mathbf{r}_p \left(\frac{D_i(\mathbf{x})}{h}\right) k_h\left(D_i(\mathbf{x})\right)\left(Y_i - \theta_{t,\mathbf{x}}^*(D_i(\mathbf{x}))\right) \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t),$$

then $\overline{\mathrm{T}}_{\mathrm{dis}}(\mathbf{x}) = \sum_{i=1}^n Z_{n,i}$. Moreover, $\mathbb{E}\left[Z_{n,i}\right] = 0$ and $\mathbb{V}[Z_{n,i}] = n^{-1}$. By Berry-Essen Theorem,

$$\begin{aligned}
&\sup_{u \in \mathbb{R}} \left|\mathbb{P}\left(\overline{\mathrm{T}}_{\mathrm{dis}}(\mathbf{x}) \leq u\right) - \Phi(u)\right| \lesssim \sum_{i=1}^n \mathbb{E}\left[\left|Z_{n,i}\right|^3\right] \\
&= \sum_{i=1}^n n^{-3} \Xi_{\mathbf{x},\mathbf{x}}^{-3/2} \mathbb{E}\left[\left|\mathbf{e}_1^\top \mathbf{\Psi}_{t,\mathbf{x}}^{-1} \mathbf{r}_p\left(\frac{D_i(\mathbf{x})}{h}\right) k_h\left(D_i(\mathbf{x})\right) \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)(Y_i - \theta_{t,\mathbf{x}}^*(D_i(\mathbf{x}))\right|^3\right] \\
&\lesssim n^{-2} \Xi_{\mathbf{x},\mathbf{x}}^{-3/2} \mathbb{E}[|k_h(D_i(\mathbf{x}))(Y_i - \theta_{t,\mathbf{x}}^*(D_i(\mathbf{x})))|^3] \\
&\lesssim n^{-2} \Xi_{\mathbf{x},\mathbf{x}}^{-3/2} \mathbb{E}[|k_h(D_i(\mathbf{x}))(\mathbb{E}[|Y_i|^3|\mathbf{X}_i] + |\theta_{t,\mathbf{x}}^*(D_i(\mathbf{x}))|^3)|] \\
&\lesssim (nh^d)^{-1/2},
\end{aligned}$$

where in the third line we used $\sup_{\mathbf{x} \in \mathscr{B}} \|\mathbf{r}_p\left(\frac{D_i(\mathbf{x})}{h}\right) k_h\left(D_i(\mathbf{x})\right)\| \lesssim 1$ holds almost surely in $\mathbf{X}_i$, and in the last line we used $\Xi_{\mathbf{x},\mathbf{x}} \gtrsim (nh^d)^{-1/2}$ from Lemma SA-3.5, Assumption SA–1(v) so that $\mathbb{E}[|Y_i|^3|\mathbf{X}_i] \lesssim 1$ and

$$\theta_{t,\mathbf{x}}^*(D_i(\mathbf{x})) = \gamma_t^*(\mathbf{x})^\top \mathbf{r}_p(D_i(\mathbf{x})) = (\mathbf{\Psi}_{t,\mathbf{x}} \mathbf{S}_{t,\mathbf{x}})^{-1} \mathbf{r}_p\left(\frac{D_i(\mathbf{x})}{h}\right),$$

implying $\max_{1 \leq i \leq n} \sup_{\mathbf{x} \in \mathscr{B}} |\theta_{t,\mathbf{x}}^*(D_i(\mathbf{x}))| \lesssim 1$ for $t \in \{0,1\}$. The counterpart of Theorem SA-3.4 gives

$$|\widehat{\mathrm{T}}_{\mathrm{dis}}(\mathbf{x}) - \overline{\mathrm{T}}_{\mathrm{dis}}(\mathbf{x})| \lesssim_{\mathbb{P}} \frac{1}{\sqrt{nh^d}} + \frac{1}{n^{\frac{v}{2+v}} h^d} + \sqrt{nh^d} \sum_{t \in \{0,1\}} |\mathfrak{B}_{n,t}(\mathbf{x})|.$$

Putting together we have

$$\mathbb{P}(\tau \in \widehat{\mathrm{I}}_{\mathrm{dis}}(\mathbf{x}, \alpha)) = \mathbb{P}(|\widehat{\mathrm{T}}_{\mathrm{dis}}(\mathbf{x})| \leq \mathfrak{c}_\alpha) = \mathbb{P}(|\overline{\mathrm{T}}_{\mathrm{dis}}(\mathbf{x})| \leq \mathfrak{c}_\alpha) + o(1) = 2(1 - \Phi(\mathfrak{c}_\alpha)) + o(1) = 1 - \alpha + o(1).$$

■

SA-6.8 Proof of Theorem SA-3.3

The statement follows from Lemma SA-3.2, Lemma SA-3.3 and the decomposition Equation (SA-3.1). ■

SA-6.9 Proof of Theorem SA-3.4

We make the decomposition based on Equation (SA-3.1) and convergence of $\widehat{\Xi}_{\mathbf{x},\mathbf{x}}$,

$$\begin{aligned}
\widehat{\mathrm{T}}_{\mathrm{dis}}(\mathbf{x}) - \overline{\mathrm{T}}_{\mathrm{dis}}(\mathbf{x}) &= \widehat{\Xi}_{\mathbf{x},\mathbf{x}}^{-1/2}\Big(\sum_{t\in\{0,1\}}(-1)^{\frac{t+1}{2}}(\widehat{\theta}_{t,\mathbf{x}}(0)-\theta_{t,\mathbf{x}}(0))\Big) - \Xi_{\mathbf{x},\mathbf{x}}^{-1/2}\Big(\sum_{t\in\{0,1\}}(-1)^{\frac{t+1}{2}}\mathbf{e}_1^\top\mathbf{\Psi}_{t,\mathbf{x}}^{-1}\mathbf{O}_{t,\mathbf{x}}\Big)\\
&= \widehat{\Xi}_{\mathbf{x},\mathbf{x}}^{-1/2}\Big(\sum_{t\in\{0,1\}}(-1)^{\frac{t+1}{2}}(\widehat{\theta}_{t,\mathbf{x}}(0)-\theta_{t,\mathbf{x}}(0)) - \sum_{t\in\{0,1\}}(-1)^{\frac{t+1}{2}}\mathbf{e}_1^\top\mathbf{\Psi}_{t,\mathbf{x}}^{-1}\mathbf{O}_{t,\mathbf{x}}\Big) \qquad (=\Delta_{1,\mathbf{x}})\\
&\quad + (\widehat{\Xi}_{\mathbf{x},\mathbf{x}}^{-1/2} - \Xi_{\mathbf{x},\mathbf{x}}^{-1/2})\sum_{t\in\{0,1\}}(-1)^{\frac{t+1}{2}}\mathbf{e}_1^\top\mathbf{\Psi}_{t,\mathbf{x}}^{-1}\mathbf{O}_{t,\mathbf{x}} \qquad (=\Delta_{2,\mathbf{x}})
\end{aligned}$$

By Lemma SA-3.2 and SA-3.3, and the decomposition Equation (SA-3.1),

$$\begin{aligned}
&\sup_{\mathbf{x}\in\mathcal{X}}\Big|\sum_{t\in\{0,1\}}(-1)^{\frac{t+1}{2}}(\widehat{\theta}_{t,\mathbf{x}}(0)-\theta_{t,\mathbf{x}}(0)) - \sum_{t\in\{0,1\}}(-1)^{\frac{t+1}{2}}\mathbf{e}_1^\top\mathbf{\Psi}_{t,\mathbf{x}}^{-1}\mathbf{O}_{t,\mathbf{x}}\Big|\\
&\lesssim_{\mathbb{P}} \sqrt{\frac{\log(1/h)}{nh^d}}\left(\sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{1+v}{2+v}}h^d}\right) + \sup_{\mathbf{x}\in\mathcal{B}}\sum_{t\in\{0,1\}}|\theta^*_{t,\mathbf{x}}(0)-\theta_{t,\mathbf{x}}(0)|.
\end{aligned}$$

Together with Lemma SA-3.5,

$$\sup_{\mathbf{x}\in\mathcal{B}}|\Delta_{1,\mathbf{x}}| \lesssim_{\mathbb{P}} \frac{\log(1/h)}{\sqrt{nh^d}} + \frac{(\log(1/h))^{\frac{3}{2}}}{n^{\frac{1+v}{2+v}}h^d} + \sqrt{nh^d}\sup_{\mathbf{x}\in\mathcal{B}}\sum_{t\in\{0,1\}}|\theta^*_{t,\mathbf{x}}(0)-\theta_{t,\mathbf{x}}(0)|. \tag{SA-6.5}$$

By Lemma SA-3.2, Lemma SA-3.3 and Lemma SA-3.5, and assume $\frac{n^{\frac{v}{2+v}}h^d}{\log(1/h)}\to\infty$, then

$$\begin{aligned}
\sup_{\mathbf{x}\in\mathcal{X}}\left|\mathbf{e}_1^\top\mathbf{\Psi}_{t,\mathbf{x}}^{-1}\mathbf{O}_{t,\mathbf{x}}\left(\Xi_{\mathbf{x},\mathbf{x}}^{-1/2}-\widehat{\Xi}_{\mathbf{x},\mathbf{x}}^{-1/2}\right)\right| &\lesssim_{\mathbb{P}} \sqrt{nh^d}\left(\sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{1+v}{2+v}}h^d}\right)\left(\sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{v}{2+v}}h^d}\right)\\
&= \sqrt{\log(1/h)}\left(1+\sqrt{\frac{\log(1/h)}{n^{\frac{v}{2+v}}h^d}}\right)\left(\sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{v}{2+v}}h^d}\right)\\
&\lesssim \sqrt{\log(1/h)}\left(\sqrt{\frac{\log(1/h)}{nh^d}} + \frac{\log(1/h)}{n^{\frac{v}{2+v}}h^d}\right).
\end{aligned}$$

Hence

$$\sup_{\mathbf{x}\in\mathcal{B}}|\Delta_{2,\mathbf{x}}| \lesssim_{\mathbb{P}} \frac{\log(1/h)}{\sqrt{nh^d}} + \frac{(\log(1/h))^{\frac{3}{2}}}{n^{\frac{1+v}{2+v}}h^d}. \tag{SA-6.6}$$

Putting together Equations (SA-6.5), (SA-6.6) give the result. ■

SA-6.10 Proof of Theorem SA-3.5

Without loss of generality, we can 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 $\mathscr{G}$, and $\phi_{\mathscr{G}} = \mathrm{Id}$ is a valid normalizing transformation (as in Lemma SA-4.1). This implies the constants $\mathtt{c}_1$ and $\mathtt{c}_2$ from Lemma SA-4.1 are all 1.

