Varying-Coefficient Fréchet Regression

Yanzhao Wang*, Jianqiang Zhang† and Wangli Xu*

**Center for Applied Statistics and School of Statistics,Renmin University of China*

† *School of Statistics and Data Science, Zhejiang Gongshang University*

Supplementary Material

This supplementary material contains additional implementation and simulation details, together with the proofs of the main results in the paper. Section S1 of the Supplementary Material collects additional implementation and simulation materials, including computational complexity and numerical optimization, small-ball probability illustrations, supplementary comparisons with flexible Fréchet regression methods, small-sample simulations, runtime summaries, as well as explanations and verifications of assumptions. We then provide the theoretical proofs. In Section S2, we show that explicit solutions of the minimization problems defining the varying-coefficient Fréchet regression are available when the response space is a Hilbert space. Section S3 contains the proofs for the varying-coefficient Fréchet regression with a Euclidean predictor U, including the asymptotic distribution results for the case where the response space is a Hilbert space. Finally, Section S4 provides the proofs for the case where U lies in a non-Euclidean metric space.

S1 Additional implementation and simulation details

S1.1 Computational complexity and numerical optimization

For computational complexity, at a fixed target point $(\boldsymbol{x}, u)$, the computation of $\widehat{s}_{\oplus}(\boldsymbol{x}, u)$ can be decomposed into two parts: the computation of the local regression weights $\{c_i(\boldsymbol{x}, u)\}_{i=1}^{n}$ and the minimization of the weighted Fréchet objective over the response space. The first part is common across response types and has complexity $O(nC_\delta + np^2 + p^3)$, where p is the dimension of $\boldsymbol{X}$ and C_δ denotes the cost of evaluating the metric $\delta(\cdot, \cdot)$ once on the predictor space of U. Specifically, evaluating the smoothing weights requires computing the n distances $\{\delta(U_i, u)\}_{i=1}^{n}$, which costs $O(nC_\delta)$. For Euclidean U with the standard Euclidean metric, $C_\delta = O(1)$. For distribution valued U under the 2-Wasserstein metric d_W, the distance is computed via discretized quantile functions. If m_U is the number of grid points used for this discretization, then $C_\delta = O(m_U)$. For SPD matrix valued U under the Cholesky metric $d_C\left(P_1, P_2\right) = \|u_1^{1/2} - u_2^{1/2}\|_F$, computing the distance requires obtaining the unique Cholesky factor $u^{1/2}$ for an $r_U \times r_U$ matrix, an operation that costs $O\left(r_U^3\right)$, yielding $C_\delta = O\left(r_U^3\right)$.Then, the formation of the weighted covariance matrix $\widehat{\Sigma}_{\boldsymbol{X}|u}$, which costs $O(np^2)$, and the inversion of this $p \times p$ matrix, which costs $O(p^3)$. Other operations, such as computing the smoothed conditional expectations and evaluating the final inner products, are of lower order and are therefore absorbed into the overall rate.

The second part, namely the minimization over the response space, depends on the response type, see the column “Minimization complexity” in Table 1. Here, m denotes the number of grid points used to discretize quantile functions for distribution valued responses, and C_m denotes the cost of the monotonicity constrained projection on this grid. In our implementation, the projection in distribution valued response is computed using the R package `OSQP` (Stellato et al. 2020), by solving a sparse quadratic program with m variables and $m-1$ monotonicity constraints under the default setting, thus one may take $C_m = O(m)$ if the number of solver iterations is treated as bounded. For SPD matrices responses, the closed form minimization in Cholesky coordinates itself costs only $O(nr_Y^2)$, since it reduces to a weighted average of $r_Y \times r_Y$ triangular matrices. The dominant $O(nr_Y^3)$ cost, however, arises from computing n Cholesky factors $\{Y_i^{1/2}\}_{i=1}^n$. For clarity, we summarize the resulting computational complexity in Table 1, and discuss the corresponding numerical optimization separately below.

For numerical optimization, Euclidean and SPD matrices responses admit closed form solutions once the local regression weights have been com-

Table 1: Computational complexity for computing $\widehat{s}_{\oplus}(\boldsymbol{x}, u)$.

Response type	Minimization complexity	Total complexity
Euclidean	$O(n)$	$O(nC_\delta + np^2 + p^3)$
Distributions	$O(nm + C_m)$	$O(nC_\delta + np^2 + p^3 + nm + C_m)$
SPD matrices	$O(nr_Y^3)$	$O(nC_\delta + np^2 + p^3 + nr_Y^3)$

puted, so no additional iterative optimization is required. For distribution responses, we work with quantile functions and their discretizations. Let Q_i denote the quantile function of Y_i, and define the weighted combination $g(t) = \sum_{i=1}^n c_i(\boldsymbol{x}, u)Q_i(t)$. On a grid $0 < t_1 < \cdots < t_m < 1$, let $\boldsymbol{g} = (g(t_1), \ldots, g(t_m))^{\mathrm{T}} \in \mathbb{R}^m$ denote its discretized version. Since the weights $c_i(\boldsymbol{x}, u)$ may take negative values, the function $g(t)$ need not be nondecreasing, and thus its discretized version $\boldsymbol{g}$ may fail to be nondecreasing. We therefore compute the estimator by projecting $\boldsymbol{g}$ onto the cone of nondecreasing sequences, namely by solving $q^* = \arg\min_{q \in \mathbb{R}^m} \|q - \boldsymbol{g}\|^2$ subject to $q_1 \le \cdots \le q_m$. In our implementation, this constrained quadratic problem is solved using `OSQP`. In addition, to improve numerical stability in the computation of the weights $\{c_i(\boldsymbol{x}, u)\}_{i=1}^n$, we add a small regularization term when $\widehat{\Sigma}_{\boldsymbol{X}|u}$ is nearly singular.

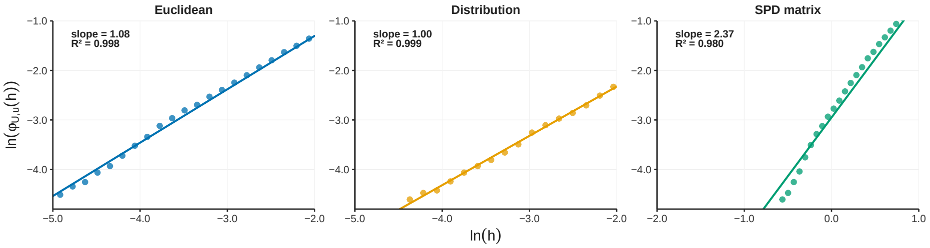
S1.2 Small-ball probability illustrations

Recall that the small-ball probability function is defined as $\varphi_{\mathcal{U},u}(h) = \mathbb{P}(U \in B_{\mathcal{U}}(u,h))$, where $B_{\mathcal{U}}(u,h) = \{u' \in \mathcal{U} : \delta(u',u) \leq h\}$. In what follows, u is taken as the Fréchet mean of U, i.e., $u = \operatorname{argmin}_{u \in \mathcal{U}} \mathbb{E}(\delta^2(U,u))$. To provide practical intuition for $\varphi_{\mathcal{U},u}(h)$, we constructed an empirical illustration for three U spaces considered in our simulations: Euclidean valued U, distribution valued U, and SPD matrix valued U. Figure 1(a) plots $\ln(\varphi_{\mathcal{U},u}(h))$ against $\ln(h)$. It can be seen that the empirical points exhibit a nearly linear relationship, with fitted slopes 1.08, 1.00, and 2.37 for Euclidean valued, distribution valued, and SPD matrix valued U considered in our simulation, respectively. This linear pattern clearly demonstrates that the small-ball probability follows an approximate power law relationship of the form $\varphi_{\mathcal{U},u}(h) \propto h^{\alpha}$, where the exponent α corresponds to the slope of the line. In addition, the corresponding R^2 values are 0.998, 0.999, and 0.980, confirming the accuracy of the power law relationship over the displayed range. Notably, the slopes for the Euclidean valued and distribution valued cases are very similar. This is reasonable because the distribution valued U is generated from a one-parameter family and therefore is confined to an one-dimensional subset of the Wasserstein space. In contrast, the larger exponent α of the SPD matrix setting implies smaller small-ball probabilities for a given neighborhood radius, indicating that samples are less concentrated around u.

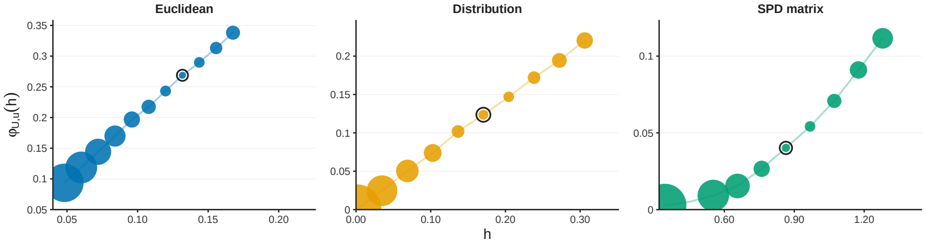
To further illustrate how the small-ball probability affects convergence and bandwidth choice, Figures 1(b)–(d) present bubble plots for $n = 50$, 100, and 200, respectively. In each panel, the bubble size represents the empirical estimation error $d(s_{\oplus}(\boldsymbol{x}, u), \widehat{s}_{\oplus}(\boldsymbol{x}, u))$, and the black circle marks the optimal bandwidth that minimizes this error. In all three settings, we observe that the estimation error initially decreases and then increases as the bandwidth h decreases, exhibiting the typical bias–variance tradeoff. Since the SPD matrix setting exhibits smaller small-ball probabilities for a given bandwidth, it requires a larger optimal bandwidth than the Euclidean and distribution valued settings.

S1.3 Bandwidth selection guidance and sensitivity analysis

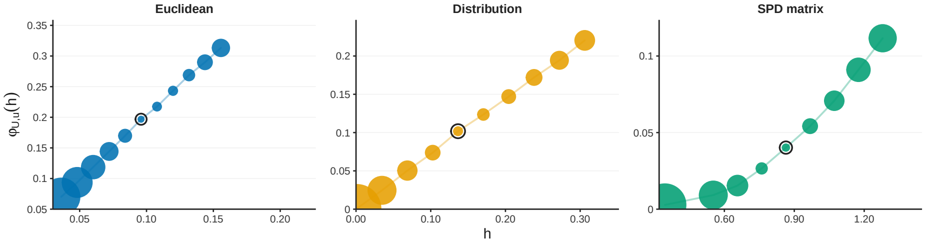
The following analysis is based on Examples 1–6 of our paper. For practical bandwidth selection, we construct the default candidate grid according to the theoretical optimal bandwidth order and the variability of the predictor U. When the predictor U is a scalar, Corollary 1 in our manuscript gives $d(s_{\oplus}(\boldsymbol{x}, u), \widehat{s}_{\oplus}(\boldsymbol{x}, u)) = O_p(h^{2/(\beta_1 - 1)} + (nh)^{-1/(2(\beta_2 - 1))})$. Balancing the bias and variance terms gives the optimal bandwidth order $n^{-\gamma}$ with $\gamma = (\beta_1 -$



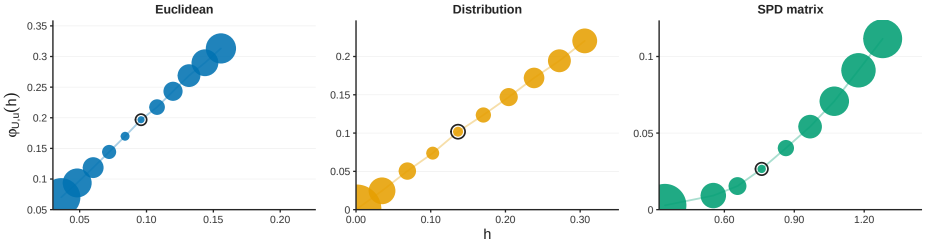
(a) Log-log small-ball probability.



(b) Error bubble plot for $n = 50$.



(c) Error bubble plot for $n = 100$.



(d) Error bubble plot for $n = 200$.

Figure 1: Empirical small-ball behavior for the three covariate spaces considered in the simulations. (a) shows $\ln(\varphi_{\mathcal{U},u}(h))$ versus $\ln(h)$, with fitted linear trends. (b)–(d) show $\varphi_{\mathcal{U},u}(h)$ versus h for $n = 50$, $n = 100$, and $n = 200$, respectively; bubble size is scaled according to $d(s_{\oplus}(\boldsymbol{x}, u), \widehat{s}_{\oplus}(\boldsymbol{x}, u))$, and black circles mark the bandwidths minimizing $d(s_{\oplus}(\boldsymbol{x}, u), \widehat{s}_{\oplus}(\boldsymbol{x}, u))$.

$1)/(4\beta_2+\beta_1-5)$. For Euclidean response, we have $\beta_1=\beta_2=2$, and hence $h=O(n^{-1/5})$, which yields the optimal convergence rate $O_p(n^{-2/5})$. When U takes values in a more general metric space, a suitable γ may be chosen using available geometric properties of that specific space, such as the rate at which the small-ball probability changes with h. In the absence of such geometric information, we adopt the scalar predictor exponent above as a practical default. The final bandwidth selected from the candidate grid via cross-validation is generally insensitive to the choice of the exponent γ. Motivated by this theoretical order, we construct the following default candidate grid,

$$h_j = c_j\,\widehat{\sigma}_U\,n^{-\gamma}, \qquad j=1,\dots,K.$$

Here $\widehat{\sigma}_U = [n^{-1}\sum_{i=1}^n \delta^2(U_i,\widehat{u}_\oplus)]^{1/2}$ is the sample Fréchet standard deviation of the predictor U, where $\widehat{u}_\oplus = \arg\min_{u\in\mathcal{U}} n^{-1}\sum_{i=1}^n \delta^2(U_i,u)$ is the sample Fréchet mean of U, and the constants $\{c_1,\dots,c_K\}$ determine the search range around $\widehat{\sigma}_U\,n^{-\gamma}$. The Fréchet standard deviation $\widehat{\sigma}_U$ rescales the bandwidth according to the variability of the predictor, in the same spirit as rule of thumb bandwidth choices in kernel smoothing (Scott 1992, p. 152), where the default bandwidth is often proportional to $\widehat{\sigma}n^{-1/5}$, and $\widehat{\sigma}$ denotes the sample standard deviation of the predictor. In our implementation, $\{c_1,\dots,c_K\}$ to be an equally spaced grid on $[0.10, 3.00]$, with

$K = 20$, a choice that balances estimation accuracy and computational efficiency. The final bandwidth is then selected by cross-validation over this grid. If the cross-validation minimum is attained at the boundary of this grid, we enlarge the range of $\{c_1, \ldots, c_K\}$ by halving the lower bound when the minimum occurs at the lower endpoint, or by doubling the upper bound when it occurs at the upper endpoint.

We next assess the sensitivity to the candidate grid from two aspects, namely the number of grid points and the search range. Firstly, we vary the number of grid points K while keeping the candidate search range fixed at $[0.10, 3.00]$ within each example. For $n = 50$ and different settings of $K \in \{10, 20, 40, 80,\ 100, 200\}$, Table 2 reports the average selected bandwidths h and the corresponding mean squared errors (MSEs), together with their standard errors, over 100 replications. The results indicate that the selected bandwidths are fairly stable across different choices of K, and the corresponding predictive performance exhibits little variation. This suggests that the bandwidth selection procedure is not sensitive to the number of grid points. Secondly, we further assess sensitivity to the search range by fixing the number of grid points at $K = 20$ and varying the range of the constants $\{c_1, \ldots, c_K\}$. Specifically, we consider the ranges $[0.10, 3.00]$, $[0.06, 4.00]$, $[0.02, 5.00]$, $[0.20, 2.00]$ and $[0.30, 1.00]$. Table 3 reports the average selected bandwidths h and MSEs, together with their standard errors for $n = 50$. The results show that, in Examples 1 and 2, the selected bandwidths and the resulting MSEs remain reasonably stable across the different ranges. An exception is the narrowest range [0.30, 1.00], where the selected bandwidth decreases substantially, and a corresponding slight increase in MSE is observed. In Examples 3–6, a similar pattern of stability is observed for the three widest grid ranges, [0.10, 3.00], [0.06, 4.00], and [0.02, 5.00], where the chosen bandwidths and the associated MSEs are almost identical. However, for the narrower ranges [0.20, 2.00] and [0.30, 1.00], the selected bandwidths h decrease noticeably, with a pronounced increase in MSE observed for [0.30,1.00]. These findings suggest that the predictive performance of the bandwidth selection procedure is fairly robust to the choice of candidate grid range, while also indicating that an overly narrow range should be avoided.

Finally, we empirically examined whether the selected bandwidths h scale plausibly with the sample size n. We conducted an additional simulation study with $n \in \{25, 50, 100, 150, 200, 250, 300\}$. For each value of n, the bandwidth was selected by cross validation over 100 repeated simulation runs, using the default grid points $K = 20$ and the search range $[0.10, 3.00]$. The average selected bandwidths are plotted in Figure 2(a) for

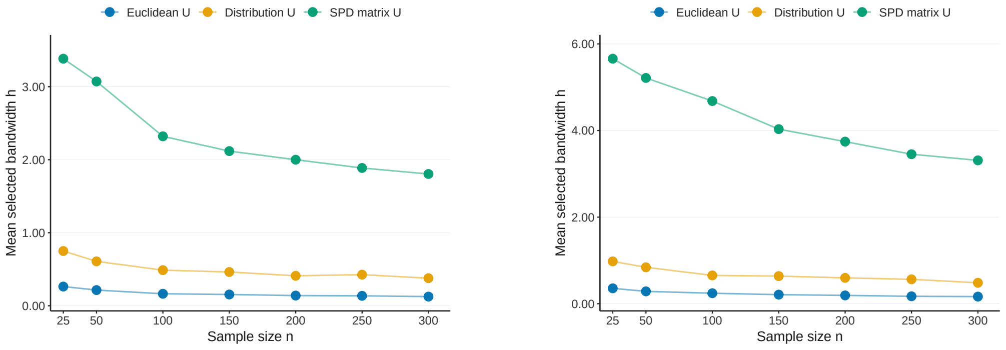
Table 2: Average selected bandwidths h and MSEs, with standard errors in parentheses, for different numbers of grid points K under the fixed range $[0.10, 3.00]$ when $n = 50$.

		$K = 10$	$K = 20$	$K = 40$	$K = 80$	$K = 100$	$K = 200$
Example 1	h	0.214 (0.007)	0.212 (0.007)	0.212 (0.007)	0.211 (0.007)	0.211 (0.007)	0.211 (0.007)
	MSE	2.041 (0.024)	2.034 (0.022)	2.038 (0.023)	2.037 (0.022)	2.039 (0.023)	2.038 (0.023)
Example 2	h	0.614 (0.019)	0.615 (0.019)	0.616 (0.019)	0.613 (0.019)	0.613 (0.019)	0.614 (0.019)
	MSE	2.093 (0.022)	2.093 (0.022)	2.090 (0.021)	2.093 (0.022)	2.095 (0.023)	2.092 (0.022)
Example 3	h	2.967 (0.088)	2.968 (0.089)	2.953 (0.089)	2.958 (0.089)	2.958 (0.089)	2.958 (0.089)
	MSE	2.410 (0.024)	2.416 (0.025)	2.417 (0.025)	2.416 (0.025)	2.417 (0.025)	2.417 (0.025)
Example 4	h	0.306 (0.008)	0.305 (0.008)	0.306 (0.008)	0.306 (0.008)	0.306 (0.008)	0.306 (0.008)
	MSE	10.025 (0.193)	10.046 (0.201)	10.020 (0.186)	10.035 (0.197)	10.029 (0.192)	10.030 (0.193)
Example 5	h	0.821 (0.024)	0.817 (0.025)	0.818 (0.024)	0.817 (0.025)	0.818 (0.025)	0.818 (0.025)
	MSE	10.040 (0.081)	10.042 (0.082)	10.039 (0.082)	10.041 (0.082)	10.039 (0.082)	10.040 (0.082)
Example 6	h	5.217 (0.142)	5.168 (0.147)	5.165 (0.147)	5.174 (0.147)	5.172 (0.147)	5.173 (0.147)
	MSE	10.211 (0.149)	10.224 (0.149)	10.232 (0.151)	10.227 (0.150)	10.227 (0.150)	10.227 (0.150)

Table 3: Average selected bandwidths h and MSEs, with standard errors in parentheses, for different ranges of the grid $\{c_1, \ldots, c_K\}$ when $n = 50$ and $K = 20$.

