Causal Treatment Effect Aggregation in Boundary Discontinuity Designs
Supplemental Appendix

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

July 4, 2025

Abstract

This supplemental appendix contains theoretical proofs and other technical results.

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

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

Contents

SA-1 Preliminary Technical Result

This section is self-contained, and provides the basis for the main analysis in the paper. We employ basic concepts and definitions from geometric measure theory. See Federer [2014] and Folland [2002] for comprehensive reviews; a more accessible source is available at https://encyclopediaofmath.org/.

Let $\mathfrak{m}$ be the Lebesgue measure, and $\mathfrak{H}^k$ be the k-dimensional Hausdorff measure. For $f : \mathbb{R}^k \to \mathbb{R}$, let $J_1 f$ denote the Jacobian of f. As in the paper, we assume throughout that $\mathscr{d} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a metric on $\mathbb{R}^d$, and define

$$d(\mathbf{x}, \mathscr{S}) = \inf_{\mathbf{s} \in \mathscr{S}} \mathscr{d}(\mathbf{x}, \mathbf{s}).$$

Assumption SA–1 (Distance)**.**

The following technical lemma gives sufficient conditions for convergence of integrals over a shrinking tubular of a manifold. Let

Lemma SA-1 (Integration over Shrinking Tubular Neighborhoods)**.** *Suppose that $\mathscr{X}$ is a compact subset of $\mathbb{R}^d$ with Hausdorff dimension d, $\mathscr{S} \subseteq \mathscr{X}$ has Hausdorff dimension $d-1$, and the following conditions hold:*

1. *There exists positive constants C_u and C_l such that $C_l \|\mathbf{x}_1 - \mathbf{x}_2\| \leq \mathscr{d}(\mathbf{x}_1, \mathbf{x}_2) \leq C_u \|\mathbf{x}_1 - \mathbf{x}_2\|$ for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathscr{X}$.*
2. *$g : \mathbb{R} \to \mathbb{R}$ is $\mathfrak{m}$-measurable, continuous over its compact support.*
3. *$m : \mathbb{R}^d \to \mathbb{R}$ is $\mathfrak{m}$-measurable.*
4. *$J_1 d(\cdot, \mathscr{S}) \neq 0$ $\mathfrak{m}$-almost everywhere on $\mathscr{X}$, and $\int_{\mathscr{X}} |\frac{m(\mathbf{x})}{J_1 d(\mathbf{x}, \mathscr{S})}| d\mathbf{x} < \infty$.*
5. *$\lim_{\epsilon \downarrow 0} M(\epsilon) = \mathsf{c}_{\mathscr{S}} M(0)$ is finite, where $M(r) = \int_{d(\mathbf{x}, \mathscr{S}) = r} \frac{m(\mathbf{x})}{J_1 d(\mathbf{x}, \mathscr{S})} d\mathfrak{H}^{d-1}(\mathbf{x})$ for all $r \geq 0$*

Then,

$$\lim_{\epsilon \downarrow 0} \int_{\mathscr{T}(\epsilon)} \frac{1}{\epsilon} g\Big(\frac{d(\mathbf{x}, \mathscr{S})}{\epsilon}\Big) m(\mathbf{x}) d\mathbf{x} = \mathsf{c}_{\mathscr{S}} \int_0^1 g(s) ds \cdot \int_{\mathscr{S}} \frac{m(\mathbf{x})}{J_1 d(\mathbf{x}, \mathscr{S})} d\mathfrak{H}^{d-1}(\mathbf{x}).$$

where

$$\mathscr{T}(\epsilon) = \Big\{ \mathbf{x} \in \mathscr{X} : d(\mathbf{x}, \mathscr{S}) \leq \epsilon \Big\}, \qquad d(\mathbf{x}, \mathscr{S}) = \inf_{\mathbf{y} \in \mathscr{S}} \mathscr{d}(\mathbf{x}, \mathbf{y}).$$

Proof. Note that $d(\cdot, \mathscr{S})$ is C_u-Lipschitz. The level sets

$$\mathscr{L}(\epsilon) = \Big\{ \mathbf{x} \in \mathscr{X} : d(\mathbf{x}, \mathscr{S}) = \epsilon \Big\}, \qquad \epsilon \in [0, \infty)$$

are $(d-1)$-dimensional rectifiable sets. Since g is continuous on a compact support, g is also bounded over $\mathscr{X}$. Hence, the function

$$\mathbf{x} \mapsto \epsilon^{-1} g(\frac{d(\mathbf{x}, \mathscr{S})}{\epsilon}) m(\mathbf{x}) J_1 d(\mathbf{x}, \mathscr{S})^{-1}$$

is $\mathfrak{m}$-summable over $\mathscr{X}$. Using the Coarea formula,

$$\begin{aligned}
\int_{\mathscr{T}(\epsilon)} \frac{1}{\epsilon} g\Big(\frac{d(\mathbf{x}, \mathscr{S})}{\epsilon}\Big) m(\mathbf{x}) d\mathbf{x} &= \int_{\mathscr{T}(\epsilon)} \frac{1}{\epsilon} g\Big(\frac{d(\mathbf{x}, \mathscr{S})}{\epsilon}\Big) m(\mathbf{x}) \frac{1}{J_1 d(\mathbf{x}, \mathscr{S})} J_1 d(\mathbf{x}, \mathscr{S}) d\mathbf{x} \\
&= \int_0^{\epsilon} \int_{d(\mathbf{x}, \mathscr{S}) = s} \frac{1}{\epsilon} g\Big(\frac{s}{\epsilon}\Big) \frac{m(\mathbf{x})}{J_1 d(\mathbf{x}, \mathscr{S})} d\mathfrak{H}^{d-1}(\mathbf{x}) ds \\
&= \int_0^{\epsilon} \frac{1}{\epsilon} g\Big(\frac{s}{\epsilon}\Big) M(s) ds \\
&= \int_0^1 g(u) M(\epsilon u) du.
\end{aligned}$$

Since g is integrable,

$$\lim_{\epsilon \to 0} \int_0^1 g(u) M(\epsilon u) du = \mathsf{c}_{\mathscr{S}} \int_0^1 g(s) ds \cdot M(0).$$

This gives the conclusion. $\square$

SA-2 Setup

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

We retain the same notation, definitions and assumptions given in the paper, whenever considering a general dimension $d \geq 2$ does not causes confusion. In particular, the natural generalization of Assumption 1 is the following.

Assumption SA–2 (Data Generating Process)**.** *Let $t \in \{0, 1\}$, $p > 0$, and $v \geq 2$.*

(i) $(Y_1(t), \mathbf{X}_1)^\top, \ldots, (Y_n(t), \mathbf{X}_n)^\top$ *are independent and identically distributed random vectors.*

(ii) *The distribution of $\mathbf{X}_i$ has a Lebesgue density $f(\mathbf{x})$ that is continuous and bounded away from zero on its support compact support $\mathscr{X} \subseteq \mathbb{R}^d$.*

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

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

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

In addition, we employ the following standard notation through the supplemental appendix.

(i) *Sample Averages.* $\mathbb{E}_n[g(\mathbf{v}_i)] = \frac{1}{n} \sum_{i=1}^n g(\mathbf{v}_i)$.

(ii) *Norms.* For a vector $\mathbf{v} \in \mathbb{R}^k$, $\|\mathbf{v}\| = (\sum_{i=1}^k \mathbf{v}_i^2)^{1/2}$ and $\|\mathbf{v}\|_\infty = \max_{1 \leq i \leq k} |\mathbf{v}_i|$. For a matrix $A \in \mathbb{R}^{m \times n}$, $\|A\|_p = \sup_{\|\mathbf{x}\|_p = 1} \|A\mathbf{x}\|_p$, $p \in \mathbb{N} \cup \{\infty\}$.

(iii) *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 $\mathfrak{p}_d = \frac{(d+p)!}{d!p!}$-dimensional vector such that $\mathbf{e}_{1+\boldsymbol{\nu}}^\top \mathbf{R}_p(\mathbf{u}) = \mathbf{u}^{\boldsymbol{\nu}}$ for all $\mathbf{u} \in \mathbb{R}^d$.

(iv) *Bounds and Asymptotics.* For reals sequences $a_n = o(b_n)$ if $\limsup \frac{a_n}{b_n} = 0$, and $a_n \lesssim b_n$ or $a_n = O(b_n)$ 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 $\operatorname{plim}_{n \to \infty} \frac{A_n}{B_n} = 0$, and $A_n \lesssim_{\mathbb{P}} B_n$ or $A_n = O_{\mathbb{P}}(B_n)$ if $\limsup_{M \to \infty} \limsup_{n \to \infty} \mathbb{P}[|\frac{A_n}{B_n}| \geq M] = 0$.

For background textbook references see van der Vaart and Wellner [1996] and Giné and Nickl [2016]. For textbook references on geometric measure theory, see Federer [2014] and Folland [2002].

SA-3 Pre-Aggregation Approach

We consider the more general local polynomial estimator described in Remark 1, with $d \geq 2$. Recall the estimator is

$$\widehat{\tau}_{\text{pre}} = \mathbf{e}_1^\top \widehat{\boldsymbol{\xi}}_1 - \mathbf{e}_1^\top \widehat{\boldsymbol{\xi}}_0$$

where

$$\widehat{\boldsymbol{\xi}}_t = \operatorname*{arg\,min}_{\boldsymbol{\xi} \in \mathbb{R}^{p+1}} \sum_{i=1}^n \big(Y_i - \mathbf{r}_p(D_i)^\top \boldsymbol{\xi}\big)^2 k_h(D_i) \mathbb{1}(T_i = t),$$

for $t \in \{0, 1\}$, and with $\mathbf{e}_1$ the conformable first unit vector, $\mathbf{r}_p(u) = (1, u, u^2, \cdots, u^p)^\top$ the usual univariate polynomial basis, $k_h(u) = k(u/h)/h$, $k(\cdot)$ a univariate kernel function, and h the bandwidth.

In addition, define

$$\widehat{\theta}_t(r) = \mathbf{r}_p(r)^\top \widehat{\boldsymbol{\xi}}_t, \qquad r \in \mathbb{R},$$

for $t \in \{0, 1\}$. The induced distance is

$$D_i = (2T_i - 1) d(\mathbf{X}_i, \mathscr{B}), \qquad d(\mathbf{x}, \mathscr{B}) = \inf_{\mathbf{b} \in \mathscr{B}} \mathscr{d}(\mathbf{x}, \mathbf{b}),$$

where recall that $\mathbf{X}_i \in \mathbb{R}^d$ and $\mathscr{B} \subset \mathbb{R}^{d-1}$. Thus, the induced conditional expectation based on distance to $\mathscr{B}$ is

$$\theta_t(r) = \mathbb{E}[Y_i | D_i = r] = \mathbb{E}[Y_i | d(\mathbf{X}_i, \mathscr{B}) = |r|, \mathbf{X}_i \in \mathscr{A}_t],$$

for $t \in \{0,1\}$, and where $r \in \mathscr{I}_t$ with $\mathscr{I}_0 = (-\infty, 0)$ and $\mathscr{I}_1 = [0, \infty)$.

The following assumption is the natural generalization of Assumption 2 in the paper for $d \geq 2$.