By similar arguments as in the proof of Theorem SA-2.7, we get properties of $\mathscr{G}$ as follows:

$$\mathtt{M}_{\mathscr{G}} \lesssim h^{-d/2}, \qquad \mathtt{E}_{\mathscr{G}} \lesssim h^{d/2}, \qquad \mathtt{TV}_{\mathscr{G}} \lesssim h^{d/2-1}, \qquad \sup_Q N(\mathscr{G}, \|\cdot\|_{Q,2}, \varepsilon(2c+1)^{d+1}\mathtt{M}_{\mathscr{G}}) \leq 2\mathbf{c}'\varepsilon^{-d-1} + 2.$$

By definition of $\theta^*(\cdot)$, for each $\mathbf{x} \in \mathscr{B}$, $t \in \{0,1\}$,

$$\theta^*_{t,\mathbf{x}}(\mathscr{d}(\mathbf{u},\mathbf{x})) = \boldsymbol{\gamma}^*_t(\mathbf{x})^\top \mathbf{r}_p(\mathscr{d}(\mathbf{u},\mathbf{x})) = (\mathbf{H}^{-1}\boldsymbol{\Psi}^{-1}_{t,\mathbf{x}}\mathbf{S}_{t,\mathbf{x}})^\top \mathbf{r}_p(\mathscr{d}(\mathbf{u},\mathbf{x})) = (\boldsymbol{\Psi}^{-1}_{t,\mathbf{x}}\mathbf{S}_{t,\mathbf{x}})^\top \mathbf{r}_p\Big(\frac{\mathscr{d}(\mathbf{u},\mathbf{x})}{h}\Big),$$

recalling

$$\boldsymbol{\Psi}_{t,\mathbf{x}} = \mathbb{E}\left[\mathbf{r}_p\left(\frac{D_i(\mathbf{x})}{h}\right)\mathbf{r}_p\left(\frac{D_i(\mathbf{x})}{h}\right)^\top k_h(D_i(\mathbf{x}))\mathbb{1}_{\mathcal{I}_t}(D_i(\mathbf{x}))\right], \quad \mathbf{S}_{t,\mathbf{x}} = \mathbb{E}\left[\mathbf{r}_p\left(\frac{D_i(\mathbf{x})}{h}\right)k_h(D_i(\mathbf{x}))Y_i\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)\right].$$

We can check that $\|\boldsymbol{\Psi}^{-1}_{t,\mathbf{x}}\| \lesssim 1$, $\|\mathbf{S}_{t,\mathbf{x}}\| \lesssim 1$ and

$$\mathtt{M}_{\mathscr{H}_t} \lesssim h^{-d/2}, \qquad \mathtt{E}_{\mathscr{H}_t} \lesssim h^{-d/2}, \qquad t \in \{0,1\}.$$

In what follows, we verify the entropy and total variation properties of $\mathscr{H}$. Using product rule we can verify

$$\sup_{\mathbf{u}\in\mathcal{X}} \sup_{\mathbf{x},\mathbf{x}'\in\mathscr{B}} \frac{|\theta^*_{t,\mathbf{x}}(\mathscr{d}(\mathbf{u},\mathbf{x})) - \theta^*_{t,\mathbf{x}}(\mathscr{d}(\mathbf{u},\mathbf{x}'))|}{\|\mathbf{x}-\mathbf{x}'\|} \lesssim h^{-1}.$$

Define $f_{t,\mathbf{x}}(\cdot) = \frac{h^{-d/2}}{\sqrt{n\Xi_{\mathbf{x},\mathbf{x}}}}\mathbf{e}_1^\top \boldsymbol{\Psi}^{-1}_{t,\mathbf{x}}\mathbf{r}_p(\cdot)\,K(\cdot)(\boldsymbol{\Psi}^{-1}_{t,\mathbf{x}}\mathbf{S}_{t,\mathbf{x}})^\top\mathbf{r}_p(\cdot)$. Then,

$$\mathfrak{K}_t(\mathbf{u};\mathbf{x})\theta^*_{t,\mathbf{x}}(\mathscr{d}(\mathbf{u},\mathbf{x})) = h^{-d/2}f_{t,\mathbf{x}}\Big(\frac{d(\mathbf{u},\mathbf{x})}{h}\Big), \qquad \mathbf{u}\in\mathcal{X}, \mathbf{x}\in\mathscr{B}, t\in\{0,1\}.$$

Take $\mathscr{H}_t = \{\mathfrak{K}_t(\cdot;\mathbf{x})\theta^*_{t,\mathbf{x}}(\mathscr{d}(\cdot,\mathbf{x})) : \mathbf{x}\in\mathscr{B}\}$, $t\in\{0,1\}$. For $t\in\{0,1\}$, $f_{t,\mathbf{x}}$ satisfies:

(i) boundedness $\quad \sup_{\mathbf{x}\in\mathscr{B}}\sup_{\mathbf{u}\in\mathcal{X}} |f_{t,\mathbf{x}}(\mathbf{u})| \leq \mathbf{c}$,

(ii) compact support $\quad \mathrm{supp}(f_{t,\mathbf{x}}(\cdot)) \subseteq [-\mathbf{c},\mathbf{c}]^d, \forall \mathbf{x}\in\mathscr{B}$,

(iii) Lipschitz continuity $\quad \sup_{\mathbf{x}\in\mathscr{B}}\sup_{\mathbf{u},\mathbf{u}'\in\mathcal{X}} \frac{|f_{\mathbf{x}}(\mathbf{u}) - f_{\mathbf{x}}(\mathbf{u}')|}{\|\mathbf{u}-\mathbf{u}'\|} \leq \mathbf{c}$

$$\sup_{\mathbf{u}\in\mathbf{X}}\sup_{\mathbf{x},\mathbf{x}'\in\mathscr{B}} \frac{|f_{\mathbf{x}}(\mathbf{u}) - f_{\mathbf{x}'}(\mathbf{u})|}{\|\mathbf{x}-\mathbf{x}'\|} \leq \mathbf{c}h^{-1},$$

for some constant $\mathbf{c}$ not depending on n. Then, by an argument similar to Cattaneo et al. (2024, Lemma 7), there exists a constant $\mathbf{c}'$ only depending on $\mathbf{c}$ and d that for any $0 \leq \varepsilon \leq 1$,

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

where supremum is taken over all finite discrete measures. Taking a constant envelope function $\mathtt{M}_{\mathscr{H}_t} = (2c+1)^{d+1}h^{-d/2}$, we have for any $0 < \varepsilon \leq 1$,

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

By Lemma SA-5.2, above implies the uniform covering number for $\mathscr{H}_t$ satisfies

$$\mathtt{N}_{\mathscr{H}_t}(\varepsilon) \leq 4\mathbf{c}'(\varepsilon/2)^{-d-1}, \qquad 0 < \varepsilon \leq 1.$$

Since $\mathscr{H} \subseteq \mathscr{H}_0 + \mathscr{H}_1$, here $+$ denotes the Minkowski sum, with $\mathtt{M}_{\mathscr{H}}$ taken to be $\mathtt{M}_{\mathscr{H}_0} + \mathtt{M}_{\mathscr{H}_1}$, a bound on the uniform covering number of $\mathscr{H}$ can be given by

$$\mathtt{N}_{\mathscr{H}}(\varepsilon) \leq 16(\mathbf{c}')^2(\varepsilon/2)^{-2d-2}, \qquad 0 < \varepsilon \leq 1.$$

With the assumption that $\mathscr{L}(E_{t,\mathbf{x}}) \leq Ch^{d-1}$ for $E_{t,\mathbf{x}} = \{\mathbf{y} \in \mathscr{A}_t : (\mathbf{y}-\mathbf{x})/h \in \mathrm{Supp}(K)\}$ for all $t \in \{0,1\}$, $\mathbf{x} \in \mathscr{B}$, and the fact that $\mathtt{TV}_{\mathscr{H}_t} \lesssim h^{d/2-1}$ for $t \in \{0,1\}$, the same argument as in the paragraph **Total Variation** in the proof of Theorem SA-2.7 shows

$$\mathtt{TV}_{\mathscr{H}} \lesssim h^{d/2-1}.$$

Now apply Lemma SA-4.2 with $\mathscr{G}$, $\mathscr{H}$ defined in Equation (SA-3.3) and $\mathscr{R} = \{\mathrm{Id}\}$, noticing that

$$(\overline{\mathrm{T}}_{\mathrm{dis}} : \mathbf{x} \in \mathscr{B}) = (A_n(g,h,r) : (g,h,r) \in \mathscr{F} \times \mathscr{R}), \qquad \mathscr{F} = \{(g_{\mathbf{x}}, h_{\mathbf{x}}) : \mathbf{x} \in \mathscr{B}\} \subseteq \mathscr{G} \times \mathscr{H},$$

the result then follows. ■

SA-6.11 Proof of Theorem SA-3.6

The result follows from Theorem SA-3.5, Theorem SA-3.4, Lemma SA-3.5 and similar arguments as the proof of Theorem SA-2.9. ■

SA-7 Proofs of Distance-Based Bias Results

SA-7.1 Proof of Lemma 2

SA-7.1.1 Upper Bound

The proof is essentially the proof for Lemma SA-3.4 with the data generating process ranging over $\mathscr{P}$. By Lemma SA-3.1 and Equation (SA-6.1), we have

$$\begin{aligned}
&\sup_{\mathbb{P}\in\mathscr{P}}\sup_{\mathbf{x}\in\mathscr{B}} |\mathfrak{B}_{n,t}(\mathbf{x})| \\
&= \sup_{\mathbb{P}\in\mathscr{P}}\sup_{\mathbf{x}\in\mathscr{B}} \left|\mathbf{e}_1^\top \mathbf{\Psi}_{t,\mathbf{x}}^{-1}\mathbf{S}_{t,\mathbf{x}} - \mu_t(\mathbf{x})\right| \\
&= \sup_{\mathbb{P}\in\mathscr{P}}\sup_{\mathbf{x}\in\mathscr{B}} \left|\mathbf{e}_1^\top \mathbf{\Psi}_{t,\mathbf{x}}^{-1}\mathbb{E}\Big[\mathbf{r}_p\left(\frac{D_i(\mathbf{x})}{h}\right) k_h(D_i(\mathbf{x}))\mathbf{r}_p(D_i(\mathbf{x}))^\top(\mu_t(\mathbf{X}_i)-\mu_t(\mathbf{x}),0,\cdots,0))\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_t)\Big]\right| \\
&\lesssim \sup_{\mathbb{P}\in\mathscr{P}}\sup_{\mathbf{x}\in\mathscr{B}}\sup_{\mathbf{z}\in\mathscr{X}} \left|\mathbf{e}_1^\top \mathbf{\Psi}_{t,\mathbf{x}}^{-1}\mathbb{E}\Big[\mathbf{r}_p\left(\frac{D_i(\mathbf{x})}{h}\right) k_h(D_i(\mathbf{x}))\mathbf{r}_p\left(\frac{D_i(\mathbf{x})}{h}\right)^\top\right| \\
&\qquad \cdot \sup_{\mathbb{P}\in\mathscr{P}}\sup_{\mathbf{x}\in\mathscr{B}}\sup_{\mathbf{z}\in\mathscr{X}} |\mu_t(\mathbf{x})-\mu_t(\mathbf{z})|\mathbb{1}(k_h(\mathscr{d}(\mathbf{z},\mathbf{x}))>0) \\
&\lesssim h.
\end{aligned}$$

SA-7.1.2 Lower Bound

The lower bound is proved by considering the following data generating process. Suppose $\mathbf{X}_i \sim \mathsf{Unif}([-2,2]^2)$, and $\mu_0(x_1,x_2) = 0$ and $\mu_1(x_1,x_2) = x_2$ for all $(x_1,x_2)\in\mathscr{X} = [-2,2]^2$. Suppose $Y_i(0)\sim\mathsf{N}(\mu_0(\mathbf{X}_i),1)$ and $Y_i(1)\sim\mathsf{N}(\mu_1(\mathbf{X}_i),1)$. Define the treatment and control region by $\mathscr{A}_1 = \{(x,y)\in\mathscr{X} : x\geq 0, y\geq 0\}$, $\mathscr{A}_0 = \mathscr{X}/\mathscr{A}_1$, $\mathscr{B} = \{(x,y)\in\mathbb{R} : 0\leq x\leq 2, y = 0 \text{ or } x = 0, 0\leq y\leq 2\}$. Suppose $Y_i = \mathbb{1}(\mathbf{X}_i\in\mathscr{A}_0)Y_i(0) + \mathbb{1}(\mathbf{X}_i\in\mathscr{A}_1)Y_i(1)$. Suppose we choose $\mathscr{d}$ to be the Euclidean distance and $D_i(\mathbf{x}) = \|\mathbf{X}_i - \mathbf{x}\|$. In this case, although the underlying conditional mean functions μ_t, $t\in\{0,1\}$ are smooth, the conditional mean given distance $\theta_{t,\mathbf{x}}$ may not even be differentiable. In this example,

$$\theta_{1,(s,0)}(r) = \begin{cases} \frac{2}{\pi r}, & \text{if } 0\leq r\leq s, \\ \frac{r+s}{\pi-\arccos(s/r)}, & \text{if } r > s. \end{cases}$$

Figure SA-1 plots $r\mapsto\theta_{1,(3/4,0)}(r)$ with the notation $\mathbf{x}_s = (s,0)$.

Under this data generating process, we can show

$$\inf_{0<h<1}\sup_{\mathbf{x}\in\mathscr{B}}\frac{|\mathfrak{B}_{n,1}(\mathbf{x}) - \mathfrak{B}_{n,0}(\mathbf{x})|}{h} > 0.$$

The proof proceeds in two steps. First, we show a scaling property of the asymptotic bias under our example, which gives a reduction to fixed-h bias calculation. Second, we prove the lower bound via the reduction from previous step.