		$[0.10, 3.00]$	$[0.06, 4.00]$	$[0.02, 5.00]$	$[0.20, 2.00]$	$[0.30, 1.00]$
Example 1	h	0.212 (0.007)	0.214 (0.008)	0.216 (0.008)	0.202 (0.005)	0.128 (0.002)
	MSE	2.034 (0.022)	2.041 (0.023)	2.038 (0.023)	2.029 (0.023)	2.086 (0.027)
Example 2	h	0.615 (0.019)	0.620 (0.021)	0.629 (0.022)	0.582 (0.015)	0.354 (0.004)
	MSE	2.093 (0.022)	2.096 (0.023)	2.087 (0.021)	2.081 (0.021)	2.145 (0.024)
Example 3	h	2.968 (0.089)	3.184 (0.123)	3.328 (0.156)	2.514 (0.043)	1.403 (0.010)
	MSE	2.416 (0.025)	2.420 (0.025)	2.426 (0.026)	2.404 (0.025)	2.638 (0.030)
Example 4	h	0.305 (0.008)	0.342 (0.012)	0.374 (0.016)	0.241 (0.004)	0.128 (0.001)
	MSE	10.046 (0.201)	10.075 (0.208)	10.059 (0.176)	10.012 (0.187)	10.684 (0.200)
Example 5	h	0.817 (0.025)	0.926 (0.038)	1.010 (0.050)	0.660 (0.012)	0.359 (0.003)
	MSE	10.042 (0.082)	10.060 (0.082)	10.068 (0.082)	10.048 (0.084)	10.915 (0.176)
Example 6	h	5.168 (0.147)	6.196 (0.232)	7.068 (0.321)	4.018 (0.064)	2.140 (0.020)
	MSE	10.224 (0.149)	10.235 (0.146)	10.225 (0.142)	10.233 (0.154)	11.901 (0.227)

distribution responses Y and in Figure 2(b) for SPD matrix responses Y. In both Figures 2(a) and 2(b), the three curves correspond to Euclidean, distribution and SPD matrix predictors U. Across the settings, the selected bandwidths generally decrease as n increases, which agrees with the theoretical expectation. At a fixed sample size, the selected bandwidths are larger when U is SPD matrix. This is consistent with Figure 1(a), where $\varphi_{\mathcal{U},u}(h)$ decays faster for SPD matrix U, indicating that a larger bandwidth is needed.



(a) Distribution response.

(b) SPD matrix response.

Figure 2: Selected bandwidth h versus sample size n for two response space with three types of predictor U.

S1.4 Comparisons with other nonparametric Fréchet regression methods

Since RFWLFR and DFR are applicable only when all predictors are Euclidean, we present comparative results for Examples 1 and 4, summarized in Table 4. It can be seen that VFR yields lower MSEs than RFWLFR and DFR across all scenarios. Thus, VFR is preferable when the effect of the predictors is expected to vary with a moderator variable. In this setting, VFR directly models the varying-coefficient structure, is easier to interpret, requires less tuning, and is computationally efficient. By contrast, more flexible methods such as RFWLFR, or other fully nonparametric Fréchet regression approaches, are recommended when the regression relationship is highly nonlinear or involves complex interactions that are not well captured by a varying-coefficient form. To further assess the performance of DFR under larger sample sizes, we additionally consider $n = 400$ and $n = 500$. These larger sample sizes are included to examine settings where the flexible nonparametric structure of DFR may benefit from more training data; nevertheless, VFR yields smaller MSEs across all scenarios reported in Table 5.

Table 4: The averaged MSEs of RFWLFR, DFR, and VFR for Examples 1 and 4, with associated standard errors in parentheses.

		$(\gamma_2, \gamma_3) = (1, 1)$			$(\gamma_2, \gamma_3) = (3, 3)$		
		RFWLFR	DFR	VFR	RFWLFR	DFR	VFR
Example 1	$n = 50$	2.454 (0.041)	2.952 (0.101)	**2.074** (0.027)	2.857 (0.061)	3.529 (0.083)	**2.312** (0.090)
	$n = 100$	2.172 (0.021)	2.442 (0.034)	**1.809** (0.013)	2.493 (0.075)	3.028 (0.209)	**2.095** (0.077)
	$n = 200$	2.039 (0.018)	2.082 (0.021)	**1.719** (0.013)	2.272 (0.110)	2.683 (0.404)	**1.831** (0.042)
Example 4	$n = 50$	11.903 (0.213)	17.332 (0.869)	**9.949** (0.107)	15.417 (0.323)	23.757 (2.087)	**12.903** (0.190)
	$n = 100$	11.250 (0.142)	13.399 (0.304)	**9.426** (0.073)	13.239 (0.152)	15.686 (0.361)	**11.240** (0.099)
	$n = 200$	10.423 (0.093)	11.230 (0.171)	**8.909** (0.071)	12.201 (0.110)	13.238 (0.399)	**10.630** (0.100)

S1.5 Small sample simulations

To assess finite-sample performance in settings comparable to the real-data example with $n = 39$, this subsection presents additional simulation experiments with sample sizes $n \in \{30, 35, 40\}$. The results are reported in Tables 6 and 7. Overall, the additional small sample results are broadly consistent with the original simulation results for $n \in \{50, 100, 200\}$. They provide a more comprehensive assessment of the finite sample behavior of the proposed method in settings comparable to the mortality application.

Table 5: The averaged MSEs of DFR and VFR for Examples 1 and 4, with associated standard errors in parentheses.

		$(\gamma_2, \gamma_3) = (1, 1)$		$(\gamma_2, \gamma_3) = (3, 3)$	
		DFR	VFR	DFR	VFR
Example 1	$n = 50$	2.952	**2.074**	3.529	**2.312**
		(0.101)	(0.027)	(0.083)	(0.090)
	$n = 100$	2.442	**1.809**	3.028	**2.095**
		(0.034)	(0.013)	(0.209)	(0.077)
	$n = 200$	2.082	**1.719**	2.683	**1.831**
		(0.021)	(0.013)	(0.404)	(0.042)
	$n = 400$	1.878	**1.643**	2.073	**1.865**
		(0.021)	(0.013)	(0.076)	(0.075)
	$n = 500$	1.824	**1.632**	1.865	**1.696**
		(0.016)	(0.013)	(0.049)	(0.045)
Example 4	$n = 50$	17.332	**9.949**	23.757	**12.903**
		(0.869)	(0.107)	(2.087)	(0.190)
	$n = 100$	13.399	**9.426**	15.686	**11.240**
		(0.304)	(0.073)	(0.361)	(0.099)
	$n = 200$	11.230	**8.909**	13.238	**10.630**
		(0.171)	(0.071)	(0.399)	(0.100)
	$n = 400$	10.010	**8.692**	11.335	**10.155**
		(0.092)	(0.066)	(0.116)	(0.097)
	$n = 500$	9.788	**8.579**	11.335	**10.285**
		(0.105)	(0.059)	(0.152)	(0.103)

Table 6: The averaged MSE of various methods and the associated standard errors (in parenthesis) when the response is probability distributions in small samples.

		$(\gamma_2, \gamma_3) = (1, 1)$				$(\gamma_2, \gamma_3) = (3, 3)$			
		GFR	LFR	PFR	VFR	GFR	LFR	PFR	VFR
Example 1	$n = 30$	2.761	2.608	2.714	**2.372**	4.638	3.489	4.425	**2.848**
		(0.028)	(0.059)	(0.029)	(0.099)	(0.117)	(0.123)	(0.125)	(0.116)
	$n = 35$	2.677	2.374	2.607	**2.208**	4.363	3.325	4.076	**2.546**
		(0.026)	(0.034)	(0.030)	(0.057)	(0.064)	(0.082)	(0.067)	(0.064)
	$n = 40$	2.665	2.362	2.596	**2.198**	4.353	3.410	4.018	**2.709**
		(0.038)	(0.048)	(0.042)	(0.090)	(0.066)	(0.197)	(0.074)	(0.255)
Example 2	$n = 30$	–	–	2.800	**2.436**	–	–	4.337	**2.991**
				(0.036)	(0.046)			(0.084)	(0.092)
	$n = 35$	–	–	2.638	**2.302**	–	–	4.171	**2.764**
				(0.029)	(0.053)			(0.074)	(0.065)
	$n = 40$	–	–	2.574	**2.150**	–	–	4.077	**2.633**
				(0.027)	(0.029)			(0.060)	(0.049)
Example 3	$n = 30$	–	–	2.793	**2.544**	–	–	4.727	**4.106**
				(0.032)	(0.026)			(0.089)	(0.085)
	$n = 35$	–	–	2.731	**2.458**	–	–	4.723	**3.845**
				(0.030)	(0.028)			(0.081)	(0.065)
	$n = 40$	–	–	2.681	**2.458**	–	–	4.518	**3.761**
				(0.024)	(0.024)			(0.088)	(0.085)

Note: "–" denotes that the corresponding method cannot be applied in these examples.

Table 7: The averaged MSE of various methods and the associated standard errors (in parenthesis) when the response is symmetric positive definite matrices in small samples.

		$(\gamma_2, \gamma_3) = (1, 1)$				$(\gamma_2, \gamma_3) = (3, 3)$			
		GFR	LFR	PFR	VFR	GFR	LFR	PFR	VFR
Example 4	$n = 30$	11.063	16.974	11.150	**10.763**	15.183	30.951	15.379	**14.743**
		(0.124)	(0.841)	(0.144)	(0.138)	(0.245)	(1.107)	(0.339)	(0.623)
	$n = 35$	10.892	17.016	10.934	**10.608**	14.757	30.318	14.553	**13.368**
		(0.118)	(0.599)	(0.138)	(0.170)	(0.196)	(1.177)	(0.191)	(0.168)
	$n = 40$	10.724	16.541	10.818	**10.428**	14.365	30.731	14.179	**13.038**
		(0.102)	(0.411)	(0.138)	(0.244)	(0.151)	(0.854)	(0.178)	(0.182)
Example 5	$n = 30$	–	–	12.563	**12.227**	–	–	18.211	**17.672**
				(0.183)	(0.154)			(0.336)	(0.415)
	$n = 35$	–	–	12.173	**11.918**	–	–	17.317	**16.782**
				(0.147)	(0.134)			(0.305)	(0.332)
	$n = 40$	–	–	12.069	**11.875**	–	–	16.576	**15.964**
				(0.175)	(0.216)			(0.182)	(0.199)
Example 6	$n = 30$	–	–	11.033	**10.871**	–	–	15.602	**15.005**
				(0.171)	(0.169)			(0.331)	(0.328)
	$n = 35$	–	–	10.767	**10.545**	–	–	15.800	**14.835**
				(0.110)	(0.114)			(0.370)	(0.343)
	$n = 40$	–	–	10.758	**10.521**	–	–	15.683	**15.225**
				(0.163)	(0.168)			(0.302)	(0.305)

Note: "–" denotes that the corresponding method cannot be applied in these examples.

S1.6 Numerical implementation and runtime summaries

We first describe the numerical implementation for each response space. For distribution responses, we work with quantile functions and their discretizations. Let Q_i denote the quantile function of Y_i, and define the weighted combination $g(t) = \sum_{i=1}^{n} c_i(\boldsymbol{x}, u) Q_i(t)$. On a grid $0 < t_1 < \cdots < t_m < 1$, let $\boldsymbol{g} = (g(t_1), \ldots, g(t_m))^{\mathrm{T}} \in \mathbb{R}^m$ denote its discretized version. Since the local regression weights $c_i(\boldsymbol{x}, u)$ may take negative values, $g(t)$ need not be nondecreasing, so $\boldsymbol{g}$ may fail to represent a valid quantile function. We therefore compute the estimated quantile vector by projecting $\boldsymbol{g}$ onto the cone of nondecreasing sequences, namely by solving $q^* = \operatorname{argmin}_{q \in \mathbb{R}^m} \|q - \boldsymbol{g}\|^2$ subject to $q_1 \leq \cdots \leq q_m$. In our implementation, this quadratic program is solved using the `OSQP` solver with sparse matrix representations of the constraint system. We employed its default stopping criteria, namely absolute and relative tolerances of 10^{-3} and a maximum iteration limit of 4000. The resulting vector q^* provides a discretized approximation to the quantile function of $\widehat{s}_{\oplus}(\boldsymbol{x}, u)$. For SPD matrix responses, we represent each SPD matrix in Cholesky coordinates, under which the weighted Fréchet minimizer admits a closed-form expression as a weighted average. Hence, no additional numerical optimization is required once the local regression weights have been obtained. To ensure numerical stability in computing these weights, when the conditional covariance matrix $\widehat{\Sigma}_{\boldsymbol{X}|u}$ is nearly singular, we invert the ridge regularized matrix $\widehat{\Sigma}_{\boldsymbol{X}|u}+\lambda I_p$, where $\lambda = 10^{-6}\max\{1, \mathrm{tr}(\widehat{\Sigma}_{\boldsymbol{X}|u})/p\}$ and p is the dimension of $\boldsymbol{X}$.

We have added a runtime summary for the simulation examples; see Table 8. For each example and each sample size $n \in \{50, 100, 200\}$, we repeated the computation 10 times and report the mean elapsed time in seconds, with standard errors in parentheses. The reported timings include bandwidth selection by cross-validation and prediction on the test sample, but exclude data generation and the evaluation of test errors. It is seen that the proposed VFR method exhibits the second-shortest runtime, only slower than the global Fréchet regression. As expected, the runtimes generally increase with the sample size n, with the largest computational burden occurring in the more complex Examples 3 and 6. For Examples 3 and 6, the repeated runs were executed in parallel using 8 workers. All computations were carried out on the Apple M2 chip, 8 CPU cores, and 16 GB memory, running macOS Darwin 25.5.0 and R version 4.3.1. The main numerical packages used include `OSQP` version 1.0.0, `quadprog` version 1.5.8, `Matrix` version 1.6.5, and `MASS` version 7.3.60.

Table 8: Average runtime in seconds, with standard errors in parentheses, across 10 repeated runs for the simulation studies.

Example	n	PFR	VFR	GFR	LFR
Example 1	50	2.18 (0.07)	1.07 (0.02)	0.19 (0.01)	3.49 (0.04)
	100	5.89 (0.04)	1.79 (0.03)	0.18 (0.00)	10.12 (0.05)
	200	23.46 (0.36)	3.23 (0.05)	0.19 (0.01)	34.55 (0.47)
Example 2	50	3.42 (0.02)	1.07 (0.01)	–	–
	100	12.53 (0.18)	1.94 (0.05)	–	–
	200	60.57 (0.47)	3.65 (0.09)	–	–
Example 3	50	224.15 (4.51)	8.13 (0.56)	–	–
	100	1009.67 (20.03)	18.27 (0.32)	–	–
	200	12537.50 (83.64)	61.72 (2.28)	–	–
Example 4	50	0.81 (0.02)	0.40 (0.01)	0.08 (0.00)	1.30 (0.02)
	100	2.00 (0.03)	0.82 (0.01)	0.14 (0.00)	2.89 (0.02)
	200	6.43 (0.09)	2.06 (0.02)	0.28 (0.00)	7.59 (0.12)
Example 5	50	3.67 (0.03)	1.86 (0.02)	–	–
	100	12.65 (0.05)	5.43 (0.06)	–	–
	200	55.02 (0.46)	18.19 (0.14)	–	–
Example 6	50	222.48 (4.46)	8.94 (0.34)	–	–
	100	1300.42 (62.70)	26.33 (1.01)	–	–
	200	6933.19 (116.63)	79.77 (6.28)	–	–

Note: "–" denotes that the corresponding method cannot be applied in these examples.

S1.7 Explanations and verifications of assumptions

We begin with the recall of Assumptions (A1)–(A6) for the predictor U in the Euclidean space.

(A1) The object $s_\oplus(\boldsymbol{x}, u)$ exists and is unique. There exist a positive integer n_0 such that for all $n \geq n_0$, $\widetilde{s}_\oplus(\boldsymbol{x}, u)$ and $\widehat{s}_\oplus(\boldsymbol{x}, u)$ exist and are unique, the latter almost surely. Additionally, for any $\epsilon > 0$,

$$\inf_{dt(y,s_\oplus(\boldsymbol{x},u))>\epsilon} \left\{S_\oplus(y;\boldsymbol{x},u) - S_\oplus\left(s_\oplus(\boldsymbol{x},u);\boldsymbol{x},u\right)\right\} > 0,$$

$$\text{and} \quad \liminf_{n\to\infty} \inf_{d(y,\tilde{s}_\oplus(\boldsymbol{x},u))>\epsilon} \left\{\widetilde{S}_n(y;\boldsymbol{x},u) - \widetilde{S}_n\left(\widetilde{s}_\oplus(\boldsymbol{x},u);\boldsymbol{x},u\right)\right\} > 0.$$

(A2) Let $B_{\mathcal{Y}}\left(s_\oplus(\boldsymbol{x},u),\delta\right) \subseteq \mathcal{Y}$ be the ball of radius δ centered at $s_\oplus(\boldsymbol{x},u)$ and $N(\epsilon,\ B_{\mathcal{Y}}(s_\oplus(\boldsymbol{x},u),\delta),d)$ be its ϵ-covering number under the metric d. Then $\int_0^1 (1+\log N\left(\delta\epsilon, B_{\mathcal{Y}}\left(s_\oplus(\boldsymbol{x},u),\delta\right),d\right))^{1/2} d\epsilon = O(1)$ as $\delta \to 0^+$.

(A3) The kernel $K(\cdot)$ is a probability density function, symmetric around zero. Furthermore, defining $K_{kj} = \int_{\mathbb{R}} K^k(u)u^j du$, K_{12}, K_{14}, K_{22}, K_{24} and K_{26} are both finite.

(A4) The densities $f_U(\cdot)$, $f_{U|Y=y}(\cdot)$, $f_{U|X_i=x_i,X_j=x_j}(\cdot)$ and $f_{U|X_i=x_i,Y=y}(\cdot)$ exist and are twice continuously differentiable. The $\sup_{y,u}|f''_{U|Y=y}(u)|$ and $\sup_{x_i,u}|f''_{U|X_i=x_i}(u)|$ are finite. For any open set $A \subseteq \mathcal{Y}$, $B \subseteq \mathbb{R}$, $C \subseteq \mathbb{R}^2$, and $D \subseteq \mathcal{Y}\times\mathbb{R}$, the integrals $\int_A d\mathbb{P}_{Y|U=u}$, $\int_B d\mathbb{P}_{X_i|U=u}$, $\int_C d\mathbb{P}_{X_i,X_j|U=u}$ and $\int_D d\mathbb{P}_{Y,X_i|U=u}$ are continuous as a function of u, with $i,j = 1,2,\ldots,p$.

(A5) There exist $\eta_1 > 0, C_1 > 0$, and $\beta_1 > 1$ such that if $d(y, s_\oplus(\boldsymbol{x},u)) < \eta_1$, $S_\oplus(y;\boldsymbol{x},u) - S_\oplus\left(s_\oplus(\boldsymbol{x},u);\boldsymbol{x},u\right) \geq C_1 d\left(y, s_\oplus(\boldsymbol{x},u)\right)^{\beta_1}$.

(A6) There exist $\eta_2 > 0, C_2 > 0$, and $\beta_2 > 1$ such that if $d\left(y, \widetilde{s}_\oplus(\boldsymbol{x},u)\right) < \eta_2$, $\liminf_{n\to\infty}\left\{\widetilde{S}_n(y;\boldsymbol{x},u) - \widetilde{S}_n\left(\widetilde{s}_\oplus(\boldsymbol{x},u);\boldsymbol{x},u\right)\right\} \geq C_2 d\left(y, \widetilde{s}_\oplus(\boldsymbol{x},u)\right)^{\beta_2}$.

Assumption (A1) forms the foundational basis for establishing the consistency of M-estimator $s_{\oplus}(\boldsymbol{x}, u)$ (see, e.g., Corollary 3.2.3 of Van Der Vaart & Wellner (1996)), as it guarantees that the weak convergence of the empirical process $\widehat{S}_{\oplus}(\boldsymbol{x}, u)$ to the population process $\widetilde{S}_{\oplus}(\boldsymbol{x}, u)$ in turn implies convergence of their minimizers. Furthermore, existence follows immediately if $\mathcal{Y}$ is compact. Building upon this, Assumptions (A2) and (A5) impose covering number requirements and curvature constraints, which originate from empirical process theory and jointly regulate the local behavior of the deviation $\widehat{S}_{\oplus}(\boldsymbol{x}, u) - \widetilde{S}_{\oplus}(\boldsymbol{x}, u)$ near the optimum to establish convergence rates. Meanwhile, the assumptions on the kernel function in (A3) and the conditional density specifications in (A4) align with standard prerequisites in nonparametric local regression frameworks.

Then we discuss Assumptions (A5) and (A6). For the varying-coefficient model in Euclidean space, direct calculation can show that $S_{\oplus}(y; \boldsymbol{x}, u) - S_{\oplus}(s_{\oplus}(\boldsymbol{x}, u); \boldsymbol{x}, u) = |y - s_{\oplus}(\boldsymbol{x}, u)|^2$ and $\widetilde{S}_n(y; \boldsymbol{x}, u) - \widetilde{S}_n(\widetilde{s}_{\oplus}(\boldsymbol{x}, u); \boldsymbol{x}, u) = |y - \widetilde{s}_{\oplus}(\boldsymbol{x}, u)|^2$. Then, Assumption (A5) becomes $(y - s_{\oplus}(\boldsymbol{x}, u))^2 \geq C_1 |y - s_{\oplus}(\boldsymbol{x}, u)|^{\beta_1}$, and (A6) becomes $\liminf_{n\to\infty} \{y - \widetilde{s}_{\oplus}(\boldsymbol{x}, u)\}^2 \geq C_2 |y - \widetilde{s}_{\oplus}(\boldsymbol{x}, u)|^{\beta_2}$. Consequently, Assumptions (A5) and (A6) hold with $\beta_1 = \beta_2 = 2$, $C_1 = C_2 = 1$, and arbitrary positive constants η_1 and η_2. This shows that the objective functions $S_{\oplus}(y; \boldsymbol{x}, u)$ and $\widetilde{S}_{\oplus}(y; \boldsymbol{x}, u)$ are quadratic in y and satisfy the twice differentiability condition. More generally, if an objective function $f(y)$ is twice continuously differentiable and is locally convex at the minimizer y^* such that $\inf_{|y-y^*|<\eta} f''(y) \geq C > 0$, then a second order Taylor expansion shows that the excess objective $f(y) - f(y^*)$ is bounded below by $C|y - y^*|^2$ for $y \in \mathbb{R}$ with $|y - y^*| < \eta$, which corresponds to the familiar strong convexity/quadratic curvature condition in Euclidean settings. Therefore, Assumptions (A5) and (A6) generalize this local quadratic growth condition from Euclidean responses to responses valued in a metric space.

