PARTIALLY-GLOBAL FRÉCHET REGRESSION

Danielle C. Tucker and Yichao Wu

University of Illinois at Chicago

Supplementary Material

S1 Definitions

Definition 1. We denote the Fréchet regression function of Y given $\mathbf{X} = \mathbf{x}$ and $Z = z$

$$m_{\oplus}(\mathbf{x}, z) = \underset{y \in \mathcal{Y}}{\arg\min}\, E(d^2(Y, y)|\mathbf{X} = \mathbf{x}, Z = z).$$

The partially-global Fréchet regression model is said to hold if

$$m_{\oplus}(\mathbf{x}, z) = s_{\oplus}(\mathbf{x}, z) \text{ for any } \mathbf{x} \in \mathcal{X}, z \in \mathcal{Z},$$

where

$$s_{\oplus}(\mathbf{x}, z) = \underset{y \in \mathcal{Y}}{\arg\min}\, S_{\oplus}(y; \mathbf{x}, z)$$

and

$$S_{\oplus}(y; \mathbf{x}, z) = E(d^2(Y, y)|Z = z) +$$

$$(\mathbf{x} - E(\mathbf{X}|Z = z))^T(\text{cov}(\mathbf{X} - E(\mathbf{X}|Z)))^{-1}\text{cov}(\mathbf{X} - E(\mathbf{X}|Z), d^2(Y, y) - E(d^2(Y, y)|Z)).$$

Definition 2. As $n \to \infty$ and $h \to 0$, the small ball probability of random objects $Z \in \mathcal{Z}$ and $Y \in \mathcal{Y}$ are defined as

$$\varphi_{\mathcal{Z},z}(h) = P(Z \in B_{\mathcal{Z}}(z, h)) \quad and \quad \varphi_{\mathcal{Y},y}(h) = P(Y \in B_{\mathcal{Y}}(y, h))$$

respectively, where

$$B_{\mathcal{Z}}(z, h) = \{z' \in \mathcal{Z}, \delta(z', z) \leq h\} \quad and \quad B_{\mathcal{Y}}(y, h) = \{z' \in \mathcal{Y}, d(y', y) \leq h\}.$$

Note: When $\mathcal{Z} = \mathcal{R}$ and is equipped with the Euclidean distance d, $\varphi_{\mathcal{Z},z}(h) = \int_{z-h}^{z+h} dF_Z$. Thus, in this case, $\varphi_{\mathcal{Z},z} = O(h)$.

Definition 3. Let Ω_1 be the set of probability distributions. The 2-Wasserstein metric distance between two distributions with CDFs $H(\cdot)$ and $G(\cdot)$ is defined as

$$d_W(H, G) = \sqrt{\int_0^1 (H^{-1}(t) - G^{-1}(t))^2 dt}.$$

We denote (Ω_1, d_W) as the metric space of probability distributions equipped with the Wasserstein distance.

Definition 4. Let Ω_2 be the set of symmetric, positive definite (SPD) matrices. Let $\mathbf{P}_1$ and $\mathbf{P}_2$ be two SPD matrices. Then, under the Cholesky decomposition, we can write $\mathbf{P}_1 = (\mathbf{P}_1^{1/2})^T\mathbf{P}_1^{1/2}$ and $\mathbf{P}_2 = (\mathbf{P}_2^{1/2})^T\mathbf{P}_2^{1/2}$, where $\mathbf{P}_1^{1/2}$ and $\mathbf{P}_2^{1/2}$ are upper triangle matrices with positive diagonal components. The Cholesky decomposition metric distance between two SPD matrices, $\mathbf{P}_1$ and $\mathbf{P}_2$, is defined as

$$d_C(\mathbf{P}_1, \mathbf{P}_2) = \sqrt{\text{trace}\left((\mathbf{P}_1^{1/2} - \mathbf{P}_2^{1/2})^T(\mathbf{P}_1^{1/2} - \mathbf{P}_2^{1/2})\right)}.$$

We denote (Ω_2, d_C) as the metric space of SPD matrices equipped with the Cholesky decomposition distance.

Definition 5. i) A function K from $\mathcal{R}$ into $\mathcal{R}^+$ such that $\int K = 1$ is called a kernel of type I if there exist two real constants $0 < C_1 < C_2 < \infty$ such that:

$$C_1 1_{[0,1]} \leq K \leq C_2 1_{[0,1]}.$$

ii) A function K from $\mathcal{R}$ into $\mathcal{R}^+$ such that $\int K = 1$ is called a kernel of type II if its support is $[0, 1]$ and if its derivative K' exists on $[0, 1]$ and satisfies for two real constants $-\infty < C_3 < C_4 < 0$:

$$C_3 \leq K' \leq C_4.$$

The following two definitions involve the concept of almost complete convergence, as defined in Ferraty and Vieu (2006).

Definition 6. One says that a sequence of real r.v.'s $(T_n)_{n\in\mathcal{N}}$ converges almost completely to some real $r.v.$ T, if and only if

$$\forall \epsilon > 0, \qquad \sum_{n\in\mathbb{N}} P(|T_n - T| > \epsilon) < \infty,$$

and the almost complete convergence of $(T_n)_{n\in\mathcal{N}}$ to T is denoted by

$$\lim_{n\to\infty} T_n = T_{a.co.}$$

It can be verified that convergence almost completely implies convergence almost surely and convergence in probability.

Definition 7. One says that the rate of almost complete convergence of $(T_n)_{n\in\mathbb{N}}$ to T is of order u_n if and only if

$$\exists\epsilon_o > 0, \qquad \sum_{n\in\mathbb{N}} P(|T_n - T| > \epsilon_0 \; u_n) < \infty,$$

and we write

$$T_n - T = O_{a.co.}(u_n).$$

S2 Assumptions

S2.1 Required for Local Linear Smoothing, Euclidean Z, and Euclidean X

Assumptions K1, L1 - L4, and P1 are analagous to Assumptions (K0), (L0) - (L3), and (P1) found in Petersen and Müller (2019), respectively.

Assumption K1. The kernel K is a probability density function, symmetric around zero. Furthermore, defining $K_{kj} = \int_{\mathfrak{R}} K^k(u)u^j du$, $|K_{14}|$ and $|K_{26}|$ are both finite.

Assumption L1. The object $s_\oplus(\mathbf{x}, z)$ exists and is unique. For all n, $\widetilde{s}_\oplus(\mathbf{x}, z)$ and $\widehat{s}_\oplus(\mathbf{x}, z)$ exist and are unique, the latter almost surely. Additionally, for any $\epsilon > 0$,

$$\inf_{d(y, s_\oplus(\mathbf{x},z))>\epsilon} \{S_\oplus(y; \mathbf{x}, z) - S_\oplus(s_\oplus(\mathbf{x}, z); \mathbf{x}, z)\} > 0$$

and

$$\liminf_{n\to\infty} \inf_{d(y, \widetilde{s}_\oplus(\mathbf{x},z))>\epsilon} \{\widetilde{S}_n(y; \mathbf{x}, z) - \widetilde{S}_n(\widetilde{s}_\oplus(\mathbf{x}, z); \mathbf{x}, z)\} > 0.$$

Assumption L2. The marginal densities $f_{\mathbf{X}}$ and f_Z, as well as the joint conditional densities g_y of $Z|Y = y$, h_y of $\mathbf{X}|Y = y$, and p_y of $\mathbf{X}, Z|Y = y$ exist and are twice continuously differentiable for $y \in \mathcal{Y}$. The joint conditional density $q_{\mathbf{x}}$ of $Z|\mathbf{X} = \mathbf{x}$ exists and is twice continuously differentiable for $\mathbf{x} \in \mathcal{X}$. For g_y and $q_{\mathbf{x}}$, $\sup_{y,z} |g_y^{''}(z)| < \infty$ and $\sup_{\mathbf{x},z} |q_{\mathbf{x}}^{''}(z)| < \infty$. Additionally, for any open $U \subset \mathcal{Y}$, $\int_U dF_{Y|Z}(y, z)$ is continuous as a function of z, and $\int_U dF_{Y|\mathbf{X}}(y, \mathbf{x})$ is continuous as a function of $\mathbf{x}$.