Assumption SA–3 (Integral Representation)**.** *Recall the density of $\mathbf{X}_i$, f, from Assumption 1. Suppose the following conditions holds.*

(i) *There exists positive constants C_u and C_l such that $C_l \|\mathbf{x}_1 - \mathbf{x}_2\| \leq \mathscr{d}(\mathbf{x}_1, \mathbf{x}_2) \leq C_u \|\mathbf{x}_1 - \mathbf{x}_2\|$ for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathscr{X}$.*

(ii) *$J_1 d(\cdot, \mathscr{B}) \neq 0$ $\mathfrak{m}$-almost everywhere on $\mathscr{X}$, and $\int_{\mathscr{X}} |\frac{f(\mathbf{x})}{J_1 d(\mathbf{x}, \mathscr{B})}| d\mathbf{x} < \infty$.*

(iii) *$\lim_{\epsilon \downarrow 0} F(s) = F(0)$ is finite, where $F(s) = \int_{d(\mathbf{x}, \mathscr{B}) = s} \frac{f(\mathbf{x})}{J_1 d(\mathbf{x}, \mathscr{B})} d\mathfrak{H}^{d-1}(\mathbf{x})$ for all $s \geq 0$.*

(iv) *θ_t is continuous at 0 for $t = 0, 1$.*

For each $t \in \{0,1\}$, the best mean square approximation based on $\mathbf{r}_p(D_i)$ is

$$\theta_t^*(D_i) = \mathbf{r}_p(D_i)^\top \boldsymbol{\gamma}_t^*$$

with

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

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

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

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

$$\begin{aligned}
\widehat{\boldsymbol{\Psi}}_t &= \mathbb{E}_n \left[\mathbf{r}_p\left(\frac{D_i}{h}\right) \mathbf{r}_p\left(\frac{D_i}{h}\right)^\top k_h(D_i) \mathbb{1}_{\mathscr{I}_t}(D_i) \right], \\
\boldsymbol{\Psi}_t &= \mathbb{E} \left[\mathbf{r}_p\left(\frac{D_i}{h}\right) \mathbf{r}_p\left(\frac{D_i}{h}\right)^\top k_h(D_i) \mathbb{1}_{\mathscr{I}_t}(D_i) \right], \\
\mathbf{O}_t &= \mathbb{E}_n \left[\mathbf{r}_p\left(\frac{D_i}{h}\right) k_h(D_i)(Y_i - \theta_t^*(D_i)) \mathbb{1}_{\mathscr{I}_t}(D_i) \right],
\end{aligned}$$

and the misspecification bias is

$$\mathfrak{B}_{n,t} = \theta_t^*(0) - \theta_t(0). \tag{SA-2}$$

In the text, $\mathfrak{B}_n = \mathfrak{B}_{n,1} - \mathfrak{B}_{n,0}$. In addition, define

$$\begin{aligned}
\widehat{\boldsymbol{\Upsilon}}_t &= h\mathbb{E}_n\bigg[\mathbf{r}_p\left(\frac{D_i}{h}\right)\mathbf{r}_p\left(\frac{D_i}{h}\right)^\top k_h\left(D_i\right)^2\left(Y_i - \widehat{\theta}_t(D_i)\right)^2 \mathbb{1}_{\mathcal{I}_t}(D_i)\bigg],\\
\boldsymbol{\Upsilon}_t &= h\mathbb{E}\bigg[\mathbf{r}_p\Big(\frac{D_i}{h}\Big)\mathbf{r}_p\Big(\frac{D_i}{h}\Big)^\top k_h(D_i)^2(Y_i - \theta_t^*(D_i))^2\mathbb{1}_{\mathcal{I}_t}(D_i)\bigg],\\
\widehat{\Xi}_t &= \frac{1}{nh}\mathbf{e}_1^\top\widehat{\boldsymbol{\Psi}}_t^{-1}\widehat{\boldsymbol{\Upsilon}}_t\widehat{\boldsymbol{\Psi}}_t^{-1}\mathbf{e}_1, \qquad \widehat{\Xi} = \widehat{\Xi}_0 + \widehat{\Xi}_{1,},\\
\Xi_t &= \frac{1}{nh}\mathbf{e}_1^\top\boldsymbol{\Psi}_t^{-1}\boldsymbol{\Upsilon}_t\boldsymbol{\Psi}_t^{-1}\mathbf{e}_1, \qquad \Xi = \Xi_0 + \Xi_1,
\end{aligned}$$

and

$$\rho_t(r) = \mathbb{E}[(Y_i - \theta_t^*(D_i))^2|D_i = r] = \mathbb{E}[Y_i^2|D_i = r] - 2\theta_t(r)\mathbf{r}_p(r)^\top\boldsymbol{\gamma}_t^* + (\mathbf{r}_p(r)^\top\boldsymbol{\gamma}_t^*)^2.$$

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

SA-3.1 Preliminary Lemmas

Lemma SA-2 (Expected Gram Matrix)**.** *Suppose Assumptions SA–2 and SA–3 hold. If $h \to 0$, then*

$$\left\|\boldsymbol{\Psi}_t - \int_{\mathscr{B}}\frac{f(\mathbf{x})}{J_1 d(\mathbf{x},\mathscr{B})}d\mathbf{x}\cdot\boldsymbol{\Gamma}\right\| = o(1), \qquad \boldsymbol{\Gamma} = \int_0^1 \mathbf{r}_p(u)\mathbf{r}_p(u)^\top k(u)du,$$

for $t \in \{0,1\}$. This implies $\|\boldsymbol{\Psi}_t\| \lesssim 1$ and $\left\|\boldsymbol{\Psi}_t^{-1}\right\| \lesssim 1$.

Proof. The typical entry of $\boldsymbol{\Psi}_t$ has the form

$$\psi_{n,t} = \mathbb{E}[(D_i/h)^v k_h(D_i)\mathbb{1}_{\mathcal{I}_t}(D_i)],$$

for some $v \in \mathbb{N}$. By Lemma SA-1, as $h \to 0$,

$$\psi_{n,t} = \int_{\mathscr{R}_t(h)}\Big(\frac{d(\mathbf{x},\mathscr{B})}{h}\Big)^v\frac{1}{h}k\Big(\frac{d(\mathbf{x},\mathscr{B})}{h}\Big)f(\mathbf{x})d\mathbf{x} = \int_0^1 s^v k(s)ds\cdot\int_{\mathscr{B}}\frac{f(\mathbf{x})}{J_1 d(\mathbf{x},\mathscr{B})}d\mathbf{x} + o(1).$$

Let $\mathfrak{p}_p = \frac{(d+p)!}{d!p!}$. Since ψ_t is $\mathfrak{p}_p \times \mathfrak{p}_p$ dimensional, the conclusion follows from a union bound. By Weyl's theorem, $|\lambda_{\min}(\boldsymbol{\Psi}_t) - \lambda_{\min}(\boldsymbol{\Gamma})| = o(1)$ and $|\lambda_{\max}(\boldsymbol{\Psi}_t) - \lambda_{\max}(\boldsymbol{\Gamma})| = o(1)$. Hence $\|\boldsymbol{\Psi}_t\| \lesssim 1$ and $\left\|\boldsymbol{\Psi}_t^{-1}\right\| \lesssim 1$. □

Lemma SA-3 (Estimated Gram Matrix)**.** *Suppose Assumptions SA–2 and SA–3 hold. If $h \to 0$ and $nh \to \infty$, then*

$$\|\widehat{\boldsymbol{\Psi}}_t - \boldsymbol{\Psi}_t\| \lesssim_{\mathbb{P}} (nh)^{-1/2},$$

for $t \in \{0,1\}$. This implies $\|\widehat{\boldsymbol{\Psi}}_t\| \lesssim_{\mathbb{P}} 1$ and $\|\widehat{\boldsymbol{\Psi}}_t^{-1}\| \lesssim_{\mathbb{P}} 1$.

Proof. The typical entry of $\widehat{\boldsymbol{\Psi}}_t$ takes the form

$$\psi_{n,t} = \frac{1}{n}\sum_{i=1}^{n}\Big(\frac{D_i}{h}\Big)^v \frac{1}{h}k\Big(\frac{D_i}{h}\Big)\mathbb{1}_{\mathscr{I}_t}(D_i).$$

By Lemma SA-1,

$$\mathbb{E}\Big[\Big(\frac{D_i}{h}\Big)^{2v}\frac{1}{h}k\Big(\frac{D_i}{h}\Big)^2\mathbb{1}_{\mathscr{I}_t}(D_i)\Big] = \int_0^1 s^{2v}k(s)^2 ds \cdot \int_{\mathscr{B}} \frac{f(\mathbf{x})}{J_1 d(\mathbf{x},\mathscr{B})}d\mathbf{x}.$$

Hence,

$$\begin{aligned}
\mathbb{V}[\psi_{n,t}] &= \frac{1}{nh^2}\mathbb{V}\Big[\Big(\frac{D_i}{h}\Big)^v k\Big(\frac{D_i}{h}\Big)\mathbb{1}_{\mathscr{I}_t}(D_i)\Big] \\
&= \frac{1}{nh}\mathbb{E}\Big[\Big(\frac{D_i}{h}\Big)^{2v}\frac{1}{h}k\Big(\frac{D_i}{h}\Big)^2\mathbb{1}_{\mathscr{I}_t}(D_i)\Big] - \frac{1}{n}\mathbb{E}\Big[\Big(\frac{D_i}{h}\Big)^v\frac{1}{h}k\Big(\frac{D_i}{h}\Big)\mathbb{1}_{\mathscr{I}_t}(D_i)\Big]^2 \\
&= \frac{1}{nh}\int_0^1 s^{2v}k(s)^2 ds \cdot \int_{\mathscr{B}}\frac{f(\mathbf{x})}{J_1 d(\mathbf{x},\mathscr{B})}d\mathbf{x}(1+o(1)) + \frac{1}{n}\int_0^1 s^v k(s)ds \cdot \int_{\mathscr{B}}\frac{f(\mathbf{x})}{J_1 d(\mathbf{x},\mathscr{B})}d\mathbf{x}(1+o(1)) \\
&= \frac{1}{nh}\int_0^1 s^{2v}k(s)^2 ds \cdot \int_{\mathscr{B}}\frac{f(\mathbf{x})}{J_1 d(\mathbf{x},\mathscr{B})}d\mathbf{x}(1+o(1)).
\end{aligned}$$

Since $\boldsymbol{\Psi}_t$ is finite dimensional, $\|\widehat{\boldsymbol{\Psi}}_t - \boldsymbol{\Psi}_t\| \lesssim_{\mathbb{P}} (nh)^{-1/2}$. The conclusion then follows from Lemma SA-3 and Weyl's theorem. □

Lemma SA-4 (Stochastic Linear Approximation)**.** *Suppose Assumptions SA–2 and SA–3 hold. If $nh \to \infty$, then*

$$\begin{aligned}
\|\mathbf{O}_t\| &\lesssim_{\mathbb{P}} (nh)^{-1/2}, \\
\Big|\mathbf{e}_1^\top \boldsymbol{\Psi}_t^{-1}\mathbf{O}_t\Big| &\lesssim_{\mathbb{P}} (nh)^{-1/2}, \\
\Big|\mathbf{e}_1^\top (\widehat{\boldsymbol{\Psi}}_t^{-1} - \boldsymbol{\Psi}_t^{-1})\mathbf{O}_t\Big| &\lesssim_{\mathbb{P}} (nh)^{-1},
\end{aligned}$$