Step 1: A Scaling Property. Let $0<h<1, 0<s<1, 0<C<1$. Define $h' = Ch$ and $s' = Cs$. Here

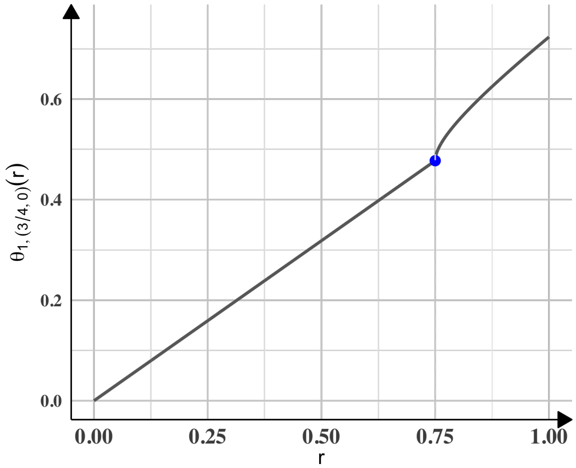

Figure SA-1. Conditional Mean Given Distance with One Kink

C is the scaling factor and denote $\mathbf{x}_s = (s, 0)$ and $\mathbf{x}_{s'} = (s', 0)$. Denote bias for $\mathbf{x}_{s'}$ under bandwidth h' to be

$$\begin{aligned}\mathrm{bias}_{n,1}(h', s') = \mathbf{e}_1^\top \mathbb{E}\Bigg[\mathbf{r}_p\left(\frac{D_i((s',0))}{h'}\right)\mathbf{r}_p\left(\frac{D_i((s',0))}{h'}\right)^\top k_{h'}(D_i((s',0)))\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_1)\Bigg]^{-1} \\ \mathbb{E}\left[\mathbf{r}_p\left(\frac{D_i((s',0))}{h'}\right)k_{h'}\left(D_i((s',0))\right)\left(\mu_1(\mathbf{X}_i - (s',0))\right)\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_1)\right], \qquad \text{(SA-7.1)}\end{aligned}$$

where we have used the fact that μ_1 is linear in our example, hence $\mu_1(\mathbf{X}_i) - \mu_1((s',0)) = \mu_1(\mathbf{X}_i - (s',0))$. We reserve the notation $\mathfrak{B}_{n,t}$, $t = 0, 1$, to the bias when bandwidth is h, that is,

$$\mathfrak{B}_{n,t}(\mathbf{x}_s) \equiv \mathrm{bias}_{n,t}(h, s), \qquad h \in (0,1), s \in (0,1), t = 0, 1.$$

Inspecting each element of the last vector, for all $l \in \mathbb{N}$,

$$\begin{aligned}&\mathbb{E}\left[\left(\frac{\|\mathbf{X}_i - (s',0)\|}{h'}\right)^l k_{h'}\left(\|\mathbf{X}_i - (s',0)\|\right)\left(\mu_1(\mathbf{X}_i - (s',0))\right)\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_1)\right] \\ &= \int_0^2\int_0^2\left(\frac{1}{h'}\right)^2\left(\frac{\|(u'-s',v')\|}{h'}\right)^l k\left(\frac{\|(u'-s',v')\|}{h'}\right)\mu_1\left((u',v')-(s',0)\right)\frac{1}{4}du'dv' \\ &\overset{(1)}{=} \int_0^{2/C}\int_0^{2/C}\left(\frac{1}{Ch}\right)^2\left(\frac{\|(Cu-Cs,Cv)\|}{Ch}\right)^l k\left(\frac{\|(Cu-Cs,Cv)\|}{Ch}\right)\mu_1\left(C(u-s,v)\right)\frac{C^2}{4}dudv \\ &= \int_0^{2/C}\int_0^{2/C}\left(\frac{1}{h}\right)^2\left(\frac{\|(u-s,v)\|}{h}\right)^l k\left(\frac{\|(u-s,v)\|}{h}\right)C\mu_1\left((u-s,v)\right)\frac{1}{4}dudv \\ &\overset{(2)}{=} \int_0^2\int_0^2\left(\frac{1}{h}\right)^2\left(\frac{\|(u-s,v)\|}{h}\right)^l k\left(\frac{\|(u,v)-(s,0)\|}{h}\right)C\mu_1\left((u,v)-(s,0)\right)\frac{1}{4}dudv \\ &= C\mathbb{E}\left[\left(\frac{\|\mathbf{X}_i - (s,0)\|}{h}\right)^l k_h\left(\|\mathbf{X}_i - (s,0)\|\right)\mu_1(\mathbf{X}_i - (s,0))\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_1)\right],\end{aligned}$$

where in (1) we have used a change of variable $(u, v) = \frac{1}{C}(u', v')$, and (2) holds since $k\left(\frac{\|\cdot-(s,0)\|}{h}\right)$ is supported in $(s, 0) + hB(0, 1)$, which is contained in $[0, 2] \times [0, 2] \subseteq [0, 2/C] \times [0, 2/C]$ for all $0 < h < 1$, $0 < s < 1$,

$0 < C < 1$. This means

$$
\begin{aligned}
&\mathbb{E}\left[\mathbf{r}_p\left(\frac{D_i((s',0))}{h'}\right) k_{h'}\left(D_i((s',0))\right)\left(\mu_1(\mathbf{X}_i-(s',0))\right)\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_1)\right]\\
&=C\mathbb{E}\left[\mathbf{r}_p\left(\frac{D_i((s,0))}{h}\right) k_{h}\left(D_i((s,0))\right)\left(\mu_1(\mathbf{X}_i-(s,0))\right)\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_1)\right].
\end{aligned}
$$

Similarly, for all $l \in \mathbb{N}$ and $0 < h < 1$, $0 < s < 1$, $0 < C < 1$,

$$
\mathbb{E}\left[\left(\frac{D_i((s',0))}{h'}\right)^l k_{h'}\left(D_i((s',0))\right)\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_1)\right]=\mathbb{E}\left[\left(\frac{D_i((s,0))}{h}\right)^l k_{h}\left(D_i((s,0))\right)\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_1)\right],
$$

implying

$$
\begin{aligned}
&\mathbb{E}\left[\mathbf{r}_p\left(\frac{D_i((s',0))}{h'}\right)\mathbf{r}_p\left(\frac{D_i((s',0))}{h'}\right)^\top k_{h'}(D_i((s',0)))\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_1)\right]\\
&=\mathbb{E}\left[\mathbf{r}_p\left(\frac{D_i((s,0))}{h}\right)\mathbf{r}_p\left(\frac{D_i((s,0))}{h}\right)^\top k_{h}(D_i((s,0)))\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_1)\right].
\end{aligned}
$$

It then follows that for all $0 < h < 1$, $0 < s < 1$, $0 < C < 1$,

$$
\mathrm{bias}_{n,1}(h',s') = C\,\mathrm{bias}_{n,1}(h,s).
$$

Moreover, for all $0 < h < 1$, $0 < s < h$,

$$
\mathfrak{B}_{n,1}(\mathbf{x}_s) = \mathrm{bias}_{n,1}(h,s) = h\,\mathrm{bias}_{n,1}\left(1,\frac{s}{h}\right). \tag{SA-7.2}
$$

Since $\mu_0 \equiv 0$, it is easy to check that

$$
\mathfrak{B}_{n,0}(\mathbf{x}_s) = \mathrm{bias}_{n,0}(h,s) \equiv 0, \qquad 0<h<1, 0<s<h.
$$

Step 2: Lower Bound on Bias. Now we want to show $\sup_{0\le s\le 1}|\mathrm{bias}_{n,1}(1,s)-\mathrm{bias}_{n,0}(1,s)|>0$. By Equation (SA-7.1),

$$
\begin{aligned}
\mathrm{bias}_{n,1}(1,s)-\mathrm{bias}_{n,0}(1,s) &= \mathbf{e}_1^\top\mathbf{\Psi}_s^{-1}\mathbf{S}_s-\mu_1(\mathbf{x}_s)-0=\mathbf{e}_1^\top\mathbf{\Psi}_s^{-1}\mathbf{S}_s,\\
\mathbf{\Psi}_s &= \mathbb{E}\left[\mathbf{r}_p\left(D_i(\mathbf{x}_s)\right)\mathbf{r}_p\left(D_i(\mathbf{x}_s)\right)^\top k(D_i(\mathbf{x}_s))\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_1)\right],\\
\mathbf{S}_s &= \mathbb{E}\left[\mathbf{r}_p\left(D_i(\mathbf{x}_s)\right)k(D_i(\mathbf{x}_s))\mu_1(\mathbf{X}_i)\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_1)\right].
\end{aligned}
$$

Changing to polar coordinates, we have

$$
\begin{aligned}
\mathbf{\Psi}_s &= \int_0^\infty\int_{\Theta_s(r)}^\pi \mathbf{r}_p(r)\mathbf{r}_p(r)^\top K(r)rd\theta dr,\\
\mathbf{S}_s &= \int_0^\infty\int_{\Theta_s(r)}^\pi \mathbf{r}_p(r)K(r)r\sin(\theta)rd\theta dr,
\end{aligned}
$$

with

$$\Theta_s(r) = \begin{cases} 0, & \text{if } 0 \le r \le s, \\ \arccos(s/r), & \text{if } r > s. \end{cases}$$

For notation simplicity, denote

$$\mathbf{A}(s) = \int_0^\infty \int_{\Theta_s(u)}^\pi \mathbf{r}_p(u)\mathbf{r}_p(u)^\top k(u) u d\theta du = \mathbf{A}_1(s) + \mathbf{A}_2(s),$$
$$\mathbf{B}(s) = \int_0^\infty \int_{\Theta_s(u)}^\pi \mathbf{r}_p(u) k(u) u \sin(\theta) u d\theta du = \mathbf{B}_1(s) + \mathbf{B}_2(s),$$

where

$$\begin{aligned}
\mathbf{A}_1(s) &= \int_0^s \int_0^\pi \mathbf{r}_p(u)\mathbf{r}_p(u)^\top k(u) u d\theta du = \pi \int_0^s \mathbf{r}_p(u)\mathbf{r}_p(u)^\top k(u) u du, \\
\mathbf{A}_2(s) &= \int_s^\infty \int_{\arccos(s/u)}^\pi \mathbf{r}_p(u)\mathbf{r}_p(u)^\top k(u) u d\theta du = \int_s^\infty (\pi - \arccos(s/u))\mathbf{r}_p(u)\mathbf{r}_p(u)^\top k(u) u du, \\
\mathbf{B}_1(s) &= \int_0^s \int_0^\pi \mathbf{r}_p(u) k(u) u \sin(\theta) u d\theta du = 2\int_0^s \mathbf{r}_p(u) k(u) u^2 du, \\
\mathbf{B}_2(s) &= \int_s^\infty \int_{\arccos(s/u)}^\pi \mathbf{r}_p(u) k(u) u \sin(\theta) u d\theta du = \int_s^\infty (1 + \frac{s}{u})\mathbf{r}_p(u) k(u) u^2 du.
\end{aligned}$$

Evaluating the above at zero gives

$$\mathbf{A}(0) = \frac{\pi}{2}\int_0^\infty u\mathbf{r}_p(u)\mathbf{r}_p(u)^\top k(u) du, \quad \mathbf{B}(0) = \int_0^\infty u^2 \mathbf{r}_p(u) k(u) du.$$

Hence

$$\text{bias}_{n,1}(1,0) - \text{bias}_{n,0}(1,0) = \mathbf{e}_1^\top \mathbf{A}(0)^{-1}\mathbf{B}(0) = \mathbf{e}_1^\top \mathbf{A}(0)^{-1}\left[\frac{2}{\pi}\mathbf{A}(0)\mathbf{e}_2\right] = 0. \tag{SA-7.3}$$

Taking derivatives with respect to s, we have

$$\begin{aligned}
\dot{\mathbf{A}}_1(s) &= \pi \mathbf{r}_p(s)\mathbf{r}_p(s)^\top k(s) s, \\
\dot{\mathbf{A}}_2(s) &= -\pi \mathbf{r}_p(s)\mathbf{r}_p(s)^\top k(s) s + \int_s^\infty \frac{1}{\sqrt{u^2 - s^2}} u\mathbf{r}_p(u)\mathbf{r}_p(u)^\top k(u) du, \\
\dot{\mathbf{B}}_1(s) &= 2\mathbf{r}_p(s) k(s) s^2, \\
\dot{\mathbf{B}}_2(s) &= -2\mathbf{r}_p(s) k(s) s^2 + \int_s^\infty u\mathbf{r}_p(u) k(u) du.
\end{aligned}$$

Evaluating the above at zero gives

$$\dot{\mathbf{A}}(0) = \int_0^\infty \mathbf{r}_p(u)\mathbf{r}_p(u)^\top k(u) du, \quad \dot{\mathbf{B}}(0) = \int_0^\infty u\mathbf{r}_p(u) k(u) du.$$