Assumption L3. There exists $\eta_1 > 0$, $C_6 > 0$, and $\beta_1 > 1$ such that

$$S_\oplus(y; \mathbf{x}, z) - S_\oplus(s_\oplus(\mathbf{x}, z); \mathbf{x}, z) \geq C_6 d(y, s_\oplus(\mathbf{x}, z))^{\beta_1},$$

provided $d(y, s_\oplus(\mathbf{x}, z)) < \eta_1$.

Assumption L4. There exists $\eta_2 > 0$, $C_7 > 0$, and $\beta_2 > 1$ such that

$$\liminf_{n\to\infty} \{\widetilde{S}_n(y; \mathbf{x}, z) - \widetilde{S}_n(\widetilde{s}_\oplus(\mathbf{x}, z); x, z)\} \geq C_7 d(y, \widetilde{s}_\oplus(\mathbf{x}, z))^{\beta_2}.$$

Note: When $\mathcal{Y} = \mathcal{R}$, with the Euclidean distance d, it can be verified that $\beta_1 = \beta_2 = 2$ (Petersen and Müller 2019).

Assumption P1. For the ball $B_{\mathcal{Y}}(s_{\oplus}(\mathbf{x}, z), h) \subset \mathcal{Y}$, let $N(\epsilon, B_{\mathcal{Y}}(s_{\oplus}(\mathbf{x}, z), h), d)$ be its covering number using balls of size ϵ. Then

$$\int_0^1 \sqrt{1 + \log\ N(h_\epsilon, B_{\mathcal{Y}}(s_{\oplus}(\mathbf{x}, z), h), d)}d\epsilon = O(1) \quad as \quad h \to 0.$$

S2.2 Required for Local Constant Smoothing, Non-Euclidean Z, and Euclidean X

Assumption N1. $\forall \epsilon > 0$, $P(Z \in B_{\mathcal{Z}}(z, \epsilon)) = \varphi_{\mathcal{Z},z}(\epsilon) > 0$.

This extends the assumption that the marginal density f of Z is strictly positive.

Assumption N2. $\lim\limits_{n\to\infty} h = 0$, $\lim\limits_{n\to\infty} \frac{\log n}{n\varphi_{\mathcal{Z},z}(h)} = 0$, and $\lim\limits_{n\to\infty} nh^2 = \infty$.

The following assumption allows us to still consider unbounded $\mathcal{Z}$.

Assumption N3. $\forall m \geq 1,\ \forall y \in \mathcal{Y},$ and $\forall z \in \mathcal{Z},\ E(d^{2m}(Y, y)|Z = z) < \sigma_{Ym}(z) < \infty$ and $E(|X_j|^m|Z = z) < \sigma_{Xjm}(z) < \infty$ for $j = 1, ..., p$, where $\sigma_{Ym}, \sigma_{X1m}, ..., \sigma_{Xpm}$ are continuous at z.

To control the effect of δ in the rate of convergence of the bias term, $d(s_{\oplus}(z), \widetilde{s}_{\oplus}(z))$, we make the following Lipschitz-type assumption.

Assumption N4. There exists $\beta_{0\mathbf{X}} > 0$ and $\beta_{0Y} > 0$ such that

$$E(X_j|Z) \in Lip_{\mathcal{Z},\beta_{0\mathbf{X}}}$$

for $j = 1, ..., p$, and for any $y \in \mathcal{Y}$,

$$E(d^2(Y, y)|Z) \in Lip_{\mathcal{Z},\beta_{0Y}},$$

where

$$Lip_{\mathcal{Z},\beta_{0\mathbf{X}}} = \{f : \mathcal{Z} \to \mathcal{R}, \exists C_0 > 0, \forall z, z' \in \mathcal{Z}, |f(z) - f(z')| < C_0\delta(z, z')^{\beta_{0\mathbf{X}}}\}$$

and

$$Lip_{\mathcal{Z},\beta_{0Y}} = \{f : \mathcal{Z}\times\mathcal{Y} \to \mathcal{R}, \exists C_0 > 0, \forall z, z' \in \mathcal{Z}, |f(y, z)-f(y, z')| < C_0\delta(z, z')^{\beta_{0Y}}\}.$$

Assumption K2. K is a kernel of type I or K is a kernel of type II and satisfies

$$\exists C_5 > 0,\ \exists\epsilon_0,\ \forall\epsilon < \epsilon_0,\ \int_0^{\epsilon} \varphi_{\mathcal{Z},z}(u)du > C_5\epsilon\varphi_{\mathcal{Z},z}(\epsilon).$$

S3 Theoretical Results and Proofs

S3.1 Case of Local Linear Smoothing, Euclidean Z, and Euclidean $\mathbf{X}$

Lemma 1. For the special case of partially linear regression model, we have $s_{\oplus}(\mathbf{x}, z) = \mathbf{x}^T\boldsymbol{\beta} + f(z)$ for any $\mathbf{x}$ and z.

Proof of Lemma 1. Recall that the partially linear model assumes that $Y = \mathbf{X}^T\boldsymbol{\beta} + f(Z) + \epsilon$ with $E(\epsilon|\mathbf{X}) = 0$ and $E(\epsilon|Z) = 0$. Consequently we have

$$\begin{aligned} E(d^2(Y,y)|Z=z) &= E((Y-y)^2|Z=z) = y^2 - 2yE(Y|Z=z) + E(Y^2|Z=z) \\ &= y^2 - 2y\left\{(E(\mathbf{X}|Z=z))^T\boldsymbol{\beta} + f(z)\right\} + E(Y^2|Z=z), \end{aligned}$$

and

$$\begin{aligned} & d^2(Y,y) - E(d^2(Y,y)|Z) \\ = \ & (Y-y)^2 - \left[y^2 - 2y\left\{(E(\mathbf{X}|Z))^T\boldsymbol{\beta} + f(Z)\right\} + E(Y^2|Z)\right] \\ = \ & -2y\left\{(\mathbf{X} - E(\mathbf{X}|Z))^T\boldsymbol{\beta} + \epsilon\right\} + Y^2 - E(Y^2|Z). \qquad \text{(S3.1)} \end{aligned}$$

It then implies that

$$\begin{aligned} & \mathrm{cov}(\mathbf{X} - E(\mathbf{X}|Z), d^2(Y,y) - E(d^2(Y,y)|Z)) \\ = \ & -2y\mathrm{cov}(\mathbf{X} - E(\mathbf{X}|Z))\boldsymbol{\beta} - 2y\mathrm{cov}(\mathbf{X} - E(\mathbf{X}|Z), \epsilon) + \mathrm{cov}(\mathbf{X} - E(\mathbf{X}|Z), Y^2 - E(Y^2|Z)) \\ = \ & -2y\mathrm{cov}(\mathbf{X} - E(\mathbf{X}|Z))\boldsymbol{\beta} + \mathrm{cov}(\mathbf{X} - E(\mathbf{X}|Z), Y^2 - E(Y^2|Z)). \end{aligned}$$

All of the above implies that the entire right hand side of (S1.1) is equal to

$$\begin{aligned} & y^2 - 2y\left\{(E(\mathbf{X}|Z=z))^T\boldsymbol{\beta} + f(z)\right\} - 2y(\mathbf{x} - E(\mathbf{X}|Z=z))^T\boldsymbol{\beta} + \text{extra terms} \\ = \ & y^2 - 2y\left\{\mathbf{x}^T\boldsymbol{\beta} + f(z)\right\} + \text{extra terms}, \qquad \text{(S3.2)} \end{aligned}$$

where the extra terms do not involve y. Therefore, the minimizer of (S3.2) is given by $\mathbf{x}^T\boldsymbol{\beta} + f(z)$, which completes the proof.