When $(\mathcal{Y}, \|\cdot\|)$ is a Hilbert space or a convex subset of a Hilbert space, by employing essentially the same argument as in Euclidean space and combining the Riesz representation theorem, the linearity of the inner product, and the Hilbert space projection theorem, it can be shown that $S_\oplus(y; \boldsymbol{x}, u) - S_\oplus(s_\oplus(\boldsymbol{x}, u); \boldsymbol{x}, u) = \|y - s_\oplus(\boldsymbol{x}, u)\|^2$ and $\widetilde{S}_n(y; \boldsymbol{x}, u) - \widetilde{S}_n(\widetilde{s}_\oplus(\boldsymbol{x}, u); \boldsymbol{x}, u) = \|y - \widetilde{s}_\oplus(\boldsymbol{x}, u)\|^2$.

For the Wasserstein response space, the 2-Wasserstein metric distance is defined as is $d_W(y_1, y_2) = (\int_0^1 (Q_{y_1}(t) - Q_{y_2}(t))^2 dt)^{1/2}$, where $Q_{y_1}(\cdot) = y_1^{-1}(\cdot)$ and $Q_{y_2}(\cdot) = y_2^{-1}(\cdot)$ denotes the quantile functions corresponding to $y_1(\cdot)$ and $y_2(\cdot)$ respectively. Accordingly, we define the quantile function space $(Q(\mathcal{Y}), \|\cdot\|_2)$, which consists of nondecreasing, left continuous functions in $L^2([0, 1])$, and equipped with the L^2 norm. Since the metric spaces $(\mathcal{Y}, d_W)$

and $(Q(\mathcal{Y}), \|\cdot\|_2)$ are bijectively and isometrically isomorphic, the two spaces can be identified, meaning that the Wasserstein space $(\mathcal{Y}, d_W)$ is essentially a convex subset of the Hilbert space $L^2([0,1])$. Therefore, Assumption (A5) and (A6) are satisfied by taking $\beta_1 = \beta_2 = 2$, $C_1 = C_2 = 1$, with η_1 and η_2 being arbitrary positive constants.

For the SPD response space, two common metrics are the Frobenius metric $d_F(y_1, y_2) = \|y_1 - y_2\|_F$ and the Cholesky decomposition metric $d_C(y_1, y_2) = \|y_1^{1/2} - y_2^{1/2}\|_F$, where $y^{1/2} = (r_{jk}(y))_{1\le j\le k\le q}$ denotes the unique upper triangular matrix with positive diagonal entries such that $y = (y^{1/2})^{\mathrm{T}} y^{1/2}$. Note that $(\mathcal{Y}, d_F)$ is a convex subset of a Hilbert space, Assumptions (A5) and (A6) are satisfied with $\beta_1 = \beta_2 = 2$, $C_1 = C_2 = 1$, and any positive positive numbers η_1, η_2. In addition, since $S_\oplus^{d_C}(y; \boldsymbol{x}, u)$ defined via the metric d_C equals $S_\oplus^{d_F}(y^{1/2}; \boldsymbol{x}, u)$ defined via the metric d_F, the map $\phi(y) = y^{1/2}$ is a bijection and an isometry between $(\mathcal{Y}, d_C)$ and $(\phi(\mathcal{Y}), \|\cdot\|_F)$. Therefore, it follows that

$$\begin{aligned} S_\oplus^{d_C}(y; \boldsymbol{x}, u) - S_\oplus^{d_C}(s_\oplus(\boldsymbol{x}, u); \boldsymbol{x}, u) &= S_\oplus^{d_F}(y^{1/2}; \boldsymbol{x}, u) - S_\oplus^{d_F}(s_\oplus^{1/2}(\boldsymbol{x}, u); \boldsymbol{x}, u) \\ &= \|y^{1/2} - s_\oplus^{1/2}(\boldsymbol{x}, u)\|_F^2 = d_C^2(y, s_\oplus(\boldsymbol{x}, u)). \end{aligned}$$

Similarly, we have $\widetilde{S}_n(y; \boldsymbol{x}, u) - \widetilde{S}_n(\widetilde{s}_\oplus(\boldsymbol{x}, u); \boldsymbol{x}, u) = d_C^2(y, \widetilde{s}_\oplus(\boldsymbol{x}, u))$. Hence, for the SPD space $(\mathcal{Y}, d_C)$, Assumptions (A5) and (A6) also hold with $\beta_1 = \beta_2 = 2$, $C_1 = C_2 = 1$, and arbitrary positive constants η_1 and η_2.

Next, we recall the Assumptions (A7)–(A10) to handle the predictor U in the metric space.

(A7) For any $\epsilon > 0$, $\mathbb{P}\left(U \in B_{\mathcal{U}}\left(u, \epsilon\right)\right) = \varphi_{\mathcal{U},u}\left(\epsilon\right) > 0$.

(A8) As $n \to \infty$, we have $h \to 0$, $\log n/(n\varphi_{\mathcal{U},u}\left(h\right)) \to 0$, and $nh \to \infty$.

(A9) There exist constants $0 < c_1 \leq c_2 < \infty$ and C, such that $c_1 \mathbb{1}_{[0,1]}(\cdot) \leq K(\cdot) \leq c_2 \mathbb{1}_{[0,1]}(\cdot)$ and $\int K_h(\delta(u', u))\delta(u', u) d\nu(u') \leq Ch \int K_h(\delta(u', u)) d\nu(u')$, where $\mathbb{1}_{[0,1]}(\cdot)$ denotes the indicator function.

(A10) The marginal density $f_U(u)$ and the conditional density $f_{U|Y=y}(u)$, $f_{U|X_i=x_i}(u)$, $f_{U|X_i=x_i,X_j=x_j}(u)$, $f_{U|X_i=x_i,Y=y}(u)$ for $i, j = 1, 2 \ldots, p$, exist (with respect to a reference measure ν) and are Lipschitz continuous. That is, there exists a constant $C > 0$ such that for all $u' \in B_{\mathcal{U}}(u, h)$ and any conditional density $f(\cdot)$ in the above, $|f(u') - f(u)| \leq C\delta(u', u)$.

Assumption (A7) ensures that local neighborhood $B_{\mathcal{U}}(u, \epsilon)$ around point u are not empty and provides sufficient sample probability mass for local smoothing methods such as kernel estimation. In Assumption (A8), as the sample size $n \to \infty$, the requirement $h \to 0$ eliminates asymptotic bias, the condition $\log n/(n\varphi_{\mathcal{U},u}(h)) \to 0$ regulates stochastic variability, and the growth condition $nh \to \infty$ guarantees that the effective number of local observations diverges. Assumption (A9) imposes mild conditions on the kernel function $K(\cdot)$. It requires that $K(\cdot)$ is bounded between two positive constants within the unit support and vanishes outside, while also satisfying an integral inequality involving the distance function $\delta(\cdot,\cdot)$. In addition, it guarantees that $\mathbb{E}[\zeta_h(U,u)|Y=y] = f_{U|Y=y}(u) f_U^{-1}(u)\,(1+O(h))$. Assumption (A10) requires the existence and Lipschitz continuity of the conditional and marginal densities. The Lipschitz condition ensures that these densities vary smoothly with respect to u, which is crucial for controlling approximation errors in estimation and establishing convergence results.

S2 Proof of Lemma 1

Proof. [**Proof of Lemma 1**] The assumption $\mathbb{E}(\|Y\|_{\mathcal{Y}}^2 \mid U=u) < \infty$, in conjunction with the Cauchy-Schwarz inequality implies that the map $y \mapsto \mathbb{E}\left(\langle Y, y\rangle \mid U=u\right)$ is a continuous linear functional on $\mathcal{Y}$. By the Riesz representation theorem, there exists a unique $\gamma_0(u) \in \mathcal{Y}$ such that

$$\mathbb{E}(\langle Y, y\rangle \mid U=u) = \langle \gamma_0(u), y\rangle \qquad \text{for all } y \in \mathcal{Y}. \tag{S2.1}$$

Next, for each $y \in \mathcal{Y}$, define the vector

$$\boldsymbol{g}(u,y) := \mathbb{E}(\langle (X-\mu(u))Y, y\rangle_p \mid U=u) \in \mathbb{R}^p.$$

Each coordinate of $\boldsymbol{g}(u,y)$ is linear and continuous in y, hence the Riesz representation theorem guarantees that there exist $\gamma_1(u) = (\gamma_{1,1}(u), \ldots, \gamma_{1,p}(u))^{\mathrm{T}} \in$

$\mathcal{Y}^p$ such that

$$(\boldsymbol{g}(u,y))_j = \mathbb{E}\left(\langle (X_j - \mu_j(u))Y, y\rangle \mid U=u\right) = \langle \gamma_{1,j}(u), y\rangle, \qquad j=1,\ldots,p. \tag{S2.2}$$

Therefore, for every $y \in \mathcal{Y}$ and $\boldsymbol{x} \in \mathbb{R}^p$,

$$\begin{aligned}\mathbb{E}(s_u(\boldsymbol{x},\boldsymbol{X})\langle Y,y\rangle \mid U=u) &= \mathbb{E}\left(\langle Y,y\rangle \mid U=u\right) + \mathbb{E}\left((\boldsymbol{x}-\mu)^{\mathrm{T}} w_2^{-1}(u)(\boldsymbol{X}-\mu)\langle Y,y\rangle \mid U=u\right)\\ &= \langle \gamma_0(u), y\rangle + (\boldsymbol{x}-\mu(u))^{\mathrm{T}} w_2^{-1}(u)\boldsymbol{g}(u,y)\\ &= \left\langle \gamma_0(u) + (\boldsymbol{x}-\mu(u))^{\mathrm{T}} w_2^{-1}(u)\gamma_1(u), y\right\rangle, \quad \text{(S2.3)}\end{aligned}$$

where we use (S2.2) in the last step. Note also that

$$\mathbb{E}(s_u(\boldsymbol{x},\boldsymbol{X}) \mid U=u) = 1 + (\boldsymbol{x}-\mu(u))^{\mathrm{T}} w_2^{-1}(u)\,\mathbb{E}\left(\boldsymbol{X}-\mu(u) \mid U=u\right) = 1. \tag{S2.4}$$

Define $\widetilde{y}(u,\boldsymbol{x}) := \gamma_0(u) + (\boldsymbol{x}-\mu(u))^{\mathrm{T}} w_2^{-1}(u)\gamma_1(u) \in \mathcal{Y}$. Recall

$$S_{\oplus}(y;\boldsymbol{x}) := \mathbb{E}(s_u(\boldsymbol{x},\boldsymbol{X})\,\|Y-y\|_{\mathcal{Y}}^2 \mid U=u)$$

in this case. For the convenience of writing, we will abbreviate $s_u(\boldsymbol{x},\boldsymbol{X})$ and $\widetilde{y}(u,\boldsymbol{x})$ to symbol s_u and $\widetilde{y}$, respectively. Then, expanding the square $S_u(y;\boldsymbol{x})$ yields

$$\begin{aligned}S_{\oplus}(y;\boldsymbol{x}) &= \mathbb{E}(s_u\,\|Y-\widetilde{y}\|_{\mathcal{Y}}^2 + 2s_u\langle Y-\widetilde{y}, \widetilde{y}-y\rangle + s_u\|\widetilde{y}-y\|_{\mathcal{Y}}^2 \mid U=u)\\ &= S_{\oplus}(\widetilde{y};\boldsymbol{x}) + 2(\mathbb{E}(s_u\,\langle Y,\widetilde{y}-y\rangle \mid U=u) - \mathbb{E}(s_u \mid U=u)\,\langle \widetilde{y},\widetilde{y}-y\rangle) + \|\widetilde{y}-y\|_{\mathcal{Y}}^2,\end{aligned}$$

where the last equality uses that s_u does not depend on y. By (S2.3) with y replaced by $\widetilde{y}-y$ and by (S2.4), the middle term vanishes, so

$$S_{\oplus}(y;\boldsymbol{x}) = S_{\oplus}(\widetilde{y};\boldsymbol{x}) + \|\widetilde{y}-y\|_{\mathcal{Y}}^2 \geq S_{\oplus}(\widetilde{y};\boldsymbol{x}),$$

with equality if and only if $y=\widetilde{y}$. Hence,

$$s_{\oplus}(\boldsymbol{x},u) = \widetilde{y}(u,\boldsymbol{x}) = \gamma_0(u) + (\boldsymbol{x}-\mu(u))^{\mathrm{T}}(w_2^{-1}(u)\gamma_1(u)).$$

Define $\boldsymbol{\beta}(u) := w_2^{-1}(u)\gamma_1(u)$ and $\beta_0(u) := \gamma_0(u) - \mu(u)^{\mathrm{T}} w_2^{-1}(u)\gamma_1(u)$. Then, we obtain

$$s_{\oplus}(\boldsymbol{x},u) = \beta_0(u) + \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}(u),$$

which completes the proof. □

S3 Proofs for Euclidean predictor U

Firstly, we establish a lemma to approximate the following quantities, for $j=0,1,2$,

$$\widetilde{\mu}_j(u) = \mathbb{E}(K_h(U-u)(U-u)^j), \tau_j(u,y) = \mathbb{E}(K_h\left(U-u\right)\left(U-u\right)^j \mid Y=y),$$
$$\widehat{\mu}_j(u) = n^{-1}\sum_{i=1}^n K_h(U_i-u)(U_i-u)^j, \gamma_j(x_i,u) = \mathbb{E}(K_h\left(U-u\right)\left(U-u\right)^j \mid X_i=x_i).$$

Lemma 1. *If (A3) and (A4) hold, then* $\widetilde{\mu}_j(u) = h^j[f_U(u)K_{1j} + hf_U'(u)K_{1(j+1)} + O\left(h^2\right)]$ *and* $\widehat{\mu}_j(u) = \widetilde{\mu}_j(u) + O_p((h^{2j-1}n^{-1})^{1/2})$ *for* $j=0,1,2$. *Additionally, uniformly over* $(u,y)\in\mathbb{R}\times\mathcal{Y}$, $\tau_j(u,y) = h^j[f_{U|Y=y}(u)K_{1j} + hf_{U|Y=y}'(u)K_{1(j+1)} +$

$O\left(h^2\right)]$, *and uniformly over* $(x_i, u) \in \mathbb{R} \times \mathbb{R}$, $\gamma_j(x_i, u) = h^j[f_{U|X_i=x_i}(u)K_{1j} + f'_{U|X_i=x_i}(u)K_{1(j+1)} + O\left(h^2\right)]$.

Proof. For $\widetilde{\mu}_j(u) = \int K_h(v-u)(v-u)^j f_U(v)dv$, we substitute $(v-u)/h = t$ and apply a second-order Taylor expansion to the density $f_U(u+ht)$ around u, then

$$\widetilde{\mu}_j(u) = h^j[f_U(u) \int t^j K(t)dt + hf'_U(u) \int t^{j+1} K(t)dt + \frac{h^2}{2} \int t^{j+2} f''_U(u+\theta(t)ht)K(t)dt],$$

for some $0 \leq \theta(t) \leq 1$. Since $f''_U(\cdot)$ is bounded by assumption (A4) and $\int t^{j+2}K(t)dt$ is finite, the result for $\widetilde{\mu}_j(u)$ can be obtained. The results for $\tau_j(u,y)$ and $\gamma_j(x_i,u)$ also can be shown by substituting $(v-u)/h = t$ and using a second order Taylor expansion of $f_{U|Y=y}(u+ht)$ and $f_{U|X_i=x_i}(u+ht)$ around u. Note that we have $\sup_{y,u} |f''_{U|Y=y}(u)| < \infty$ and $\sup_{x_i,u} |f''_{U|X_i=x_i}(u)| < \infty$ by assumption (A4), the $O(h^2)$ terms are uniform over $(u,y) \in \mathbb{R} \times \mathcal{Y}$ and $(x_i,u) \in \mathbb{R} \times \mathbb{R}$.

Since $\widehat{\mu}_j(u)$ is an average of i.i.d. terms, we have $\mathbb{E}(\widehat{\mu}_j(u)) = \widetilde{\mu}_j(u)$, then compute its variance. Substituting $(v-u)/h = t$ and apply a second-order Taylor expansion, then

$$\mathbb{E}[K_h^2(U-u)(U-u)^{2j}] = h^{2j-1}[f_U(u) \int K^2(t)t^{2j}dt + hf'_U(u) \int K^2(t)t^{2j+1}dt + O(h^2)].$$

Thus, $\mathrm{Var}(K_h(U-u)(U-u)^j) = O(h^{2j-1})$, it follows from Chebyshev's inequality that $\widehat{\mu}_j(u) = \widetilde{\mu}_j(u) + O_p((h^{2j-1}n^{-1})^{1/2})$. □

Based on this lemma, we can get the convergence rate of the bias of our model in Theorem 1. The proof is provided below. From here onwards, for simplicity of notation, the dependence of objects such as $S_\oplus$, $s_\oplus$, w_i, etc. on $(y, \boldsymbol{x}, u)$ will be dropped.

Proof of Theorem 1. Firstly, we will show that $\frac{\mathbb{P}_{Y|U}}{d\mathbb{P}_Y}(y \mid u) = f_{U|Y=y}(u)/f_U(u)$ for all u such that $f_U(u) > 0$, where $\frac{d\mathbb{P}_{Y|U}}{d\mathbb{P}_Y}(y \mid u)$ denotes the Radon-Nikodym derivative of the conditional probability measure $\mathbb{P}_{Y|U}$ with respect to the marginal probability measure $\mathbb{P}_Y$. For any open set $A \subseteq \mathcal{Y}$, let

$$a(u) := \int_A \frac{f_{U|Y}(u \mid y)}{f_U(u)} d\mathbb{P}_Y(y), \quad b(u) := \int_A d\mathbb{P}_{Y|U=u}.$$

By assumption (A4), both $a(\cdot)$ and $b(\cdot)$ are continuous. Since $a(\cdot)$, $f_U(\cdot)$ and $f_{U|Y=y}(\cdot)$ are both nonnegative, Fubini's theorem implies that for any $z \in \mathbb{R}$,

$$\begin{aligned}\int_{-\infty}^{z} a(u) f_U(u) du &= \int_A \Big(\int_{-\infty}^{z} f_{U|Y}(u \mid y) du \Big) d\mathbb{P}_Y(y) \\ &= \int_A \Big(\int_{-\infty}^{z} d\mathbb{P}_{U|Y=y}(u) \Big) d\mathbb{P}_Y(y) = \int_{A \times (-\infty, z]} d\mathbb{P}_{Y,U}(y, u) \\ &= \int_{-\infty}^{z} \Big(\int_A d\mathbb{P}_{Y|U=u}(y) \Big) f_U(u) du = \int_{-\infty}^{z} b(u) f_U(u) du.\end{aligned}$$

Then the claim follows by Dynkin's π-λ theorem. Next, we can get the following equation using Lemma 1,

$$\begin{aligned}\int \zeta_h(u', u) d\mathbb{P}_{U|Y=y}(u') &= \frac{\widetilde{\mu}_2(u)\tau_0(u,y) - \widetilde{\mu}_1(u)\tau_1(u,y)}{\widetilde{\sigma}_0^2(u)} \\ &= \frac{f_{U|Y=y}(u)\big[K_{10}K_{12} - K_{11}^2 + h(f_U'/f_U)(K_{10}K_{13} - K_{11}K_{12}) + O(h^2)\big]}{f_U(u)\big[K_{10}K_{12} - K_{11}^2 + h(f_U'/f_U)(K_{10}K_{13} - K_{11}K_{12}) + O(h^2)\big]} \\ &= \frac{f_{U|Y=y}(u)}{f_U(u)}\big[1 + O(h^2)\big] = \frac{f_{U|Y=y}(u)}{f_U(u)} + O(h^2),\end{aligned}$$

where the error term is uniform over $y \in \mathcal{Y}$. Hence, using the previously established fact that $\frac{d\mathbb{P}_{Y|U}}{d\mathbb{P}_Y}(y \mid u) = f_{U|Y=y}(u)/f_U(u)$, we can get

$$\begin{aligned}\mathbb{E}\left(\zeta_h(U,u)d^2\left(Y,y\right)\right) &= \int \zeta_h\left(u',u\right) d^2\left(y',y\right) d\mathbb{P}_{U,Y}(u',y') \\ &= \int d^2\left(y',y\right) \frac{f_{U|Y=y}(u)}{f_U(u)} d\mathbb{P}_Y(y) + O\left(h^2\right) \\ &= \int d^2\left(y',y\right) d\mathbb{P}_{Y|U=u}(y') + O\left(h^2\right) \\ &= \mathbb{E}\left(d^2(Y,y) \mid U = u\right) + O\left(h^2\right).\end{aligned}$$

Using the notion in equation (3.1) and (4.1), it shows that $\widetilde{w}_0\left(y;u\right) = w_0\left(y;u\right) + O\left(h^2\right)$.

By the same techniques, we can find that $\frac{d\mathbb{P}_{X_i|U}}{d\mathbb{P}_{X_i}}(x_i \mid u) = f_{U|X_i=x_i}(u)/f_U(u)$ for $i = 1, \ldots, p$. It follows that $\mathbb{E}\left(\zeta_h(U,u)X_i\right) = \mathbb{E}\left(X_i \mid U = u\right) + O\left(h^2\right)$. Then we will get $\mathbb{E}\left(\zeta_h(U,u)\boldsymbol{X}\right) = \mathbb{E}\left(\boldsymbol{X} \mid U = u\right) + \boldsymbol{O}\left(h^2\right)$, where $\boldsymbol{O}(h^2)$ is a p-dimensional vector whose each component is $O(h^2)$. Thus we can obtain that $\widetilde{w}_1\left(\boldsymbol{x},u\right) = w_1\left(\boldsymbol{x},u\right) + \boldsymbol{O}(h^2)^{\mathrm{T}}$.