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

Proof. By the definition of the best linear approximation θ_t^*, we have $\mathbb{E}[\mathbf{O}_t] = 0$. A typical entry of $\mathbf{O}_t$ has the form

$$o_t = \mathbb{E}_n\left[\left(\frac{D_i}{h}\right)^v k_h(D_i)(Y_i - \theta_t^*(D_i))\mathbb{1}_{\mathscr{I}_t}(D_i)\right]$$

for $v \in \mathbb{N}$. Since $\sup_{r\in\mathbb{R}} \mathbb{1}(k(r/h) > 0)\rho_t(r) \lesssim 1$, by Lemma SA-1,

$$\begin{aligned}
\mathbb{V}[o_t] &= \frac{1}{nh}\mathbb{E}\left[\left(\frac{D_i}{h}\right)^{2v}\frac{1}{h}k\left(\frac{D_i}{h}\right)^2\rho_t(D_i)\mathbb{1}_{\mathscr{I}_t}(D_i)\right] \\
&\lesssim \frac{1}{nh}\mathbb{E}\left[\left(\frac{D_i}{h}\right)^{2v}\frac{1}{h}k\left(\frac{D_i}{h}\right)^2\mathbb{1}_{\mathscr{I}_t}(D_i)\right]
\end{aligned}$$

$$\lesssim \frac{1}{nh} \int_0^1 s^{2v} k(s)^2 ds.$$

Since $\mathbf{O}_t$ is finite dimensional, $\|O_t\| \lesssim (nh)^{-1/2}$. The other results follow from Lemma SA-2 and Lemma SA-3. □

Lemma SA-5 (Expected Meat Matrix)**.** *Suppose Assumptions SA–2 and SA–3 hold. If* $h \to 0$*, then*

$$\left\| \boldsymbol{\Upsilon}_t - \iota_t \int_{\mathscr{B}} \frac{f(\mathbf{x})}{J_1 d(\mathbf{x}, \mathscr{B})} d\mathbf{x} \cdot \boldsymbol{\Sigma} \right\| = o(1), \qquad \boldsymbol{\Sigma} = \int_0^1 \mathbf{r}_p(u) \mathbf{r}_p(u)^\top k(u)^2 du,$$

and

$$\iota_t = \int_{\mathscr{B}} \sigma_t^2(\mathbf{b}) \frac{f(\mathbf{b})}{J_1 d(\mathbf{b}, \mathscr{B})} d\mathfrak{H}^{d-1}(\mathbf{b}) + \int_{\mathscr{B}} \Big(\mu_t(\mathbf{b}) - \frac{\int_{\mathscr{B}} \mu_t(\mathbf{x}) \frac{f(\mathbf{x})}{J_1 d(\mathbf{x},\mathscr{B})} d\mathbf{x}}{\int_{\mathscr{B}} \frac{f(\mathbf{u})}{J_1 d(\mathbf{u},\mathscr{B})} d\mathbf{u}} \Big)^2 \frac{f(\mathbf{b})}{J_1 d(\mathbf{b}, \mathscr{B})} d\mathfrak{H}^{d-1}(\mathbf{b}),$$

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

Proof. Consider $s_t(\mathbf{x}) = \mathbb{E}[Y_i(t)^2 | \mathbf{X}_i = \mathbf{x}] = \mu_t(\mathbf{x})^2 + \sigma_t(\mathbf{x})^2$, $\mathbf{x} \in \mathscr{X}$. A typical element of $\boldsymbol{\Upsilon}_t$ has the form

$$\begin{aligned}
\upsilon_t &= h\mathbb{E}\Big[\Big(\frac{D_i}{h} \Big)^v k_h(D_i)^2 (Y_i - \theta_t^*(D_i))^2 \mathbb{1}_{\mathcal{I}_t}(D_i) \Big] \\
&= \int_{\mathscr{R}_t(h)} \Big(\frac{d(\mathbf{x}, \mathscr{B})}{h} \Big)^v \frac{1}{h} k\Big(\frac{d(\mathbf{x}, \mathscr{B})}{h} \Big)^2 (s_t(\mathbf{x}) - 2\mu_t(\mathbf{x}) \theta_t^*(d(\mathbf{x}, \mathscr{B})) + \theta_t^*(d(\mathbf{x}, \mathscr{B}))^2) f(\mathbf{x}) d\mathbf{x} \\
&= \int_0^1 s^v k(s)^2 ds \Big[\int_{\mathscr{B}} \frac{s_t(\mathbf{x}) f(\mathbf{x})}{J_1 d(\mathbf{x}, \mathscr{B})} d\mathbf{x} - 2 \int_{\mathscr{B}} \frac{\mu_t(\mathbf{x}) f(\mathbf{x})}{J_1 d(\mathbf{x}, \mathscr{B})} d\mathbf{x} \cdot \theta_t^*(0) + \int_{\mathscr{B}} \frac{f(\mathbf{x})}{J_1 d(\mathbf{x}, \mathscr{B})} d\mathbf{x} \theta_t^*(0)^2 \Big] + o(1),
\end{aligned}$$

where in the last line we have used Lemma SA-1, and $\theta_t^*(r) = \mathbf{R}_p(r)^\top \boldsymbol{\gamma}_t^*$ is continuous in r. Moreover,

$$\theta_t^*(0) = \mathbf{e}_1^\top \boldsymbol{\Psi}_t^{-1} \mathbb{E}\Big[\mathbf{r}_p \Big(\frac{D_i}{h} \Big) k_h(D_i) Y_i \mathbb{1}_{\mathcal{I}_t}(D_i) \Big].$$

Using Lemma SA-1 again, we have

$$\begin{aligned}
\mathbb{E}\Big[\mathbf{r}_p \Big(\frac{D_i}{h} \Big) k_h(D_i) Y_i \mathbb{1}_{\mathcal{I}_t}(D_i) \Big] &= \int_{R_t(h)} \mathbf{r}_p \Big(\frac{d(\mathbf{x}, \mathscr{B})}{h} \Big) k_h(d(\mathbf{x}, \mathscr{B})) \mu_t(\mathbf{x}) \mathbb{1}_{\mathcal{I}_t}(D_i) d\mathfrak{H}^{d-1}(\mathbf{x}) \\
&= \boldsymbol{\Lambda} \int_{\mathscr{B}} \frac{\mu_t(\mathbf{x}) f(\mathbf{x})}{J_1 d(\mathbf{x}, \mathscr{B})} d\mathfrak{H}^{d-1}(\mathbf{x}),
\end{aligned}$$

where $\boldsymbol{\Lambda} = \int_0^1 \mathbf{r}_p(u) k(u) du$. Together with Lemma SA-2, we have

$$\theta_t^*(0) = \mathbf{e}_1^\top \boldsymbol{\Gamma}^{-1} \boldsymbol{\Lambda} \frac{\int_{\mathscr{B}} \mu_t(\mathbf{x}) f(\mathbf{x}) (J_1 d(\mathbf{x}, \mathscr{B}))^{-1} d\mathfrak{H}^{d-1}(\mathbf{x})}{\int_{\mathscr{B}} f(\mathbf{x}) (J_1 d(\mathbf{x}, \mathscr{B}))^{-1} d\mathfrak{H}^{d-1}(\mathbf{x})},$$

where the definition of $\boldsymbol{\Gamma}$ from Lemma SA-2 implies $\mathbf{e}_1^\top \boldsymbol{\Gamma}^{-1}\boldsymbol{\Lambda} = 1$. Finite dimensionality and the union bound over the entries give the result. $\square$

Lemma SA-6 (Covariance)**.** *Suppose Assumptions SA–2 and SA–3 hold. If $h \to 0$ and $nh \to \infty$, then*

$$\left\|\widehat{\boldsymbol{\Upsilon}}_t - \boldsymbol{\Upsilon}_t\right\| \lesssim_{\mathbb{P}} (nh)^{-1/2},$$
$$nh|\widehat{\Xi}_t - \Xi_t| \lesssim_{\mathbb{P}} (nh)^{-1/2},$$

for $t \in \{0,1\}$. This implies $1 \lesssim_{\mathbb{P}} \lambda_{\min}(\widehat{\boldsymbol{\Upsilon}}_t) \lesssim_{\mathbb{P}} 1$ and $(nh)^{-1} \lesssim_{\mathbb{P}} \widehat{\Xi}_t \lesssim_{\mathbb{P}} (nh)^{-1}$.

Proof. Denote $\eta_{i,t} = Y_i - \theta_t^*(D_i)$ and $\xi_{i,t} = \theta_t^*(D_i) - \widehat{\theta}_t(D_i)$. Then

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

and we decompose the error into

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

$k_h(D_i) \neq 0$ implies $\|\mathbf{r}_p(D_i/h)\|_2 \lesssim 1$. Define

$$\mathbf{U}_t = \mathbb{E}_n\left[\mathbf{r}_p\left(\frac{D_i}{h}\right)k_h(D_i)\theta_t^*(D_i)\mathbb{1}_{\mathscr{I}_t}(D_i)\right],$$
$$\mathbf{H} = \operatorname{diag}(1, h, \cdots, h^p).$$

Hence, by Lemma SA-3 and SA-4,

$$\begin{aligned}
&\max_{t\in\{0,1\}}\max_{1\le i\le n}|\xi_{i,t}|\\
&= \max_{t\in\{0,1\}}\max_{1\le i\le n}|\mathbf{r}_p(D_i)^\top(\widehat{\boldsymbol{\gamma}}_t - \boldsymbol{\gamma}_t^*)|\mathbb{1}(k_h(D_i)\ge 0)\\
&= \max_{t\in\{0,1\}}\max_{1\le i\le n}|\mathbf{r}_p(D_i)^\top\mathbf{H}^{-1}(\widehat{\boldsymbol{\Psi}}_t^{-1}\mathbf{O}_t + (\widehat{\boldsymbol{\Psi}}_t^{-1} - \boldsymbol{\Psi}_t^{-1})\mathbf{U}_t)|\mathbb{1}(k_h(D_i)\ge 0)\\
&\le \max_{t\in\{0,1\}}\left\|\widehat{\boldsymbol{\Psi}}_t^{-1}\mathbf{O}_t\right\|_2 + \max_{t\in\{0,1\}}\left\|(\widehat{\boldsymbol{\Psi}}_t^{-1} - \boldsymbol{\Psi}_t^{-1})\mathbf{U}_t\right\|_2
\end{aligned}$$

$$\lesssim_{\mathbb{P}} (nh)^{-1/2},$$

Assume $nh \to \infty$, similar maximal inequality as in the proof of Lemma SA-3 shows

$$\max_{t=0,1} \|\Delta_{1,t}\| \lesssim_{\mathbb{P}} \max_{t\in\{0,1\}} \max_{1\leq i\leq n} |\xi_{i,t}|^2 \lesssim_{\mathbb{P}} (nh)^{-1},$$
$$\max_{t=0,1} \|\Delta_{2,t}\| \lesssim_{\mathbb{P}} \max_{t\in\{0,1\}} \max_{1\leq i\leq n} |\xi_{i,t}| \lesssim_{\mathbb{P}} (nh)^{-1/2}. \tag{SA-3}$$