□

The following lemma is simply an extension of Lemma 1 found in Petersen and Müller (2019).

Lemma 2. Let $\tilde{\mu}_j(z) = E(K_h(Z-z)(Z-z)^j)$ and $\hat{\mu}_j(z) = \frac{1}{n}\sum_{i=1}^n K_h(z_i - z)(z_i - z)^j$, as well as $\tau_j(z,y) = E(K_h(Z-z)(Z-z)^j|Y=y)$ and $\gamma_j(\mathbf{x},z) = E(K_h(Z-z)(Z-z)^j|\mathbf{X}=\mathbf{x})$, for $j = 0,1,2$. If Assumptions K1 and L3 hold, then

$$\tilde{\mu}_j(z) = h^j[f_Z(z)K_{1j} + hf'_Z(z)K_{1(j+1)} + O(h^2)]$$

and $\hat{\mu}_j(z) = \tilde{\mu}_j(z) + O_p((h^{2j-1}n^{-1})^{1/2})$ for $j = 0,1,2$. Further,

$$\tau_j(z,y) = h^j[g_y(z)K_{1j} + hg'_y(z)K_{1(j+1)} + O(h^2)]$$

and

$$\gamma_j(\mathbf{x},z) = h^j[q_{\mathbf{x}}(z)K_{1j} + hq'_x(z)K_{1(j+1)} + O(h^2)].$$

Proof of Lemma 2. The results for $\tilde{\mu}_j(z)$, $\tau_j(z,y)$, and $\gamma_j(\mathbf{x},z)$ can be shown using a 2nd order Taylor expansion of $f_Z(z)$, $g_y(z)$, and $q_{\mathbf{x}}(z)$. The proof for $\hat{\mu}_j(z)$ can be found in Petersen and Müller (2019). □

Theorem 1. If assumptions P1, K1, L2, L1, and L3 hold, then

$$d(s_\oplus(\mathbf{x},z), \widetilde{s}_\oplus(\mathbf{x},z)) = O(h^{2/(\beta_1-1)})$$

as $h \to 0$.

Proof of Theorem 1. Recall equations (8), (9), (12), (13), and (14) in the main manuscript. Using the same proof technique as Theorem 3 in Petersen and Müller (2019), we can show that $\frac{dF_{Y|Z}(z,y)}{dF_Y(y)} = \frac{g_y(z)}{f_Z(z)}$ for all z such that $f_Z(z) > 0$. Then, by applying Lemma 2 and again following the proof of Theorem 3 in Petersen and Müller (2019), one can show that $E(\zeta_h(Z,z)|Y) = \frac{g_y(z)}{f_Z(z)} + O(h^2)$, where the error term is uniform over $y \in \mathcal{Y}$. Then, using the fact that $\frac{dF_{Y|Z}(z,y)}{dF_Y(y)} = \frac{g_y(z)}{f_Z(z)}$,

$$
\begin{aligned}
E\left(\zeta_h(Z,z)d^2(Y_i,y)\right) &= \int d^2(y',y)\zeta_h(z',z)dF_{Z,Y}(z,y) \\
&= \int d^2(y',y)\frac{g_y(z)}{f_Z(z)}dF_Y(y) + O(h^2) \\
&= \int d^2(y',y)dF_{Z|Y}(z,y) + O(h^2) \\
&= E(d^2(Y,y)|Z=z) + O(h^2).
\end{aligned}
$$

That is, using the notation in (10) and (12) in the main manuscript, we have that $\tilde{w}_0(y;z) = w_0(y;z) + O(h^2)$. Now, take a look at the second piece of (12). Once again using the techniques of Theorem 3 and applying Lemma 2 in Petersen and Müller (2019), we have that $E(\zeta_h(Z,z)\mathbf{X}) = \frac{q_{\mathbf{x}}(z)}{f_Z(z)} + O(h^2)$. Further, one can show that $\frac{dF_{X|Z}}{dF_X} = \frac{q_{\mathbf{x}}(z)}{f_Z(z)}$ using the same approach as above. Thus,

$$
E(\zeta_h(Z,z)\mathbf{X}) = E(\mathbf{X}|Z=z) + O(h^2).
$$

Putting this together, we get that $\tilde{w}_1(\mathbf{x},z)w_2^{-1}w_3(y) = w_1(\mathbf{x},z)w_2^{-1}w_3(y) +$

$O(h^2)$, and furthermore,

$$\begin{aligned}\widetilde{S}_n(y;\mathbf{x},z) &= w_0(y;z)+w_1(\mathbf{x},z)w_2^{-1}w_3(y)+O(h^2)\\ &= S_\oplus(y;\mathbf{x},z)+O(h^2).\end{aligned}$$

Then, by L1, we have that $d(s_\oplus(\mathbf{x},z),\widetilde{s}_\oplus(\mathbf{x},z))=o(1)$ as $h=h_n\to 0$. Next, define $r_h=h^{-\frac{\beta_1}{\beta_1-1}}$ and set $C_{j,n}=\{y: 2^{j-1}<r_h d(y,s_\oplus(\mathbf{x},z))^{\beta_1/2}\le 2^j\}$. Then, for any $M>0$, using similar arguments as Theorem 2 and Theorem 3 in Petersen and Müller (2019) and using L3, there exists $a>0$ such that, for large n,

$$\begin{aligned}I(r_h d(\tilde{s}(\mathbf{x},z),s_\oplus(\mathbf{x},z))^{\beta_1/2}>2^M) &\le a\sum_{j\ge M}\frac{2^{2j(1-\beta_1)/\beta_1}}{r_h^{2(1-\beta_1)/\beta_1}h^{-2}}\\ &\le a\sum_{j\ge M}\left(\frac{1}{4^{(\beta_1-1)/\beta_1}}\right)^j, \quad \text{(S3.3)}\end{aligned}$$

which converges because $\beta_1>1$. Thus, for some $M>0$, we have

$$d(\tilde{s}(\mathbf{x},z),s_\oplus(\mathbf{x},z))\le 2^{2M/\beta_1}h^{2/(\beta_1-1)}.$$

□

Lemma 3. Suppose assumptions K1 and L1 hold, $\mathcal{Y}$ is bounded and that $h\to 0$ and $nh^2\to\infty$ as $n\to\infty$. Then

$$d(\widetilde{s}_\oplus(\mathbf{x},z),\widehat{s}_\oplus(\mathbf{x},z))=o_p(1).$$

Proof of Lemma 3. First, as in Petersen and Müller (2019), we will show that $\widetilde{S}_n - \widehat{S}_n$ converges weakly to 0. Combining this with Assumption L1, we will have the result.