Furthermore, we will verify that $\widetilde{w}_2\left(u\right) = w_2\left(u\right) + \boldsymbol{O}(h^2)_{p\times p}$ holds, where

$\boldsymbol{O}(h^2)_{p\times p}$ is a $p\times p$ matrix composed with $O(h^2)$. Note that $\mathbb{E}\{\zeta_h(U,u)\}=1$, we have

$$\begin{aligned}\widetilde{w}_2(u) &= \mathbb{E}\{\zeta_h(U,u)[\boldsymbol{X}-\mathbb{E}(\zeta_h(U,u)\boldsymbol{X})][\boldsymbol{X}-\mathbb{E}(\zeta_h(U,u)\boldsymbol{X})]^{\mathrm{T}}\}\\ &= \mathbb{E}(\zeta_h(U,u)\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}})-\mathbb{E}(\zeta_h(U,u)\boldsymbol{X})\mathbb{E}(\zeta_h(U,u)\boldsymbol{X})^{\mathrm{T}}.\end{aligned}$$

A similar argument also yields $\mathbb{E}(\zeta_h(U,u)\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}})=\mathbb{E}(\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\mid U=u)+\boldsymbol{O}(h^2)_{p\times p}$. In addition, $\mathbb{E}(\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\mid U=u)=\mathrm{Cov}(\boldsymbol{X}\mid U=u)+\mathbb{E}(\boldsymbol{X}\mid U=u)\mathbb{E}(\boldsymbol{X}\mid U=u)^{\mathrm{T}}$. This, together with $\mathbb{E}\left(\zeta_h(U,u)\boldsymbol{X}\right)=\mathbb{E}\left(\boldsymbol{X}\mid U=u\right)+\boldsymbol{O}\left(h^2\right)$ ensures that

$$\widetilde{w}_2(u)=\mathrm{Cov}(\boldsymbol{X}\mid U=u)+\boldsymbol{O}(h^2)_{p\times p}=w_2(u)+\boldsymbol{O}(h^2)_{p\times p}.$$

Similarly, we have $\widetilde{w}_3\left(y;u\right)=w_3\left(y;u\right)+\boldsymbol{O}\left(h^2\right)$.

For the ease of presentation , we denote $\delta_i:=\widetilde{w}_i-w_i$, $i=0,1,2,3$, and $\Delta_2:=\widetilde{w}_2^{-1}-w_2^{-1}=-w_2^{-1}\delta_2\widetilde{w}_2^{-1}$. Due to $\|\delta_2\|=O(h^2)$, we have $\|\Delta_2\|=O(h^2)$. Then

$$\begin{aligned}\widetilde{S}_n\left(y;\boldsymbol{x},u\right)-S_{\oplus}\left(y;\boldsymbol{x},u\right) &= (\widetilde{w}_0(y;u)-w_0(y;u))+(\widetilde{w}_1\widetilde{w}_2^{-1}\widetilde{w}_3-w_1w_2^{-1}w_3)\\ &= \delta_0+(w_1+\delta_1)(w_2^{-1}+\Delta_2)(w_3+\delta_3)-w_1w_2^{-1}w_3\\ &= \delta_0+w_1\Delta_2w_3+w_1w_2^{-1}\delta_3+\delta_1w_2^{-1}w_3+O(h^4)=O(h^2).\end{aligned}$$

With this and the assumption (A1), we get that $d\left(s_{\oplus}\left(\boldsymbol{x},u\right),\widetilde{s}_{\oplus}\left(\boldsymbol{x},u\right)\right)=o\left(1\right)$, as $h\to 0$.

To derive the convergence rates of $\widetilde{s}_{\oplus}(\boldsymbol{x}, u)$, set $r_h = h^{-\frac{\beta_1}{(\beta_1 - 1)}}$ and define $C_{j,n} = \{y : 2^{j-1} < r_h d(y, s_{\oplus}(\boldsymbol{x}, u))^{\beta_1/2} \leq 2^j\}$. Choose η_1 satisfying the assumption (A5). Set $\widetilde{\eta}_1 := \eta_1^{\beta_1/2}$. For any integer $M > 0$,

$$
\begin{aligned}
&\mathbb{1}\left\{r_h d\left(s_{\oplus}(\boldsymbol{x}, u), \widetilde{s}_{\oplus}(\boldsymbol{x}, u)\right)^{\beta_1/2} > 2^M\right\} \\
&:= \sum_{\substack{j>M \\ 2^j < r_h \widetilde{\eta}_1}} \mathbb{1}\left\{2^{j-1} < r_h d\left(s_{\oplus}, \widetilde{s}_{\oplus}\right)^{\beta_1/2} \leq 2^j\right\} + \sum_{\substack{j>M \\ 2^j \geq r_h \widetilde{\eta}_1}} \mathbb{1}\left\{2^{j-1} < r_h d\left(s_{\oplus}, \widetilde{s}_{\oplus}\right)^{\beta_1/2} \leq 2^j\right\} \\
&\leq \sum_{\substack{j>M \\ 2^j < r_h \widetilde{\eta}_1}} \mathbb{1}\left\{2^{j-1} < r_h d\left(s_{\oplus}, \widetilde{s}_{\oplus}\right)^{\beta_1/2} \leq 2^j\right\} + \mathbb{1}\left\{2 d\left(s_{\oplus}, \widetilde{s}_{\oplus}\right)^{\beta_1/2} > \widetilde{\eta}_1\right\}.
\end{aligned}
$$

By the definition of $C_{j,n}$, we have

$$
\mathbb{1}\left\{2^{j-1} < r_h d\left(s_{\oplus}, \widetilde{s}_{\oplus}\right)^{\beta_1/2} \leq 2^j\right\} = \mathbb{1}\{\widetilde{s}_{\oplus} \in C_{j,n}\} \leq \mathbb{1}\left\{\inf_{y \in C_{j,n}} (\widetilde{S}_n(y) - \widetilde{S}_n(s_{\oplus})) \leq 0\right\}.
$$

In addition, we notice that when $y \in C_{j,n}$, $d(y, s_{\oplus}(\boldsymbol{x}, u)) \leq (2^j / r_h)^{2/\beta_1}$. If $2^j < r_h \widetilde{\eta}_1$, we have $d(y, s_{\oplus}(\boldsymbol{x}, u)) \leq \eta_1$. Then according to Assumption (A5), we have

$$
\inf_{y \in C_{j,n}} (S_{\oplus}(y; \boldsymbol{x}, u) - S_{\oplus}(s_{\oplus}(\boldsymbol{x}, u); \boldsymbol{x}, u) \geq C_1 d(y, s_{\oplus}(\boldsymbol{x}, u))^{\beta_1}) > C_1 (2^{2(j-1)} / r_h^2).
$$

Let $T_n(y; \boldsymbol{x}, u) = \widetilde{S}_n(y; \boldsymbol{x}, u) - S_{\oplus}(y; \boldsymbol{x}, u)$. In the following, we will show that, for some constant a, we have

$$
\sup_{d(y, s_{\oplus}) < (2^j / r_h)^{2/\beta_1}} |T_n(y) - T_n(s_{\oplus})| \leq a (2^j / r_h)^{2/\beta_1} h^2.
$$

In fact, for $y_1, y_2 \in \mathcal{Y}$, we have the following decomposition

$$T_n(y_1;\boldsymbol{x},u)-T_n(y_2;\boldsymbol{x},u) = \Delta_0(y_1,y_2)+\widetilde{w}_1\widetilde{w}_2^{-1}\Delta_1(y_1,y_2)-w_1w_2^{-1}\Delta_2(y_1,y_2), \tag{S3.1}$$

where $\Delta_0(y_1,y_2) = (\widetilde{w}_0(y_1)-w_0(y_1))-(\widetilde{w}_0(y_2)-w_0(y_2))$, $\Delta_1(y_1,y_2) = \widetilde{w}_3(y_1)-\widetilde{w}_3(y_2)$ and $\Delta_2(y_1,y_2) = w_3(y_1)-w_3(y_2)$. Based on the identity established in the proof of Theorem 1,

$$\mathbb{E}\left[\zeta_h(U,u)d^2(Y,y)\right] = \int d^2(y',y)(1+O(h^2))\, d\mathbb{P}_{Y|U=u}(y'),$$

we can derive that

$$\begin{aligned} |\Delta_0(y_1,y_2)| &= |\mathbb{E}(\zeta_h(U,u)(d^2(Y,y_1)-d^2(Y,y_2)))-\mathbb{E}(d^2(Y,y_1)-d^2(Y,y_2)\mid U=u)| \\ &\le \int |d^2(y,y_1)-d^2(y,y_2)| d\mathbb{P}_{Y|U=u}(y)O(h^2) = O(d(y_1,y_2)h^2). \end{aligned}$$

Using $\widetilde{w}_2^{-1}-w_2^{-1} = -\widetilde{w}_2^{-1}(\widetilde{w}_2-w_2)w_2^{-1}$, we obtain $\widetilde{w}_1\widetilde{w}_2^{-1} = w_1w_2^{-1}+\delta_1\widetilde{w}_2^{-1}-w_1w_2^{-1}\delta_2\widetilde{w}_2^{-1}+O(h^4)$. Then, it follows that

$$\begin{aligned} &\widetilde{w}_1\widetilde{w}_2^{-1}\Delta_1(y_1,y_2)-w_1w_2^{-1}\Delta_2(y_1,y_2) \\ &= w_1w_2^{-1}(\Delta_1(y_1,y_2)-\Delta_2(y_1,y_2))+\delta_1\widetilde{w}_2^{-1}\Delta_1(y_1,y_2)-w_1w_2^{-1}\delta_2\widetilde{w}_2^{-1}\Delta_1(y_1,y_2). \end{aligned}$$

Let $\Delta = d^2(Y,y_1)-d^2(Y,y_2)$. Elementary calculus then shows that

$$\begin{aligned} &\Delta_1(y_1,y_2)-\Delta_2(y_1,y_2) = \mathbb{E}[\zeta_h(\boldsymbol{X}-\mathbb{E}(\zeta_h\boldsymbol{X}))\Delta]-\mathbb{E}[(\boldsymbol{X}-\mathbb{E}(\boldsymbol{X}\mid U=u))\Delta\mid U=u] \\ &= \mathbb{E}[\zeta_h\boldsymbol{X}\Delta]-\mathbb{E}(\boldsymbol{X}\Delta\mid U=u)+\mathbb{E}(\boldsymbol{X}\mid U=u)\mathbb{E}(\Delta\mid U=u)-\mathbb{E}(\zeta_h\boldsymbol{X})\mathbb{E}(\zeta_h\Delta), \end{aligned}$$

where $\mathbb{E}[\zeta_h \boldsymbol{X}\Delta] - \mathbb{E}(\boldsymbol{X}\Delta \mid U=u) = \boldsymbol{O}(h^2 d(y_1,y_2))$ since $\boldsymbol{X}\Delta$ is bounded.

The second term

$$\begin{aligned}
&\mathbb{E}(\boldsymbol{X} \mid U=u)\mathbb{E}(\Delta \mid U=u) - \mathbb{E}(\zeta_h \boldsymbol{X})\mathbb{E}(\zeta_h \Delta) \\
&= (\mathbb{E}(\boldsymbol{X} \mid U=u) - \mathbb{E}(\zeta_h \boldsymbol{X}))\mathbb{E}(\Delta \mid U=u) + \mathbb{E}(\zeta_h \boldsymbol{X})(\mathbb{E}(\Delta \mid U=u) - \mathbb{E}(\zeta_h \Delta)) \\
&= \boldsymbol{O}(h^2)O(d(y_1,y_2)) + \boldsymbol{O}(1)O(h^2 d(y_1,y_2)) = \boldsymbol{O}(h^2 d(y_1,y_2)).
\end{aligned}$$

Hence $\Delta_1(y_1,y_2) - \Delta_2(y_1,y_2) = \boldsymbol{O}(h^2 d(y_1,y_2))$. Together with $\Delta_1(y_1,y_2) = \boldsymbol{O}(d(y_1,y_2))$ and $\Delta_2(y_1,y_2) = \boldsymbol{O}(d(y_1,y_2))$, we can get $|T_n(y_1) - T_n(y_2)| \leq Ch^2 d(y_1,y_2)$. Taking $y_2 = s_\oplus(\boldsymbol{x},u)$ and $d(y,s_\oplus) < (2^j/r_h)^{2/\beta_1}$ gives $\sup_{d(y,s_\oplus)<(2^j/r_h)^{2/\beta_1}} |T_n(y) - T_n(s_\oplus)| \leq a(2^j/r_h)^{2/\beta_1} h^2$.

If $2^j < r_h \widetilde{\eta}_1$, then

$$\begin{aligned}
\mathbb{1}\Big\{\inf_{y\in C_{j,n}} (\widetilde{S}_n(y) - \widetilde{S}_n(s_\oplus)) \leq 0\Big\} &\leq \mathbb{1}\Big\{\inf_{y\in C_{j,n}} (S_\oplus(y) - S_\oplus(s_\oplus)) + \sup_{y\in C_{j,n}} (\widetilde{S}_n(s_\oplus) - \widetilde{S}_n(y)) \leq 0\Big\} \\
&\leq \mathbb{1}\Big\{\sup_{y\in C_{j,n}} |T_n(y) - T_n(s_\oplus)| > C_1 \frac{2^{2(j-1)}}{r_h^2}\Big\}
\end{aligned}$$

by noticing that

$$\sup_{y\in C_{j,n}} |T_n(y) - T_n(s_\oplus)| \geq \sup_{y\in C_{j,n}} (T_n(y) - T_n(s_\oplus)) = \sup_{y\in C_{j,n}} [\widetilde{S}_n(s_\oplus) - \widetilde{S}_n(y;\boldsymbol{x},u) + S_\oplus(y;\boldsymbol{x},u) - S_\oplus(s_\oplus)]$$

and

$$\sup_{y\in C_{j,n}} (\widetilde{S}_n(s_\oplus) - \widetilde{S}_n(y)) = -\inf_{y\in C_{j,n}} (\widetilde{S}_n(y) - \widetilde{S}_n(s_\oplus)).$$

With this, we can get $\mathbb{1}\big\{r_h d(s_\oplus(\boldsymbol{x},u), \widetilde{s}_\oplus(\boldsymbol{x},u))^{\beta_1/2} > 2^M\big\}$ is less or equal to

$$\sum_{\substack{j>M \\ 2^j<r_h\widetilde{\eta}_1}} \mathbb{1}\big\{\sup_{y\in C_{j,n}} |T_n(y)-T_n(s_\oplus)| > C_1\frac{2^{2(j-1)}}{r_h^2}\big\} + \mathbb{1}\big\{2d\left(s_\oplus,\widetilde{s}_\oplus\right)^{\beta_1/2} > \widetilde{\eta}_1\big\},$$

where the second term of the right-hand side goes to zero for any $\widetilde{\eta}_1 > 0$ since $d\left(s_\oplus,\widetilde{s}_\oplus\right) = o\left(1\right)$. Now we focus on the first term. Since $\mathbb{1}\{\sup_{y\in C_{j,n}} |T_n(y)-T_n(s_\oplus)| > C_1\frac{2^{2(j-1)}}{r_h^2}\} \le \sup_{y\in C_{j,n}} |T_n(y)-T_n(s_\oplus)|\frac{r_h^2}{C_1 2^{2(j-1)}}$, we have

$$\begin{aligned}\sum_{\substack{j>M \\ 2^j<r_h\widetilde{\eta}_1}} \mathbb{1}\big\{\sup_{y\in C_{j,n}} |T_n(y)-T_n(s_\oplus)| > C_1\frac{2^{2(j-1)}}{r_h^2}\big\} &\le \sum_{\substack{j>M \\ 2^j<r_h\widetilde{\eta}_1}} \sup_{y\in C_{j,n}} |T_n(y)-T_n(s_\oplus)|\frac{r_h^2}{C_1 2^{2(j-1)}} \\ &\le 4aC_1^{-1} \sum_{\substack{j>M \\ 2^j<r_h\widetilde{\eta}_1}} \frac{h^2 2^{2j(1-\beta_1)/\beta_1}}{r_h^{2(1-\beta_1)/\beta_1}} \\ &\le 4aC_1^{-1} \sum_{j>M} \left(\frac{1}{4^{(\beta_1-1)/\beta_1}}\right)^j,\end{aligned}$$

which converges since $\beta_1 > 1$. It follows that $\sup_{j\ge M}\sup_{y\in C_{j,n}} |T_n(y) - T_n(s_\oplus)| \le C_1\frac{2^{2(j-1)}}{r_h^2}$ for sufficiently large M. Thus, $d\left(s_\oplus\left(\boldsymbol{x},u\right),\widetilde{s}_\oplus\left(\boldsymbol{x},u\right)\right) = O(r_h^{-2/\beta_1}) = O(h^{2/(\beta_1-1)})$. This completes the proof of Theorem 1. □

Before getting the convergence rate of the stochastic element $d\left(\widetilde{s}_\oplus\left(\boldsymbol{x},u\right),\widehat{s}_\oplus\left(\boldsymbol{x},u\right)\right)$, we need the following lemmas,

Lemma 2. *Let $\mathbb{M}_n$ be stochastic processes indexed by a metric space Θ, and let $M_n : \Theta \mapsto \mathbb{R}$ be deterministic functions. Suppose that $\|\mathbb{M}_n - M_n\|_\Theta \to 0$ in probability and that there exists a sequence θ_n such that*

$$\liminf_{n\to\infty} \inf_{d(\theta,\theta_n)>\epsilon} (M_n(\theta) - M_n(\theta_n)) > 0,$$

for every $\epsilon > 0$. *Then any sequence* $\widehat{\theta}_n$, *such that* $\mathbb{M}_n(\widehat{\theta}_n) \leq \inf_{\theta \in \Theta} \mathbb{M}_n(\theta) + o_p(1)$, *satisfies* $d(\widehat{\theta}_n, \theta_n) \to 0$ *in probability.*

Proof. Since $\|\mathbb{M}_n - M_n\|_\Theta \to 0$ in probability, it follows that $\mathbb{M}_n(\theta_n) - M_n(\theta_n) \to 0$ in probability. So we have

$$\mathbb{M}_n(\widehat{\theta}_n) \leq \inf_{\theta \in \Theta} \mathbb{M}_n(\theta) + o_p(1) \leq \mathbb{M}_n(\theta_n) + o_p(1) \leq M_n(\theta_n) + o_p(1).$$

Therefore, again by the uniform convergence, we have

$$\begin{aligned} M_n(\widehat{\theta}_n) - M_n(\theta_n) &\leq M_n(\widehat{\theta}_n) - \mathbb{M}_n(\widehat{\theta}_n) + o_p(1) \\ &\leq \|\mathbb{M}_n - M_n\|_\Theta + o_p(1) \xrightarrow{\mathbb{P}} 0. \end{aligned}$$

Additionally, given $\epsilon > 0$, there exists $\delta > 0$, $N \in \mathbb{N}^+$, when $n > N$ and $d(\theta, \theta_n) \geq \epsilon$, we have

$$M_n(\theta) - M_n(\theta_n) > \delta.$$

Thus for $n > N$, we can obtain $\{d(\widehat{\theta}_n, \theta_n) \geq \epsilon\} \subseteq \{M_n(\widehat{\theta}_n) - M_n(\theta_n) > \delta\}$. Therefore, $\mathbb{P}(d(\widehat{\theta}_n, \theta_n) \geq \epsilon) \to 0$. □

Lemma 3. *Assume* $\mathcal{Y}$ *and* $\boldsymbol{X}$ *are bounded. Under Assumptions (A1) and (A3), if* $h \to 0$ *and* $nh \to \infty$, *then*

$$d\left(\widetilde{s}_{\oplus}\left(\boldsymbol{x}, u\right), \widehat{s}_{\oplus}\left(\boldsymbol{x}, u\right)\right) = o_p(1).$$

Proof. According to the Lemma 2 and assumption (A1), we only need to show that $\sup_{y \in \mathcal{Y}} |\widetilde{S}_n\left(y; \boldsymbol{x}, u\right) - \widehat{S}_n\left(y; \boldsymbol{x}, u\right)| \xrightarrow{\mathbb{P}} 0$, then this lemma can be obtained. Denote by $l^\infty(\mathcal{Y})$ the space of bounded functions on $\mathcal{Y}$. To achieve this, we can demonstrate that $\widetilde{S}_n(\cdot, \boldsymbol{x}, u) - \widehat{S}_n(\cdot, \boldsymbol{x}, u) \xrightarrow{W} 0$ in $l^\infty(\mathcal{Y})$ and then apply Theorem 1.3.6 from Van Der Vaart & Wellner (1996). By Theorem 1.5.4 of Van Der Vaart & Wellner (1996), it is sufficient to show that $\widetilde{S}_n(\cdot, \boldsymbol{x}, u) - \widehat{S}_n(\cdot, \boldsymbol{x}, u)$ is asymptotically tight and the marginals converge weakly. Combined with Theorem 1.5.7 of Van Der Vaart & Wellner (1996), it is sufficient to show

(i) $\widetilde{S}_n(y; \boldsymbol{x}, u) - \widehat{S}_n(y; \boldsymbol{x}, u) = o_p(1)$ for each $y \in \mathcal{Y}$,