Using matrix calculus, we know

$$
\begin{aligned}
&\frac{d}{ds}\operatorname{bias}_{n,1}(1,s) - \operatorname{bias}_{n,0}(1,s)\Big|_{s=0} \\
&= \frac{d}{ds}\mathbf{e}_1^\top \mathbf{A}(s)^{-1}\mathbf{B}(s)\Big|_{s=0} && \text{(SA-7.4)}\\
&= -\mathbf{e}_1^\top \mathbf{A}(0)^{-1}\dot{\mathbf{A}}(0)\left[\mathbf{A}(0)^{-1}\mathbf{B}(0)\right] + \mathbf{e}_1^\top \mathbf{A}(0)^{-1}\dot{\mathbf{B}}(0) && \text{(SA-7.5)}\\
&= -\mathbf{e}_1^\top \mathbf{A}(0)^{-1}\dot{\mathbf{A}}(0)\left[\frac{2}{\pi}\mathbf{e}_2\right] + \mathbf{e}_1^\top\left[\frac{2}{\pi}\mathbf{e}_1\right] \\
&= -\frac{2}{\pi}\mathbf{e}_1^\top \mathbf{A}(0)^{-1}\int_0^\infty \begin{bmatrix} u \\ u^2 \\ \cdots \\ u^{p+1}\end{bmatrix} k(u)du + \mathbf{e}_1^\top\left[\frac{2}{\pi}\mathbf{e}_1\right] \\
&= -\frac{4}{\pi^2} + \frac{2}{\pi}. && \text{(SA-7.6)}
\end{aligned}
$$

Combining Equations (SA-7.3) and (SA-7.4), and the fact that $\frac{d}{ds}\operatorname{bias}_{n,1}(1,s) - \operatorname{bias}_{n,0}(1,s)$ is continuous in s, we can show $\sup_{0\le s\le 1}|\operatorname{bias}_{n,1}(1,s) - \operatorname{bias}_{n,0}(1,s)| > 0$. Combining with Equation (SA-7.2), we have

$$
\begin{aligned}
\inf_{0<h<1}\sup_{\mathbf{x}\in\mathscr{B}} \frac{|\mathfrak{B}_{n,1}(\mathbf{x}) - \mathfrak{B}_{n,0}(\mathbf{x})|}{h} &\geq \inf_{0<h<1}\sup_{0<s<h}\frac{|\operatorname{bias}_{n,1}(s,h) - \operatorname{bias}_{n,0}(s,h)|}{h} \\
&= \inf_{0<h<1}\sup_{0<s<h}\left|\operatorname{bias}_{n,1}\left(1,\frac{s}{h}\right)\right| \\
&> 0.
\end{aligned}
$$

■

SA-7.2 Proof of Lemma 3

The proof of part (i) follows from part (ii) with $\mathscr{B}\cap B(\mathbf{x},\varepsilon)$ as the boundary. To prove part (ii), without loss of generality, we assume that $\iota = p+1$, and want to show $\sup_{\mathbf{x}\in\mathscr{B}^o}|\mathfrak{B}_{n,t}(\mathbf{x})| \lesssim h^{p+1}$. This means we have assumed that $\mathscr{B}$ has a one-to-one curve length parametrization γ that is C^{p+3} with curve length L, there exists $\varepsilon,\delta>0$ such that for all $\mathbf{x}\in\gamma([\delta, L-\delta])$ and $0<r<\varepsilon$, $S(\mathbf{x},r)$ intersects $\mathscr{B}$ with two points, $s(\mathbf{x},r)$ and $t(\mathbf{x},r)$. Define $a(\mathbf{x},r)$ and $b(\mathbf{x},r)$ to be the number in $[0,2\pi]$ such that

$$[a(\mathbf{x},r), b(\mathbf{x},r)] = \{\theta : \mathbf{x} + r(\cos\theta,\sin\theta)\in\mathscr{A}_1\}.$$

Then, for $\mathbf{x}\in\mathscr{B}$ and $0<r<\varepsilon$, $\theta_{1,\mathbf{x}}(r)$ has the following explicit representation:

$$\theta_{1,\mathbf{x}}(r) = \frac{\int_{a(\mathbf{x},r)}^{b(\mathbf{x},r)}\mu_1(\mathbf{x}+r(\cos\theta,\sin\theta))f_X(\mathbf{x}+r(\cos\theta,\sin\theta))d\theta}{\int_{a(\mathbf{x},r)}^{b(\mathbf{x},r)} f_X(\mathbf{x}+r(\cos\theta,\sin\theta))d\theta}.$$

Step 1: Curve length v.s. Distance to $\gamma(0)$

W.l.o.g., assume $\gamma(0) = \mathbf{x}$ and $\gamma'(0) = (1, 0)$. Let $T : [0, \infty) \to [0, \infty)$ to be a continuous increasing function that satisfies

$$\|\gamma \circ T(r)\|^2 = r^2, \qquad \forall r \in [0, h].$$

Initial Case: $l = 1, 2, 3$. We will show that T is C^l on $(0, h)$. For notational simplicity, define another function $\phi : [0, \infty) \to [0, \infty)$ by $\phi(t) = \|\gamma(t)\|^2$. Using implicit derivations iteratively,

$$\begin{aligned}
&\phi \circ T(r) = r^2,\\
&\phi'(T(r))T'(r) = 2r,\\
&\phi''(T(r))(T'(r))^2 + \phi'(T(r))T''(r) = 2,\\
&\phi'''(T(r))(T'(r))^3 + 3\phi''(T(r))T'(r)T''(r) + \phi'(T(r))T'''(r) = 0.
\end{aligned} \tag{1}$$

From the above equalities, we get

$$\begin{aligned}
T'(r) &= \frac{2r}{\phi'(T(r))},\\
T''(r) &= \frac{2 - \phi''(T(r))\,(T'(r))^2}{\phi'(T(r))},\\
T'''(r) &= -\frac{\phi'''(T(r))(T'(r))^3 + 3\phi''(T(r))T'(r)T''(r)}{\phi'(T(r))}.
\end{aligned}$$

Since we have assumed γ is C^{p+3} on $(0, h)$, ϕ is also C^{p+1} on $(0, h)$. It follows from the above calculation that T is C^{p+3} on $(0, h)$. In order to find the limit of derivatives of T at 0, we need

$$\begin{aligned}
&\phi(t) = \gamma_1(t)^2 + \gamma_2(t)^2, && \phi(0) = 0,\\
&\phi'(t) = 2\gamma_1(t)\gamma_1'(t) + 2\gamma_2(t)\gamma_2'(t), && \phi'(0) = 0,\\
&\phi''(t) = 2\gamma_1'(t)\gamma_1'(t) + 2\gamma_1(t)\gamma_1''(t) + 2\gamma_2'(t)\gamma_2'(t) + 2\gamma_2(t)\gamma_2''(t), && \phi''(0) = 2,\\
&\phi'''(t) = 6\gamma_1'(t)\gamma_1''(t) + 2\gamma_1(t)\gamma_1'''(t) + 6\gamma_2'(t)\gamma_2''(t) + 2\gamma_2(t)\gamma_2'''(t).
\end{aligned}$$

Using L'Hôpital's rule

$$\begin{aligned}
\lim_{r\downarrow 0} T'(r) &= \lim_{r\downarrow 0} \frac{2}{\phi''(T(r))T'(r)} = \frac{2}{2\lim_{r\downarrow 0} T'(r)} \implies \lim_{r\downarrow 0} T'(r) = 1,\\
\lim_{r\downarrow 0} T''(r) &= \lim_{r\downarrow 0} \frac{-\phi'''(T(r))(T'(r))^3 - \phi''(T(r))2T'(r)T''(r)}{\phi''(T(r))T'(r)}\\
&= \frac{-\phi^{(3)}(0) - 4\lim_{r\downarrow 0} T''(r)}{2}\\
&= \frac{-\phi^{(3)}(0)}{6}
\end{aligned}$$

$$
\begin{aligned}
\lim_{r\downarrow 0} T^{(3)}(r) = &-\lim_{r\downarrow 0} \frac{\phi^{(4)}(T(r))(T'(r))^4 + \phi^{(3)}(T(r))3(T'(r))^2T''(r) + 3\phi^{(3)}(T(r))(T'(r))^2T''(r)}{\phi''(T(r))T'(r)} \\
&+ \lim_{r\downarrow 0} \frac{3\phi''(T(r))T'(r)T^{(3)}(r)}{\phi''(T(r))T'(r)} \\
&= -\frac{\phi^{(4)}(0) - (\phi^{(3)}(0))^2 + 6\lim_{r\downarrow 0} T^{(3)}(r)}{2} \\
&= -\frac{\phi^{(4)}(0) - (\phi^{(3)}(0))^2}{8}.
\end{aligned}
$$

Induction Step: $l \geq 4$. Assume $\lim_{r\downarrow 0} T^{(i)}(r)$ exists and is finite for $0 \leq i \leq l-2$ and there exists a function $q(r)$ such that (i) $q(r)$ is a polynomial of $\phi^{(j)}(T(r)), 1 \leq j \leq l-1$ and $T^{(k)}(r), 1 \leq k \leq l-2$, (ii) $\lim_{r\downarrow 0} q(r) = 0$ and (iii)

$$
q(r) + \phi'(T(r))T^{(l-1)}(r) = 0. \tag{2}
$$

For $l = 4$, this assumption can be verified from Equation (1). Using L'hopital's rule,

$$
\begin{aligned}
\lim_{r\downarrow 0} T^{(l-1)}(r) &= \lim_{r\downarrow 0} -\frac{q(r)}{\phi'(T(r)} \\
&\overset{L'h}{=} \lim_{r\downarrow 0} -\frac{q'(r)}{\phi''(T(r))T'(r)}.
\end{aligned}
$$

From the previous paragraph, $\lim_{r\downarrow 0} \phi''(T(r))T'(r)$ exists and is finite. And $q'(r)$ is a polynomial of $\phi^{(j)}(T(r)), 1 \leq j \leq l$ and $T^{(k)}(r), 1 \leq k \leq l-1$. Hence $\lim_{r\downarrow 0} T^{(l-1)}(r)$ can be solved from the following equation and is finite:

$$
\lim_{r\downarrow 0} q'(r) + \lim_{r\downarrow 0} \phi''(T(r))T'(r) \cdot \lim_{r\downarrow 0} T^{(l-1)}(r) = 0. \tag{3}
$$

Taking derivatives on both sides of Equation (2),

$$
q'(r) + \phi''(T(r))T'(r)T^{(l-1)}(r) + \phi'(T(r))T^{(l)}(r) = 0.
$$

Take $q_2(r) = q'(r) + \phi''(T(r))T'(r)T^{(l-1)}(r)$. Then, (i) $q_2(r)$ is a polynomial of $\phi^{(j)}(T(r)), 1 \leq j \leq l$ and $T^{(k)}(r), 1 \leq k \leq l-1$, (ii) $\lim_{r\downarrow 0} q_2(r) = 0$, and (iii)

$$
q_2(r) + \phi'(T(r))T^{(l)}(r) = 0.
$$

Continue this argument till $l = p+3$, $\lim_{r\downarrow 0} T^{(j)}(r)$ exists and is a polynomial of $\phi^{(0)}(0), \ldots, \phi^{(j+1)}(0)$, which implies that it is bounded by a constant only depending on γ.