For $\Delta_{3,t}$, notice that a typical entry of has the form

$$\mathfrak{g}_n - \mathbb{E}[\mathfrak{g}_n],$$

where

$$\mathfrak{g}_n = \frac{1}{n}\sum_{i=1}^n \Big(\frac{D_i}{h}\Big)^v \frac{1}{h} k\Big(\frac{D_i}{h}\Big)^2 (Y_i - \theta_t^*(D_i))^2 \mathbb{1}_{\mathscr{I}_t}(D_i).$$

Since we have assumed $\sup_{\mathbf{x}\in\mathcal{X}} \mathbb{E}[Y_t^4|\mathbf{X}_i = \mathbf{x}] < \infty$, and by Jensen's inequality $\mathbb{E}[\theta_t^*(D_i)^4] < \infty$, a similar argument as for Lemma SA-3 implies

$$\mathbb{V}[\mathfrak{g}_n] \lesssim (nh)^{-1},$$

and hence

$$\max_{t=0,1} \|\Delta_{3,t}\| \lesssim_{\mathbb{P}} (nh)^{-1/2}. \tag{SA-4}$$

Putting together Equation (SA-3) and (SA-4), we get $\|\widehat{\boldsymbol{\Upsilon}}_t - \boldsymbol{\Upsilon}_t\| \lesssim_{\mathbb{P}} (nh)^{-1/2}$. By Lemma SA-5 and Weyl's theorem, we have $1 \lesssim_{\mathbb{P}} \lambda_{\min}(\widehat{\boldsymbol{\Upsilon}}_t) \lesssim_{\mathbb{P}} 1$. Using Lemma SA-3 in addition, we can show

$$nh|\widehat{\Xi}_t - \Xi_t| \lesssim_{\mathbb{P}} |\mathbf{e}_1^\top \widehat{\boldsymbol{\Psi}}_t^{-1} \widehat{\boldsymbol{\Upsilon}}_t \widehat{\boldsymbol{\Psi}}_t^{-1} \mathbf{e}_1 - \mathbf{e}_1^\top \boldsymbol{\Psi}_t^{-1} \boldsymbol{\Upsilon}_t \boldsymbol{\Psi}_t^{-1} \mathbf{e}_1| \lesssim_{\mathbb{P}} (nh)^{-1/2}.$$

Lemma SA-2 and SA-5 imply $(nh)^{-1} \lesssim \Xi_t \lesssim (nh)^{-1}$, hence $(nh)^{-1} \lesssim_{\mathbb{P}} \widehat{\Xi}_t \lesssim_{\mathbb{P}} (nh)^{-1}$. □

SA-3.2 Mean Square Error Approximation

Theorem SA-1 (MSE Expansion)**.** *Suppose Assumptions SA-2 and SA-3 hold. If $h \to 0$ and $nh \to \infty$, then*

$$\mathbb{E}\Big[\Big(\sum_{t=0,1} (-1)^{t+1}(\mathbf{e}_1^\top \boldsymbol{\Psi}_t^{-1} \mathbf{O}_t + \theta_t^*(0) - \theta_t(0))\Big)^2\Big] = \mathfrak{B}_n^2 + \Xi + O_{\mathbb{P}}((nh)^{-3/2}).$$

Proof. By definition of $\mathbf{\Upsilon}_t$ we have,

$$\begin{aligned}\mathbb{E}[\mathbf{O}_t\mathbf{O}_t^\top] &= \frac{1}{nh}\mathbb{E}\left[\mathbf{r}_p\left(\frac{D_i}{h}\right)\mathbf{r}_p\left(\frac{D_i}{h}\right)^\top hk_h(D_i)^2(Y_i-\theta_t^*(D_i))^2\mathbb{1}_{\mathcal{I}_t}(D_i)\right]\\ &= \frac{1}{nh}\mathbf{\Upsilon}_t.\end{aligned}$$

Hence

$$\begin{aligned}&\mathbb{E}\Big[\Big(\sum_{t=0,1}(-1)^{t+1}(\mathbf{e}_1^\top\mathbf{\Psi}_t^{-1}\mathbf{O}_t+(\theta_t^*(0)-\theta_t(0)))\Big)^2\Big]\\ &=2\mathbb{E}\Big[\Big(\sum_{t=0,1}(-1)^{t+1}(\mathbf{e}_1^\top\mathbf{\Psi}_t^{-1}\mathbf{O}_t)\Big)\Big(\sum_{t=0,1}(-1)^{t+1}(\theta_t^*(0)-\theta_t(0))\Big)\Big]\\ &\quad+\mathbb{E}\Big[\Big(\sum_{t=0,1}(-1)^{t+1}(\mathbf{e}_1^\top\mathbf{\Psi}_t^{-1}\mathbf{O}_t)\Big)^2\Big]+\Big(\sum_{t=0,1}(-1)^{t+1}(\theta_t^*(0)-\theta_t(0))\Big)^2\\ &=\Xi_0+\Xi_1+\Big(\sum_{t=0,1}(-1)^{t+1}(\theta_t^*(0)-\theta_t(0))\Big)^2,\end{aligned}$$

where in the third line we have used $\mathbb{E}[\mathbf{O}_t]=\mathbf{0}$ for $t=0,1$, and the independence between $\mathbf{O}_0$ and $\mathbf{O}_1$. The conclusion follows. $\square$

Corollary SA-1 (Convergence Rate)**.** *Suppose Assumption 1, SA–3 hold, and $nh\to\infty$. Then*

$$|\widehat{\tau}_{\mathrm{pre}}-\tau_f|\lesssim_{\mathbb{P}}(nh)^{-1/2}+|\mathfrak{B}_n|.$$

Proof. The conclusion follows from Lemma SA-2, SA-5 and Theorem SA-1. $\square$

SA-3.3 Central Limit Theorem

The feasible t-statistics by

$$\widehat{\mathrm{T}}_{\mathrm{pre}}=\frac{\widehat{\tau}_{\mathrm{pre}}-\tau_f}{\sqrt{\widehat{\Xi}}}.$$

Theorem SA-2 (Asymptotic Normality)**.** *Suppose Assumptions SA–2 and SA–3 hold. If $nh\to\infty$ and $\sqrt{nh}|\mathfrak{B}_n|\to0$, then*

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

Proof. Define the stochastic linearization of $\widehat{\mathrm{T}}_{\mathrm{pre}}$ to be $\overline{\mathrm{T}}_{\mathrm{pre}}=\Xi^{-1/2}\mathbf{e}_1^\top\mathbf{\Psi}_t^{-1}\mathbf{O}_t$. First, we bound the stochastic linearization error. Using the decomposition (SA-1) and convergence of $\widehat{\Xi}$,

$$\begin{aligned}
\widehat{\mathrm{T}}_{\mathrm{pre}} - \overline{\mathrm{T}}_{\mathrm{pre}} &= \widehat{\Xi}^{-1/2}\Bigg(\sum_{t\in\{0,1\}} (-1)^{\frac{t+1}{2}}(\widehat{\theta}_t(0)-\theta_t(0))\Bigg) - \Xi^{-1/2}\Bigg(\sum_{t\in\{0,1\}} (-1)^{\frac{t+1}{2}}\mathbf{e}_1^\top \mathbf{\Psi}_t^{-1}\mathbf{O}_t\Bigg) \\
&= \widehat{\Xi}^{-1/2}\Bigg(\sum_{t\in\{0,1\}} (-1)^{\frac{t+1}{2}}(\widehat{\theta}_t(0)-\theta_t(0)) - \sum_{t\in\{0,1\}} (-1)^{\frac{t+1}{2}}\mathbf{e}_1^\top \mathbf{\Psi}_t^{-1}\mathbf{O}_t\Bigg) && (=\Delta_1) \\
&\quad + (\widehat{\Xi}^{-1/2} - \Xi^{-1/2}) \sum_{t\in\{0,1\}} (-1)^{\frac{t+1}{2}}\mathbf{e}_1^\top \mathbf{\Psi}_t^{-1}\mathbf{O}_t && (=\Delta_2)
\end{aligned}$$

By Lemma SA-3 and SA-4, and the decomposition Equation (SA-1),

$$\sup_{\mathbf{x}\in\mathscr{X}}\Bigg|\sum_{t\in\{0,1\}} (-1)^{\frac{t+1}{2}}(\widehat{\theta}_t(0)-\theta_t(0)) - \sum_{t\in\{0,1\}} (-1)^{\frac{t+1}{2}}\mathbf{e}_1^\top \mathbf{\Psi}_t^{-1}\mathbf{O}_t\Bigg| \lesssim_{\mathbb{P}} (nh)^{-1} + |\mathfrak{B}_n|.$$

Together with Lemma SA-6,

$$|\Delta_1| \lesssim_{\mathbb{P}} (nh)^{-1/2} + \sqrt{nh}|\mathfrak{B}_n|. \tag{SA-5}$$

By Lemma SA-3, Lemma SA-4 and Lemma SA-6, and assume $nh \to \infty$, then

$$|\Delta_2| = \sum_{t\in\{0,1\}}\left|\mathbf{e}_1^\top \mathbf{\Psi}_t^{-1}\mathbf{O}_t\left(\Xi^{-1/2} - \widehat{\Xi}^{-1/2}\right)\right| \lesssim_{\mathbb{P}} (nh)^{-1/2}. \tag{SA-6}$$

Putting together Equations (SA-5), (SA-6) gives

$$|\widehat{\mathrm{T}}_{\mathrm{pre}} - \overline{\mathrm{T}}_{\mathrm{pre}}| \lesssim_{\mathbb{P}} (nh)^{-1/2} + \sqrt{nh}|\mathfrak{B}_n|.$$

Next, we consider the convergence of $\overline{\mathrm{T}}_{\mathrm{pre}}$. Notice that if we define

$$Z_{n,i} = \frac{1}{n}\Xi^{-1/2}\mathbf{e}_1^\top \mathbf{\Psi}_t^{-1}\mathbf{r}_p\left(\frac{D_i}{h}\right)k_h(D_i)(Y_i - \theta_t^*(D_i))\mathbb{1}_{\mathcal{I}_t}(D_i),$$

then $\overline{\mathrm{T}}_{\mathrm{pre}} = \sum_{i=1}^n Z_{n,i}$. Moreover, $\mathbb{E}[Z_{n,i}] = 0$ and $\mathbb{V}[Z_{n,i}] = n^{-1}$. By Berry-Essen Theorem,

$$\begin{aligned}
&\sup_{u\in\mathbb{R}}\left|\mathbb{P}\left(\overline{\mathrm{T}}_{\mathrm{pre}} \le u\right) - \Phi(u)\right| \lesssim \sum_{i=1}^n \mathbb{E}\left[|Z_{n,i}|^3\right] \\
&= \sum_{i=1}^n n^{-3}\Xi^{-3/2}\mathbb{E}\left[\left|\mathbf{e}_1^\top \mathbf{\Psi}_t^{-1}\mathbf{r}_p\left(\frac{D_i}{h}\right)k_h(D_i)\mathbb{1}_{\mathcal{I}_t}(D_i)(Y_i - \theta_t^*(D_i)\right|^3\right] \\
&\lesssim n^{-2}\Xi^{-3/2}\mathbb{E}[|k_h(D_i)(Y_i - \theta_t^*(D_i))|^3] \\
&\lesssim n^{-2}\Xi^{-3/2}\mathbb{E}[|k_h(D_i)^3(\mathbb{E}[|Y_i|^3|\mathbf{X}_i] + |\theta_t^*(D_i)|^3)|] \\
&\lesssim (nh)^{-1/2},
\end{aligned}$$

where in the third line we used $\|\mathbf{r}_p\left(\frac{D_i}{h}\right)k_h\left(D_i\right)\| \lesssim 1$ holds almost surely, in last line we have use Lemma SA-1 with Assumptions 1 and SA–3 to get $\mathbb{E}[|k_h(D_i)^3(\mathbb{E}[|Y_i|^3|\mathbf{X}_i]+|\theta_t^*(D_i)|^3)|] \lesssim h^{-2}$, and the fact that $\Xi \gtrsim (nh)^{-1/2}$ from Lemma SA-6. □