(ii) For all $\epsilon, \eta > 0$, there exists $\delta > 0$ such that

$$\limsup_{n\to\infty} \mathbb{P}\left(\sup_{d(y_1,y_2)<\delta} \left| \left(\widetilde{S}_n - \widehat{S}_n\right)(y_1; \boldsymbol{x}, u) - \left(\widetilde{S}_n - \widehat{S}_n\right)(y_2; \boldsymbol{x}, u) \right| > \epsilon \right) < \eta.$$

First, we prove (i). Let $\zeta_{hi} = \zeta_h(U_i, u)$, then we have

$$\widehat{w}_0 - \widetilde{w}_0 = \frac{1}{n}\sum_{i=1}^n (s_{in} - \zeta_{hi}) d^2(Y_i, y) + \frac{1}{n}\sum_{i=1}^n (\zeta_{hi} d^2(Y_i, y) - \mathbb{E}(\zeta_{hi} d^2(Y_i, y))). \tag{S3.2}$$

Observe that $s_{in} - \zeta_{hi} = W_{0n} K_h(U_i - u) + W_{1n} K_h(U_i - u)(U_i - u)$, where

$$W_{0n} = \frac{\widehat{\mu}_2}{\widehat{\sigma}_0^2} - \frac{\widetilde{\mu}_2}{\widetilde{\sigma}_0^2}, \quad W_{1n} = \frac{\widehat{\mu}_1}{\widehat{\sigma}_0^2} - \frac{\widetilde{\mu}_1}{\widetilde{\sigma}_0^2}.$$

We will determine the order of W_{0n}, which can be written as $(\widehat{\mu}_2 \widetilde{\sigma}_0^2 - \widehat{\sigma}_0^2 \widetilde{\mu}_2)/\widehat{\sigma}_0^2 \widetilde{\sigma}_0^2$. From the lemma 1, we have $\widehat{\mu}_j(u) = \widetilde{\mu}_j(u) + O_p((h^{2j-1} n^{-1})^{1/2})$ for $j = 0, 1, 2$, and $\widetilde{\mu}_2 = O(h^2)$. Moreover, the symmetry condition of K in assumption (A3) ensures that $\widetilde{\mu}_1 = h f_U(u) K_{11} + O(h^2) = O(h^2)$. Let

$\epsilon_j := \widehat{\mu}_j(u) - \widetilde{\mu}_j(u)$, then we can get that

$$\begin{aligned}\widehat{\mu}_2\widetilde{\sigma}_0^2 - \widehat{\sigma}_0^2\widetilde{\mu}_2 &= (\widetilde{\mu}_2 + \epsilon_2)(\widetilde{\mu}_0\widetilde{\mu}_2 - \widetilde{\mu}_1^2) - \widetilde{\mu}_2((\widetilde{\mu}_0 + \epsilon_0)(\widetilde{\mu}_2 + \epsilon_2) - (\widetilde{\mu}_1 + \epsilon_1)^2)\\ &= -\widetilde{\mu}_2^2\epsilon_0 - \widetilde{\mu}_2\epsilon_0\epsilon_2 - \widetilde{\mu}_1^2\epsilon_2 + 2\widetilde{\mu}_1\widetilde{\mu}_2\epsilon_1 + \widetilde{\mu}_2\epsilon_1^2\\ &= O(h^4)(O_p(h^{-1/2}n^{-1/2}) + O_p(h^{3/2}n^{-1/2}) + O_p(h^{1/2}n^{-1/2})) + O(h^2)O_p(hn^{-1}).\end{aligned}$$

Since $nh \to \infty$ and $h \to 0$ as $n \to \infty$, we have $\widehat{\mu}_2\widetilde{\sigma}_0^2 - \widehat{\sigma}_0^2\widetilde{\mu}_2 = O_p(h^{7/2}n^{-1/2})$. We also can find that $\widetilde{\sigma}_0^2 = \widetilde{\mu}_0\,\widetilde{\mu}_2 - \widetilde{\mu}_1^2 = h^2 f_U^2(K_{10}K_{12} - K_{11}^2) + h^3 f_U f_U'(K_{10}K_{13} - K_{11}K_{12}) + O(h^4)$. Then, $\widehat{\sigma}_0^2 - \widetilde{\sigma}_0^2 = \epsilon_0\widetilde{\mu}_2 + \widetilde{\mu}_0\epsilon_2 - 2\widetilde{\mu}_1\epsilon_1 + \epsilon_0\epsilon_2 - \epsilon_1^2$. Hence, if $nh \to \infty$,

$$\widehat{\sigma}_0^2 - \widetilde{\sigma}_0^2 = O_p(h^{3/2}/\sqrt{n}) + O_p(h^{3/2}/\sqrt{n}) + O_p(h/n) + O_p(h^{5/2}/\sqrt{n}) = o_p(h^2).$$

Therefore $\widehat{\sigma}_0^2\,\widetilde{\sigma}_0^2 = (\widetilde{\sigma}_0^2 + o_p(h^2))\,\widetilde{\sigma}_0^2 = (\widetilde{\sigma}_0^2)^2 + o_p(h^4)$. It follows that

$$\begin{aligned}(\widetilde{\sigma}_0^2)^2 &= (h^2 f_U^2(K_{10}K_{12} - K_{11}^2) + h^3 f_U f_U'(K_{10}K_{13} - K_{11}K_{12}) + O(h^4))^2\\ &= h^4 f_U^4(K_{10}K_{12} - K_{11}^2)^2 + 2h^5 f_U^3 f_U'(K_{10}K_{12} - K_{11}^2)(K_{10}K_{13} - K_{11}K_{12}) + O(h^6).\end{aligned}$$

Consequently, $\widehat{\sigma}_0^2\,\widetilde{\sigma}_0^2 = h^4\, f_U(u)^4(K_{10}K_{12} - K_{11}^2)^2 + o_p(h^4)$. It yields that $W_{0n} = O_p((nh)^{-1/2})$. We also can get $W_{1n} = O_p((nh^3)^{-1/2})$ through analogous arguments. Since the kernel K is symmetric around zero in assumption (A3), $\mathbb{E}(K_h(U_i - u)(U_i - u)^j d^2(Y_i, y))$ is $O(h^j)$. Using a similar approach to that of Lemma 1, it follows that $\mathbb{E}(K_h^2(U_i - u)(U_i - u)^{2j} d^4(Y_i, y)) = O(h^{2j-1})$.

Then the first term of (S3.2) can be written as

$$W_{0n}\frac{1}{n}\sum_{i=1}^{n}K_h(U_i-u)d^2(Y_i,y)+W_{1n}\frac{1}{n}\sum_{i=1}^{n}K_h(U_i-u)(U_i-u)d^2(Y_i,y)$$

$$=W_{0n}O_p(1)+W_{1n}[O(h)+O_p((h/n)^{1/2})]=O_p((nh)^{-1/2})+O_p((nh)^{-1/2}))+O_p((nh)^{-1}),$$

which is $O_p((nh)^{-1/2})$, since $nh\to\infty$ and $h\to 0$. For the second term of (S3.2), it is a centralized sum with variance $n^{-1}\operatorname{Var}(\zeta_{hi}d^2(Y_i,y))$. To derive the order of $\mathbb{E}(\zeta_{hi}^2)$, recall from Lemma 1 that $\widetilde{\mu}_0(u)=f_U(u)+O(h^2)$, $\widetilde{\mu}_1(u)=h^2f_U'(u)K_{12}+O(h^3)$, $\widetilde{\mu}_2(u)=h^2f_U(u)K_{12}+O(h^4)$ and $\widetilde{\sigma}_0^2(u)=h^2f_U(u)^2K_{12}+O(h^4)$. Let $W_i:=K_h(U_i-u)$, $\Delta(U_i):=\widetilde{\mu}_2(u)-\widetilde{\mu}_1(u)\,(U_i-u)$ and $T_i=(U_i-u)/h$. Then $K_h(U_i-u)=K(T_i)/h$ and $\Delta(U_i)=h^2(f_U(u)(K_{12}-K_{11}T_i)+hf_U'(u)(K_{13}-K_{12}T_i)+O(h^2))$. Hence

$$\begin{aligned}W_i^2\,\Delta(U_i)^2&=\frac{K(T_i)^2}{h^2}\,h^4\left\{f_U^2(u)\,(K_{12}-K_{11}T_i)^2+O(h)\right\}\\&=h^2K(T_i)^2\left\{f_U^2(u)\,(K_{12}-K_{11}T_i)^2+O(h)\right\}.\end{aligned}$$

Taking expectation gives that

$$\begin{aligned}\mathbb{E}(W_i^2\,\Delta(U_i)^2)&=h\int h^2K(t)^2(f_U^2(u)(K_{12}-K_{11}t)^2+O(h))(f_U(u)+O(h)t)dt\\&=h^3f_U^3(u)\int K^2(t)(K_{12}-K_{11}t)^2dt\;+O(h^4).\end{aligned}$$

Moreover, $\{\widetilde{\sigma}_0^2(u)\}^{-2}=\{h^4f_U^4(u)(K_{10}K_{12}-K_{11}^2)^2\}^{-1}(1+O(h))$. Therefore,

$$\frac{\mathbb{E}(W_i^2\Delta(U_i)^2)}{\{\widetilde{\sigma}_0^2(u)\}^2}=\frac{1}{h\,f_U(u)}\,\frac{\int K(t)^2\,(K_{12}-K_{11}t)^2\,dt}{(K_{10}K_{12}-K_{11}^2)^2}\;+\;O(1).$$

For the symmetric kernel $K_{11} = 0$,

$$\mathbb{E}(\zeta_{hi}^2) = \frac{1}{h\, f_U(u)} \frac{K_{12}^2 \int K(t)^2\, dt}{(K_{10}K_{12})^2} + O(1) = \frac{1}{h\, f_U(u)} \frac{\int K(t)^2\, dt}{K_{10}^2} + O(1).$$

It follows that the second term of (S3.2) is $O_p((nh)^{-1/2})$. Thus, $\widehat{w}_0\,(y;u) = \widetilde{w}_0\,(y;u) + O_p((nh)^{-1/2})$. The difference between $\widehat{w}_1\,(\boldsymbol{x},u) - \widetilde{w}_1\,(\boldsymbol{x},u)$ can be written as

$$\widehat{w}_1 - \widetilde{w}_1 = \frac{1}{n}\sum_{i=1}^{n}(\zeta_{hi}\boldsymbol{X}_i^{\mathrm{T}} - \mathbb{E}(\zeta_{hi}\boldsymbol{X}_i)^{\mathrm{T}}) + \frac{1}{n}\sum_{i=1}^{n}(s_{in} - \zeta_{hi})\boldsymbol{X}_i^{\mathrm{T}},$$

which the first term is $\boldsymbol{O}_p((nh)^{-1/2})^{\mathrm{T}}$ due to $\mathbb{E}(\zeta_{hi}^2) = O(h^{-1})$ and $\boldsymbol{X}$ is bounded. A similar argument also yields that the second term is $\boldsymbol{O}_p((nh)^{-1/2})^{\mathrm{T}}$, where $\boldsymbol{O}_p((nh)^{-1/2})$ is the p-dimensional vector whose each component is $O_p((nh)^{-1/2})$. Thus, $\widehat{w}_1\,(\boldsymbol{x},u) = \widetilde{w}_1\,(\boldsymbol{x},u) + \boldsymbol{O}_p((nh)^{-1/2})^{\mathrm{T}}$.

Let $\bar{\boldsymbol{X}}_h = n^{-1}\sum_{i=1}^{n} s_{in}\boldsymbol{X}_i$ and $\mu_{\boldsymbol{X}_h} = \mathbb{E}(\zeta_{hi}\boldsymbol{X}_i)$, we can get that $\|\mu_{\boldsymbol{X}_h}\| \leq \|\boldsymbol{X}_i\| \cdot \mathbb{E}(|\zeta_{hi}|) \leq \|\boldsymbol{X}_i\| \cdot O(h^{-2})[O(h^2) + O(h^3)] = \boldsymbol{O}(1)$. It follows that $\boldsymbol{X}_i - \mu_{\boldsymbol{X}_h} = \boldsymbol{O}_p(1)$. We pay attention to the difference between $\widehat{w}_2\,(u)$ and $\widetilde{w}_2\,(u)$,

$$\widehat{w}_2 - \widetilde{w}_2 = \underbrace{\frac{1}{n}\sum_{i=1}^{n}(s_{in} - \zeta_{hi})(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_h)(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_h)^{\mathrm{T}}}_{A}$$
$$+ \underbrace{\frac{1}{n}\sum_{i=1}^{n}\{\zeta_{hi}(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_h)(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_h)^{\mathrm{T}} - \mathbb{E}[\zeta_{hi}(\boldsymbol{X}_i - \mu_{\boldsymbol{X}_h})(\boldsymbol{X}_i - \mu_{\boldsymbol{X}_h})^{\mathrm{T}}]\}}_{B},$$

We rewrite $\boldsymbol{X}_i - \bar{\boldsymbol{X}}_h = \boldsymbol{X}_i - \mu_{\boldsymbol{X}_h} - (\bar{\boldsymbol{X}}_h - \mu_{\boldsymbol{X}_h}) = \boldsymbol{X}_i - \mu_{\boldsymbol{X}_h} + \boldsymbol{O}_p((nh)^{-1/2})$ and denote $(\boldsymbol{X}_i - \mu_{\boldsymbol{X}_h})(\boldsymbol{X}_i - \mu_{\boldsymbol{X}_h})^{\mathrm{T}}$ as $\Sigma_{\boldsymbol{X}_h}$. We expand the product:

$(\boldsymbol{X}_i-\bar{\boldsymbol{X}}_h)(\boldsymbol{X}_i-\bar{\boldsymbol{X}}_h)^{\mathrm{T}}=\Sigma_{\boldsymbol{X}_h}-(\boldsymbol{X}_i-\mu_{\boldsymbol{X}_h})(\bar{\boldsymbol{X}}_h-\mu_{\boldsymbol{X}_h})^{\mathrm{T}}-(\bar{\boldsymbol{X}}_h-\mu_{\boldsymbol{X}_h})(\boldsymbol{X}_i-\mu_{\boldsymbol{X}_h})^{\mathrm{T}}+(\bar{\boldsymbol{X}}_h-\mu_{\boldsymbol{X}_h})(\bar{\boldsymbol{X}}_h-\mu_{\boldsymbol{X}_h})^{\mathrm{T}}$. For every element (k,l) in $\Sigma_{\boldsymbol{X}_h}$, we have $\mathbb{E}(\frac{1}{n}\sum_{i=1}^n K_h(U_i-u)\Sigma_{\boldsymbol{X}_{h,kl}})=O(1)$, $\mathrm{Var}(\frac{1}{n}\sum_{i=1}^n K_h(U_i-u)\Sigma_{\boldsymbol{X}_{h,kl}})=O((nh)^{-1})$, $\mathbb{E}(\frac{1}{n}\sum_{i=1}^n K_h(U_i-u)(U_i-u)\Sigma_{\boldsymbol{X}_{h,kl}})=O(h)$ and $\mathrm{Var}(\frac{1}{n}\sum_{i=1}^n K_h(U_i-u)(U_i-u)\Sigma_{\boldsymbol{X}_{h,kl}})=O(h/n)$. Then break A into a equation containing W_{0n} and W_{1n} as before, for every element (k,l) in A, we have

$$\begin{aligned}A_{kl}&=W_{0n}\frac{1}{n}\sum_{i=1}^n K_h(U_i-u)\Sigma_{\boldsymbol{X}_{h,kl}}+W_{1n}\frac{1}{n}\sum_{i=1}^n K_h(U_i-u)(U_i-u)\Sigma_{\boldsymbol{X}_{h,kl}}+O_p((nh)^{-1})\\&=O_p((nh)^{-1/2})O(1)+O_p((nh^3)^{-1/2})(O(h)+O_p((h/n)^{1/2}))+O_p((nh)^{-1}),\end{aligned}$$

thus $A=\boldsymbol{O}_p((nh)^{-1/2})_{p\times p}$, which is a $p\times p$ matrix composed of $O_p((nh)^{-1/2})$. We also can get $B=B_{\text{main}}-B_{\text{cross}}$, where $B_{\text{main}}=n^{-1}\sum_{i=1}^n\zeta_{hi}\Sigma_{\boldsymbol{X}_h}-\mathbb{E}\left[\zeta_{hi}\Sigma_{\boldsymbol{X}_h}\right]$ and $B_{\text{cross}}=\frac{1}{n}\sum_{i=1}^n\zeta_{hi}[(\boldsymbol{X}_i-\mu_{\boldsymbol{X}_h})(\bar{\boldsymbol{X}}_h-\mu_{\boldsymbol{X}_h})^{\mathrm{T}}+(\bar{\boldsymbol{X}}_h-\mu_{\boldsymbol{X}_h})(\boldsymbol{X}_i-\mu_{\boldsymbol{X}_h})^{\mathrm{T}}-(\bar{\boldsymbol{X}}_h-\mu_{\boldsymbol{X}_h})(\bar{\boldsymbol{X}}_h-\mu_{\boldsymbol{X}_h})^{\mathrm{T}}]$. Define $Z_i=\zeta_{hi}\Sigma_{\boldsymbol{X}_h}$. Since $\|\boldsymbol{X}_i\|\le C$ and $\mathbb{E}[\zeta_{hi}^2]=O(h^{-1})$, for each element (k,l): $\mathrm{Var}(Z_{i,kl})\le(4C)^2\mathbb{E}[\zeta_{hi}^2]=O(h^{-1})$, then $\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n Z_{i,kl}\right)=\frac{1}{n}\mathrm{Var}(Z_{i,kl})=O\left((nh)^{-1}\right)$. Thus $B_{\text{main}}=\boldsymbol{O}_p((nh)^{-1/2})$. We decompose $B_{\text{cross}}=B_2+B_3+B_4$, where $B_2=-\frac{1}{n}\sum_{i=1}^n\zeta_{hi}(\boldsymbol{X}_i-\mu_{\boldsymbol{X}_h})(\bar{\boldsymbol{X}}_h-\mu_{\boldsymbol{X}_h})^{\mathrm{T}}$, $B_3=-\frac{1}{n}\sum_{i=1}^n\zeta_{hi}(\bar{\boldsymbol{X}}_h-\mu_{\boldsymbol{X}_h})(\boldsymbol{X}_i-\mu_{\boldsymbol{X}_h})^{\mathrm{T}}$, $B_4=\frac{1}{n}\sum_{i=1}^n\zeta_{hi}(\bar{\boldsymbol{X}}_h-\mu_{\boldsymbol{X}_h})(\bar{\boldsymbol{X}}_h-\mu_{\boldsymbol{X}_h})^{\mathrm{T}}$. We have $\bar{\boldsymbol{X}}_h-\mu_{\boldsymbol{X}_h}=\boldsymbol{O}_p((nh)^{-1/2})$, $n^{-1}\sum_{i=1}^n|\zeta_{hi}|=O_p(\mathbb{E}|\zeta_{hi}|+\sqrt{\mathbb{E}[\zeta_{hi}^2]/n})=O_p(1)$, and $n^{-1}\sum_{i=1}^n\zeta_{hi}(\boldsymbol{X}_i-\mu_{\boldsymbol{X}_h})=\boldsymbol{O}_p((nh)^{-1/2})$. Thus, for each element (k,l), $B_{2,kl}=O_p((nh)^{-1/2})$, $B_{3,kl}=O_p((nh)^{-1/2})$,

$B_{4,kl} = O_p(1) \cdot O_p((nh)^{-1}) = O_p((nh)^{-1})$. Consequently, $B_{\text{cross}} = \boldsymbol{O}_p((nh)^{-1/2})_{p\times p}$, and thus $B = \boldsymbol{O}_p((nh)^{-1/2})_{p\times p}$ as $nh \to \infty$. Thus, we have $\widehat{w}_2\,(u) = \widetilde{w}_2\,(u) + \boldsymbol{O}_p((nh)^{-1/2})_{p\times p}$. For the difference between $\widehat{w}_3\,(y;u)$ and $\widetilde{w}_3\,(y;u)$, let $\bar{d_h^2} = n^{-1}\sum_{i=1}^n s_{in} d^2(Y_i, y)$ and $\mu_{d_h^2} = \mathbb{E}(\zeta_{hi} d^2(Y_i, y))$, then we have

$$\widehat{w}_3 - \widetilde{w}_3 = \frac{1}{n}\sum_{i=1}^n (s_{in} - \zeta_{hi})(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_h)(d^2(Y_i, y) - \bar{d_h^2})$$
$$+ \frac{1}{n}\sum_{i=1}^n \{\zeta_{hi}(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_h)(d^2(Y_i, y) - \bar{d_h^2}) - \mathbb{E}[\zeta_{hi}(\boldsymbol{X}_i - \mu_{\boldsymbol{X}_h})(d^2(Y_i, y) - \mu_{d_h^2})]\},$$

using the same method, we can get that $\widehat{w}_3\,(y;u) = \widetilde{w}_3\,(y;u) + \boldsymbol{O}_p((nh)^{-1/2})$. Then, it can be verified that $\widehat{S}_n\,(y;\boldsymbol{x},u) - \widetilde{S}_n\,(y;\boldsymbol{x},u) = O_p((nh)^{-1/2})$ as we do before. When $h \to 0$ and $nh \to \infty$, we have that $\widetilde{S}_n\,(y;\boldsymbol{x},u) - \widehat{S}_n\,(y;\boldsymbol{x},u) = o_p\,(1)$.