Step 2: $(p+1)$**-times continuously differentiable** S_r

We use the notation $\gamma(t) = (\gamma_1(t), \gamma_2(t))$. Define

$$
A(t) = \angle\gamma(t) - \gamma(0), \gamma'(0) = \arcsin\left(\frac{\gamma_2(t)}{\|\gamma(t)\|}\right).
$$

Since γ is C^{p+3}, we can Taylor expand γ at 0 to get

$$\gamma(t) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} t + \begin{pmatrix} u_2 \\ v_2 \end{pmatrix} t^2 + \cdots \begin{pmatrix} u_{p+2} \\ v_{p+2} \end{pmatrix} t^{p+2} + \begin{pmatrix} R_1(t) \\ R_2(t) \end{pmatrix},$$

where we have used the fact that $\gamma_2'(0) = 0$ and $\|\gamma'(0)\| = 1$ and

$$R_1(t) = \int_0^t \frac{\gamma_1^{(p+3)}(s)(t-s)^{p+2}}{(p+2)!} ds, \qquad R_2(t) = \int_0^t \frac{\gamma_2^{(p+3)}(s)(t-s)^{p+2}}{(p+2)!} ds.$$

Since γ is C^{p+3}, $R_1(t)/t$ and $R_2(t)/t$ are C^{p+3} on $(0,\infty)$. We *claim* that $\lim_{t\downarrow 0} \frac{d^v}{dt^v}(R_1(t)/t)$ exists and is uniformly bounded for all $\mathbf{x} \in \mathscr{B}$, for all $0 \le v \le p+1$. Define $\varphi(t) = R_1(t)/t$. Then

$$\begin{aligned}
\varphi'(t) &= -\frac{R_1(t)}{t^2} + \frac{R_1'(t)}{t}, \\
\varphi''(t) &= \frac{2R_1(t)}{t^3} - \frac{2R_1'(t)}{t^2} + \frac{R_1''(t)}{t}, \\
\varphi^{(3)}(t) &= -\frac{6R_1(t)}{t^4} + \frac{6R_1'(t)}{t^3} - \frac{3R_1^{(2)}(t)}{t^2} + \frac{R_1^{(3)}(t)}{t} \qquad \cdots
\end{aligned}$$

where

$$R_1'(t) = \int_0^t \frac{\gamma_1^{(p+1)}(s)(t-s)^{p-1}}{(p-1)!} ds, \qquad R_1''(t) = \int_0^t \frac{\gamma_1^{(p+1)}(s)(t-s)^{p-2}}{(p-2)!} ds, \qquad \cdots$$

Since γ_1 is C^{p+3}, there exists $C_1 > 0$ only depending on γ such that for all $0 \le v \le p+3$,$\left|\frac{d^v}{dt^v} R_1(t)\right| \le C_1 t^{p+1-v}$. Hence

$$\lim_{r\downarrow 0} \varphi^{(j)}(r) = 0, \qquad \forall 0 \le j \le p+1.$$

Similarly, $\lim_{r\downarrow 0} \frac{d^v}{dt^v}(R_2(t)/t)$ exists and is uniformly bounded for all $0 \le v \le p+1$. Then

$$\frac{\gamma_2(t)}{\|\gamma(t)\|} = \frac{v_2 t + \cdots + v_{p+2} t^{p+2} + R_2(t)/t}{\sqrt{(1 + u_2 t + \cdots u_{p+2} t^{p+2} + R_1(t)/t)^2 + (v_2 t + \cdots + v_{p+2} t^{p+2} + R_2(t)/t)^2}}, t > 0.$$

Notice that $\gamma_2(t)/\|\gamma(t)\|$ is of the form

$$p(t)(1+q(t))^\alpha,$$

where $\alpha < 0$ and $p(t), q(t)$ are C^{p+1} on $(0,\infty)$ with $\lim_{r\downarrow 0} d^v/dt^v p(t)$ and $\lim_{r\downarrow 0} d^v/dt^v q(t)$ finite. Since the derivative of $p(t)(1+q(t))^\alpha$ is

$$p'(t)(1+q(t))^\alpha + p(t)\alpha(1+q(t))^{\alpha-1} q'(t),$$

which is the sum of two terms of the form $p_2(t)(1+q_2(t))^\alpha$ with p_2 and q_2 functions that are C^p with finite limits at 0. Continue this argument, we see that $\frac{\gamma_2(\cdot)}{\|\gamma(\cdot)\|}$ is C^{p+1} on $(0,\infty)$ and $\lim_{r\downarrow 0} \frac{d^v}{dt^v}(\gamma_2(t)/\|\gamma(t)\|)$ exist and are uniformly bounded for all $\mathbf{x} \in \mathscr{B}$ and for all $0 \le v \le p+1$.

Since arcsin is C^{p+1} with bounded (higher order derivatives) on $[-1/2, 1/2]$, A is C^{p+1} on $(0,\delta)$ and for all $0 \le v \le p+1$, $\lim_{r\downarrow 0} A^{(v)}(t)$ exist and are uniformly bounded for all $\mathbf{x} \in \mathscr{B}$.

Step 3: $(p+1)$**-times continuously differentiable conditional density**

By the previous two steps, $a(\mathbf{x}, r) = A \circ T(r)$ is C^{p+1} on $(0, \infty)$ with $|\lim_{r \downarrow 0} \frac{d^v}{dr^v} a(\mathbf{x}, r)| < \infty$. Similarly, we can show that $b(\mathbf{x}, r)$ is C^{p+1} in r with finite limits at $r = 0$. By the assumption that f_X is C^{p+1} and bounded below by $\underline{f}$, $\theta_{1,\mathbf{x}}$ is C^{p+1} with $\lim_{r \downarrow 0} \frac{d^v}{dr^v} \theta_{1,\mathbf{x}}(r)$ uniformly bounded for all $\mathbf{x} \in \mathscr{B}$ and for all $0 \leq v \leq p+1$.

This completes the proof. ■

SA-7.3 Proof of Theorem 6

Let $s > 0$ be a parameter that is chosen later. Consider the following two data generating processes.

Data Generating Process $\mathbb{P}_0$. Let $\mathscr{X} = \{r(\cos\theta, \sin\theta) : 0 \leq r \leq 1, 0 \leq \theta \leq \Theta(r)\}$, where

$$\Theta(r) = \begin{cases} \pi, & 0 \leq r < s, \\ \theta_k, & s + ks^2 \leq r < s + (k+1)s^2, 0 \leq k < K, \\ \theta_K, & s + Ks^2 \leq r < 1, \end{cases}$$

with $K = \lfloor \frac{1-s}{s^2} \rfloor$ and θ_k is the unique zero of

$$\frac{\sin(\theta)}{\theta} = \frac{(k+\frac{1}{2})s^2}{s + (k+\frac{1}{2})s^2}$$

over $\theta \in [0, \pi]$, and θ_K is the unique zero of

$$\frac{\sin(\theta)}{\theta} = \frac{Ks^2 + 1 - s}{s + Ks^2 + 1}$$

over $\theta \in [0, \pi]$. Suppose $\mathbf{X}_i$ has density f_X given by

$$f_X(r(\cos\theta, \sin\theta)) = \frac{1}{2\Theta(r)}, \qquad 0 \leq r \leq 1, 0 \leq \theta \leq \Theta(r).$$

Suppose

$$\mu_0(x_1, x_2) = \frac{1}{2} + \frac{1}{100} x_1, \qquad (x_1, x_2) \in \mathbb{R}^2.$$

Suppose $Y_i = \mathbb{1}(\eta_i \leq \mu(\mathbf{X}_i))$ where $(\eta_i : i : 1, \cdots, n)$ are i.i.d. random variables independent of $(\mathbf{X}_i : 1, \cdots, n)$. Let $\eta_0(r) = \mathbb{E}_{\mathbb{P}_0}[Y_i | \|\mathbf{X}_i - (0,0)\| = r]$, for $r \geq 0$. In particular, $\mathtt{bd}(\mathscr{X})$ has length $\pi + 2$. Hence, $\mathtt{bd}(\mathscr{X})$ is a rectifiable curve.

Data Generating Process $\mathbb{P}_1$. Let $\mathscr{X} = \{r(\cos\theta, \sin\theta) : 0 \leq r \leq 1, 0 \leq \theta \leq \pi/2\}$, $\mathbf{X}_i$ is uniformly distributed on $\mathscr{X}$, and

$$\mu_1(x_1, x_2) = \frac{1}{2} + \frac{1}{100}(x_1 - s), \qquad (x_1, x_2) \in \mathbb{R}^2.$$

Suppose $Y_i = \mathbb{1}(\eta_i \leq \mu(\mathbf{X}_i))$ where $(\eta_i : 1, \cdots, n)$ are i.i.d random variables independent to $(\mathbf{X}_i : 1, \cdots, n)$. Let $\eta_1(r) = \mathbb{E}_{\mathbb{P}_1}[Y_i | \|\mathbf{X}_i - (0,0)\| = r]$, for $r \geq 0$. In particular, $\mathtt{bd}(\mathscr{X})$ has length $\pi/2 + 2$. Hence, $\mathtt{bd}(\mathscr{X})$ is a rectifiable curve.

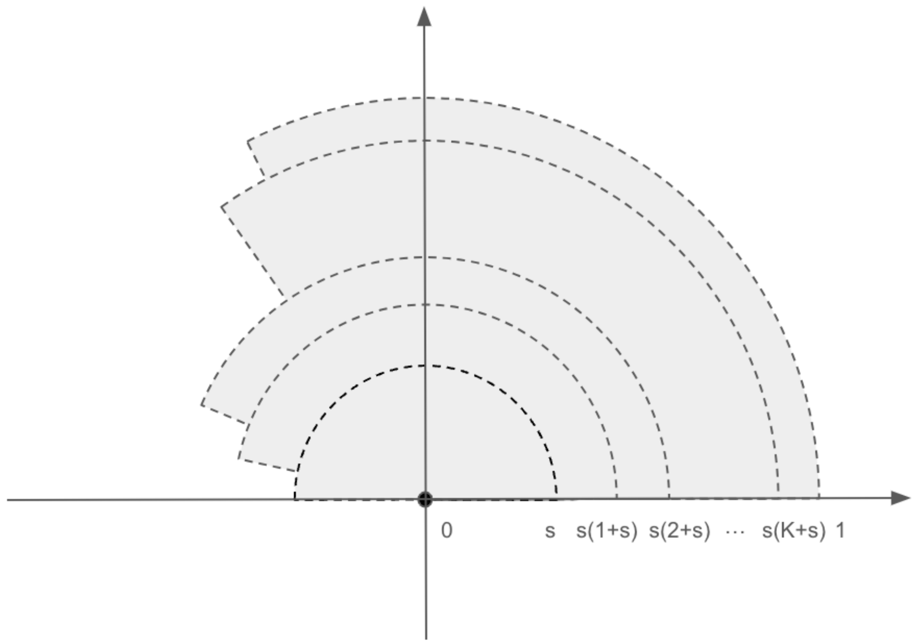

Figure SA-2. $\mathscr{X}$ from DGP $\mathbb{P}_0$

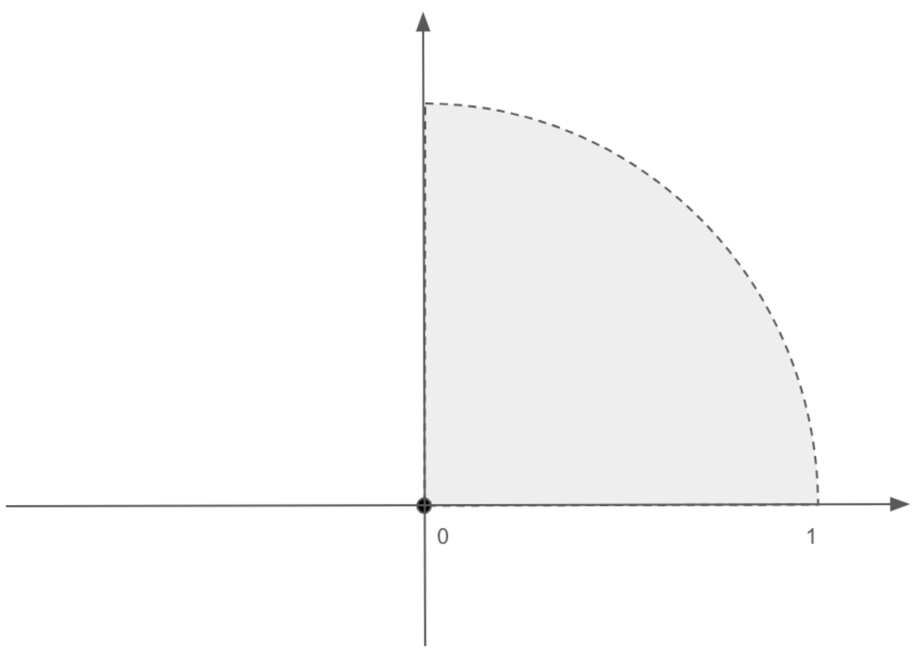

Figure SA-3. $\mathscr{X}$ from DGP $\mathbb{P}_1$

Minimax Lower Bound. First, we show under the previous two models, $\mathbb{P}_0(\|\mathbf{X}_i\| \le r) = \mathbb{P}_1(\|\mathbf{X}_i\| \le r)$ for all $r \ge 0$. Since in $\mathbb{P}_1$, $\mathbf{X}_i$ is uniform distributed on $\mathbb{R}$, we know $\mathbb{P}_1(\|\mathbf{X}_i\| \le r) = r^2$, $0 \le r \le 1$.

$$\mathbb{P}_0(\|\mathbf{X}_i\| \le r) = \int_0^r \int_0^{\Theta(s)} \frac{1}{2\Theta(s)} s d\theta ds = r^2, \quad 0 \le r \le 1.$$