Let $\zeta_{hi}(z) \equiv \zeta_h(Z_i, z)$ for all $i = 1, ..., n$. We can write

$$\widehat{S}_n(y; \mathbf{x}, z) - \widetilde{S}_n(y; \mathbf{x}, z) = \hat{w}_0(y; z) - \tilde{w}_0(y; z) \tag{S3.4}$$

$$+ \hat{w}_1(\mathbf{x}, z)\hat{w}_2^{-1}\hat{w}_3(y) - \tilde{w}_1(\mathbf{x}, z)w_2^{-1}w_3(y) \tag{S3.5}$$

In Lemma 2 of Petersen and Müller (2019), it is shown that the term on the right hand side of (S3.4) is $O_p((nh)^{-1/2})$. Then we just need to show that (S3.5) is $O_p((nh)^{-1/2})$.

First of all, notice that we can easily extend the result of Lemma 2 to $\hat{w}_1(\mathbf{x}, z)$ and $\tilde{w}_1(\mathbf{x}, z)$. That is, we have that $\hat{w}_1(\mathbf{x}, z) = \tilde{w}_1(\mathbf{x}, z) + O_p((nh)^{-1/2})$.

Then, using Theorem 1 from Speckman (1988) and keeping our assumptions, we have that each element in the vector $\hat{w}_2^{-1}\hat{w}_3(y)$ converges in probability to the corresponding element in $w_2^{-1}w_3(y)$ at a rate of $O(hn^{-1/2}) + O_p(n^{-1/2})$.

Thus, the rate of convergence is dominated by (S3.4), and since convergence almost completely implies convergence in probability, we have that

$$\widehat{S}_n(y; \mathbf{x}, z) - \widetilde{S}_n(y; \mathbf{x}, z) = O_p((nh)^{-1/2}).$$

Therefore, we have that $\widetilde{S}_n(y;\mathbf{x},z) - \widehat{S}_n(y;\mathbf{x},z) = o_p(1)$ when $h \to 0$, $nh^2 \to \infty$ and $n \to \infty$. According to Theorem 1.5.4 in van der Vaart and Wellner (1996), we lastly need to show that for any $\eta > 0$,

$$\limsup_n P\left(\sup_{d(y_1,y_2)<\delta} |(\widetilde{S}_n - \widehat{S}_n)(y_1;\mathbf{x},z) - (\widetilde{S}_n - \widehat{S}_n)(y_2;\mathbf{x},z)| > \eta\right) \to 0 \tag{S3.6}$$

as $\delta \to 0$.

Since $E(|\zeta_{hi}(z)|) = O(1)$ and $E(\zeta_{hi}^2(z)) = O(h^{-1})$, we have that $n^{-1}\sum_{i=1}^n |s_{hi}(z)| = O_p(1)$. Then, $|\widehat{S}_n(y_1;\mathbf{x},z) - \widehat{S}_n(y_2;\mathbf{x},z)| \le 2\text{diam}(\mathcal{Y})d(y_1,y_2)n^{-1}\sum_{i=1}^n |s_{hi}(z)| = O_p(d(y_1,y_2))$. Similarly, $|\widetilde{S}_n(y_1;\mathbf{x},z) - \widetilde{S}_n(y_2;\mathbf{x},z)| = O(d(y_1,y_2))$. Thus, (S3.6) is verified. □

Theorem 2. If assumptions P1, K1, L1, and L4 hold, and if $h \to 0$ and $nh^2 \to \infty$, then

$$d(\widetilde{s}_\oplus(\mathbf{x},z), \widehat{s}_\oplus(\mathbf{x},z)) = O_p((nh)^{-\frac{1}{2(\beta_2-1)}}).$$

Proof of Theorem 2. We will follow similar arguments as Theorem 2 and Theorem 4 in Petersen and Müller (2019). Set $T_{n,h}(y;\mathbf{x},z) = \widehat{S}_n(y;\mathbf{x},z) - \widetilde{S}_n(y;\mathbf{x},z)$. Further, let

$$D_i = d^2(Y_i, y) - d^2(Y_i, \widetilde{s}_\oplus(\mathbf{x},z)). \tag{S3.7}$$

Then we have $|T_{n,h}(y;\mathbf{x},z) - T_{n,h}(\widetilde{s}_{\oplus}(\mathbf{x},z);\mathbf{x},z)|$ is less than or equal to

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{i=1}^{n}\left[s_{hi}(z)-\zeta_{hi}(z)\right]D_i\right| \quad \text{(S3.8)}\\
&+\left|\frac{1}{n}\sum_{i=1}^{n}\left[\zeta_{hi}(z)D_i - E(\zeta_{hi}(z)D_i)\right]\right|\\
&+\Bigg|\hat{w}_1(\mathbf{x},z)\hat{w}_2^{-1}\frac{1}{n}\sum_{i=1}^{n}\left[X_i - \frac{1}{n}\sum_{j=1}^{n}s_{hj}(Z_i)\mathbf{X}_j\right]\left[D_i - \frac{1}{n}\sum_{j=1}^{n}s_{hj}(Z_i)D_j\right]\\
&\qquad - \tilde{w}_1(\mathbf{x},z)w_2^{-1}E\left[(X_i - E(X_i|Z_i))(D_i - E(D_i|Z_i))\right]\Bigg|
\end{aligned}$$

Because $|D_i| \leq 2diam(\mathcal{Y})d(y,\widetilde{s}_{\oplus}(\mathbf{x},z))$, it is shown in Petersen and Müller (2019) that the first term of (S3.8) is $O_p(d(y,\widetilde{s}_{\oplus}(\mathbf{x},z)))$, which is independent of y and $\widetilde{s}_{\oplus}(\mathbf{x},z)$. To handle the second term of (S3.8), they further show that for small δ,

$$E\left(\sup_{d(y,\widetilde{s}_{\oplus}(\mathbf{x},z))<\delta}\left|\frac{1}{n}\zeta_{hi}(z)D_i - E[\zeta_{hi}(z)D_i]\right|\right) \;=\; O(\delta(nh)^{-1/2}).$$

Thus, we must work with the last absolute value term of (S3.8). Although this term is very messy, it is actually simpler than the first two terms, as it does not involve an expectation with $\zeta_{hi}(z)$. Therefore, we can directly again use the fact that $|D_i| \leq 2diam(\mathcal{Y})d(y,\widetilde{s}_{\oplus}(\mathbf{x},z))$ and combine this with our result from Lemma 3, we get that final term of (S3.8) is in fact also $O_p((nh)^{-1/2})$. This simplifies the rest of the proof to match that of Theorem 4 in Petersen and Müller (2019). That is, we can define

$$\begin{aligned}
B_R &= \Bigg\{ \sup_{d(y,\widetilde{s}_\oplus(\mathbf{x},z))<\delta} \left| \frac{1}{n} \sum_{i=1}^{n} \left[s_{hi}(z) - \zeta_{hi}(z) \right] D_i \right| \\
&+ \Bigg| \hat{w}_1(\mathbf{x}, z)\hat{w}_2^{-1} \frac{1}{n} \sum_{i=1}^{n} \left[X_i - \frac{1}{n} \sum_{j=1}^{n} s_{hj}(Z_i)\mathbf{X}_j \right] \left[D_i - \frac{1}{n} \sum_{j=1}^{n} s_{hj}(Z_i) D_j \right] \\
&- \tilde{w}_1(\mathbf{x}, z) w_2^{-1} E\left[(X_i - E(X_i|Z_i))(D_i - E(D_i|Z_i)) \right] \Bigg| \leq R\delta(nh)^{-1/2} \Bigg\},
\end{aligned}$$

where $\delta > 0$, $R > 0$ and so that $P(B_R^C) \to 0$. Therefore, as in Petersen and Müller (2019), we have

$$E\left(I_{B_R} \sup_{d(y,\widetilde{s}_\oplus(\mathbf{x},z))<\delta} |T_{n,h}(y; \mathbf{x}, z) - T_{n,h}(\widetilde{s}_\oplus(\mathbf{x}, z); x, z)| \right) \leq \frac{a\delta}{(nh)^{1/2}},$$

where a depends on R and Assumption L4. The rest of this proof aligns with Petersen and Müller (2019). Thus, we have that $d(\widehat{s}_\oplus(\mathbf{x}, z), \widetilde{s}_\oplus(\mathbf{x}, z)) = O_p((nh)^{-\frac{1}{2(\beta_2-1)}})$.