SA-3.4 Approximation Error

Lemma SA-7 (Approximation Error)**.** *Suppose Assumptions SA–2 and SA–3 hold. If θ_t is $(s+1)$-times continuously differentiable at 0, and $h \to 0$, then*

$$\mathfrak{B}_{n,t} \lesssim h^{\min\{s,p\}+1}$$

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

Proof. Consider

$$\mathbf{S}_t = \mathbb{E}\left[\mathbf{r}_p\left(\frac{D_i}{h}\right)k_h(D_i)Y_i\mathbb{1}_{\mathscr{I}_t}(D_i)\right] = \mathbb{E}\left[\mathbf{r}_p\left(\frac{D_i}{h}\right)k_h(D_i)\theta_t(D_i)\mathbb{1}_{\mathscr{I}_t}(D_i)\right].$$

Define $\eta_t(h)$ to be the p-dimensional vector that is $(\theta_t(0), h\theta_t^{(1)}(0), \cdots, h^p\theta_t^{(p)}(0))$ if $p \leq s$, and $(\theta_t(0), h\theta_t^{(1)}(0), \cdots, h^s\theta_t^{(s)}(0), 0, \cdots, 0)$ otherwise.

$$\begin{aligned}
|\mathfrak{B}_{n,t}| &= \left|\mathbf{e}_1^\top\mathbf{\Psi}_t^{-1}\mathbf{S}_t - \theta_t(0)\right| \\
&= \left|\mathbf{e}_1^\top\mathbf{\Psi}_t^{-1}\mathbb{E}\Big[\mathbf{r}_p\left(\frac{D_i}{h}\right)k_h(D_i)(\theta_t(D_i) - \mathbf{r}_p\left(\frac{D_i}{h}\right)^\top\eta_t(h))\mathbb{1}_{\mathscr{I}_t}(D_i)\Big]\right|.
\end{aligned}$$

Assuming θ_t is s-times continuously differentiable, we have almost surely

$$\max_{1\leq i\leq n}\left|\theta_t(D_i) - \mathbf{r}_p\left(\frac{D_i}{h}\right)^\top\eta_t(h)\right|\mathbb{1}(k_h(D_i)) \lesssim h^{\min\{s,p\}+1}.$$

By Lemma SA-2, $\|\mathbf{\Psi}_t^{-1}\| \lesssim 1$, and the same argument on each entry and finite dimensionality of the basis implies

$$\mathbb{E}\Big[\Big\|\mathbf{r}_p\left(\frac{D_i}{h}\right)k_h(D_i)\mathbb{1}_{\mathscr{I}_t}(D_i)\Big\|\Big] \lesssim 1.$$

The conclusion follows. □

SA-3.5 Verification of High Level Conditions

This section verifies part (ii)-(iv) of Assumption SA–3, and the additional smoothness assumed in Lemma SA-7, for the special case when $\mathscr{B}$ is a finite collection of connected linear segments.

Lemma SA-8 (Piecewise Linear Boundary)**.** *Suppose Assumption SA–2 holds with $d = 2$, and $\mathscr{d}(\cdot,\cdot)$ is the Euclidean distance. If $\mathscr{B}$ is piecewise linear with finite many pieces, then*

(i) $J_1 d(\cdot, \mathscr{B}) = 1$ $\mathfrak{m}$*-almost everywhere on* $\mathscr{R}_t(\epsilon)$*, and* $\int_{\mathscr{R}_t(\epsilon)} |\frac{f(\mathbf{x})}{J_1 d(\mathbf{x}, \mathscr{B})}| d\mathbf{x} < \infty$*,*

(ii) $\lim_{\epsilon \downarrow 0} F(s) = F(0)$ *is finite, where* $F(s) = \int_{d(\mathbf{x}, \mathscr{B}) = s} \frac{f(\mathbf{x})}{J_1 d(\mathbf{x}, \mathscr{B})} d\mathfrak{H}^1(\mathbf{x})$ *for all* $s \geq 0$*, and*

(iii) θ_t *is continuous at* 0*,*

for small enough $\epsilon > 0$*, and* $t \in \{0, 1\}$*.*

If, in addition, μ_0*,* μ_1 *and* f *are* s*-times continuously differentiable on* $\mathscr{X}$*, then*

(iv) θ_0 *and* θ_1 *are* s*-times continuously differentiable on* $[0, \epsilon]$*.*

Proof. First, we verify the Jacobian condition (i). Since $\mathscr{B}$ is piecewise linear, we denote the linear pieces by $\mathscr{B}_1, \cdots, \mathscr{B}_M$ with $\mathscr{B} = \sqcup_{j=1}^M \mathscr{B}_j$. For small enough ϵ, for any $\mathbf{x} \in \mathscr{R}_t(\epsilon)$, the distance to the boundary depends on at most two linear pieces. That is, there exists $1 \leq j_1, j_2 \leq M$ such that

$$d(\mathbf{x}, \mathscr{B}) = d(\mathbf{x}, \mathscr{B}_{j_1} \cup \mathscr{B}_{j_2}),$$

as in Figure SA-1. For any point $\mathbf{x}$ on line segment $\overline{ce}$, the directional derivative of $d(\cdot, \mathscr{B})$ in the direction normal to $\overline{ce}$ is 1, and the directional derivative of $d(\cdot, \mathscr{B})$ in the direction parallel to $\overline{ce}$ is 0. Hence with $\nabla d(\cdot, \mathscr{B})|_{\mathbf{x}} = \mathbf{O}[1, 0]^\top$, where $\mathbf{O}$ is some orthonormal change of coordinate matrix. This shows $J_1 d(\cdot, \mathscr{B})|_{\mathbf{x}} = 1$. The same argument applies for $\mathbf{x}$ on the line segment $\overline{df}$ and on the segment $\overline{vp}$, $\overline{vq}$.

For any $\mathbf{x}$ on the arc $\overline{cd}$, the directional derivative of $d(\cdot, \mathscr{B})$ at $\mathbf{x}$ is 1 in the direction $\overrightarrow{O\mathbf{x}}$, and the directional derivative of $d(\cdot, \mathscr{B})$ at $\mathbf{x}$ is 1 in the direction orthogonal to $\overrightarrow{O\mathbf{x}}$. Thus $\nabla d(\cdot, \mathscr{B})|_{\mathbf{x}} = \mathbf{O}[1, 0]^\top$, where $\mathbf{O}$ is some orthonormal change of coordinate matrix, and hence $J_1 d(\cdot, \mathscr{B})|_{\mathbf{x}} = 1$.

This shows $J_1 d(\cdot, \mathscr{B}) = 1$ $\mathfrak{m}$-almost everywhere on $\mathscr{R}_t(\epsilon)$, and $\int_{\mathscr{R}_t(\epsilon)} |\frac{f(\mathbf{x})}{J_1 d(\mathbf{x}, \mathscr{B})}| d\mathbf{x} < \infty$ follows from continuity of f on compact support $\mathscr{X}$.

Next, we want to show continuity of F at 0 from (ii). It suffices to show $|F(s) - F(0)| \leq C\epsilon$ for all $s \in (0, \epsilon)$. Again, since $\mathscr{B}$ has finitely many pieces, with ϵ small enough, we can localize to Figure SA-1, and suppose $\mathscr{B} = \mathscr{B}_i \sqcup \mathscr{B}_j$. Consider $t = 0$, and the level set $\{\mathbf{x} \in \mathscr{A}_0 : d(\mathbf{x}, \mathscr{B}) = r\} = \overline{ce} \sqcup \overline{cd} \sqcup \overline{df}$. Define the projection on $\mathscr{B}$ by

$$\mathbf{P}(\mathbf{x}) = \arg\min_{\mathbf{y} \in \mathscr{B}} \|\mathbf{x} - \mathbf{y}\|.$$

Then

$$\begin{aligned}
F(r) - F(0) &= \int_{\overline{ce} \sqcup \overline{df}} f(\mathbf{x}) d\mathfrak{H}^{d-1}(\mathbf{x}) + \int_{\overline{cd}} f(\mathbf{x}) d\mathfrak{H}^{d-1}(\mathbf{x}) - \int_{\mathscr{B}_i \sqcup \mathscr{B}_j} f(\mathbf{x}) d\mathfrak{H}^{d-1}(\mathbf{x}) \\
&= \int_{\overline{ce} \sqcup \overline{df}} (f(\mathbf{x}) - f(\mathbf{P}(\mathbf{x}))) d\mathfrak{H}^{d-1}(\mathbf{x}) + \int_{\overline{cd}} f(\mathbf{x}) d\mathfrak{H}^{d-1}(\mathbf{x}).
\end{aligned}$$

The definition of $\mathbf{P}$ implies $\|\mathbf{x} - \mathbf{P}(\mathbf{x})\| \leq \epsilon$. Moreover, $\mathfrak{H}^{d-1}(\overline{cd}) \leq 2\pi\epsilon$. Since f is continuous on compact support $\mathscr{X}$, f is also uniform continuous on $\mathscr{X}$. Moreover, since $\mathscr{X}$ is compact, the pieces

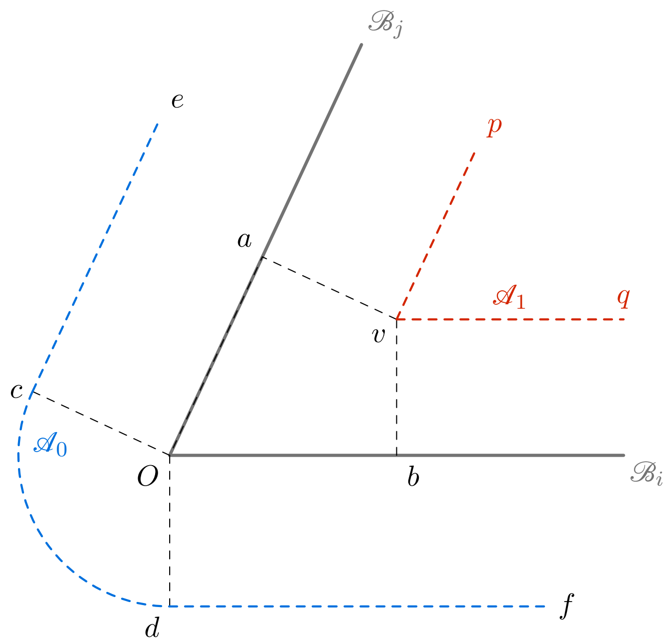

Figure SA-1: Level sets to piecewise linear boundary.