Next, we prove (ii). For any $y_1, y_2 \in \mathcal{Y}$, $|(\widetilde{S}_n - \widehat{S}_n)(y_1;\boldsymbol{x},u) - (\widetilde{S}_n - \widehat{S}_n)(y_2;\boldsymbol{x},u)|$ is less than or equal to $|\widehat{S}_n(y_1;\boldsymbol{x},u) - \widehat{S}_n(y_2;\boldsymbol{x},u)| + |\widetilde{S}_n(y_1;\boldsymbol{x},u) - \widetilde{S}_n(y_2;\boldsymbol{x},u)|$, which is equal to $|\widehat{w}_0(y_1) - \widehat{w}_0(y_2) + \widehat{w}_1\widehat{w}_2^{-1}(\widehat{w}_3(y_1) - \widehat{w}_3(y_2))| + |\widetilde{w}_0(y_1) - \widetilde{w}_0(y_2) + \widetilde{w}_1\widetilde{w}_2^{-1}(\widetilde{w}_3(y_1) - \widetilde{w}_3(y_2))|$. Since $\widehat{w}_1\widehat{w}_2^{-1} = \boldsymbol{O}_p(1)^{\mathrm{T}}$ and $n^{-1}\sum_{i=1}^n |s_{in}| = O_p(1)$, there exists $M_n = O_p(1)$ such that

$$\left|\widehat{S}_n(y_1;\boldsymbol{x},u) - \widehat{S}_n(y_2;\boldsymbol{x},u)\right| \le M_n |d^2(Y_i, y_1) - d^2(Y_i, y_2)|$$
$$\le M_n |d(Y_i, y_1) - d(Y_i, y_2)||d(Y_i, y_1) + d(Y_i, y_2)|$$
$$\le M_n 2\,\mathrm{diam}(\mathcal{Y}) d(y_1, y_2) = O_p(d(y_1, y_2)).$$

Due to $\widetilde{w}_1\widetilde{w}_2^{-1} = \boldsymbol{O}(1)^{\mathrm{T}}$ and $\mathbb{E}(\zeta_h(U,u)) = O(1)$, it follows that $\left|\widetilde{S}_n(y_1;\boldsymbol{x},u) - \widetilde{S}_n(y_2;\boldsymbol{x},u)\right| \le$

$O(d(y_1,y_2))$. Therefore, we have $\sup_{d(y_1,y_2)<\delta} |(\widetilde{S}_n - \widehat{S}_n)(y_1;\boldsymbol{x},u) - (\widetilde{S}_n - \widehat{S}_n)(y_2;\boldsymbol{x},u)| = O_p(\delta)$, which can deduce (ii). Hence, $d\left(\widetilde{s}_\oplus\left(\boldsymbol{x},u\right),\widehat{s}_\oplus\left(\boldsymbol{x},u\right)\right) = o_p(1)$. □

Next, we establish the convergence rate of $d(\widetilde{s}_\oplus(\boldsymbol{x},u),\widehat{s}_\oplus(\boldsymbol{x},u))$.

Proof of Theorem 2. For notational simplicity, we set $T_n(y;\boldsymbol{x},u) = \widehat{S}_n(y;\boldsymbol{x},u) - \widetilde{S}_n(y;\boldsymbol{x},u)$, $D_i = d^2(Y_i,y) - d^2(Y_i,\widetilde{s}_\oplus(\boldsymbol{x},u))$, $\bar{\boldsymbol{X}}_h = n^{-1}\sum_{i=1}^n s_{in}\boldsymbol{X}_i$, $\bar{D}_h = n^{-1}\sum_{i=1}^n s_{in}D_i$, $\mu_{\boldsymbol{X}_h} = \mathbb{E}(\zeta_{hi}\boldsymbol{X}_i)$, and $\mu_{D_h} = \mathbb{E}(\zeta_{hi}D_i)$. Then, we have the follow inequality,

$$\begin{aligned} |T_n\left(y;\boldsymbol{x},u\right) - T_n\left(\widetilde{s}_\oplus\left(\boldsymbol{x},u\right);\boldsymbol{x},u\right)| &\le |\frac{1}{n}\sum_{i=1}^n (s_{in}-\zeta_{hi})D_i| + |\frac{1}{n}\sum_{i=1}^n(\zeta_{hi}D_i - \mathbb{E}(\zeta_{hi}D_i))| + \\ &|\widehat{w}_1\widehat{w}_2^{-1}\frac{1}{n}\sum_{i=1}^n s_{in}(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_h)(D_i - \bar{D}_h) - \widetilde{w}_1\widetilde{w}_2^{-1}\mathbb{E}\{\zeta_{hi}(\boldsymbol{X}_i - \mu_{\boldsymbol{X}_h})(D_i - \mu_{D_h})\}|. \end{aligned} \tag{S3.3}$$

Since $|D_i| \le 2\,\mathrm{diam}\,(\mathcal{Y})\,d\left(y,\widetilde{s}_\oplus\left(\boldsymbol{x},u\right)\right)$, we can rewrite the first term on the right side of (S3.3) into a equation containing W_{0n} and W_{1n} as

$$\begin{aligned} |\frac{1}{n}\sum_{i=1}^n(s_{in}-\zeta_{hi})D_i| &= W_{0n}\frac{1}{n}\sum_{i=1}^n K_h(U_i-u)D_i + W_{1n}\frac{1}{n}\sum_{i=1}^n K_h(U_i-u)(U_i-u)D_i \\ &= O_p((nh)^{-1/2})O_p(d(y,\widetilde{s}_\oplus)) + O_p((nh^3)^{-1/2})O_p((h+(h/n)^{1/2}))d(y,\widetilde{s}_\oplus)) \\ &= O_p((nh)^{-1/2}d(y,\widetilde{s}_\oplus)), \end{aligned}$$

where the $O_p(\cdot)$ is independent of y and $\widetilde{s}_\oplus$.

Next, we control the second term uniformly over $B_{\mathcal{Y}}(\widetilde{s}_\oplus(\boldsymbol{x},u),\delta)$. Define the function $g_y : \mathbb{R}\times\mathcal{Y}\to\mathbb{R}$ by $g_y(U,Y) = \zeta_h(U,u)d^2(Y,y)$, and the associated function class $\mathcal{M}_\delta = \{g_y - g_{\widetilde{s}_\oplus(\boldsymbol{x},u)} : d(y, \widetilde{s}_\oplus(\boldsymbol{x}, u)) < \delta\}$. Then, $\mathcal{M}_\delta$ possesses an envelope function $G_\delta(U, Y) = 2\,\mathrm{diam}(\mathcal{Y})\delta|\zeta_h(U, u)|$ with $\mathbb{E}(G_\delta^2(U, Y)) = O(\delta^2 h^{-1})$. Theorem 2.14.2 of Van Der Vaart & Wellner (1996) yields bound of the second term in terms of the bracketing integral as

$$\mathbb{E}\left(\sup_{d(y,\widetilde{s}_\oplus(\boldsymbol{x},u))<\delta}\left|\frac{1}{n}\sum_{i=1}^{n}\zeta_{hi}D_i - \mathbb{E}(\zeta_{hi}D_i)\right|\right) \leq CJ_{[]}(1, \mathcal{M}_\delta, L_2(P))\sqrt{\mathbb{E}(G_\delta^2(U, Y))/n}.$$

Therefore, $J_{[]}(1, \mathcal{M}_\delta, L_2(P))$ needs to be controlled. Since

$$|g_{y_1}(U, Y) - g_{y_2}(U, Y)| \leq d(y_1, y_2)2\,\mathrm{diam}(\mathcal{Y})\delta|\zeta_h(U, u)|,$$

Theorem 2.7.11 of Van Der Vaart & Wellner (1996) implies

$$N_{[]}(2M\epsilon, \mathcal{M}_\delta, L_2(P)) \leq N(\epsilon, B_{\mathcal{Y}}(\widetilde{s}_\oplus(\boldsymbol{x}, u), \delta), d),$$

where $M = 2\,\mathrm{diam}(\mathcal{Y})\|\zeta_h(U, u)\|_{P,2}$. Note that $\|G_\delta\|_{P,2} = M\delta$, we have

$$\begin{aligned}
J_{[]}(1, \mathcal{M}_\delta, L_2(P)) &= \int_0^1 [1 + \log N_{[]}(\epsilon M\delta, M_\delta, L_2(P)]^{1/2} d\epsilon \\
&\leq \int_0^1 [1 + \log N(\delta\epsilon/2, B_{\mathcal{Y}}(\widetilde{s}_\oplus(\boldsymbol{x}, u), \delta), d)]^{1/2} d\epsilon \\
&= 2\int_0^{1/2} [1 + \log N(\delta\epsilon, B_{\mathcal{Y}}(\widetilde{s}_\oplus(\boldsymbol{x}, u), \delta), d)]^{1/2} d\epsilon \\
&\leq 2\int_0^{1} [1 + \log N(\delta\epsilon, B_{\mathcal{Y}}(\widetilde{s}_\oplus(\boldsymbol{x}, u), \delta), d)]^{1/2} d\epsilon.
\end{aligned}$$

With Assumption (A2), it gives

$$\mathbb{E}\big(\sup_{d(y,\widetilde{s}_\oplus(\boldsymbol{x},u))<\delta}\big|\frac{1}{n}\sum_{i=1}^{n}\zeta_{hi}D_i - \mathbb{E}(\zeta_{hi}D_i)\big|\big) = O(\delta(nh)^{-1/2}).$$

In the same way as we proved $\widehat{S}_n\left(y;\boldsymbol{x},u\right)-\widetilde{S}_n\left(y;\boldsymbol{x},u\right)=O_p((nh)^{-1/2})$ in Lemma 3, we can get that $\sup_{d(y,\widetilde{s}_\oplus(\boldsymbol{x},u))<\delta'}[|\widehat{w}_1\widehat{w}_2^{-1}\frac{1}{n}\sum_{i=1}^n s_{in}(\boldsymbol{X}_i-\bar{\boldsymbol{X}}_h)(D_i-\bar{D}_h)-\widetilde{w}_1\widetilde{w}_2^{-1}\mathbb{E}\{\zeta_{hi}(\boldsymbol{X}_i-\mu_{\boldsymbol{X}_h})(D_i-\mu_{D_h})\}|]$ is $O_p((nh)^{-1/2}d(y,\widetilde{s}_\oplus))$ since $\max_{i=1,\dots,n}D_i=O(d(y,\widetilde{s}_\oplus))$. Then, for $R>0$ and $\delta>0$, define event

$$B(R')=\Big\{\sup_{d(y,\widetilde{s}_\oplus(\boldsymbol{x},u))<\delta'}[|\tfrac{1}{n}\textstyle\sum_{i=1}^n(s_{in}-\zeta_{hi})D_i|+|\tfrac{1}{n}\sum_{i=1}^n(s_{in}-\zeta_{hi})D_i|+$$
$$|\widehat{w}_1\widehat{w}_2^{-1}\tfrac{1}{n}\textstyle\sum_{i=1}^n s_{in}(\boldsymbol{X}_i-\bar{\boldsymbol{X}}_h)(D_i-\bar{D}_h)-\widetilde{w}_1\widetilde{w}_2^{-1}\mathbb{E}\{\zeta_{hi}(\boldsymbol{X}_i-\mu_{\boldsymbol{X}_h})(D_i-\mu_{D_h})\}|]\le R'\delta'(nh)^{-1/2}\Big\}.$$

It follows that $\mathbb{P}(B^c(R'))\to 0$ for sufficiently large R. Therefore, we have

$$\mathbb{E}\Big(\mathbb{1}_{B(R')}\sup_{d(y,\widetilde{s}_\oplus(\boldsymbol{x},u))<\delta'}\left|T_n\left(y;\boldsymbol{x},u\right)-T_n\left(\widetilde{s}_\oplus\left(\boldsymbol{x},u\right);\boldsymbol{x},u\right)\right|\Big)\le C\delta'(nh)^{-1/2},$$

where $\mathbb{1}_{B(R')}$ is the indicator function for the set $B(R')$ and C is a constant depending on R' and the assumption (A6).

To finish the proof, let $t_n=(nh)^{\frac{\beta_2}{4(\beta_2-1)}}$ and define

$$S_{j,n}=\{y:2^{j-1}<t_nd(y,\widetilde{s}_\oplus(\boldsymbol{x},u))^{\beta_2/2}\le 2^j\}.$$

Choose η_2 satisfying both Assumption (A6) and the requirement that condition (A2) holds any $h<\eta_2$. Let $\tilde{\eta}:=(\eta_2/2)^{\beta_2/2}$. For any integer M,

$$\mathbb{P}\left(t_nd(\widetilde{s}_\oplus(\boldsymbol{x},u),\widehat{s}_\oplus(\boldsymbol{x},u))^{\beta_2/2}>2^M\right)\le\mathbb{P}\left(B^c(R')\right)+\mathbb{P}\left(2d(\widetilde{s}_\oplus(\boldsymbol{x},u),\widehat{s}_\oplus(\boldsymbol{x},u))>\tilde{\eta}\right)$$
$$+\sum_{\substack{j\ge M\\2^j\le t_n\tilde{\eta}}}\mathbb{P}[(\mathbb{1}_{B(R')}\sup_{y\in S_{j,n}}\left|T_n\left(y;\boldsymbol{x},u\right)-T_n\left(\widetilde{s}_\oplus\left(\boldsymbol{x},u\right);\boldsymbol{x},u\right)\right|)\ge C\frac{2^{2(j-1)}}{t_n^2}],\tag{S3.4}$$

where the second term tends to 0 for any $\tilde{\eta}>0$ by Lemma 3. By definition of $S_{j,n}$, we have $d(y,\widetilde{s}_\oplus(\boldsymbol{x},u))<(2^j/t_n)^{2/\beta_2}$. Therefore, applying Markov's inequality, one can show that the third term on the right hand side of inequality (S3.4) is bounded by

$$4aC^{-1}\sum_{\substack{j\geq M\\ 2^j\leq t_n\tilde{\eta}}}\frac{2^{2j(1-\beta_2)/\beta_2}}{t_n^{2(1-\beta_2)/\beta_2}\sqrt{nh}}\leq 4aC^{-1}\sum_{j\geq M}\left(\frac{1}{4^{(\beta_2-1)/\beta_2}}\right)^j,$$

which converges since $\beta_2>1$. Combining these ingredients, we conclude that

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

□

Corollary 1 is obtained by applying of the triangle inequality to Theorems 1–2.

Proof of Theorem 3. Define the smoothed expectation operator $\widetilde{\mathbb{E}}_u[\cdot]=\mathbb{E}[\zeta_h(U,u)\cdot]$ and write $\widetilde{\mu}(u)=\widetilde{\mathbb{E}}_u[\boldsymbol{X}]$. By essentially the same arguments in the proof of proof of Lemma 1, we conclude that there exist unique elements $\widetilde{\gamma}_0(u)\in\mathcal{Y}$ and $\widetilde{\gamma}_1(u)\in\mathcal{Y}^p$, such that for all $y\in\mathcal{Y}$ and $\boldsymbol{y}\in\mathcal{Y}^p$, $\widetilde{\mathbb{E}}_u(\langle Y,y\rangle)=\langle\widetilde{\gamma}_0(u),y\rangle$ and $\widetilde{\mathbb{E}}_u(\langle(\boldsymbol{X}-\widetilde{\mu}(u))Y,\boldsymbol{y}\rangle_p)=\langle\widetilde{\gamma}_1(u),\boldsymbol{y}\rangle_p$. Furthermore, $\widetilde{s}_{\oplus}(\boldsymbol{x},u)$ admits the explicit linear form $\widetilde{s}_{\oplus}(\boldsymbol{x},u)=\widetilde{\beta}_0(u)+\boldsymbol{x}^{\mathrm{T}}\widetilde{\beta}_1(u)$, where $\widetilde{\beta}_1(u)=\widetilde{w}_2^{-1}(u)\widetilde{\gamma}_1(u),\widetilde{\beta}_0(u)=\widetilde{\gamma}_0(u)-\widetilde{\mu}(u)^{\mathrm{T}}\widetilde{\beta}_1(u)$. We consider the local linear smoothing scheme defined by the weights $\zeta_h(U,u)$ satisfying $\mathbb{E}[\zeta_h(U,u)]=1$. First, we derive the convergence rate of *bias* term. The goal is to establish that $\|\widetilde{\boldsymbol{\beta}}_0(u)-\boldsymbol{\beta}_0(u)\|_{\mathcal{Y}}^2 = O(h^2)$ and $\|\widetilde{\boldsymbol{\beta}}_1(u)-\boldsymbol{\beta}_1(u)\|_{\mathcal{Y}^p}^2 = O(h^2)$.

Analogous to the proof of Theorem 1, by Assumption (A4), one can show that

$$\mathbb{E}\left(\zeta_h(U,u)f(\boldsymbol{X},Y)\right) = \mathbb{E}\left(f(\boldsymbol{X},Y) \mid U=u\right) + O\left(h^2\right)$$

for any bounded measurable $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{S}$ with $\mathcal{S}$ a Banach space, where the $O\left(h^2\right)$ term holds with respect to the norm of $\mathcal{S}$. Applying this result yields $\widetilde{\mu}(u)-\mu(u) = O(h^2)$, $\widetilde{w}_2(u)-w_2(u) = O(h^2)$, $\widetilde{\gamma}_0(u)-\gamma_0(u) = O(h^2)$, $\widetilde{\gamma}_1(u)-\gamma_1(u) = O(h^2)$, where $O\left(h^2\right)$ terms hold with respect to the corresponding norm. Recall the elements $\gamma_0(u)$, $\widetilde{\gamma}_0(u)$, $\gamma_1(u)$, and $\widetilde{\gamma}_1(u)$ are defined via the Riesz representation theorem. Note that in a Hilbert space, the Bochner integral commutes with the inner product. Therefore, it is equivalent to directly defining the above four terms as the Bochner (conditional) expectations. In the subsequent analysis, we will adopt the Bochner integral definition.

Now we analyse the difference $\widetilde{\boldsymbol{\beta}}_1(u)-\boldsymbol{\beta}_1(u)$. Adding and subtracting terms gives

$$\widetilde{\boldsymbol{\beta}}_1(u)-\boldsymbol{\beta}_1(u) = \widetilde{w}_2^{-1}(u)\big(\widetilde{\gamma}_1(u)-\gamma_1(u)\big) + \big(\widetilde{w}_2^{-1}(u)-w_2^{-1}(u)\big)\gamma_1(u).$$

For the second term, we have shown that

$$\|\widetilde{w}_2^{-1}(u) - w_2^{-1}(u)\| = O(h^2).$$

Consequently, using the boundedness of $\gamma_1(u)$ (which follows from the boundedness of X and Y), we obtain

$$\begin{aligned}\|\widetilde{\boldsymbol{\beta}}_1(u) - \boldsymbol{\beta}_1(u)\|_{\mathcal{Y}^p} &\le \|\widetilde{w}_2^{-1}(u)\|\|\widetilde{\gamma}_1(u) - \gamma_1(u)\|_{\mathcal{Y}^p} + \|\widetilde{w}_2^{-1}(u) - w_2^{-1}(u)\|\|\gamma_1(u)\|_{\mathcal{Y}^p} \\ &= O(1)\cdot O(h^2) + O(h^2)\cdot O(1) = O(h^2).\end{aligned}$$

Next we turn to the $\beta_0(u) - \widetilde{\beta}_0(u)$. Recall $\beta_0(u) = \gamma_0(u) - \mu(u)^{\mathrm{T}}\boldsymbol{\beta}_1(u)$ and $\widetilde{\beta}_0(u) = \widetilde{\gamma}_0(u) - \widetilde{\mu}(u)^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}_1(u)$. Hence

$$\begin{aligned}\widetilde{\beta}_0(u) - \beta_0(u) &= \big(\widetilde{\gamma}_0(u) - \gamma_0(u)\big) - \big(\widetilde{\mu}(u)^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}_1(u) - \mu(u)^{\mathrm{T}}\boldsymbol{\beta}_1(u)\big) \\ &= \big(\widetilde{\gamma}_0(u) - \gamma_0(u)\big) - \Big[\big(\widetilde{\mu}(u) - \mu(u)\big)^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}_1(u) + \mu(u)^{\mathrm{T}}\big(\widetilde{\boldsymbol{\beta}}_1(u) - \boldsymbol{\beta}_1(u)\big)\Big].\end{aligned}$$

It follows from the boundedness of $\widetilde{\boldsymbol{\beta}}_1(u)$ and $\mu(u)$ that

$$\begin{aligned}\|\widetilde{\beta}_0(u) - \beta_0(u)\|_{\mathcal{Y}} &\le \|\widetilde{\gamma}_0(u) - \gamma_0(u)\|_{\mathcal{Y}} + \|\widetilde{\mu}(u) - \mu(u)\|\|\widetilde{\boldsymbol{\beta}}_1(u)\|_{\mathcal{Y}^p} + \|\mu(u)\|\|\widetilde{\boldsymbol{\beta}}_1(u) - \boldsymbol{\beta}_1(u)\|_{\mathcal{Y}^p} \\ &= O(h^2) + O(h^2)\cdot O(1) + O(1)\cdot O(h^2) = O(h^2).\end{aligned}$$

Thus, we have shown that $\|\widetilde{\beta}_0(u) - \beta_0(u)\|_{\mathcal{Y}} = O(h^2)$ and $\|\widetilde{\boldsymbol{\beta}}_1(u) - \boldsymbol{\beta}_1(u)\|_{\mathcal{Y}^p} = O(h^2)$.