□

Corollary 1. Under the assumptions of Theorem 1 and Theorem 2,

$$d(s_\oplus(\mathbf{x}, z), \widehat{s}_\oplus(\mathbf{x}, z)) = O(h^{2/(\beta_1-1)}) + O_p((nh)^{-\frac{1}{2(\beta_2-1)}}).$$

Proof of Corollary 1. This follows from applying the triangle inequality to the results of Theorem 1 and Theorem 2. □

S3.2 Case of Local Constant Smoothing and Non-Euclidean Z

Theorem 3. Suppose assumptions P1, K2, L1, L3, and N2 - N4 hold. If $h \to 0$, then

$$d(s_\oplus(z), \widetilde{s}_\oplus(z)) = O(h^{\beta_{0Y}/(\beta_1-1)}).$$

Proof of Theorem 3. First, notice that the distance d is just a map from $\mathcal{Y} \times \mathcal{Y}$ to $\mathcal{R}$. Thus, utilizing Lemma 6.12 in Ferraty and Vieu (2006), we have that under assumptions K2, N3, and N4,

$$\widetilde{S}_n(y; z) \quad = \quad E(d^2(Y, y)|Z = z) + O(h^{\beta_{0Y}})$$

for any $y \in \mathcal{Y}$ because $K \in [0, 1]$ and $E(\zeta_h(Z, z)) = 1$. By Assumption L1, this means that $d(s_\oplus(z), \widetilde{s}_\oplus(z)) = o(1)$ as $h \to 0$, since $\beta_{0Y} > 0$.

Next, define $r_h = h^{-\frac{\beta_{0Y}\beta_1}{2(\beta_1-1)}}$ and set $D_{j,n} = \{y : 2^{j-1} < r_h d(y, s_\oplus(z))^{\beta_1/2} \leq 2^j\}$. Then following the same arguments as Theorem 3 in Petersen and Müller (2019) and using Assumption L3, we have that for some $M > 0$,

$$d(\widetilde{s}_\oplus(z), s_\oplus(z)) \leq 2^{2M/\beta_1} h^{\beta_{0Y}/(\beta_1-1)}$$

for large n. Thus, $d(\widetilde{s}_\oplus(z), s_\oplus(z)) = O(h^{\beta_{0Y}/(\beta_1-1)})$. □

Lemma 4. If assumptions K2, L1, and N1-N3 hold, and $\mathcal{Y}$ is bounded, then

$$d(\widetilde{s}_\oplus(z), \widehat{s}_\oplus(z)) = o_p(1).$$

Proof of Lemma 4. Let $\widehat{S}_n(y;z) = \frac{1}{n}\sum_{i=1}^n s_{hi}(z)d^2(Y_i,y)$. We first look at the difference $\widehat{S}_n(y;z) - \widetilde{S}_n(y;z)$

$$= n^{-1}\sum_{i=1}^n \frac{K_h(\delta(Z_i,z))d^2(Y_i,y)}{n^{-1}\sum_{i=1}^n K_h(\delta(Z_i,z))} - \frac{E(K_h(\delta(Z_i,z))d^2(Y_i,y))}{E(K_h(\delta(Z_i,z)))}. \quad \text{(S3.9)}$$

Then, we can use the proof and results of Lemma 6.3 in Ferraty and Vieu (2006). By noting that (S3.9) can be written as

$$\frac{n^{-1}\sum_{i=1}^n K_h(\delta(Z_i,z))d^2(Y_i,y) \times E(K_h(\delta(Z_i,z)))}{E(K_h(\delta(Z_i,z))) \times n^{-1}\sum_{i=1}^n K_h(\delta(Z_i,z))} - \frac{E(K_h(\delta(Z_i,z))d^2(Y_i,y))}{E(K_h(\delta(Z_i,z)))},$$

we find that (S3.9) is $O_{a.co}\left(\sqrt{\frac{\log n}{n\varphi_{\mathcal{Z},z}(h)}}\right)$. Because convergence almost completely implies convergence in probability, we have then shown that $\widehat{S}_n(y;z) - \widetilde{S}_n(y;z) = o_p(1)$ for any $y \in \mathcal{Y}$, since we have assumed that $\lim_{n\to\infty} \frac{\log n}{n\varphi_{\mathcal{Z},z}(h)} = 0$. Then, according to Van der Vaart and Wellner (1996), the last thing we need to show is that for any $\eta > 0$,

$$\limsup_n P\left(\sup_{d(y_1,y_2)<h'} |(\widetilde{S}_n - \widehat{S}_n)(y_1;z) - (\widetilde{S}_n - \widehat{S}_n)(y_2;z)| > \eta\right) \to 0 \text{ as } h' \to 0.$$

Since the kernel function is assumed to be non-negative, we have that $n^{-1}\sum_{i=1}^n |s_{hi}(z)| = 1$. Then,

$|\widehat{S}_n(y_1;z) - \widehat{S}_n(y_2;z)| \leq 2\text{diam}(\mathcal{Y})d(y_1,y_2)n^{-1}\sum_{i=1}^n |s_{hi}(z)| = O_p(d(y_1,y_2))$.

Similarly, $|\widetilde{S}_n(y_1;z) - \widetilde{S}_n(y_2;z)| = O(d(y_1,y_2))$, which verifies the above. □

Theorem 4. If assumptions P1, K2 , L1, L4, and N1-N3 hold, then

$$d(\widetilde{s}_\oplus(z), \widehat{s}_\oplus(z)) = O_p((n\varphi_{\mathcal{Z},z}(h))^{\frac{-1}{2(\beta_2-1)}}).$$

Proof of Theorem 4. We utilize similar arguments as the proof of Theorem 4 in Petersen and Müller (2019). Define $T_{n,h}(y;z) = \widehat{S}_n(y;z) - \widetilde{S}_n(y;z)$. Letting

$$D_i(y,z) = d^2(Y_i, y) - d^2(Y_i, \widetilde{s}_\oplus(z)),$$

we have

$$\begin{aligned} |T_{n,h}(y;z) - T_{n,h}(\widetilde{s}_\oplus(z);z)| &\leq \left| n^{-1} \sum_{i=1}^{n} [s_{hi}(z) - \zeta_{hi}(z)] D_i(y,z) \right| \quad \text{(S3.10)} \\ &+ \left| n^{-1} \sum_{i=1}^{n} (\zeta_{hi}(z) D_i - E[\zeta_{hi}(z) D_i(y)]) \right|, \end{aligned}$$

where we notate $\zeta_{hi} = K_h(\delta(Z_i, z))/E[K_h(\delta(Z_i, z))]$.