Hence, choosing $(0,0)$ as the point of evaluation in both $\mathbb{P}_0$ and $\mathbb{P}_1$, we have

$$\begin{aligned}
&d_{\mathrm{KL}}(\mathbb{P}_0(\|\mathbf{X}_i - (0,0)\|, Y_i), \mathbb{P}_1(\|\mathbf{X}_i - (0,0)\|, Y_i)) \\
&= \int_0^\infty \int_{-\infty}^\infty d\mathbb{P}_0(r,y) \log \frac{d\mathbb{P}_0(r,y)}{d\mathbb{P}_1(r,y)} \\
&= \int_0^\infty \int_{-\infty}^\infty d\mathbb{P}_0(r) d\mathbb{P}_0(y|r) \log \frac{d\mathbb{P}_0(r) d\mathbb{P}_0(y|r)}{d\mathbb{P}_1(r) d\mathbb{P}_1(y|r)} \\
&= \int_0^\infty d\mathbb{P}_0(r) \int_{-\infty}^\infty d\mathbb{P}_0(y|r) \log \frac{d\mathbb{P}_0(y|r)}{d\mathbb{P}_1(y|r)} \\
&= 2\int_0^1 d_{\mathrm{KL}}(\mathsf{Bern}(\eta_0(r)), \mathsf{Bern}(\eta_1(r))) r dr.
\end{aligned}$$

Under $\mathbb{P}_0$, $\mathbf{X}_i$ is uniformly distributed on $\{r(\cos\theta, \sin\theta) : 0 \le \theta \le \Theta(r)\}$ for each $0 < r \le 1$. Hence

$$\eta_0(r) = \frac{1}{2} + \frac{1}{100}\frac{1}{\Theta(r)} \int_0^{\Theta(r)} r\cos(u) du - \frac{s}{100} = \frac{1}{2} + \frac{1}{100} r \frac{\sin(\Theta(r))}{\Theta(r)}.$$

Thus, for $0 \le k < K$,

$$\begin{aligned}
\eta_0\Big(s + (k + \frac{1}{2})s^2\Big) &= \frac{1}{2} + \frac{1}{100}\Big((s + (k+\frac{1}{2})s^2)\frac{\sin(\Theta_k)}{\Theta_k}\Big) \\
&= \frac{1}{2} + \frac{1}{100}\Big((s + (k+\frac{1}{2})s^2)\frac{(k+\frac{1}{2})s^2}{s + (k+\frac{1}{2})s^2}\Big) \\
&= \eta_1\Big(s + (k+\frac{1}{2})s^2\Big).
\end{aligned}$$

Since both η_0 and η_1 are 1-Lipschitz on all intervals $[s + ks^2, s + (k+1)s^2]$ for all $0 \le k < K$, we know $|\eta_0(r) - \eta_1(r)| \le 2s^2$ for all $r \in [s, 1]$. Moreover, $\eta_0(r) = \frac{1}{2}$ for all $0 \le r \le s$ and $\eta_1(r) = \frac{1}{2} + \frac{1}{100}(r\frac{2}{\pi} - s)$. Hence $|\eta_0(r) - \eta_1(r)| \le s$ for all $0 \le r \le s$. Hence,

$$\begin{aligned}
\int_0^1 d_{\mathrm{KL}}(\mathsf{Bern}(\eta_0(r)), \mathsf{Bern}(\eta_1(r))) r dr &\le \int_0^1 d_{\chi^2}(\mathsf{Bern}(\eta_0(r)), \mathsf{Bern}(\eta_1(r))) r dr \\
&= \int_0^1 (\eta_1(r)\Big(\frac{\eta_0(r) - \eta_1(r)}{\eta_1(r)}\Big)^2 + (1 - \eta_1(r))\Big(\frac{\eta_0(r) - \eta_1(r)}{1 - \eta_1(r)}\Big)^2) r dr \\
&\le \frac{1}{\frac{1}{2} - \frac{3}{100}} \int_0^1 (\eta_0(r) - \eta_1(r))^2 r dr \\
&\le \frac{1}{\frac{1}{2} - \frac{3}{100}} \int_0^s s^2 r dr + \frac{1}{\frac{1}{2} - \frac{3}{100}} \int_s^1 (2s^2)^2 r dr \\
&\le \frac{5}{\frac{1}{2} - \frac{3}{100}} s^4.
\end{aligned}$$

Moreover, $|\mu_0(0,0) - \mu_1(0,0)| = \frac{1}{100}s$. Hence, by Tsybakov (2008, Theorem 2.2 (iii)), take $\frac{5}{\frac{1}{2} - \frac{3}{100}} s_*^4 = \frac{\log 2}{n}$, and conclude that

$$\inf_{T_n \in \mathscr{T}} \sup_{\mathbb{P} \in \mathscr{P}} \sup_{\mathbf{x} \in \mathscr{B}(\mathbb{P})} \mathbb{E}_{\mathbb{P}}[|T_n(\mathbf{U}_n(\mathbf{x})) - \mu(\mathbf{x})|] \geq \frac{1}{1600} s_* \gtrsim n^{-\frac{1}{4}}.$$

This concludes the proof. ■

SA-8 Proofs for Section SA-4

SA-8.1 Proof of Lemma SA-4.1

We will use a truncation argument. Let $\tau_n \gtrsim 1$ be the level of truncation. For each $r \in \mathscr{R}$, define

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

and define the class $\tilde{\mathscr{R}} = \{\tilde{r} : r \in \mathscr{R}\}$. For an overview of our argument, suppose Z_n^R is a 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: Gaussian strong approximation for the truncated process — $\|R_n(g,\tilde{r}) - Z_n^R(g,\tilde{r})\|_{\mathscr{G}\times\mathscr{R}}$ Observe that $\mathtt{M}_{\tilde{\mathscr{R}},\mathscr{Y}} \lesssim \tau_n$ and $\mathtt{pTV}_{\tilde{\mathscr{R}},\mathscr{Y}} \lesssim \tau_n$, and $\tilde{\mathscr{R}}$ is a VC-type class with envelope $M_{\tilde{\mathscr{R}},\mathscr{Y}} = M_{\mathscr{R},\mathscr{Y}}\mathbb{1}(|\cdot| \leq \tau_n)$ 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 $\mathtt{v} = \tau_n$ 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

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

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 $0 < \varepsilon < 1$,

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

where sup is over all finite discrete measure on $\mathbb{R}^{d+1}$, and $M_{\tilde{\mathscr{R}},\mathscr{Y}}(y) = M_{\mathscr{R},\mathscr{Y}}(y)\mathbb{1}(|y| \leq \tau_n)$. 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 $[-\tau_n, \tau_n]$. Denote $\mathbf{X} = (\mathbf{x}_i)_{1\le i\le n}$,

$$\mathbb{E}\Big[\max_{1\le i\le n} \mathtt{M}_{\mathscr{G}}^2 (M_{\mathscr{R},\mathscr{Y}}(y_i) - M_{\tilde{\mathscr{R}},\mathscr{Y}}(y_i))^2 \Big|\mathbf{X}\Big]^{\frac{1}{2}} \lesssim \mathtt{M}_{\mathscr{G}}\mathbb{E}\Big[\big(\max_{1\le i\le n} M_{\mathscr{R},\mathscr{Y}}(y_i)\big)^2\Big|\mathbf{X}\Big]^{\frac{1}{2}} \lesssim \mathtt{M}_{\mathscr{G}} n^{\frac{1}{2+v}},$$

$$\begin{aligned}\sup_{(g,r)\in\mathscr{G}\times\mathscr{R}} \mathbb{E}[g(\mathbf{x}_i)^2 r(y_i)^2 \mathbb{1}(|y_i| \ge \tau_n^{1/\alpha})]^{\frac{1}{2}} &\lesssim \sup_{(g,r)\in\mathscr{G}\times\mathscr{R}} \mathbb{E}\left[g(\mathbf{x}_i)^2 \mathbb{E}[r(y_i)^{2+v}|\mathbf{x}_i]^{\frac{2}{2+v}} \mathbb{P}(|y_i| \ge \tau_n|\mathbf{x}_i)^{\frac{v}{2+v}}\right] \\ &\lesssim \sqrt{\mathtt{M}_{\mathscr{G}}\mathtt{E}_{\mathscr{G}}\tau_n}.\end{aligned}$$

By Jensen's inequality, we also have

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

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

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

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 $0 < \varepsilon < 1$,

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

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

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

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

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

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

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

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 τ_n such that

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

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 $\mathtt{c} = \mathtt{c}_{\mathscr{G}}\mathtt{c}_{\mathscr{R},\mathscr{Y}}\mathtt{k}$. ■

SA-8.2 Proof of Lemma SA-4.2

Since A_n is the addition of two M_n processes, indexed by $\mathscr{G} \times \mathscr{R}$ and $\mathscr{H} \times \mathscr{S}$ respectively, the Gaussian strong approximation error essentially depends on the *worst case scenario* between $\mathscr{G}$ and $\mathscr{H}$, and between $\mathscr{R}$ and $\mathscr{S}$. Hence (1) taking maximums $\mathtt{E} = \max\{\mathtt{E}_{\mathscr{G}}, \mathtt{E}_{\mathscr{H}}\}$, $\mathtt{M} = \max\{\mathtt{M}_{\mathscr{G}}, \mathtt{M}_{\mathscr{H}}\}$ and $\mathtt{TV} = \max\{\mathtt{TV}_{\mathscr{G}}, \mathtt{TV}_{\mathscr{H}}\}$; (2) noticing that A_n is still indexed by a VC-type class of functions, we can get the claimed result.

For a more rigor proof, we can not apply Cattaneo and Yu (2025, Theorem SA.1) on $(M_n(g,r) : g \in \mathscr{G}, r \in \mathscr{R})$ and $(M_n(h,s) : h \in \mathscr{H}, s \in \mathscr{S})$ directly, since this ignores the dependence structure between the two empirical processes. However, we can still project the functions onto a Haar basis, and control the *strong approximation error for projected process* and the *projection error* as in the proof of Cattaneo and Yu (2025, Theorem SA.1) and show both errors can be controlled via *worst case scenario* between $\mathscr{G}$ and $\mathscr{H}$, and between $\mathscr{R}$ and $\mathscr{S}$.

Reductions: Here we present some reductions to our problem. By the same argument as in Section SA-II.3 (Proofs of Theorem 1) in the supplemental appendix of Cattaneo and Yu (2025), we can show there exists $\mathbf{u}_i, 1 \leq i \leq n$ i.i.d $\mathsf{Uniform}([0,1]^d)$ on a possibly enlarged probability space, such that

$$f(\mathbf{x}_i) = f(\phi_{\mathscr{G}\cup\mathscr{H}}^{-1}(\mathbf{u}_i)), \qquad \forall f \in \mathscr{G}\cup\mathscr{H}, \forall 1 \leq i \leq n.$$

With the help of Cattaneo and Yu (2025, Lemma SA.10), we can assume w.l.o.g. that $\mathbf{x}_i$'s are i.i.d $\mathsf{Uniform}(\mathcal{X})$ with $\mathcal{X} = [0,1]^d$, and $\phi_{\mathscr{G}\cup\mathscr{H}} : [0,1]^d \to [0,1]^d$ is the identity function. Although we assume $\sup_{\mathbf{x}\in\mathcal{X}} \mathbb{E}[|Y_i|^{2+v}|\mathbf{X}_i = \mathbf{x}] < \infty$, we first present the result under the assumption $\sup_{\mathbf{x}\in\mathcal{X}} \mathbb{E}[\exp(|y_i|)|\mathbf{x}_i = \mathbf{x}] \leq 2$, which is the same as in Cattaneo and Yu (2025, Theorem 2). Also in correspondence to the notations in Cattaneo and Yu (2025, Theorem 2), we set $\alpha = 1$ throughout this proof.

Cell Constructions and Projections: The constructions here are the same as those in Cattaneo and Yu (2025), and we present them here for completeness. Let $\mathscr{A}_{M,N}(\mathbb{P},1) = \{\mathscr{C}_{j,k} : 0 \leq k < 2^{M+N-j}, 0 \leq j \leq M+N\}$ be an axis-aligned cylindered quasi-dyadic expansion of $\mathbb{R}^{d+1}$, with depth M for the main subspace $\mathbb{R}^d$ and depth N for the multiplier subspace $\mathbb{R}$, with respect to $\mathbb{P}$, the joint distribution of $(\mathbf{x}_i, y_i)$ taking values in $\mathbb{R}^d \times \mathbb{R}$, as in Cattaneo and Yu (2025, Definition SA.4). To see what $\mathscr{A}_{M,N}(\mathbb{P},1)$ is, it can be given by the following iterative partition procedure:

1. *Initialization ($q = 0$)*: Take $\mathscr{C}_{M+N-q,0} = \mathcal{X} \times \mathbb{R}$ where $\mathcal{X} = [0,1]^d$.