We now present the proof of the asymptotic distribution for $\sqrt{nh}(\widehat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}(u))$. By Theorem 1.8.4 of Van Der Vaart & Wellner (1996), the proof proceeds by establishing: the asymptotic normality of the scalar projection $\langle \boldsymbol{y}, \widehat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}(u)\rangle_{\mathcal{Y}^{p+1}}$ for any fixed direction $\boldsymbol{y} = (y_0, \boldsymbol{y}_1^{\mathrm{T}})^{\mathrm{T}} \in \mathcal{Y}^{p+1}$; additionally, $\widehat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}(u)$ is asymptotically finite-dimensional.

Define $\boldsymbol{Z}_i = (1, \boldsymbol{X}_i^{\mathrm{T}})^{\mathrm{T}}$. Redefine $\widehat{\boldsymbol{\beta}}(u) = (\widehat{\beta}_0(u), \widehat{\boldsymbol{\beta}}_1(u)^{\mathrm{T}})^{\mathrm{T}}$, where $\widehat{\beta}_0(u)$ is the intercept and $\widehat{\boldsymbol{\beta}}_1(u)$ is the vector of slope coefficients. The definition for $\widetilde{\boldsymbol{\beta}}(u)$ is analogous to that given for $\widehat{\boldsymbol{\beta}}(u)$. Note that $(\widetilde{\beta}_0(u), \widetilde{\boldsymbol{\beta}}_1(u)^{\mathrm{T}})^{\mathrm{T}}$ is the solution to a locally weighted least squares problem

$$\underset{\beta_0 \in \mathcal{Y}, \beta_1 \in \mathcal{Y}^p}{\arg\min}\ \widetilde{\mathbb{E}}_u \left[\| Y - (\beta_0 + \boldsymbol{X}^{\mathrm{T}} \boldsymbol{\beta}_1) \|_{\mathcal{Y}}^2 \right].$$

and satisfies $\widetilde{\mathbb{E}}_u\big[\boldsymbol{Z}(Y - \boldsymbol{Z}^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}(u))\big] = 0$. Similarly, $\widehat{\boldsymbol{\beta}}(u)$ satisfies

$$(\widehat{\beta}_0(u), \widehat{\boldsymbol{\beta}}_1(u)^{\mathrm{T}})^{\mathrm{T}} = \underset{\beta_0 \in \mathcal{Y}, \boldsymbol{\beta}_1 \in \mathcal{Y}^p}{\arg\min} \sum_{i=1}^{n} s_{in}(u) \| Y_i - \beta_0 - \boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{\beta}_1 \|_{\mathcal{Y}}^2$$

and $\sum_{i=1}^n s_{in}(u)\boldsymbol{Z}_i\big(Y_i - \boldsymbol{Z}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}(u)\big) = 0$. Consequently, we have the following equation

$$\sum_{i=1}^{n} s_{in}(u)\boldsymbol{Z}_i\big(Y_i - \boldsymbol{Z}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}(u)\big) - \Big(\sum_{i=1}^{n} s_{in}(u)\boldsymbol{Z}_i\boldsymbol{Z}_i^{\mathrm{T}}\Big)\big(\widehat{\boldsymbol{\beta}}(u) - \widetilde{\boldsymbol{\beta}}(u)\big) = 0.$$

Set $\widehat{\Sigma}_{\boldsymbol{Z}}(u) = \frac{1}{n}\sum_{i=1}^n s_{in}(u)\boldsymbol{Z}_i\boldsymbol{Z}_i^{\mathrm{T}}$. Then we have the representation

$$\widehat{\boldsymbol{\beta}}(u) - \widetilde{\boldsymbol{\beta}}(u) = \widehat{\Sigma}_{\boldsymbol{Z}}^{-1}(u) \cdot \frac{1}{n}\sum_{i=1}^{n} s_{in}(u)\boldsymbol{Z}_i\big(Y_i - \boldsymbol{Z}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}(u)\big). \tag{S3.5}$$

Introduce the residual based on the smoothed coefficients as $\widetilde{\epsilon}_i = Y_i - \widetilde{\beta}_0(U_i) - \boldsymbol{X}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}_1(U_i)$. From the model $Y_i = \beta_0(U_i) + \boldsymbol{\beta}_1(U_i)^{\mathrm{T}}\boldsymbol{X}_i + \epsilon_i$ with $\mathbb{E}[\epsilon_i \mid U_i, \boldsymbol{X}_i] = 0$, and using the bias result $\|\widetilde{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}(u)\| = O(h^2)$, we have

$$\widetilde{\epsilon}_i = \epsilon_i + \big[\beta_0(U_i) - \widetilde{\beta}_0(U_i)\big] + \big[\boldsymbol{\beta}_1(U_i) - \boldsymbol{X}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}_1(U_i)\big] = \epsilon_i + O(h^2).$$

Write the sample weight as $s_{in}(u) = \zeta_{hi}(u) + r_i(u)$, where $r_i(u) := s_{in}(u) - \zeta_{hi}(u)$. From the proof of Lemma 3, $r_i(u)$ admits the decomposition

$$r_i(u) = W_{0n}K_h(U_i - u) + W_{1n}K_h(U_i - u)(U_i - u),$$

with $W_{0n} = O_p((nh)^{-1/2})$ and $W_{1n} = O_p((nh^3)^{-1/2})$.

Now expand the right-hand side of (S3.5) as

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} s_{in}(u)\boldsymbol{Z}_i\big(Y_i - \boldsymbol{Z}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}(u)\big) &= \frac{1}{n}\sum_{i=1}^{n}\zeta_{hi}(u)\boldsymbol{Z}_i\widetilde{\epsilon}_i + \frac{1}{n}\sum_{i=1}^{n} r_i(u)\boldsymbol{Z}_i\widetilde{\epsilon}_i \\ &\quad + \frac{1}{n}\sum_{i=1}^{n} s_{in}(u)\boldsymbol{Z}_i\boldsymbol{Z}_i^{\mathrm{T}}\big[\widetilde{\boldsymbol{\beta}}(U_i) - \widetilde{\boldsymbol{\beta}}(u)\big] \\ &:= I_1 + I_2 + I_3.\end{aligned}$$

We will show that $\langle \boldsymbol{y}, I_2\rangle$ and $\langle \boldsymbol{y}, I_3\rangle$ are $o_p((nh)^{-1/2})$.

Write $I_2 = W_{0n}I_{21} + W_{1n}I_{22}$ with $I_{21} = n^{-1}\sum_{i=1}^{n} K_h(U_i - u)\boldsymbol{Z}_i\widetilde{\epsilon}_i$ and $I_{22} = n^{-1}\sum_{i=1}^{n} K_h(U_i - u)(U_i - u)\boldsymbol{Z}_i\widetilde{\epsilon}_i$. Following the same line of reasoning as in Lemma 3, we have

$$\mathbb{E}[I_{21}] = \mathbb{E}[K_h(U-u)\boldsymbol{Z}\widetilde{\epsilon}] = \mathbb{E}[K_h(U-u)\boldsymbol{Z}\epsilon] + \mathbb{E}[K_h(U-u)\boldsymbol{Z}]O(h^2) = O(h^2)$$

since $\mathbb{E}[K_h(U-u)\boldsymbol{Z}\epsilon] = \mathbb{E}_{\boldsymbol{X},U}\mathbb{E}[K_h(U-u)\boldsymbol{Z}\epsilon \mid \boldsymbol{X}, U] = 0$. In addition, $\mathbb{E}[I_{22}] = O(h^3)$. The variances satisfy $\mathrm{Var}(I_{21}) = O((nh)^{-1})$ and $\mathrm{Var}(I_{22}) =$

$O(n^{-1}h)$. Hence $I_{21} = O_p(h^2 + (nh)^{-1/2})$ and $I_{22} = O_p(h^3 + n^{-1/2}h^{1/2})$. With $W_{0n} = O_p((nh)^{-1/2})$ and $W_{1n} = O_p((nh^3)^{-1/2})$, we obtain $\|I_2\|_{\mathcal{Y}^{p+1}} = o_p((nh)^{-1/2})$, provided $nh \to \infty$ and $h \to 0$.

For a fixed direction $\boldsymbol{y} = (y_0, \boldsymbol{y}_1^{\mathrm{T}})^{\mathrm{T}} \in \mathcal{Y}^{p+1}$, define the real-valued function $\phi_{\boldsymbol{y}}(u) := \langle \boldsymbol{y}, \boldsymbol{\beta}(u)\rangle_{\mathcal{Y}^{p+1}}$. To analyze its smoothness, we express its components through conditional expectations. Note that

$$\langle \gamma_0(u), y_0\rangle_{\mathcal{Y}} = \mathbb{E}\left[\langle Y, y_0\rangle_{\mathcal{Y}} \mid U = u\right] = \int_{\mathcal{Y}} \langle z, y_0\rangle_{\mathcal{Y}} \frac{f_{U|Y=z}(u)}{f_U(u)}\, d\mathbb{P}_Y(z),$$

$$\begin{aligned}\langle \boldsymbol{\gamma}_1(u), \boldsymbol{y}_1\rangle_{\mathcal{Y}^p} &= \mathbb{E}\left[\langle (\boldsymbol{X} - \boldsymbol{\mu}(u))Y, \boldsymbol{y}_1\rangle_{\mathcal{Y}^p} \mid U = u\right] \\ &= \int_{\mathcal{X}\times\mathcal{Y}} \langle (\boldsymbol{x} - \boldsymbol{\mu}(u))y, \boldsymbol{y}_1\rangle_{\mathcal{Y}^p} \frac{f_{U|\boldsymbol{X}=\boldsymbol{x},Y=y}(u)}{f_U(u)} d\mathbb{P}_{\boldsymbol{X},Y}(\boldsymbol{x}, y).\end{aligned}$$

By Assumption (A4), the conditional densities are twice continuously differentiable in u, and their second derivatives are uniformly bounded. Under the additional condition that $f_U(u) > 0$ and that X and Y are bounded, we can apply the dominated convergence theorem to differentiate under the integral sign in the above expressions. Consequently, the real-valued function $\phi_{\boldsymbol{y}}(u)$ is twice continuously differentiable. Let $\phi'_{\boldsymbol{y}}(u)$ and $\phi''_{\boldsymbol{y}}(u)$ denote its first and second derivatives. For each fixed u, the maps

$$\boldsymbol{y} \mapsto \phi'_{\boldsymbol{y}}(u) \quad \text{and} \quad \boldsymbol{y} \mapsto \phi''_{\boldsymbol{y}}(u)$$

are continuous linear functionals on $\mathcal{Y}^{p+1}$. By the Riesz representation theorem, there exist unique elements, which we denote by $\boldsymbol{\beta}'(u) \in \mathcal{Y}^{p+1}$

and $\boldsymbol{\beta}''(u) \in \mathcal{Y}^{p+1}$, such that

$$\phi_{\boldsymbol{y}}'(u) = \langle \boldsymbol{y}, \boldsymbol{\beta}'(u) \rangle_{\mathcal{Y}^{p+1}}, \qquad \phi_{\boldsymbol{y}}''(u) = \langle \boldsymbol{y}, \boldsymbol{\beta}''(u) \rangle_{\mathcal{Y}^{p+1}},$$

for all $\boldsymbol{y} \in \mathcal{Y}^{p+1}$. Consequently, for each i we have the Taylor expansion

$$\boldsymbol{\beta}\left(U_i\right) - \boldsymbol{\beta}(u) = \boldsymbol{\beta}'(u)\left(U_i - u\right) + \frac{1}{2}\boldsymbol{\beta}''\left(\tilde{u}_i\right)\left(U_i - u\right)^2,$$

where $\tilde{u}_i$ lies between u and U_i.

Now consider $\langle \boldsymbol{y}, I_3 \rangle$. Substituting the above expansion gives

$$\begin{aligned} \langle \boldsymbol{y}, I_3 \rangle = &\frac{1}{n}\sum_{i=1}^{n} s_{in}(u) \langle \boldsymbol{y}, \boldsymbol{Z}_i \boldsymbol{Z}_i^{\mathrm{T}} \boldsymbol{\beta}'(u)(U_i - u) \rangle \\ &+ \frac{1}{2n}\sum_{i=1}^{n} s_{in}(u) \langle \boldsymbol{y}, \boldsymbol{Z}_i \boldsymbol{Z}_i^{\mathrm{T}} \boldsymbol{\beta}''(\tilde{u}_i)(U_i - u)^2 \rangle + O(h^2). \end{aligned}$$

To derive the order of $n^{-1}\sum_{i=1}^{n} s_{in}(u)\boldsymbol{Z}_i\boldsymbol{Z}_i^{\mathrm{T}}(U_i - u)$, we consider the following decomposition

$$\begin{aligned} n^{-1}\sum_{i=1}^{n} s_{in}(u)\boldsymbol{Z}_i\boldsymbol{Z}_i^{\mathrm{T}}(U_i - u) &= n^{-1}\sum_{i=1}^{n} \zeta_{hi}(u)\boldsymbol{Z}_i\boldsymbol{Z}_i^{\mathrm{T}}(U_i - u) + n^{-1}\sum_{i=1}^{n} r_i(u)\boldsymbol{Z}_i\boldsymbol{Z}_i^{\mathrm{T}}(U_i - u) \\ &:= I_{31} + I_{32}. \end{aligned}$$

By Chebyshev's inequality, we have $I_{31} = O_p(\mathbb{E}[\zeta_h(u)\boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}}(U-u)] + (\mathbb{E}(\zeta_h^2(u)(U-u)^2)/n)^{1/2})$. Applying the law of iterated expectations yields

$$\mathbb{E}[\zeta_h(u) Z_i Z_j (U - u)] = \mathbb{E}[Z_i Z_j \mathbb{E}[\zeta_h(u)(U - u) \mid Z_i, Z_j].$$

Recall we have established in the proof of Lemma 3 that $\widetilde{\sigma}_0^2 = h^2 f_U^2(K_{10}K_{12} - K_{11}^2) + O(h^3)$. This, combined with Lemma 1 implies $\mathbb{E}[\zeta_h(u) Z_i Z_j (U - u)] =$

$O(h^2)$. To upper bound the I_{32}, we further expand it as

$$I_{32} = W_{0n} n^{-1} \sum_{i=1}^{n} \boldsymbol{Z}_i \boldsymbol{Z}_i^{\mathrm{T}} K_h(U_i - u)(U_i - u) + W_{1n} n^{-1} \sum_{i=1}^{n} \boldsymbol{Z}_i \boldsymbol{Z}_i^{\mathrm{T}} K_h(U_i - u)(U_i - u)^2$$

$$:= W_{0n} I_{321} + W_{1n} I_{322}.$$

Applying a similar argument to that of Lemma 1, it can be shown that $\mathbb{E}(K_h^2(U_i - u)(U_i - u)^2 = O(h)$ and $\mathbb{E}(K_h^2(U_i - u)(U_i - u)^4) = O(h^3)$. Thus, we have $\mathbb{E}(I_{321}) \leq O((\mathbb{E}(K_h^2(U_i - u)(U_i - u)^2)^{1/2}) = O(h^{1/2})$ and $\operatorname{Var}(I_{321}) \leq \mathbb{E}(K_h^2(U_i - u)(U_i - u)^2)/n = O(h/n)$. Also, $\mathbb{E}(I_{322}) \leq O((\mathbb{E}(K_h^2(U_i - u)(U_i - u)^4)^{1/2}) = O(h^{3/2})$ and $\operatorname{Var}(I_{323}) \leq \mathbb{E}(K_h^2(U_i - u)(U_i - u)^4)/n = O(h^3/n)$. Therefore, $I_{32} = O_p((nh)^{-1/2})h^{1/2} + (nh^3)^{-1/2})h^{3/2}) = o_p((nh)^{-1/2})$. Similarly, we have $n^{-1} \sum_{i=1}^{n} s_{in}(u) \langle \boldsymbol{y}, \boldsymbol{Z}_i \boldsymbol{Z}_i^{\mathrm{T}} \boldsymbol{\beta}''(\tilde{u}_i)(U_i - u)^2 \rangle = O_p(h^2) + o_p((nh)^{-1/2})$, yielding that

$$\langle \boldsymbol{y}, I_3 \rangle = O_p(h^2) + o_p((nh)^{-1/2}).$$

Under the assumption $nh^5 \to 0$, we have $\|I_3\|_{\mathcal{Y}^{p+1}} = o_p((nh)^{-1/2})$ and $\langle \boldsymbol{y}, I_3 \rangle = o_p((nh)^{-1/2})$. In addition, the matrix $\widehat{\Sigma}_{\boldsymbol{Z}}(u)$ converges to $\widetilde{\mathbb{E}}_u(\boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}})$. Indeed, it can be shown that $\|\widehat{\Sigma}_{\boldsymbol{Z}}(u) - \widetilde{\mathbb{E}}_u(\boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}})\| = O_p((nh)^{-1/2})$. Combining these ingredients and Slutsky's theorem, we conclude that

$$\sqrt{nh} \langle \boldsymbol{y}, \widehat{\boldsymbol{\beta}}(u) - \widetilde{\boldsymbol{\beta}}(u) \rangle_{\mathcal{Y}^{p+1}} \xrightarrow{w} \sqrt{h/n}\, \widetilde{\mathbb{E}}_u^{-1}(\boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}}) \sum_{i=1}^{n} \langle \boldsymbol{y}, \zeta_{hi}(u) \boldsymbol{Z}_i \widetilde{\epsilon}_i \rangle.$$

Now replace $\widetilde{\epsilon}_i = \epsilon_i + O(h^2)$. The bias term is

$$\sqrt{\frac{h}{n}} \sum_{i=1}^{n} \zeta_h(U_i, u) \boldsymbol{Z}_i [\boldsymbol{\beta}(U_i) - \widetilde{\boldsymbol{\beta}}(U_i)]^{\mathrm{T}} \boldsymbol{Z}_i = O_p(\sqrt{n} h^{5/2}) = o_p(1),$$

under $nh^5 \to 0$. Write $\widetilde{\Sigma}_{\boldsymbol{Z}}(u) := \widetilde{\mathbb{E}}_u(\boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}})$. Then,

$$\sqrt{nh}(\widehat{\boldsymbol{\beta}}(u) - \widetilde{\boldsymbol{\beta}}(u)) = \widetilde{\Sigma}_{\boldsymbol{Z}}(u)^{-1} \cdot \sqrt{\frac{h}{n}} \sum_{i=1}^{n} \zeta_h(U_i, u) \boldsymbol{Z}_i \epsilon_i + o_p(1).$$

Fix an arbitrary direction $\boldsymbol{y} \in \mathcal{Y}^{p+1}$. Define $\boldsymbol{c} = \widetilde{\Sigma}_{\boldsymbol{Z}}(u)^{-1}\boldsymbol{y} \in \mathcal{Y}^{p+1}$ and

$$T_n := \sqrt{\frac{h}{n}} \sum_{i=1}^{n} \zeta_h(U_i, u) \langle \boldsymbol{c}, \boldsymbol{Z}_i \epsilon_i \rangle.$$

We show T_n converges to a normal distribution. Write $T_n = \sum_{i=1}^n \xi_{ni}$ with $\xi_{ni} = \sqrt{h/n} \cdot \zeta_h(U_i, u)\langle \boldsymbol{c}, \boldsymbol{Z}_i\epsilon_i \rangle$. Since $\mathbb{E}[\epsilon_i \mid U_i, \boldsymbol{X}_i] = 0$, we have $\mathbb{E}[\xi_{ni}] = 0$. The conditional variance is

$$\mathrm{Var}(\xi_{ni} \mid U_i, \boldsymbol{X}_i) = \frac{h}{n} \zeta_h^2(U_i, u) \sum_{j,k} Z_{ij} Z_{ik} \langle c_j, \Sigma_\epsilon(U_i, \boldsymbol{X}_i) c_k \rangle,$$

and thus $\sigma_n^2 = h\mathbb{E}\Big[\zeta_h^2(U, u) \sum_{j,k} Z_j Z_k \langle c_j, \Sigma_\epsilon(U, \boldsymbol{X}) c_k \rangle\Big]$. Here, the conditional covariance operator $\Sigma_\epsilon(U, \boldsymbol{X}) : \mathcal{Y} \to \mathcal{Y}$ is defined as the linear operator such that for all $a, b \in \mathcal{Y}$, $\mathbb{E}\left[\langle a, \epsilon \rangle_{\mathcal{Y}} \langle b, \epsilon \rangle_{\mathcal{Y}} \mid U, \boldsymbol{X}\right] = \langle a, \Sigma_\epsilon(U, \boldsymbol{X}) b \rangle_{\mathcal{Y}}$. Using the linear property of the inner product, algebraic manipulation shows that the double sum equals

$$\sum_{j,k} Z_j Z_k \langle c_j, \Sigma_\epsilon c_k \rangle = \langle \boldsymbol{y}, \big(\widetilde{\Sigma}_{\boldsymbol{Z}}(u)^{-1}(\boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}})\widetilde{\Sigma}_{\boldsymbol{Z}}(u)^{-1}\big) \otimes \Sigma_\epsilon(U, \boldsymbol{X})\, \boldsymbol{y} \rangle.$$

Here, the (i, j)th entry of the operator $A \otimes \Sigma_\epsilon(U, \boldsymbol{X})$ is $A_{ij}\Sigma_\epsilon(U, \boldsymbol{X})$. Thus,

$$\sigma_n^2 = h\,\mathbb{E}\Big[\zeta_h^2(U, u) \langle \boldsymbol{y}, \big(\widetilde{\Sigma}_{\boldsymbol{Z}}(u)^{-1}(\boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}})\widetilde{\Sigma}_{\boldsymbol{Z}}(u)^{-1}\big) \otimes \Sigma_\epsilon(U, \boldsymbol{X})\, \boldsymbol{y} \rangle\Big].$$