Note that $|D_i(y,z)| \leq 2\text{diam}(\mathcal{Y}) d(y, \widetilde{s}_\oplus(z))$. We have

$E[n^{-1} \sum_{i=1}^n K_h(\delta(Z_i, z))] = E[K_h(\delta(Z, z))]$ and

$E(K^2(h^{-1}\delta(Z,z))) \leq C_2 \varphi_{\mathcal{Z},z}(h)$ by applying Lemma 4.3 and 4.4 from Ferraty and Vieu (2006). Thus, $n^{-1} \sum_{i=1}^n K_h(\delta(Z_i, z)) = E[K_h(\delta(Z, z))] + O_p((n^{-1}\varphi_{\mathcal{Z},z}(h))^{1/2})$. These results imply that the first term on the right hand side of (S3.10) is

$O_p(d(y, \widetilde{s}_\oplus(z))(n\varphi_{\mathcal{Z},z}(h))^{-1/2})$, where the O_p term is independent of y and $\widetilde{s}_\oplus(z)$. Then, we can define

$$B_R \left\{ \sup_{d(y, \widetilde{s}_\oplus(z)) < h'} \left| n^{-1} \sum_{i=1}^{n} [s_{hi}(z) - \zeta_{hi}(z)] D_i(y,z) \right| \leq R h' (n\varphi_{\mathcal{Z},z}(h))^{-1/2} \right\}$$

for $R > 0$, so that $P(B_R^C) \to 0$. Next, to control the second term on the right hand side of (S3.10), define functions $m_Y : \mathcal{Z} \times \mathcal{Y} \to \mathcal{R}$ by

$$m_Y(z, y) = \frac{K_h(\delta(Z, z)) d^2(Y, y)}{E(K_h(\delta(Z, z)))}$$

and the corresponding function class $\mathcal{M}_{n,h'} = \{m_y - m_{\widetilde{s}_\oplus(z)} : d(y, \widetilde{s}_\oplus(z)) < h'\}$. An envelope function for $\mathcal{M}_{nh'}$ is

$$M_{nh'}(z) = \frac{2\text{diam}(\mathcal{Y}) h' K_h(\delta(Z, z))}{E(K_h(\delta(Z, z)))},$$

And $E(M_{nh'}^2(z)) = O(h'^2 \varphi_{\mathcal{Z},z}^{-1}(h))$. Using this fact together with Theorems 2.7.11 and 2.14.2 of Van der Vaart and Wellner (1996) and Assumption P1, for small h',

$$E\left(\sup_{d(y, \widetilde{s}_\oplus(z)) < h'} \left| n^{-1} \zeta_{hi}(z) D_i(y, z) - E[\zeta_{hi}(z) D_i(y, z)] \right| \right) = O(h'(n\varphi_{\mathcal{Z},z}(h))^{-1/2}).$$

Combining this with (S3.10) and the definition of B_R,

$$E\left(I_{B_R} \sup_{d(y, \widetilde{s}_\oplus(z)) < h'} |T_{n,h}(y; z) - T_{n,h}(\widetilde{s}_\oplus(z); z)| \right) \leq \frac{a h'}{(n\varphi_{\mathcal{Z},z}(h))^{1/2}},$$

where I_{B_R} is the indicator function for the set B_R and a is a constant depending on R and the entropy integral in Assumption P1. To complete the proof, set $T_{n,h} = (n\varphi_{\mathcal{Z},z}(h))^{\frac{\beta_2}{4(\beta_2 - 1)}}$ and define

$$Q_{j,n} = \{y : 2^{j-1} < T_{n,h} d(y, \widetilde{s}_\oplus(z))^{\beta_2/2} \leq 2^j\}.$$

Choose η_2 satisfying Assumption L4 and such that Assumption N1 is satisfied for any $h' < \eta_2$. Set $\tilde{\eta} := (\eta_2/2)^{\beta_2/2}$. For any integer M,

$$P(T_{n,h} d(\widetilde{s}_\oplus(z), \widehat{s}_\oplus(z))^{\beta_2/2} > 2^M)$$

$$\leq P(B_R^C) + P(2d(\widetilde{s}_\oplus(z), \widehat{s}_\oplus(z)) > \eta) \quad \text{(S3.11)}$$
$$+ \sum_{j \geq M, 2^j \leq T_{n,h}\tilde{\eta}} P\left(\left\{ \sup_{y \in Q_{j,n}} |T_{n,h}(y;z) - T_{n,h}(\widetilde{s}_\oplus(z);z)| \geq C \frac{2^{2(j-1)}}{t_n^2} \right\} \cap B_R \right),$$

where the last term goes to 0 by Lemma 4. Since $d(y, \widetilde{s}_\oplus(z)) < (2^j/T_{n,h})^{2/\beta_2}$ on $Q_{j,n}(z)$, it is implied that the sum on the right hand side of (S3.11) is bounded by

$$4aC^{-1} \sum_{j \geq M, 2^j \leq T_{n,h}\tilde{\eta}} \frac{2^{2j(1-\beta_2)/\beta_2}}{T_{n,h}^{2(1-\beta_2)/\beta_2} (n\varphi_{\mathcal{Z},z}(h))^{1/2}} \leq 4aC^{-1} \sum_{j \geq M} \left(\frac{1}{4^{(\beta_2-1)/\beta_2}} \right)^j,$$

which converges since $\beta_2 > 1$. Thus,

$$d(\widetilde{s}_\oplus(z), \widehat{s}_\oplus(z)) = O_p(T_{n,h}^{2/\beta_2}) = O_p((n\varphi_{\mathcal{Z},z}(h))^{\frac{-1}{2(\beta_2-1)}}).$$

□

Corollary 2. Under the assumptions of Theorem 3 and Theorem 4, we have

$$d(s_\oplus(z), \widehat{s}_\oplus(z)) = O(h^{\frac{\beta_{0Y}}{(\beta_1-1)}}) + O_p((n\varphi_{\mathcal{Z},z}(h))^{\frac{-1}{2(\beta_2-1)}}).$$

Proof of Corollary 2. This follows from applying the triangle inequality to the results of Theorem 3 and Theorem 4. □

S3.3 Case of Local Constant Smoothing, Non-Euclidean Z, and Euclidean $\mathbf{X}$

Theorem 5. Suppose assumptions P1, K2, L1, L3, N2 - N4 hold. Then

$$d(s_{\oplus}(\mathbf{x}, z), \widetilde{s}_{\oplus}(\mathbf{x}, z)) = O(h^{\beta_0/(\beta_1-1)}),$$

where $\beta_0 = \min\{\beta_{0\mathbf{X}}, \beta_{0Y}\}$.

Proof of Theorem 5. Let $\beta_0 = \min\{\beta_{0\mathbf{X}}, \beta_{0Y}\} > 0$. Our first goal will be to show that $\widetilde{S}_n(y; \mathbf{x}, z) = S_{\oplus}(y; \mathbf{x}, z) + O(h^{\beta_0})$. That is, we need to show that

$$\tilde{w}_0(y; z) + \tilde{w}_1(\mathbf{x}, z) w_2^{-1} w_3(y) = w_0(y; z) + w_1(\mathbf{x}, z) w_2^{-1} w_3(y) + O(h^{\beta_0}).$$

Recall from Theorem 3, we have that $\tilde{w}_0(y; z) = w_0(y; z) + O(h^{\beta_{0Y}})$. Now, let us write $\tilde{w}_1(\mathbf{x}, z) = (\tilde{w}_{1,1}(x_1, z), ..., \tilde{w}_{1,p}(x_p, z))^T$ and $w_1(\mathbf{x}, z) = (w_{1,1}(x_1, z), ..., w_{1,p}(x_p, z))^T$, as each element of $\tilde{w}_1(\mathbf{x}, z)$ and $w_1(\mathbf{x}, z)$ corresponds to an element in the random vector $\mathbf{X} \in \mathcal{R}^p$. Then, using Lemma 6.12 from Ferraty and Vieu (2006) and Assumptions K2, N3, and N4, we have that $\tilde{w}_{1,j}(x_j, z) = w_{1,j}(x_j, z) + O(h^{\beta_{0\mathbf{X}}})$ for each $j = 1, ..., p$.