2. *Iteration ($q = 1, \ldots, M$)*: Given $\mathscr{C}_{K-l,k}$ for $0 \leq l \leq q-1, 0 \leq k < 2^l$, take $s = (q \mod d) + 1$, and construct $\mathscr{C}_{K-q,2k} = \mathscr{C}_{K-q+1,k} \cap \{(\mathbf{x}, y) \in [0,1]^d \times \mathbb{R} : \mathbf{e}_s^\top \mathbf{x} \leq c_{K-q+1,k}\}$ and $\mathscr{C}_{K-q,2k+1} = \mathscr{C}_{K-q+1,k} \cap \{(\mathbf{x}, y) \in [0,1]^d \times \mathbb{R} : \mathbf{e}_s^\top \mathbf{x} > c_{K-j+1,k}\}$ such that $\mathbb{P}(\mathscr{C}_{K-q,2k})/\mathbb{P}(\mathscr{C}_{K-q+1,k}) \in [\frac{1}{1+\rho}, \frac{\rho}{1+\rho}]$ for all $0 \leq k < 2^{q-1}$. Continue until $(\mathscr{C}_{N,k} : 0 \leq k < 2^M)$ has been constructed. By construction, for each $0 \leq l < M$, $\mathscr{C}_{N,l} = \mathcal{X}_{0,l} \times \mathcal{Y}_{0,N,0}$, with $\mathcal{Y}_{0,N,0} = \mathbb{R}$.

3. *Iteration ($q = M+1, \cdots, M+N$)*: Given $\mathscr{C}_{K-l,k}$ for $0 \leq l \leq q-1, 0 \leq k < 2^l$, each $\mathscr{C}_{M+N-q,k}$ can be written as $\mathcal{X}_{0,l} \times \mathcal{Y}_{l,M+N-q,m}$ with $k = 2^{q-M}l + m$. Construct $\mathscr{C}_{M+N-q-1,2k} = \mathcal{X}_{0,l} \times \mathcal{Y}_{l,M+N-q-1,2m}$ and $\mathscr{C}_{M+N-q-1,2k+1} = \mathcal{X}_{0,l} \times \mathcal{Y}_{l,M+N-q-1,2m+1}$, such that there exists some $\mathfrak{c}_{M+N-q,k} \in \mathbb{R}$ with $\mathcal{Y}_{l,M+N-q-1,2m} = \mathcal{Y}_{l,M+N-q,m} \cap (-\infty, \mathfrak{c}_{M+N-q,k})$ and $\mathcal{Y}_{l,M+N-q-1,2m+1} = \mathcal{Y}_{l,M+N-q,m} \cap (\mathfrak{c}_{M+N-q,k}, \infty)$, $\mathbb{P}(y_i \in \mathcal{Y}_{l,M+N-q-1,2m}|\mathbf{x}_i \in \mathcal{X}_{0,l}) = \mathbb{P}(y_i \in \mathcal{Y}_{l,M+N-q-1,2m+1}|\mathbf{x}_i \in \mathcal{X}_{0,l}) = \frac{1}{2}\mathbb{P}(y_i \in \mathcal{Y}_{l,M+N-q-1,m}|\mathbf{x}_i \in \mathcal{X}_{0,l})$.

Consider the projection $\Pi_1(\mathscr{A}_{M,n}(\mathbb{P},1))$ given in Equation (SA-7) in Cattaneo and Yu (2025), noticing that $\mathscr{A}_{M,N}(\mathbb{P},1)$ is one special instance of $\mathscr{C}_{M,N}(\mathbb{P},\rho)$. That is, define $e_{j,k} = \mathbb{1}_{\mathscr{C}_{j,k}}$ and $\widetilde{e}_{j,k} = e_{j-1,2k} - e_{j-1,2k+1}$,

$$\Pi_1(\mathscr{C}_{M,N}(\mathbb{P},\rho))[g,r] = \gamma_{M+N,0}(g,r)e_{M+N,0} + \sum_{1\leq j\leq M+N} \sum_{0\leq k<2^{M+N-j}} \widetilde{\gamma}_{j,k}(g,r)\widetilde{e}_{j,k}, \tag{SA-8.1}$$

where $e_{j,k} = \mathbb{1}(\mathscr{C}_{j,k})$ and $\widetilde{e}_{j,k} = \mathbb{1}(\mathscr{C}_{j-1,2k}) - \mathbb{1}(\mathscr{C}_{j-1,2k+1})$, and

$$\gamma_{j,k}(g,r) = \begin{cases} \mathbb{E}[g(X)r(Y)|X \in \mathcal{X}_{j-N,k}], & \text{if } N \leq j \leq M+N, \\ \mathbb{E}[g(X)|X \in \mathcal{X}_{0,l}] \cdot \mathbb{E}[r(Y)|X \in \mathcal{X}_{0,l}, Y \in \mathcal{Y}_{l,0,m}], & \text{if } j < N, k = 2^{N-j}l + m, \end{cases}$$

and $\widetilde{\gamma}_{j,k}(g,r) = \gamma_{j-1,2k}(g,r) - \gamma_{j-1,2k+1}(g,r)$. We will use Π_1 as a shorthand for $\Pi_1(\mathscr{C}_{M,N}(\mathbb{P},\rho))$.

For simplicity, we denote $\Pi_1(\mathscr{A}_{M,n}(\mathbb{P},1))$ by Π_1 instead. Now define the projected empirical process

$$\Pi_1 A_n(g,h,r,s) = \Pi_1 M_n(g,r) + \Pi_1 M_n(h,s), \qquad g \in \mathscr{G}, h \in \mathscr{H}, r \in \mathscr{R}, s \in \mathscr{S},$$

where $\Pi_1 M_n(g,r)$ and $\Pi_1 M_n(h,s)$ are given in Equation (SA-10) in Cattaneo and Yu (2025), that is,

$$\Pi_1 M_n(g,r) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big(\Pi_1[g,r](\mathbf{x}_i, y_i) - \mathbb{E}[\Pi_1[g,r](\mathbf{x}_i, y_i)]\big),$$
$$\Pi_1 M_n(h,s) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big(\Pi_1[h,s](\mathbf{x}_i, y_i) - \mathbb{E}[\Pi_1[h,s](\mathbf{x}_i, y_i)]\big).$$

Construction of Gaussian Process Suppose $(\widetilde{\xi}_{j,k} : 0 \leq k < 2^{M+N-j}, 1 \leq j \leq M+N)$ are i.i.d. standard Gaussian random variables. Take $F_{(j,k),m}$ to be the cumulative distribution function of $(S_{j,k} - mp_{j,k})/\sqrt{mp_{j,k}(1-p_{j,k})}$, where $p_{j,k} = \mathbb{P}(\mathscr{C}_{j-1,2k})/\mathbb{P}(\mathscr{C}_{j,k})$ and $S_{j,k}$ is a $\mathrm{Bin}(m, p_{j,k})$ random variable, and $G_{(j,k),m}(t) = \sup\{x : F_{(j,k),m}(x) \leq t\}$. We define $U_{j,k}, \widetilde{U}_{j,k}$'s via the following iterative scheme:

1. *Initialization:* Take $U_{M+N,0} = n$.

2. *Iteration:* Suppose we've defined $U_{l,k}$ for $j < l \leq M+N, 0 \leq k < 2^{M+N-l}$, then solve for $U_{j,k}$'s s.t.

$$\begin{aligned}
\widetilde{U}_{j,k} &= \sqrt{U_{j,k} p_{j,k}(1-p_{j,k})}\, G_{(j,k),U_{j,k}} \circ \Phi(\widetilde{\xi}_{j,k}), \\
\widetilde{U}_{j,k} &= (1-p_{j,k}) U_{j-1,2k} - p_{j,k} U_{j-1,2k+1} = U_{j-1,2k} - p_{j,k} U_{j,k}, \\
U_{j-1,2k} &+ U_{j-1,2k+1} = U_{j,k}, \quad 0 \leq k < 2^{M+N-j}.
\end{aligned}$$

Continue till we have defined $U_{0,k}$ for $0 \leq k < 2^{M+N}$.

Then, $\{U_{j,k} : 0 \leq j \leq K, 0 \leq k < 2^{M+N-j}\}$ have the same joint distribution as $\{\sum_{i=1}^{n} e_{j,k}(\mathbf{x}_i, y_i) : 0 \leq j \leq K, 0 \leq k < 2^{M+N-j}\}$. By Vorob'ev–Berkes–Philipp theorem (Dudley, 2014, Theorem 1.31), $\{\widetilde{\xi}_{j,k} : 0 \leq k < 2^{M+N-j}, 1 \leq j \leq M+N\}$ can be constructed on a possibly enlarged probability space such that the previously constructed $U_{j,k}$ satisfies $U_{j,k} = \sum_{i=1}^{n} e_{j,k}(\mathbf{x}_i)$ almost surely for all $0 \leq j \leq M+N, 0 \leq k < 2^{M+N-j}$. We will show $\widetilde{\xi}_{j,k}$'s can be given as a Brownian bridge indexed by $\widetilde{e}_{j,k}$'s.

Since all of $\mathscr{G}$, $\mathscr{H}$, $\mathscr{R}$ and $\mathscr{S}$ are VC-type, we can show $\mathscr{G} \times \mathscr{H} + \mathscr{R} \times \mathscr{S}$ is also VC-type, here $+$ is the Minkowski sum. Hence $\mathscr{F} = \mathscr{G} \times \mathscr{H} + \mathscr{R} \times \mathscr{S} \cup \Pi_1[G \times \mathscr{H} + \mathscr{R} \times \mathscr{S}]$ is pre-Gaussian.

Then, by Skorohod Embedding lemma (Dudley, 2014, Lemma 3.35), on a possibly enlarged probability space, we can construct a Brownian bridge $(Z_n(f) : f \in \mathscr{F})$ that satisfies

$$\widetilde{\xi}_{j,k} = \frac{\mathbb{P}(\mathscr{C}_{j,k})}{\sqrt{\mathbb{P}(\mathscr{C}_{j-1,2k})\mathbb{P}(\mathscr{C}_{j-1,2k+1})}} Z_n(\widetilde{e}_{j,k}),$$

for $0 \leq k < 2^{M+N-j}, 1 \leq j \leq M+N$. Moreover, call

$$V_{j,k} = \sqrt{n} Z_n(e_{j,k}), \qquad \widetilde{V}_{j,k} = \sqrt{n} Z_n(\widetilde{e}_{j,k}), \qquad \widetilde{\xi}_{j,k} = \frac{\mathbb{P}(\mathscr{C}_{j,k})}{\sqrt{n\mathbb{P}(\mathscr{C}_{j-1,2k})\mathbb{P}(\mathscr{C}_{j-1,2k+1})}} \widetilde{V}_{j,k}.$$

for $0 \leq k < 2^{K-j}, 1 \leq j \leq K$. We have for $g \in \mathscr{G}, h \in \mathscr{H}, r \in \mathscr{R}, s \in \mathscr{S}$,

$$\sqrt{n}\Pi_1 A_n(g,h,r,s) = \sum_{j=1}^{M+N} \sum_{0 \leq k < 2^{M+N-j}} (\widetilde{\gamma}_{j,k}[g,r] + \widetilde{\gamma}_{j,k}[h,s]) \widetilde{U}_{j,k},$$
$$\sqrt{n}\Pi_1 Z_n(g,h,r,s) = \sum_{j=1}^{M+N} \sum_{0 \leq k < 2^{M+N-j}} (\widetilde{\gamma}_{j,k}[g,r] + \widetilde{\gamma}_{j,k}[h,s]) \widetilde{V}_{j,k}.$$

Decomposition Fix one $(g,h,r,s) \in \mathscr{G} \times \mathscr{H} \times \mathscr{R} \times \mathscr{S}$, we decompose by

$$\begin{aligned} & A_n(g,h,r,s) - Z_n(g,h,r,s) \\ = \; & \underbrace{\Pi_1 A_n(g,h,r,s) - \Pi_1 Z_n(g,h,r,s)}_{\text{strong approximation (SA) error for projected}} + \underbrace{A_n(g,h,r,s) - \Pi_1 A_n(g,h,r,s) + \Pi_1 Z_n(g,h,r,s) - Z_n(g,h,r,s)}_{\text{projection error}}. \end{aligned}$$

SA error for Projected Process The strong approximation error essentially depends on the Hilbertian pseudo norm

$$\sum_{j=1}^{M+N} \sum_{0 \le k < 2^{M+N-j}} (\widetilde{\gamma}_{j,k}[g,r] + \widetilde{\gamma}_{j,k}[h,s])^2 \le 2 \sum_{j=1}^{M+N} \sum_{0 \le k < 2^{M+N-j}} (\widetilde{\gamma}_{j,k}[g,r])^2 + 2 \sum_{j=1}^{M+N} \sum_{0 \le k < 2^{M+N-j}} (\widetilde{\gamma}_{j,k}[h,s])^2.$$

Hence, Cattaneo and Yu (2025, Lemma SA.19) gives with probability at least $1 - 2e^{-t}$,

$$|\Pi_1 A_n(g,h,r,s) - \Pi_1 Z_n(g,h,r,s)| \le C_1 C_\alpha \sqrt{\frac{N^{2\alpha+1} 2^M \mathtt{EM}}{n}} t + C_1 C_\alpha \sqrt{\frac{(\|\Pi_1[g,r]\|_\infty + \|\Pi_1[h,s]\|_\infty)^2 (M+N)}{n}} t,$$

where $C_1 > 0$ is a universal constant and $C_\alpha = 1 + (2\alpha)^{\alpha/2}$.