$\mathscr{B}_i$ and $\mathscr{B}_j$ has finite length, say $\mathfrak{H}^{d-1}(\mathscr{B}) \leq L$. Then

$$|F(r) - F(0)| \leq \sup_{\|\mathbf{x}-\mathbf{y}\| \leq \epsilon} |f(\mathbf{x}) - f(\mathbf{y})| L + 2\pi\epsilon \sup_{\mathbf{x} \in \mathscr{X}} f(\mathbf{x}) \to 0, \qquad \text{as } \epsilon \to 0.$$

This proves (ii).

To show (iii) holds, notice that the same argument as above implies that for $t = 0, 1$, the function

$$s \mapsto \int_{\{\mathbf{x} \in \mathscr{A}_t : d(\mathbf{x}, \mathscr{B}) = s\}} f(\mathbf{x}) \mu_t(\mathbf{x}) d\mathfrak{H}^1(\mathbf{x})$$

is also continuously at 0. Since f is continouous on compact support $\mathscr{X}$, f is also bounded from below on $\mathscr{X}$. This implies for $t = 0, 1$,

$$\theta_t(r) = \frac{\int_{\{\mathbf{x} \in A_t : d(\mathbf{x}, \mathscr{B}) = r\}} f(\mathbf{x}) \mu_t(\mathbf{x}) d\mathfrak{H}^{d-1}(\mathbf{x})}{\int_{\{\mathbf{x} \in A_t : d(\mathbf{x}, \mathscr{B}) = r\}} f(\mathbf{x}) d\mathfrak{H}^{d-1}(\mathbf{x})}$$

is also continuous at zero.

Now, we want to show (iv) holds. W.l.o.g., we assume $\mathscr{B}$ is composed with two linear pieces as in Figure SA-1. Suppose $\gamma_{r,t}$ is a curve length parametrization of the level set $\{\mathbf{x} \in \mathscr{A}_t : d(\mathbf{x}, \mathscr{B}) = r\}$.

First, consider $t = 0$. Denote $L_i = \mathfrak{H}^1(\mathscr{B}_i)$ and $L_j = \mathfrak{H}^1(\mathscr{B}_j)$. Denote α to be the angle of $\angle aob$. Then we can define the curve length parametrization to be

$$\gamma_{r,0}(s) =$$

$$\begin{cases} (L_i - s, -r), & s \in [0, L_i], \\ r(\cos(\frac{3}{2}\pi - s/r), \sin(3/2\pi - s/r)), & s \in [L_i, L_i + \theta r], \\ (s - L_i - \theta r)(\cos(\alpha), \sin(\alpha)) + r(\sin(\alpha), \cos(\alpha)), & s \in [L_i + \theta r, L_i + L_j + \theta r]. \end{cases}$$

Hence for any function $\psi : \mathbb{R}^2 \to \mathbb{R}$, we have

$$\begin{aligned} &\int_{\{\mathbf{x} \in \mathscr{A}_0 : d(\mathbf{x}, \mathscr{B}) = r\}} \psi(\mathbf{x}) d\mathfrak{H}^1(\mathbf{x}) \\ &= \int_0^{L_i + L_j + \theta r} \psi(\gamma_{r,0}(s)) \|\gamma'_{r,t}(s)\| ds \\ &= \int_0^{L_i + L_j + \theta r} \psi(\gamma_{r,0}(s)) ds \\ &= I_1(r) + I_2(r) + I_3(r), \end{aligned}$$

where

$$I_1(r) = \int_0^{L_i} \psi((L_i - s, -r)) ds,$$

and using a change variable with $u = s/r$,

$$\begin{aligned} I_2(r) &= \int_{L_i}^{L_i + \theta r} \psi(r(\cos(\frac{3}{2}\pi - s/r), \sin(3/2\pi - s/r))) ds \\ &= \int_{L_i}^{L_i + \theta} \psi(r(\cos(\frac{3}{2}\pi - u), \sin(3/2\pi - u))) r du, \end{aligned}$$

and using a change of variable with $v = s - L_i - \theta r$,

$$\begin{aligned} I_3(r) &= \int_{L_i + \theta r}^{L_i + L_j + \theta r} \psi((s - L_i - \theta r)(\cos(\alpha), \sin(\alpha)) + r(\sin(\alpha), \cos(\alpha))) ds \\ &= \int_0^{L_j} \psi(v(\cos(\alpha), \sin(\alpha)) + r(\sin(\alpha), \cos(\alpha))) dv. \end{aligned}$$

It follows that if ψ is s times continuously differentiable, then $r \mapsto \int_{\{\mathbf{x} \in \mathscr{A}_0 : d(\mathbf{x}, \mathscr{B}) = r\}} \psi(\mathbf{x}) d\mathfrak{H}^1(\mathbf{x})$ is also s times continuously differentiable.

Similarly, for $t = 1$, a curve length parametrization is given by

$$\gamma_{r,1}(s) =$$
$$\begin{cases} (L_j - s, s), & s \in [0, L_j - r\cot(\frac{\alpha}{2})], \\ (r\cot(\alpha/2), s) + (s - L_j + r\cot(\alpha/2))(\cos(\alpha), \sin(\alpha)), & s \in [L_j - r\cot(\frac{\alpha}{2}), L_i + L_j - 2r\cot(\frac{\alpha}{2})]. \end{cases}$$

The smoothness of $L_j - r\cot(\frac{\alpha}{2})$ and $L_i + L_j - 2r\cot(\frac{\alpha}{2})$ in terms of r, smoothness of $\gamma_{r,1}(s)$ in r on each piece implies that for any function $\psi : \mathbb{R}^2 \to \mathbb{R}$ that is s times continuously differentiable, we have $r \mapsto \int_{\{\mathbf{x}\in\mathscr{A}_1: d(\mathbf{x},\mathscr{B})=r\}} \psi(\mathbf{x}) d\mathfrak{H}^1(\mathbf{x})$ is also s times continuously differentiable.

Hence if both μ_t and f are s times continuously differentiable, then both

$$r \to \int_{\{\mathbf{x}\in A_t: d(\mathbf{x},\mathscr{B})=r\}} f(\mathbf{x}) d\mathfrak{H}^{d-1}(\mathbf{x})$$

and

$$r \to \int_{\{\mathbf{x}\in A_t: d(\mathbf{x},\mathscr{B})=r\}} f(\mathbf{x})\mu_t(\mathbf{x}) d\mathfrak{H}^{d-1}(\mathbf{x})$$

are s times continuously differentiable. Since we have assumed f continuous and positive on $\mathscr{X}$, it is bounded from below on $\mathscr{X}$. Hence

$$\theta_t(r) = \frac{\int_{\{\mathbf{x}\in A_t: d(\mathbf{x},\mathscr{B})=r\}} f(\mathbf{x})\mu_t(\mathbf{x}) d\mathfrak{H}^{d-1}(\mathbf{x})}{\int_{\{\mathbf{x}\in A_t: d(\mathbf{x},\mathscr{B})=r\}} f(\mathbf{x}) d\mathfrak{H}^{d-1}(\mathbf{x})}$$

is also s times continuously differentiable, for $t = 0, 1$. □

Corollary SA-2 (Bias for Piecewise Linear Boundary)**.** *Suppose Assumption SA–2 holds with $d = 2$, and $\mathscr{d}(\cdot,\cdot)$ is the Euclidean distance. If $\mathscr{B}$ is piecewise linear with finite many pieces, and f is $(p+1)$-times continuously differentiable, then*

$$\mathfrak{B}_{n,t} \lesssim h^{p+1},$$

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

Proof. The conclusion follows from Lemma SA-7 and Lemma SA-8. □

SA-4 Post-Aggregation Approach

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

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

The following assumption is the natural generalization of Assumption 3 in the paper to the case $d \ge 2$.

Assumption SA–4 (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) $\liminf_{h\downarrow 0}\inf_{\mathbf{x}\in\mathscr{B}}\int_{\mathscr{A}_t} K_h(\mathbf{u}-\mathbf{x})d\mathbf{u}\gtrsim 1$.

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

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

where $\widehat{\varepsilon}_i(\mathbf{x})=Y_i-\sum_{t\in\{0,1\}}\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_t)\widehat{\boldsymbol{\beta}}_t(\mathbf{x})^\top\mathbf{R}_p(\mathbf{X}_i-\mathbf{x})$ and $\sigma_t^2(\mathbf{x})=\mathbb{V}[Y_i(t)|\mathbf{X}_i=\mathbf{x}]$.

In addition, we have

$$\begin{aligned}
\bar{B}_{t,\mathbf{x}} &= \mathbf{e}_1^\top\widehat{\mathbf{\Gamma}}_{t,\mathbf{x}}^{-1}\sum_{|\boldsymbol{\omega}|=p+1}\frac{\mu_t^{(\boldsymbol{\omega})}(\mathbf{x})}{\boldsymbol{\omega}!}\mathbb{E}_n\Bigg[\mathbf{R}_p\left(\frac{\mathbf{X}_i-\mathbf{x}}{h}\right)\left(\frac{\mathbf{X}_i-\mathbf{x}}{h}\right)^{\boldsymbol{\omega}}K_h(\mathbf{X}_i-\mathbf{x})\Bigg], \qquad \bar{B}_{\mathbf{x}}=\bar{B}_{1,\mathbf{x}}-\bar{B}_{0,\mathbf{x}},\\
B_{t,\mathbf{x}} &= \mathbf{e}_1^\top\mathbf{\Gamma}_{t,\mathbf{x}}^{-1}\sum_{|\boldsymbol{\omega}|=p+1}\frac{\mu_t^{(\boldsymbol{\omega})}(\mathbf{x})}{\boldsymbol{\omega}!}\mathbb{E}\Bigg[\mathbf{R}_p\left(\frac{\mathbf{X}_i-\mathbf{x}}{h}\right)\left(\frac{\mathbf{X}_i-\mathbf{x}}{h}\right)^{\boldsymbol{\omega}}K_h(\mathbf{X}_i-\mathbf{x})\Bigg], \qquad B_{\mathbf{x}}=B_{1,\mathbf{x}}-B_{0,\mathbf{x}},\\
\mathbf{Q}_{t,\mathbf{x}} &= \mathbb{E}_n\left[\mathbf{R}_p\left(\frac{\mathbf{X}_i-\mathbf{x}}{h}\right)K_h(\mathbf{X}_i-\mathbf{x})\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_t)u_i\right],
\end{aligned}$$

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

$$\begin{aligned}
\widehat{V}_{t,\mathbf{x}} &= h^{-1}\mathbf{e}_1^\top\widehat{\mathbf{\Gamma}}_{t,\mathbf{x}}^{-1}\widehat{\mathbf{\Sigma}}_{t,\mathbf{x},\mathbf{x}}\widehat{\mathbf{\Gamma}}_{t,\mathbf{x}}^{-1}\mathbf{e}_1, \qquad \widehat{V}_{\mathbf{x}}=\widehat{V}_{0,\mathbf{x}}+\widehat{V}_{1,\mathbf{x}},\\
V_{t,\mathbf{x}} &= h^{-1}\mathbf{e}_1^\top\mathbf{\Gamma}_{t,\mathbf{x}}^{-1}\mathbf{\Sigma}_{t,\mathbf{x},\mathbf{x}}\mathbf{\Gamma}_{t,\mathbf{x}}^{-1}\mathbf{e}_1, \qquad V_{\mathbf{x}}=V_{0,\mathbf{x}}+V_{1,\mathbf{x}},
\end{aligned}$$