Therefore, $\widetilde{S}_n(y; \mathbf{x}, z) - S_{\oplus}(y; \mathbf{x}, z) = O(h^{\beta_{0Y}}) + O(h^{\beta_{0\mathbf{X}}}) = O(h^{\beta_0})$. Then, by L1, we have $d(s_{\oplus}(\mathbf{x}, z), \widetilde{s}_{\oplus}(\mathbf{x}, z)) = o(1)$ as $h = h_n \to 0$, since $\beta_0 > 0$.

Next, define $r_h = h^{-\frac{\beta_0 \beta_1}{2(\beta_1-1)}}$ and set $D_{j,n} = \{y : 2^{j-1} < r_h d(y, s_{\oplus}(z))^{\beta_1/2} \leq 2^j\}$. Then following the same arguments as Theorem 3 in Petersen and

Müller (2019) and using Assumption L3, we have that for some $M > 0$,

$$d(\widetilde{s}_{\oplus}(z), s_{\oplus}(z)) \leq 2^{2M/\beta_1} h^{\beta_0/(\beta_1-1)}$$

for large n. Thus, $d(\widetilde{s}_{\oplus}(z), s_{\oplus}(z)) = O(h^{\beta_0/(\beta_1-1)})$. □

Lemma 5. If assumptions K2, L1, and N1-N3 hold, and $\mathcal{Y}$ is bounded, then

$$d(\widetilde{s}_{\oplus}(\mathbf{x}, z), \widehat{s}_{\oplus}(\mathbf{x}, z)) = o_p(1).$$

Proof of Lemma 5. Recall equations (15) - (21) in the main manuscript.We can write

$$\widehat{S}_n(y; \mathbf{x}, z) - \widetilde{S}_n(y; \mathbf{x}, z) = \hat{w}_0(\mathbf{x}, z) - \tilde{w}_0(\mathbf{x}, z) \tag{S3.12}$$

$$+ \hat{w}_1(\mathbf{x}, z)\hat{w}_2^{-1}\hat{w}_3(y) - \tilde{w}_1(\mathbf{x}, z)w_2^{-1}w_3(y) \tag{S3.13}$$

In Lemma 4, we showed that $\hat{w}_0(y; z) = \tilde{w}_0(y; z) + O_{a.co.}\left(\sqrt{\frac{\log n}{n\varphi_{\mathcal{Z},z}(h)}}\right)$. Now, let us write $\hat{w}_1(\mathbf{x}, z) = (\hat{w}_{1,1}(x_1, z), ..., \hat{w}_{1,p}(x_p, z))^T$ and $\tilde{w}_1(\mathbf{x}, z) = (\tilde{w}_{1,1}(x_1, z), ..., \tilde{w}_{1,p}(x_p, z))^T$. It can similarly be shown that $\hat{w}_{1,j}(x_j, z) = \tilde{w}_{1,j}(x_j, z) + O_{a.co.}\left(\sqrt{\frac{\log n}{n\varphi_{\mathcal{Z},z}(h)}}\right)$ for $j = 1, .., p$.

Then, using Theorem 1 from Speckman (1988) and keeping our assumptions, we have that each element in the vector $\hat{w}_2^{-1}\hat{w}_3(y)$ converges in probability to the corresponding element in $w_2^{-1}w_3(y)$ at a rate of $O(n^{-1/2}) + O_p(n^{-1/2})$.

Thus, the rate of convergence is dominated by (S3.12), and since convergence almost completely implies convergence in probability, we have that

$$\widehat{S}_n(y;\mathbf{x},z) - \widetilde{S}_n(y;\mathbf{x},z) = O_p\left(\sqrt{\frac{\log n}{n\varphi_{\mathcal{Z},z}(h)}}\right),$$

which does not differ from the result of Lemma 4. Therefore, we have that $\widehat{S}_n(y;\mathbf{x},z) - \widetilde{S}_n(y;\mathbf{x},z) = o_p(1)$ for any $y \in \mathcal{Y}$. Then, according to Van der Vaart and Wellner (1996), the last thing we need to show is that for any $\eta > 0$,

$$\limsup_{n\to\infty} P\left(\sup_{d(y_1,y_2)<h'} |(\widetilde{S}_n - \widehat{S}_n)(y_1;\mathbf{x},z) - (\widetilde{S}_n - \widehat{S}_n)(y_2;\mathbf{x},z)| > \eta\right) \to 0 \text{ as } h' \to 0.$$

We can write

$$\begin{aligned}
|\widehat{S}_n(y_1;\mathbf{x},z) - \widehat{S}_n(y_2;\mathbf{x},z)| &= |\hat{w}_0(y_1,z) - \hat{w}_0(y_2,z) + \hat{w}_1(\mathbf{x},z)\hat{w}_2^{-1}[\hat{w}_3(y_1) - \hat{w}_3(y_2)]| \\
&\leq |\hat{w}_0(y_1,z) - \hat{w}_0(y_2,z)| \\
&+ \left|\hat{w}_1(\mathbf{x},z)\hat{w}_2^{-1}\right| |[\hat{w}_3(y_1) - \hat{w}_3(y_2)]|. \qquad \text{(S3.14)}
\end{aligned}$$

From Lemma 4, we have that the first term of (S3.14) is $O_p(d(y_1,y_2))$, because the kernel function is assumed to be non-negative and so $n^{-1}\sum_{i=1}^n |s_{hi}(z)| = 1$. Further, then, $\left|\hat{w}_1(\mathbf{x},z)\hat{w}_2^{-1}\right| |[\hat{w}_3(y_1) - \hat{w}_3(y_2)]| = O_p(d(y_1,y_2))$. Thus, $|\widehat{S}_n(y_1;\mathbf{x},z) - \widehat{S}_n(y_2;\mathbf{x},z)| \leq 2\text{diam}(\mathcal{Y})d(y_1,y_2)n^{-1}\sum_{i=1}^n |s_{hi}(z)| = O_p(d(y_1,y_2))$. Similarly, $|\widetilde{S}_n(y_1;\mathbf{x},z) - \widetilde{S}_n(y_2;\mathbf{x},z)| = O(d(y_1,y_2))$, which verifies the lemma.