Projection Error For the projection error, we use the simple observation that

$$|A_n(g,h,r,s) - \Pi_1 A_n(g,h,r,s)| \le |M_n(g,r) - \Pi_1 M_n(g,r)| + |M_n(h,s) - \Pi_1 M_n(h,s)|,$$

and Cattaneo and Yu (2025, Lemma SA.23) to get for all $t > N$,

$$\begin{aligned} & \mathbb{P}\Big[|A_n(g,h,r,s) - \Pi_1 A_n(g,h,r,s)| > C_2 \sqrt{C_{2\alpha}} \sqrt{N^2 \mathtt{V} + 2^{-N} \mathtt{M}^2} t^{\alpha + \frac{1}{2}} + C_2 C_\alpha \frac{\mathtt{M}}{\sqrt{n}} t^{\alpha+1}\Big] \le 4ne^{-t}, \\ & \mathbb{P}\Big[|Z_n(g,h,r,s) - \Pi_1 Z_n(g,h,r,s)| > C_2 \sqrt{C_{2\alpha}} \sqrt{N^2 \mathtt{V} + C_2 C_\alpha 2^{-N} \mathtt{M}^2} t^{\frac{1}{2}} + C_2 C_\alpha \frac{\mathtt{M}}{\sqrt{n}} t\Big] \le 4ne^{-t}, \end{aligned}$$

where $C_\alpha = 1 + (2\alpha)^{\frac{\alpha}{2}}$ and $C_{2\alpha} = 1 + (4\alpha)^\alpha$ and C_2 is a constant that only depends on the distribution of $(\mathbf{x}_1, y_1)$, with

$$\mathtt{V} = \min\{2\mathtt{M}, \sqrt{d}\mathtt{L} 2^{-M/d}\} 2^{-M/d} \mathtt{TV}_{\mathscr{H}}.$$

Uniform SA Error: Since all of $\mathscr{G}$, $\mathscr{H}$, $\mathscr{R}$ and $\mathscr{S}$ are VC-type class, from a union bound argument and the same control over fluctuation error as in Cattaneo and Yu (2025, Lemma SA.18), denoting $\mathscr{F} = \mathscr{G} \times \mathscr{H} \times \mathscr{R} \times \mathscr{S}$, we get for all $t > 0$ and $0 < \delta < 1$,

$$\mathbb{P}\big[\|A_n - A_n \circ \pi_{\mathscr{F}_\delta}\|_{\mathscr{F}} + \|Z_n - Z_n \circ \pi_{\mathscr{F}_\delta}\|_{\mathscr{F}} > C_1 C_\alpha \mathtt{F}_n(t,\delta)\big] \le \exp(-t),$$

where $C_\alpha = 1 + (2\alpha)^{\frac{\alpha}{2}}$ and

$$\mathtt{F}_n(t,\delta) = J(\delta)\mathtt{M} + \frac{(\log n)^{\alpha/2} \mathtt{M} J^2(\delta)}{\delta^2 \sqrt{n}} + \frac{\mathtt{M}}{\sqrt{n}} t + (\log n)^\alpha \frac{\mathtt{M}}{\sqrt{n}} t^\alpha.$$

where

$$\begin{aligned}J(\delta) &= 3\delta\big(\sqrt{\mathtt{d}_{\mathscr{G}}\log(\frac{2\mathtt{c}_{\mathscr{G}}}{\delta})} + \sqrt{\mathtt{d}_{\mathscr{H}}\log(\frac{2\mathtt{c}_{\mathscr{H}}}{\delta})} + \sqrt{\mathtt{d}_{\mathscr{R}}\log(\frac{2\mathtt{c}_{\mathscr{R}}}{\delta})} + \sqrt{\mathtt{d}_{\mathscr{S}}\log(\frac{2\mathtt{c}_{\mathscr{S}}}{\delta})}\big)\\ &\lesssim \sqrt{\mathtt{d}\log(\mathtt{c}/\delta)},\end{aligned}$$

recalling $\mathtt{c} = \mathtt{c}_{\mathscr{G},\mathscr{Q}_{\mathscr{G}\cup\mathscr{H}}} + \mathtt{c}_{\mathscr{H},\mathscr{Q}_{\mathscr{G}\cup\mathscr{H}}} + \mathtt{c}_{\mathscr{R},\mathscr{Y}} + \mathtt{c}_{\mathscr{S},\mathscr{Y}} + \mathtt{k}$, $\mathtt{d} = \mathtt{d}_{\mathscr{G},\mathscr{Q}_{\mathscr{G}\cup\mathscr{H}}}\mathtt{d}_{\mathscr{H},\mathscr{Q}_{\mathscr{G}\cup\mathscr{H}}}\mathtt{d}_{\mathscr{R},\mathscr{Y}}\mathtt{d}_{\mathscr{S},\mathscr{Y}}\mathtt{k}$. Choosing the optimal M^*, N^* gives $\mathbb{P}\big[\|A_n - Z_n^A\|_{\mathscr{F}} > C_1\mathtt{v}\mathsf{T}_n(t)\big] \leq C_2 e^{-t}$ for all $t > 0$, where

$$\mathsf{T}_n(t) = \min_{\delta\in(0,1)}\{\mathsf{A}_n(t,\delta) + \mathsf{F}_n(t,\delta)\},$$

with

$$\begin{aligned}\mathsf{A}_n(t,\delta) &= \sqrt{d}\min\Big\{\Big(\frac{\mathtt{c}_1^d\mathtt{ETV}^d\mathtt{M}^{d+1}}{n}\Big)^{\frac{1}{2(d+1)}}, \Big(\frac{\mathtt{c}_1^d\mathtt{c}_2^d\mathtt{E}^2\mathtt{M}^2\mathtt{TV}^d\mathtt{L}^d}{n^2}\Big)^{\frac{1}{2(d+2)}}\Big\}(t+\log(n\mathtt{N}(\delta)N^*))^{\alpha+1}\\ &\quad + \sqrt{\frac{\mathtt{M}^2(M^*+N^*)}{n}}(\log n)^\alpha(t+\log(n\mathtt{N}(\delta)N^*))^{\alpha+1},\\ \mathsf{F}_n(t,\delta) &= J(\delta)\mathtt{M} + \frac{(\log n)^{\alpha/2}\mathtt{M}J^2(\delta)}{\delta^2\sqrt{n}} + \frac{\mathtt{M}}{\sqrt{n}}\sqrt{t} + (\log n)^\alpha\frac{\mathtt{M}}{\sqrt{n}}t^\alpha,\end{aligned}$$

where

$$\begin{aligned}\mathscr{V}_{\mathscr{R}} &= \{\theta(\cdot,r) : r\in\mathscr{R}\},\\ \mathtt{N}(\delta) &= \mathtt{N}_{\mathscr{G},\mathscr{Q}_{\mathscr{G}\cup\mathscr{H}}}(\delta/2,\mathtt{M}_{\mathscr{G},\mathscr{Q}_{\mathscr{G}\cup\mathscr{H}}})\mathtt{N}_{\mathscr{H},\mathscr{Q}_{\mathscr{G}\cup\mathscr{H}}}(\delta/2,\mathtt{M}_{\mathscr{H},\mathscr{Q}_{\mathscr{G}\cup\mathscr{H}}})\mathtt{N}_{\mathscr{R},\mathscr{Y}}(\delta/2,M_{\mathscr{R}})\mathtt{N}_{\mathscr{S},\mathscr{Y}}(\delta/2,M_{\mathscr{S},\mathscr{Y}}),\\ J(\delta) &= 2J_{\mathscr{Q}_{\mathscr{G}\cup\mathscr{H}}}(\mathscr{G},\mathtt{M}_{\mathscr{G},\mathscr{Q}_{\mathscr{G}\cup\mathscr{H}}},\delta/2) + 2J_{\mathscr{Q}_{\mathscr{G}\cup\mathscr{H}}}(\mathscr{H},\mathtt{M}_{\mathscr{H},\mathscr{Q}_{\mathscr{G}\cup\mathscr{H}}},\delta/2) + 2J_{\mathscr{Y}}(\mathscr{R},M_{\mathscr{R},\mathscr{Y}},\delta/2) + 2J_{\mathscr{Y}}(\mathscr{S},M_{\mathscr{S},\mathscr{Y}},\delta/2),\\ M^* &= \Big\lfloor\log_2\min\Big\{\Big(\frac{\mathtt{c}_1 n\mathtt{TV}}{\mathtt{E}}\Big)^{\frac{d}{d+1}}, \Big(\frac{\mathtt{c}_1\mathtt{c}_2 n\mathtt{LTV}}{\mathtt{EM}}\Big)^{\frac{d}{d+2}}\Big\}\Big\rfloor,\\ N^* &= \Big\lceil\log_2\max\Big\{\Big(\frac{n\mathtt{M}^{d+1}}{\mathtt{c}_1^d\mathtt{ETV}^d}\Big)^{\frac{1}{d+1}}, \Big(\frac{n^2\mathtt{M}^{2d+2}}{\mathtt{c}_1^d\mathtt{c}_2^d\mathtt{TV}^d\mathtt{L}^d\mathtt{E}^2}\Big)^{\frac{1}{d+2}}\Big\}\Big\rceil.\end{aligned}$$

Truncation Argument for y_i's with Finite Moments The above result is derived under the assumption that $\sup_{\mathbf{x}\in\mathscr{X}}\mathbb{E}[\exp(|y_i|)|\mathbf{x}_i = \mathbf{x}] < \infty$. For the result under the condition $\sup_{\mathbf{x}\in\mathscr{X}}\mathbb{E}[|y_i|^{2+v}|\mathbf{x}_i = \mathbf{x}] < \infty$, we can use the same truncation argument as in Section SA-8.1 (proof of Lemma SA-4.1) and the VC-type conditions for $\mathscr{G},\mathscr{H},\mathscr{R},\mathscr{S}$ to get the stated conclusions. ■

References

Cattaneo, M. D., Chandak, R., Jansson, M., and Ma, X. (2024), "Boundary Adaptive Local Polynomial Conditional Density Estimators," *Bernoulli*, 30, 3193–3223.

Cattaneo, M. D., and Yu, R. R. (2025), "Strong Approximations for Empirical Processes Indexed by Lipschitz Functions," *Annals of Statistics*.

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

——— (2014b), "Gaussian Approximation of Suprema of Empirical Processes," *Annals of Statistics*, 42, 1564–1597.

Chernozhuokov, V., Chetverikov, D., Kato, K., and Koike, Y. (2022), "Improved central limit theorem and bootstrap approximations in high dimensions," *Annals of Statistics*, 50, 2562–2586.

Dudley, R. M. (2014), *Uniform central limit theorems*, Vol. 142, Cambridge university press.

Federer, H. (2014), *Geometric measure theory*, Springer.

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

Giné, E., and Nickl, R. (2016), *Mathematical Foundations of Infinite-dimensional Statistical Models*, New York: Cambridge University Press.

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

Tsybakov, A. (2008), *Introduction to Nonparametric Estimation*, Springer.

van der Vaart, A. W., and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, Springer.