Recall that we assumed, only for simplicity, that $\int_{\mathscr{B}} w(\mathbf{b})dH^1(\mathbf{b})=1$. Then, the post-aggregation estimator becomes

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

Finally, we define the aggregated bias and variance quantities:

$$\begin{aligned}
B_{\mathscr{B}} &= B_{1,\mathscr{B}}-B_{0,\mathscr{B}}, \qquad B_{t,\mathscr{B}}=\int_{\mathscr{B}}B_{t,\mathbf{b}}w(\mathbf{b})d\mathfrak{H}^{d-1}(\mathbf{b}),\\
\bar{B}_{\mathscr{B}} &= \bar{B}_{1,\mathscr{B}}-\bar{B}_{0,\mathscr{B}}, \qquad \bar{B}_{t,\mathscr{B}}=\int_{\mathscr{B}}\bar{B}_{t,\mathbf{b}}w(\mathbf{b})d\mathfrak{H}^{d-1}(\mathbf{b}),\\
\Omega_{\mathscr{B}} &= \Omega_{1,\mathscr{B}}+\Omega_{0,\mathscr{B}}, \qquad \Omega_{t,\mathscr{B}}=\int_{\mathscr{B}}\int_{\mathscr{B}}\Omega_{t,\mathbf{b}_1,\mathbf{b}_2}w(\mathbf{b}_1)w(\mathbf{b}_2)d\mathfrak{H}^{d-1}(\mathbf{b}_1)d\mathfrak{H}^{d-1}(\mathbf{b}_2),
\end{aligned}$$

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

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

SA-4.1 Preliminary Lemmas

This section states the preliminary results on matrix convergence. 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-9 (Gram)**.** *Suppose Assumption SA–2(i)-(ii) and Assumption SA–4 hold. If* $\frac{\log(1/h)}{nh^d} = o(1)$*, then*

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

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

Proof. See Lemma 2.1 in the supplemental appendix of Cattaneo et al. [2025]. □

Lemma SA-10 (Bias)**.** *Suppose Assumption SA–2(i)-(iii) and Assumption SA–4 hold. If* $\frac{\log(1/h)}{nh^d} = o(1)$*, then*

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

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

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

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

Proof. See Lemma 2.2 in the supplemental appendix of Cattaneo et al. [2025]. □

Lemma SA-11 (Aggregated Bias Along $\mathscr{B}$)**.** *Suppose Assumption SA–2(i)-(iii) and Assumption SA–4 hold. If* $\frac{\log(1/h)}{nh^d} = o(1)$ *and* $h = o(1)$*, then*

$$\begin{aligned}\int_{\mathscr{B}} (\mathbb{E}[\widehat{\mu}_t(\mathbf{x})|\mathbf{X}] - \mu_t(\mathbf{x})) w(\mathbf{x}) d\mathfrak{H}^{d-1}(\mathbf{x}) &= h^{p+1} B_{\mathscr{B}} + o_{\mathbb{P}}(h^{p+1}) \\ &= h^{p+1} \bar{B}_{\mathscr{B}} + o_{\mathbb{P}}(h^{p+1}).\end{aligned}$$

Proof. The conclusion follows from Lemma SA-10 and the assumption that $\int_{\mathscr{B}} w(\mathbf{x}) d\mathfrak{H}^{d-1}(\mathbf{x}) = 1$. □

Lemma SA-12 (Stochastic Linear Approximation)**.** *Suppose Assumption SA–2(i)-(v) and Assumption SA–4 hold. If* $\frac{\log(1/h)}{nh^d} = o(1)$*, then*

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

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

Proof. See Lemma 2.3 in the supplemental appendix of Cattaneo et al. [2025]. □

Lemma SA-13 (Covariance)**.** *Suppose Assumptions SA–2 and SA–4 hold. If* $\frac{\log(1/h)}{nh^d} = o(1)$*, then*

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

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

Proof. See Lemma 2.4 from the supplemental appendix of Cattaneo et al. [2025]. □

Lemma SA-14 (Variance of Aggregated Estimator)**.** *Suppose Assumptions SA–2 and SA–4 hold. If* $\frac{\log(1/h)}{nh^d} = o(1)$ *and* $h = o(1)$*, then*

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

where

$$(nh)^{-1} \lesssim \Omega_{\mathcal{B}} \lesssim (nh)^{-1}.$$

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

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

Proof. Observe that $\mathbb{V}[\widehat{\tau}_w|\mathbf{X}] = \sum_{t=0,1} \mathbb{V}[\int_{\mathcal{B}} \widehat{\mu}_t(\mathbf{b}) w(\mathbf{b}) d\mathfrak{H}^{d-1}(\mathbf{b})]$, where for $t = 0, 1$,

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

Consider

$$\overline{\mathbf{\Sigma}}_{t,\mathbf{b}_1,\mathbf{b}_2} = h^d \mathbb{E}_n\Big[\mathbf{R}_p\left(\frac{\mathbf{X}_i - \mathbf{b}_1}{h}\right)\mathbf{R}_p\left(\frac{\mathbf{X}_i - \mathbf{b}_2}{h}\right)^\top K_h(\mathbf{X}_i - \mathbf{b}_1)K_h(\mathbf{X}_i - \mathbf{b}_2)\sigma_t^2(\mathbf{X}_i)\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)\Big]. \tag{SA-7}$$

Then the same argument as the proof of Lemma SA-9 implies if $\frac{\log(1/h)}{nh^d} = o(1)$, then

$$\sup_{\mathbf{b}_1,\mathbf{b}_2 \in \mathscr{B}} \|\overline{\mathbf{\Sigma}}_{t,\mathbf{b}_1,\mathbf{b}_2} - \mathbf{\Sigma}_{t,\mathbf{b}_1,\mathbf{b}_2}\| = O_{\mathbb{P}}\Big(\sqrt{\frac{\log(1/h)}{nh^d}}\Big). \tag{SA-8}$$

Together with Lemma SA-9, we have when $\frac{\log(1/h)}{nh^d} = o(1)$,

$$\begin{aligned}
&\sup_{\mathbf{b}_1,\mathbf{b}_2 \in \mathscr{B}} \|\mathbb{C}\mathrm{ov}\Big[\widehat{\mu}_t(\mathbf{b}_1), \widehat{\mu}_t(\mathbf{b}_2)\Big|\mathbf{X}\Big] - \Omega_{t,\mathbf{b}_1,\mathbf{b}_2}\| \\
&= \sup_{\mathbf{b}_1,\mathbf{b}_2 \in \mathscr{B}} \|(nh^d)^{-1}\mathbf{e}_1^\top \widehat{\mathbf{\Gamma}}_t^{-1} \overline{\mathbf{\Sigma}}_{t,\mathbf{b}_1,\mathbf{b}_2} \widehat{\mathbf{\Gamma}}_t^{-1}\mathbf{e}_1 - \Omega_{t,\mathbf{b}_1,\mathbf{b}_2}\| \\
&= O_{\mathbb{P}}\Big(\frac{\log(1/h)^{1/2}}{(nh^d)^{3/2}}\Big).
\end{aligned}$$

We have assumed that K is supported on a compact set. W.l.o.g, assume the support of K has a diameter no greater than R, $0 < R < \infty$. Consider the set

$$\mathscr{E}(h) = \{(\mathbf{x}, \mathbf{y}) \in \mathscr{B} \times \mathscr{B} : \|\mathbf{x} - \mathbf{y}\| \le hR\}. \tag{SA-9}$$

Since $\mathscr{B}$ is $d-1$ dimensional, we have $\mathfrak{m}(\mathscr{E}(h)) \lesssim h^{d-1}$, where $\mathfrak{m}$ is the $2(d-1)$ dimensional Lebesgue measure. Hence

$$\begin{aligned}
&\mathbb{V}\Big[\int_{\mathscr{B}} \widehat{\mu}_t(\mathbf{b}) w(\mathbf{b}) d\mathfrak{H}^{d-1}(\mathbf{b})\Big|\mathbf{X}\Big] - \Omega_t \\
&= \int_{\mathscr{B}}\int_{\mathscr{B}} \Big(\mathbb{C}\mathrm{ov}\Big[\widehat{\mu}_t(\mathbf{b}_1), \widehat{\mu}_t(\mathbf{b}_2)\Big|\mathbf{X}\Big] - \Omega_{t,\mathbf{b}_1,\mathbf{b}_2}\Big) w(\mathbf{b}_1) w(\mathbf{b}_2) d\mathfrak{H}^{d-1}(\mathbf{b}_1) d\mathfrak{H}^{d-1}(\mathbf{b}_2) \\
&\lesssim \sup_{\mathbf{b}_1,\mathbf{b}_2 \in \mathscr{B}} \|\mathbb{C}\mathrm{ov}\Big[\widehat{\mu}_t(\mathbf{b}_1), \widehat{\mu}_t(\mathbf{b}_2)\Big|\mathbf{X}\Big] - \Omega_{t,\mathbf{b}_1,\mathbf{b}_2}\| \cdot \\
&\qquad\qquad \int_{\mathscr{B}}\int_{\mathscr{B}} \mathbb{1}((\mathbf{b}_1, \mathbf{b}_2) \in \mathscr{E}(h)) w(\mathbf{b}_1) w(\mathbf{b}_2) d\mathfrak{H}^{d-1}(\mathbf{b}_1) d\mathfrak{H}^{d-1}(\mathbf{b}_2) \\
&= O_{\mathbb{P}}\Big(h^{d-1}\frac{\log(1/h)^{1/2}}{(nh^d)^{3/2}}\Big) \\
&= o_{\mathbb{P}}((nh)^{-1}),
\end{aligned} \tag{SA-10}$$

where in the last line we have used the assumption that $\frac{\log(1/h)}{nh^d} = o(1)$. This proves the first claim.

For the second argument,

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

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

This shows the upper bound. For the lower bound, let $\mathbf{b}_1 \in \mathscr{B}$ and $\mathbf{b}_2 = \mathbf{b}_1 + h\delta$ for some vector delta such that $\sup_{\mathbf{x}\in\mathscr{X}} K_h(\mathbf{x}-\mathbf{b}_1)K_h(\mathbf{x}-\mathbf{b}_2) > 0$. A change of variable then implies a typical element of $\mathbf{\Sigma}_{t,\mathbf{b}_1,\mathbf{b}_2}$ has the form

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

It follows that $|\Omega_{t,\mathbf{b}_1,\mathbf{b}_2}| \gtrsim (nh^d)^{-1}$ for $(\mathbf{b}_1,\mathbf{b}_2)$ on a set $\mathscr{E}'(h)$ such that $\mathfrak{m}(\mathscr{E}'(h)) \gtrsim h^{d-1}$. This leads to the lower bound in the second claim.