□

Theorem 6. If assumptions P1, K2, L1, L4, and N1-N3 hold, then

$$d(\widetilde{s}_\oplus(\mathbf{x}, z), \widehat{s}_\oplus(\mathbf{x}, z),) = O_p((n\varphi_{\mathcal{Z},z}(h))^{\frac{-1}{2(\beta_2-1)}}).$$

Proof of Theorem 6. We will follow similar arguments as Theorem 2 and Theorem 4 in Petersen and Müller (2019). Set $T_{n,h}(y; \mathbf{x}, z) = \widehat{S}_n(y; \mathbf{x}, z) - \widetilde{S}_n(y; \mathbf{x}, z)$. Further, let

$$D_i = d^2(Y_i, y) - d^2(Y_i, \widetilde{s}_\oplus(\mathbf{x}, z)) \tag{S3.15}$$

Denote $\zeta_{hi}(z) \equiv \zeta_h(Z_i, z)$. Then we have $|T_{n,h}(y; \mathbf{x}, z) - T_{n,h}(\widetilde{s}_\oplus(\mathbf{x}, z); \mathbf{x}, z)|$ is less than or equal to

$$\begin{aligned} & \left| \frac{1}{n} \sum_{i=1}^{n} \left[s_{hi}(z) - \zeta_{hi}(z) \right] D_i \right| \qquad \text{(S3.16)} \\ + & \left| \frac{1}{n} \sum_{i=1}^{n} \left[\zeta_{hi}(z) D_i - E(\zeta_{hi}(z) D_i) \right] \right| \\ + & \left| \hat{w}_1(\mathbf{x}, z) \hat{w}_2^{-1} \frac{1}{n} \sum_{i=1}^{n} \left[X_i - \frac{1}{n} \sum_{j=1}^{n} s_{hj}(Z_i) \mathbf{X}_j \right] \left[D_i - \frac{1}{n} \sum_{j=1}^{n} s_{hj}(Z_i) D_j \right] \right. \\ - & \left. \tilde{w}_1(\mathbf{x}, z) w_2^{-1} E\left[(X_i - E(X_i|Z_i))(D_i - E(D_i|Z_i)) \right] \right| \end{aligned}$$

From Theorem 4, we have that the first term is $O_p(d(y, \widetilde{s}_\oplus(z))(n\varphi_{\mathcal{Z},z}(h))^{-1/2})$ and is independent of y and $\widetilde{s}_\oplus(\mathbf{x}, z)$. Further, we can define

$$B_R \left\{ \sup_{d(y, \widetilde{s}_\oplus(z)) < h'} \left| n^{-1} \sum_{i=1}^{n} [s_{hi}(z) - \zeta_{hi}(z)] D_i(y, z) \right| \leq R h' (n\varphi_{\mathcal{Z},z}(h))^{-1/2} \right\}$$

for $R > 0$, so that $P(B_R^C) \to 0$. Next, to control the second term on the right hand side of (S3.16), define functions $m_Y : \mathcal{Z} \times \mathcal{Y} \to \mathcal{R}$ by

$$m_Y(z, y) = \frac{K_h(\delta(Z, z))d^2(Y, y)}{E(K_h(\delta(Z, z)))}$$

and the corresponding function class $\mathcal{M}_{n,h'} = \{m_y - m_{\widetilde{s}_\oplus(z)} : d(y, \widetilde{s}_\oplus(z)) < h'\}$. An envelope function for $\mathcal{M}_{nh'}$ is

$$M_{nh'}(z) = \frac{2\mathrm{diam}(\mathcal{Y})h' K_h(\delta(Z, z))}{E(K_h(\delta(Z, z)))},$$

and $E(M_{nh'}^2(z)) = O(h'^2 \varphi_{\mathcal{Z},z}^{-1}(h))$. Using this fact together with Theorems 2.7.11 and 2.14.2 of Van der Vaart and Wellner (1996) and Assumption P1, for small h',

$$E\left(\sup_{d(y,\widetilde{s}_\oplus(z))<h'} \left| n^{-1}\zeta_{hi}(z)D_i - E[\zeta_{hi}(z)D_i] \right| \right) = O(h'(n\varphi_{\mathcal{Z},z}(h))^{-1/2}).$$

Combining this with (S3.10) and the definition of B_R,

$$E\left(I_{B_R} \sup_{d(y,\widetilde{s}_\oplus(z))<h'} |T_{n,h}(y; \mathbf{x}, z) - T_{n,h}(\widetilde{s}_\oplus(z); \mathbf{x}, z)| \right) \leq \frac{ah'}{(n\varphi_{\mathcal{Z},z}(h))^{1/2}},$$

where I_{B_R} is the indicator function for the set B_R and a is a constant depending on R and the entropy integral in Assumption P1. To complete the proof, set $T_{n,h} = (n\varphi_{\mathcal{Z},z}(h))^{\frac{\beta_2}{4(\beta_2-1)}}$ and define

$$Q_{j,n} = \{y : 2^{j-1} < T_{n,h} d(y, \widetilde{s}_\oplus(z))^{\beta_2/2} \leq 2^j\}.$$

Choose η_2 satisfying Assumption L4 and such that Assumption N1 is satisfied for any $h' < \eta_2$. Set $\tilde{\eta} := (\eta_2/2)^{\beta_2/2}$. For any integer M,

$$P(T_{n,h} d(\widetilde{s}_\oplus(z), \widehat{s}_\oplus(z))^{\beta_2/2} > 2^M) \leq P(B_R^C) + P(2d(\widetilde{s}_\oplus(z), \widehat{s}_\oplus(z)) > \eta) \quad \text{(S3.17)}$$
$$+ \sum_{j \geq M, 2^j \leq T_{n,h}\tilde{\eta}} P\left(\left\{\sup_{y \in Q_{j,n}} |T_{n,h}(y; \mathbf{x}, z) - T_{n,h}(\widetilde{s}_\oplus(z); \mathbf{x}, z)| \geq C\frac{2^{2(j-1)}}{t_n^2}\right\} \cap B_R\right),$$

where the last term goes to 0 by Lemma 4. Since $d(y, \widetilde{s}_\oplus(z)) < (2^j/T_{n,h})^{2/\beta_2}$ on $Q_{j,n}(z)$, it is implied that the sum on the right hand side of (S3.17) is bounded by

$$4aC^{-1} \sum_{j \geq M, 2^j \leq T_{n,h}\tilde{\eta}} \frac{2^{2j(1-\beta_2)/\beta_2}}{T_{n,h}^{2(1-\beta_2)/\beta_2}(n\varphi_{\mathcal{Z},z}(h))^{1/2}} \leq 4aC^{-1} \sum_{j \geq M} \left(\frac{1}{4^{(\beta_2-1)/\beta_2}}\right)^j,$$

which converges since $\beta_2 > 1$ The rest of this proof aligns with Petersen and Müller (2019). Thus, we have that $d(\widehat{s}_\oplus(\mathbf{x}, z), \widetilde{s}_\oplus(\mathbf{x}, z)) = O_p((n\varphi_{\mathcal{Z},z}(h))^{-\frac{1}{2(\beta_2-1)}})$.

□

Corollary 3. If the assumptions of Theorems 5 and 6 hold, then

$$d(s_\oplus(\mathbf{x}, z), \widehat{s}_\oplus(\mathbf{x}, z),) = O(h^{\frac{\beta_0}{(\beta_1-1)}}) + O_p((n\varphi_{\mathcal{Z},z}(h))^{\frac{-1}{2(\beta_2-1)}}),$$

where $\beta_0 = \min\{\beta_{0\mathbf{X}}, \beta_{0Y}\}$.

Proof of Corollary 3. This follows from applying the triangle inequality to the results of Theorem 5 and Theorem 6. □

References

Ferraty, F. and P. Vieu (2006). *Nonparametric Functional Data Analysis*. New York: Springer.

Petersen, A. and H.-G. Müller (2019). Fréchet regression for random objects with euclidean predictors. *The Annals of Statistics 47*(2), 691–719.

Speckman, P. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society. Series B (Methodological) 50*(3), 413–436.

Van der Vaart, A. and J. Wellner (1996). *Weak Convergence and Empirical Processes.* New York: Springer.