The third claim follows from Lemma SA-13 and the same analysis as Equation (SA-10). □

SA-4.2 Mean Square Error Expansion

Theorem SA-3 (MSE Expansions)**.** *Suppose Assumptions SA–2 and SA–4 hold. If* $\frac{\log(1/h)}{n^{\frac{v}{2+v}}h^d} = o(1)$ *and* $h = o(1)$*, then*

$$
\mathbb{E}[(\widehat{\tau}_w-\tau_w)^2|\mathbf{X}] = \Omega_{\mathscr{B}} + h^{2p+2}B_{\mathscr{B}}^2 + o_{\mathbb{P}}((nh)^{-1}) + o_{\mathbb{P}}(h^{2p+2}).
$$

Proof. Using the decomposition

$$
\mathbb{E}[(\widehat{\tau}_w-\tau_w)^2|\mathbf{X}] = \mathbb{V}[\widehat{\tau}_w|\mathbf{X}] + (\mathbb{E}[\widehat{\tau}_w|\mathbf{X}]-\tau_w)^2,
$$

the conclusion then follows from Lemmas SA-11 and SA-14. □

SA-4.3 Central Limit Theorem

The feasible t-statistics is

$$
\widehat{\mathrm{T}}_w = \frac{\widehat{\tau}_w-\tau_w}{\sqrt{\widehat{\Omega}_{\mathscr{B}}}}.
$$

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

$o(1)$ *and* $(nh)h^{2p+2} = o(1)$*, then*

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

Proof. We decompose the error $\widehat{\tau}_w - \tau_w = (\widehat{\mu}_{1,w} - \mu_{1,w}) - (\widehat{\mu}_{0,w} - \mu_{0,w})$, where

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

for $t \in \{0,1\}$, and using Lemma SA-11 to bound the approximation error

For the non-linearity error, using $\overline{\boldsymbol{\Sigma}}_{t,\mathbf{b}_1,\mathbf{b}_2}$ in Equation (SA-7),

$$\begin{aligned}
&\mathbb{E}\left[\left(\int_{\mathscr{B}} \mathbf{e}_1^\top (\widehat{\boldsymbol{\Gamma}}_{t,\mathbf{b}}^{-1} - \boldsymbol{\Gamma}_{t,\mathbf{b}}^{-1}) \mathbf{Q}_{t,\mathbf{b}} d\mathfrak{H}^{d-1}(\mathbf{b}) \right)^2 \middle| \mathbf{X} \right] \\
&= \int_{\mathscr{B}} \int_{\mathscr{B}} \mathbf{e}_1^\top (\widehat{\boldsymbol{\Gamma}}_{t,\mathbf{b}_1}^{-1} - \boldsymbol{\Gamma}_{t,\mathbf{b}_1}^{-1}) (nh^d)^{-1} \overline{\boldsymbol{\Sigma}}_{t,\mathbf{b}_1,\mathbf{b}_2} (\widehat{\boldsymbol{\Gamma}}_{t,\mathbf{b}_2}^{-1} - \boldsymbol{\Gamma}_{t,\mathbf{b}_2}^{-1}) \mathbf{e}_1 w(\mathbf{b}_1) w(\mathbf{b}_2) d\mathfrak{H}^{d-1}(\mathbf{b}_1) d\mathfrak{H}^{d-1}(\mathbf{b}_2).
\end{aligned}$$

Since $\overline{\boldsymbol{\Sigma}}_{t,\mathbf{b}_1,\mathbf{b}_2} = 0$ if $\mathbf{b}_1$ and $\mathbf{b}_2$ are farther away form each other than the diameter of $\operatorname{Supp}(K)$, we can use Equation (SA-9) to get

$$\begin{aligned}
&\int_{\mathscr{B}} \int_{\mathscr{B}} \mathbf{e}_1^\top (\widehat{\boldsymbol{\Gamma}}_{t,\mathbf{b}_1}^{-1} - \boldsymbol{\Gamma}_{t,\mathbf{b}_1}^{-1}) (nh^d)^{-1} \overline{\boldsymbol{\Sigma}}_{t,\mathbf{b}_1,\mathbf{b}_2} (\widehat{\boldsymbol{\Gamma}}_{t,\mathbf{b}_2}^{-1} - \boldsymbol{\Gamma}_{t,\mathbf{b}_2}^{-1}) \mathbf{e}_1 d\mathfrak{H}^{d-1}(\mathbf{b}_1) d\mathfrak{H}^{d-1}(\mathbf{b}_2) \\
&\le \sup_{\mathbf{b} \in \mathscr{B}} (\widehat{\boldsymbol{\Gamma}}_{t,\mathbf{b}}^{-1} - \boldsymbol{\Gamma}_{t,\mathbf{b}}^{-1})^2 \sup_{\mathbf{b}_1,\mathbf{b}_2 \in \mathscr{B}} \|\overline{\boldsymbol{\Sigma}}_{t,\mathbf{b}_1,\mathbf{b}_2}\| \sup_{\mathbf{b} \in \mathscr{B}} |w(\mathbf{b})| (nh^d)^{-1} \mathfrak{m}(\mathscr{E}(h)) \\
&\lesssim \sup_{\mathbf{b} \in \mathscr{B}} (\widehat{\boldsymbol{\Gamma}}_{t,\mathbf{b}}^{-1} - \boldsymbol{\Gamma}_{t,\mathbf{b}}^{-1})^2 \sup_{\mathbf{b}_1,\mathbf{b}_2 \in \mathscr{B}} \|\overline{\boldsymbol{\Sigma}}_{t,\mathbf{b}_1,\mathbf{b}_2}\| (nh)^{-1},
\end{aligned}$$

where the constant does not depend on n or $\mathbf{X}$. Then, by Lemma SA-9 and Equation (SA-8),

$$\begin{aligned}
&\int_{\mathscr{B}} \mathbf{e}_1^\top (\widehat{\boldsymbol{\Gamma}}_{t,\mathbf{b}}^{-1} - \boldsymbol{\Gamma}_{t,\mathbf{b}}^{-1}) \mathbf{Q}_{t,\mathbf{b}} d\mathfrak{H}^{d-1}(\mathbf{b}) \\
&\lesssim_{\mathbb{P}} \left[\sup_{\mathbf{b} \in \mathscr{B}} (\widehat{\boldsymbol{\Gamma}}_{t,\mathbf{b}}^{-1} - \boldsymbol{\Gamma}_{t,\mathbf{b}}^{-1})^2 \sup_{\mathbf{b}_1,\mathbf{b}_2 \in \mathscr{B}} \|\overline{\boldsymbol{\Sigma}}_{t,\mathbf{b}_1,\mathbf{b}_2}\| (nh)^{-1} \right]^{1/2} \\
&\lesssim_{\mathbb{P}} o_{\mathbb{P}}((nh)^{-1/2})
\end{aligned}$$

because $\frac{\log(1/h)}{nh^d} = o(1)$.

Consider now the stochastic linearized T-statistic

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

Then, by Lemma SA-14 and the previous bounds,

$$\widehat{\mathrm{T}}_w - \overline{\mathrm{T}}_w = (\widehat{\Omega}_{\mathscr{B}}^{-1/2} - \Omega_{\mathscr{B}}^{-1/2}) \int_{\mathscr{B}} \mathbf{e}_1^\top \mathbf{\Gamma}_{t,\mathbf{b}}^{-1} \mathbf{Q}_{t,\mathbf{b}} d\mathfrak{H}^{d-1}(\mathbf{b}) + o_{\mathbb{P}}(1) = o_{\mathbb{P}}(1) \tag{SA-11}$$

because $\frac{\log(1/h)}{n^{\frac{v}{2+v}} h^d} = o(1)$ and $(nh)h^{2p+2} = o(1)$.

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

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

which satisfies $\mathbb{E}[Z_i] = 0$. The definition of $\Omega_{\mathscr{B}}$ implies that $\sum_{i=1}^n \mathbb{V}[Z_i] = \Omega_{\mathscr{B}}^{-1/2} \Omega_{\mathscr{B}} \Omega_{\mathscr{B}}^{-1/2} = 1$. Hence,

$$\begin{aligned}
&\sum_{i=1}^n \mathbb{E}[|Z_i^3|] \\
&= n^{-3} \Omega_{\mathscr{B}}^{-3/2} \sum_{i=1}^n \mathbb{E}\Bigg[\bigg(\int_{\mathscr{B}} \mathbf{e}_1^\top \mathbf{\Gamma}_{t,\mathbf{b}}^{-1} \mathbf{R}_p \left(\frac{\mathbf{X}_i - \mathbf{b}}{h} \right) K_h(\mathbf{X}_i - \mathbf{b}) \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t) u_i w(\mathbf{b}) d\mathfrak{H}^{d-1}(\mathbf{b}) \bigg)^3 \Bigg].
\end{aligned}$$

Consider

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

where

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

The same argument as Lemma SA-9 shows that

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

when $\frac{\log(1/n)}{nh^d} = o(1)$. Suppose the support of K has diameter less than R, and consider

$$\mathscr{F}(h) = \{(\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3) \in \mathscr{B}^3 : \|\mathbf{b}_i - \mathbf{b}_j\| \leq R, j = 1, 2, 3\}.$$

Since $\mathscr{B}$ is $d-1$ dimensional, $\mathfrak{m}(\mathscr{F}(h)) \lesssim h^{2d-2}$. It follows that

$$\begin{aligned}
&\mathbb{E}\Bigg[\bigg(\int_{\mathscr{B}} \mathbf{e}_1^\top \mathbf{\Gamma}_{t,\mathbf{b}}^{-1} \mathbf{R}_p \left(\frac{\mathbf{X}_i - \mathbf{b}}{h} \right) K_h(\mathbf{X}_i - \mathbf{b}) \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t) u_i w(\mathbf{b}) d\mathfrak{H}^{d-1}(\mathbf{b}) \bigg)^3 \Bigg] \\
&= \mathbb{E}\left[\int_{\mathbf{b}_1 \in \mathscr{B}} \int_{\mathbf{b}_2 \in \mathscr{B}} \int_{\mathbf{b}_3 \in \mathscr{B}} G(\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3) w(\mathbf{b}_1) w(\mathbf{b}_2) w(\mathbf{b}_3) d\mathfrak{H}^{d-1}(\mathbf{b}_1) d\mathfrak{H}^{d-1}(\mathbf{b}_2) d\mathfrak{H}^{d-1}(\mathbf{b}_3) \right] \\
&\lesssim \mathfrak{m}(\mathscr{F}(h)) \sup_{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3 \in \mathscr{B}} \mathbb{E}[|G(\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)|] \lesssim nh^{-2}.
\end{aligned}$$

Together with the rate of $\Omega_{\mathscr{B}}$ from Lemma SA-14, we have

$$\sum_{i=1}^{n} \mathbb{E}[|Z_i^3|] \lesssim (nh)^{-1/2}.$$

By Berry-Esseen lemma, we have

$$\sup_{u \in \mathbb{R}} |\mathbb{P}(\overline{\mathrm{T}}_w \leq u) - \Phi(u)| = O((nh)^{-1/2}),$$

and the final conclusion then follows from Equation (SA-11). □

References

Matias D Cattaneo, Rocio Titiunik, and Ruiqi Rae Yu. Estimation and inference in boundary discontinuity designs. *arXiv preprint arXiv:2505.05670*, 2025.

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

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

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

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