Next, we compute the limit of σ_n^2. Define $g(U, \boldsymbol{X}) := \langle \boldsymbol{y}, \widetilde{\Sigma}_{\boldsymbol{Z}}(u)^{-1} \left(\boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}}\right) \widetilde{\Sigma}_{\boldsymbol{Z}}(u)^{-1} \otimes$

$\Sigma_\epsilon(U, \boldsymbol{X})\, \boldsymbol{y}\rangle$. Then, we have

$$\sigma_n^2 = h\, \mathbb{E}\left[\zeta_h^2(U, u)\, g(U, \boldsymbol{X})\right] = h \int\int \zeta_h^2(v, u)\, g(v, \boldsymbol{x})\, f_{U,\boldsymbol{X}}(v, \boldsymbol{x})\, dv\, d\boldsymbol{x}.$$

Let $t = (v-u)/h$. Then, we have $v = u + ht$ and $dv = hdt$. It follows from Lemma 1 that $\widetilde{\mu}_0(u) = f_U(u)K_{10} + \frac{h^2}{2} f_U''(u)K_{12} + O(h^3), \widetilde{\mu}_1(u) = h^2 f_U'(u)K_{12} + O(h^3)$, and $\widetilde{\mu}_2(u) = h^2 f_U(u)K_{12} + O(h^4)$. Therefore, $\widetilde{\sigma}_0^2(u) = h^2 f_U^2(u)K_{12} + \frac{h^4}{2} f_U(u) f_U''(u) K_{12}^2 - h^4[f_U'(u)]^2 K_{12}^2 + O(h^4)$. Define $B(u) = C(u)/f_U^2(u)$ and $C(u) = \frac{1}{2} f_U(u) f_U''(u) K_{12} - [f_U'(u)]^2 K_{12}$. We have

$$\widetilde{\sigma}_0^2(u) = h^2 f_U^2(u) K_{12}\left[1 + h^2 B(u) + O(h^2)\right].$$

In addition,

$$\widetilde{\mu}_2(u) - \widetilde{\mu}_1(u)(v-u) = h^2 f_U(u)K_{12} - h^3 t f_U'(u) K_{12} + O(h^3) = h^2 f_U(u) K_{12}\left[1 - htA(u) + O(h^2)\right],$$

where $A(u) = f_U'(u)/f_U(u)$. Therefore, the weight function expands as

$$\begin{aligned}\zeta_h(v, u) &= \frac{K(t)}{h} \cdot \frac{h^2 f_U(u) K_{12}\left[1 - htA(u) + O(h^2)\right]}{h^2 f_U^2(u) K_{12}\left[1 + h^2 B(u) + O(h^2)\right]} \\ &= \frac{1}{h} \cdot \frac{K(t)}{f_U(u)} \cdot \frac{1 - htA(u) + O(h^2)}{1 + h^2 B(u) + O(h^2)}.\end{aligned}$$

A Taylor expansion of $(1 + h^2 B(u) + O(h^2))^{-1}$ shows that

$$\frac{1 - htA(u) + O(h^2)}{1 + h^2 B(u) + O(h^2)} = 1 - htA(u) - h^2 B(u) + O(h^2).$$

Consequently, we have

$$\zeta_h(v, u) = \frac{1}{h} \cdot \frac{K(t)}{f_U(u)}\left[1 - htA(u) - h^2 B(u) + O(h^2)\right].$$

and

$$\zeta_h^2(v,u) = \frac{1}{h^2}\cdot\frac{K^2(t)}{f_U^2(u)}\left[1-2htA(u)+h^2(t^2A^2(u)-2B(u))+O(h^2)\right].$$

Now substitute into the integral for σ_n^2, we have

$$\sigma_n^2 = \int\int\frac{K^2(t)}{f_U^2(u)}\left[1-2htA(u)+h^2(t^2A^2(u)-2B(u))+O(h^2)\right]H(u+ht,\boldsymbol{x})\,dtd\boldsymbol{x},$$

where $H(v,\boldsymbol{x}) := g(v,\boldsymbol{x})f_{U,\boldsymbol{X}}(v,\boldsymbol{x})$. Perform a Taylor expansion of $H(u+ht,\boldsymbol{x})$ around u gives

$$H(u+ht,\boldsymbol{x}) = H(u,\boldsymbol{x}) + ht\,\partial_u H(u,\boldsymbol{x}) + \frac{h^2t^2}{2}\partial_{uu}H(u,\boldsymbol{x}) + O(h^3t^3).$$

Multiplying the two expansions inside the integral:

$$\begin{aligned}&\frac{K^2(t)}{f_U^2(u)}\Big[1-2htA(u)+h^2(t^2A^2(u)-2B(u))\Big]\cdot\Big[H(u,\boldsymbol{x})+ht\,\partial_uH(u,\boldsymbol{x})+\frac{h^2t^2}{2}\partial_{uu}H(u,\boldsymbol{x})\Big]\\&+O(h^2) = \frac{K^2(t)}{f_U^2(u)}\Big\{H(u,\boldsymbol{x})+ht\partial_uH(u,\boldsymbol{x})-2htA(u)H(u,\boldsymbol{x})+\frac{h^2t^2}{2}\partial_{uu}H(u,\boldsymbol{x})\\&-2h^2t^2A(u)\partial_uH(u,\boldsymbol{x})+h^2t^2A^2(u)H(u,\boldsymbol{x})-2h^2B(u)H(u,\boldsymbol{x})\Big\}+O(h^2).\end{aligned}$$

Thus,

$$\begin{aligned}\sigma_n^2 = \frac{1}{f_U^2(u)}\int\Big[&K_{20}H(u,\boldsymbol{x})+h^2K_{22}\Big(\frac{1}{2}\partial_{uu}H(u,\boldsymbol{x})-2A(u)\partial_uH(u,\boldsymbol{x})+A^2(u)H(u,\boldsymbol{x})\Big)\\&-2h^2B(u)K_{20}H(u,\boldsymbol{x})\Big]d\boldsymbol{x}+O(h^2).\end{aligned}$$

Note that $\int H(u,\boldsymbol{x})d\boldsymbol{x} = f_U(u)\,\mathbb{E}[g(U,\boldsymbol{X})\mid U=u]$. Therefore, the leading term is

$$\sigma_n^2 = \frac{K_{20}}{f_U(u)}\mathbb{E}[g(U,\boldsymbol{X})\mid U=u]+O(h^2).$$

Recalling the definition of $g(U, \boldsymbol{X})$, we obtain the asymptotic variance:

$$\sigma^2 := \lim_{n\to\infty} \sigma_n^2 = \frac{K_{20}}{f_U(u)} \lim_{n\to\infty} \left\langle \boldsymbol{y}, \widetilde{\Sigma}_{\boldsymbol{Z}}(u)^{-1} \mathbb{E}\left[(\boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}}) \otimes \Sigma_\epsilon(U, \boldsymbol{X}) \mid U = u\right] \widetilde{\Sigma}_{\boldsymbol{Z}}(u)^{-1} \boldsymbol{y} \right\rangle_{\mathcal{Y}^{p+1}}$$
$$= \frac{K_{20}}{f_U(u)} \left\langle \boldsymbol{y}, \left(\mathbb{E}\left[\boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}} \mid U = u\right]\right)^{-1} \mathbb{E}\left[(\boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}}) \otimes \Sigma_\epsilon(U, \boldsymbol{X}) \mid U = u\right] \left(\mathbb{E}\left[\boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}} \mid U = u\right]\right)^{-1} \boldsymbol{y} \right\rangle_{\mathcal{Y}^{p+1}}.$$

To verify the Lyapunov condition, note that $|\xi_{ni}| \le C\sqrt{h/n}$ for some constant C. For any $\delta > 0$,

$$\sum_{i=1}^{n} \mathbb{E}|\xi_{ni}|^{2+\delta} \le n \cdot C^{2+\delta}(h/n)^{1+\delta/2} = C^{2+\delta} n^{-\delta/2} h^{1+\delta/2}.$$

Since $\sigma_n^{2+\delta} = O(1)$, we have $\frac{1}{\sigma_n^{2+\delta}} \sum_{i=1}^{n} \mathbb{E}|\xi_{ni}|^{2+\delta} = O(n^{-\delta/2} h^{1+\delta/2}) = o(1)$. Therefore, by the Lindeberg-Feller central limit theorem,

$$\frac{T_n}{\sigma_n} \xrightarrow{d} N(0, 1),$$

and thus $T_n \xrightarrow{d} N(0, \sigma^2)$. It follows from $\|\widetilde{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}(u)\| = O\left(h^2\right)$ and $nh^5 \to 0$ that $\sqrt{nh}(\widehat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}(u))$ converges weakly to a zero-mean Gaussian random object $\mathcal{G} \in \mathcal{Y}^{p+1}$ whose covariance operator is

$$\Sigma_{\mathcal{G}}(u) = \frac{K_{20}}{f_U(u)} \Big(\mathbb{E}[\boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}} \mid U = u]\Big)^{-1} \mathbb{E}\Big[(\boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}}) \otimes \Sigma_\epsilon(U, \boldsymbol{X}) \mid U = u\Big] \Big(\mathbb{E}[\boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}} \mid U = u]\Big)^{-1}.$$

Next, we establish the asymptotic finite-dimensionality of $\sqrt{nh}(\widehat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}(u))$. Recall the linear representation

$$\sqrt{nh}\big(\widehat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}(u)\big) = \widetilde{\Sigma}_{\boldsymbol{Z}}^{-1}(u)\sqrt{h/n} \sum_{i=1}^{n} \zeta_h(U_i, u) \boldsymbol{Z}_i \epsilon_i + o_p(1) := M_n + R_n.$$

Note that $\operatorname{tr}\left(\Sigma_\epsilon(U, \boldsymbol{X})\right) = \mathbb{E}\left[\|\epsilon\|_{\mathcal{Y}}^2 \mid U, \boldsymbol{X}\right]$, where $\operatorname{tr}(\cdot)$ denotes the trace of an operator. Since $\boldsymbol{Z}$ and ϵ are bounded, the covariance operator $\Sigma_{\mathcal{G}}(u)$

is a trace-class, hence compact, self-adjoint operator on $\mathcal{Y}^{p+1}$. Let $\{e_j\}_{j=1}^{\infty}$ be a complete orthonormal basis of $\mathcal{Y}^{p+1}$ consisting of eigenvectors of the limit covariance operator $\Sigma_{\mathcal{G}}(u)$, whose existence is guaranteed by Hilbert-Schmidt theorem. Write $\Sigma_{\mathcal{G}}(u)e_j = \lambda_j e_j$ with $\lambda_j \geq 0$. The trace-class property gives

$$\sum_{j=1}^{\infty} \lambda_j = \mathrm{tr}\big(\Sigma_{\mathcal{G}}(u)\big) < \infty.$$

For a fixed finite index set $I \subset \mathbb{N}$ denote by P_I the orthogonal projection onto $\mathrm{span}\{e_j : j \in I\}$. Because $R_n = o_p(1)$, for any $\eta > 0$ we have $\mathbb{P}(\|R_n\| > \eta) \to 0$. Hence, given $\delta, \varepsilon > 0$ we can choose $N_1 \in \mathbb{N}$ such that for all $n \geq N_1$

$$\mathbb{P}\Big(\|P_I R_n\|^2 > \frac{\delta}{2}\Big) \leq \mathbb{P}\Big(\|R_n\|^2 > \frac{\delta}{2}\Big) < \frac{\varepsilon}{2}.$$

Using a Taylor expansion and $K_{44} < \infty$ assumed in Condition (A3), one can show that $\mathbb{E}\left(\zeta_{hi}^4\right) = O\left(h^{-3}\right)$, and thus $\sup_n \mathbb{E}[\|M_n\|^4] < \infty$. It then follows that $\{\|M_n\|^2\}$ is uniformly integrable. Then, under the boundness of $\boldsymbol{Z}$ and ϵ, uniform integrability of $\{\|M_n\|^2\}$, we have

$$\lim_{n\to\infty} \mathbb{E}\big[|\langle M_n, e_j\rangle|^2\big] = \mathbb{E}\big[|\langle \mathcal{G}, e_j\rangle|^2\big] = \lambda_j$$

for every j. Given $\delta, \varepsilon > 0$, one can pick a finite set $I = \{1, \ldots, K\}$ such that

$$\sum_{j>K} \lambda_j < \frac{\delta\varepsilon}{8}.$$

Also, for any fixed K and any $\epsilon > 0$, there exists an integer $N_0 > K$ such that $\sum_{j>N_0} \lambda_j < \frac{\epsilon}{3}$. Define $S_n(N) = \sum_{j=K+1}^{N} |\langle M_n, e_j\rangle|^2$ for $N > K$. Using $\mathbb{E}[\|M_n\|^2] = \sum_{j=1}^{\infty} \mathbb{E}[|\langle M_n, e_j\rangle|^2]$, we have

$$\mathbb{E}\Big[\sum_{j>N} |\langle M_n, e_j\rangle|^2\Big] = \mathbb{E}[\|M_n\|^2] - \sum_{j=1}^{N} \mathbb{E}[|\langle M_n, e_j\rangle|^2].$$

Using the fact that uniform integrability of $\{\|M_n\|^2\}$ guaranteed by $\sup_n \mathbb{E}[\|M_n\|^4] < \infty$, together with $M_n \xrightarrow{w} \mathcal{G}$ implies $\lim_{n\to\infty} \mathbb{E}[\|M_n\|^2] = \mathbb{E}[\|\mathcal{G}\|^2] = \sum_{j=1}^{\infty} \lambda_j$, we obtain

$$\limsup_{n\to\infty} \mathbb{E}\Big[\sum_{j>N} |\langle M_n, e_j\rangle|^2\Big] = \sum_{j=1}^{\infty} \lambda_j - \sum_{j=1}^{N} \lambda_j = \sum_{j>N} \lambda_j.$$

Now choose $N_1 \geq N_0$ such that $\sum_{j>N_1} \lambda_j < \epsilon/3$. Then, $\limsup_{n\to\infty} \mathbb{E}\big[\sum_{j>N_1} |\langle M_n, e_j\rangle|^2\big] < \epsilon/3$, and hence there exists an integer n_0 such that for all $n \geq n_0$, $\mathbb{E}[\sum_{j>N_1} |\langle M_n, e_j\rangle|^2] < \frac{\epsilon}{3}$. Consequently, we have $\lim_{n\to\infty} \mathbb{E}[S_n(N_1)] = \sum_{j=K+1}^{N_1} \lambda_j$. Thus, there exists $n_1 \geq n_0$ such that for all $n \geq n_1$, $|\mathbb{E}[S_n(N_1)] - \sum_{j=K+1}^{N_1} \lambda_j| < \frac{\epsilon}{3}$. Now write

$$\mathbb{E}\Big[\sum_{j>K} |\langle M_n, e_j\rangle|^2\Big] = \mathbb{E}[S_n(N_1)] + \mathbb{E}\Big[\sum_{j>N_1} |\langle M_n, e_j\rangle|^2\Big].$$

For $n \geq n_1$, one can obtain

$$\begin{aligned}\Big|\mathbb{E}\Big[\sum_{j>K} |\langle M_n, e_j\rangle|^2\Big] - \sum_{j>K} \lambda_j\Big| &\leq \Big|\mathbb{E}[S_n(N_1)] - \sum_{j=K+1}^{N_1} \lambda_j\Big| + \mathbb{E}\Big[\sum_{j>N_1} |\langle M_n, e_j\rangle|^2\Big] + \Big|\sum_{j>N_1} \lambda_j\Big| \\ &< \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon,\end{aligned}$$

meaning that $\lim_{n\to\infty} \mathbb{E}[\sum_{j>K} |\langle M_n, e_j\rangle|^2] = \sum_{j>K} \lambda_j$. Thus, there exists

N_2 such that for all $n \geq N_2$

$$\mathbb{E}\Big[\sum_{j>K} |\langle M_n, e_j\rangle|^2\Big] < \sum_{j>K} \lambda_j + \frac{\delta\varepsilon}{8} < \frac{\delta\varepsilon}{4}.$$

Applying Markov's inequality we obtain, for $n \geq N_2$,

$$\mathbb{P}\Big(\sum_{j>K} |\langle M_n, e_j\rangle|^2 > \frac{\delta}{2}\Big) \leq \frac{2}{\delta}\,\mathbb{E}\Big[\sum_{j>K} |\langle M_n, e_j\rangle|^2\Big] < \frac{2}{\delta}\cdot\frac{\delta\varepsilon}{4} = \frac{\varepsilon}{2},$$

yielding that

$$\limsup_{n\to\infty} \mathbb{P}\big(\|M_n - P_I M_n\|^2 > \tfrac{\delta}{2}\big) \leq \frac{\varepsilon}{2},$$

which establishes the asymptotic finite-dimensionality of $\{M_n\}$. Using the elementary bound $\|a+b\|^2 \leq 2\|a\|^2 + 2\|b\|^2$,

$$\|X_n - P_I X_n\|^2 \leq 2\|M_n - P_I M_n\|^2 + 2\|R_n - P_I R_n\|^2 \leq 2\|M_n - P_I M_n\|^2 + 2\|R_n\|^2.$$

Consequently,

$$\mathbb{P}\big(\|X_n - P_I X_n\|^2 > \delta\big) \leq \mathbb{P}\Big(\|M_n - P_I M_n\|^2 > \frac{\delta}{4}\Big) + \mathbb{P}\Big(\|R_n\|^2 > \frac{\delta}{4}\Big).$$

Therefore,

$$\limsup_{n\to\infty} \mathbb{P}\big(\|X_n - P_I X_n\|^2 > \delta\big) < \varepsilon.$$

Since $\delta, \varepsilon > 0$ were arbitrary, the sequence $\big\{\sqrt{nh}(\widehat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}(u))\big\}$ is asymptotically finite-dimensional. This completes the proof of Theorem 3. $\square$

Proof of Corollary 2. Define the linear prediction operator $A : \mathcal{Y}^{p+1} \to \ell^\infty_{\mathcal{Y}}(\mathcal{X})$ by $(A\boldsymbol{y})(x) = y_0 + x^{\mathrm{T}}\boldsymbol{y}_1$, where $\boldsymbol{y} = (y_0, \boldsymbol{y}_1^{\mathrm{T}})^{\mathrm{T}} \in \mathcal{Y}^{p+1}$ and $x \in$

$\mathcal{X}$. Clearly, A is a linear operator. We verify that A is bounded, hence continuous. For any $\boldsymbol{y} \in \mathcal{Y}^{p+1}$,

$$\begin{aligned}\|A\boldsymbol{y}\|_{\mathcal{X}} = \sup_{x\in\mathcal{X}} \|y_0 + x^{\mathrm{T}}\boldsymbol{y}_1\|_{\mathcal{Y}} &\leq \|y_0\|_{\mathcal{Y}} + \sup_{x\in\mathcal{X}} \sum_{j=1}^{p} |x_j|\,\|y_{1j}\|_{\mathcal{Y}} \\ &\leq \|y_0\|_{\mathcal{Y}} + C\sum_{j=1}^{p} \|y_{1j}\|_{\mathcal{Y}} \leq (1 + \operatorname{diam}(\mathcal{X})p)\,\|\boldsymbol{y}\|_{\mathcal{Y}^{p+1}}.\end{aligned}$$

Thus A is a bounded linear operator.

Observe that $\mathcal{M}_n = A(\sqrt{nh}(\widehat{s}_{\oplus}(x,u) - s_{\oplus}(x,u)))$. Because $\sqrt{nh}(\widehat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}(u)) \xrightarrow{w} \mathcal{G}$ in $\mathcal{Y}^{p+1}$ and $A : \mathcal{Y}^{p+1} \to \ell^{\infty}_{\mathcal{Y}}(\mathcal{X})$ is continuous, Theorem 1.3.6 of Van Der Vaart & Wellner (1996) yields $\mathcal{M}_n \xrightarrow{w} A(\mathcal{G}) =: \mathcal{M}$ in $\ell^{\infty}_{\mathcal{Y}}(\mathcal{X})$. The limit $\mathcal{M}$ is a zero-mean Gaussian process, since $\mathcal{G}$ is a zero-mean Gaussian random object and A is linear. Its covariance structure follows from the definition of A and the covariance operator of $\mathcal{G}$: for any $x, y \in \mathcal{X}$ and $a, b \in \mathcal{Y}$,

$$\begin{aligned}&\operatorname{Cov}\big(\langle a, \mathcal{M}(x)\rangle_{\mathcal{Y}}, \langle b, \mathcal{M}(y)\rangle_{\mathcal{Y}}\big) = \operatorname{Cov}\big(\langle a, A(\mathcal{G})(x)\rangle_{\mathcal{Y}}, \langle b, A(\mathcal{G})(y)\rangle_{\mathcal{Y}}\big) \\ &= \operatorname{Cov}\big(\langle (a, x^{\mathrm{T}}a), \mathcal{G}\rangle_{\mathcal{Y}^{p+1}}, \langle (b, y^{\mathrm{T}}b), \mathcal{G}\rangle_{\mathcal{Y}^{p+1}}\big) = \big\langle (a, x^{\mathrm{T}}a),\ \Sigma_{\mathcal{G}}(u)\ (b, y^{\mathrm{T}}b)\big\rangle_{\mathcal{Y}^{p+1}}.\end{aligned}$$

Because $\mathcal{X}$ is bounded,

$$\begin{aligned}\big\|\widehat{s}_{\oplus}(\boldsymbol{x},u) - s_{\oplus}(\boldsymbol{x},u)\big\|_{\mathcal{Y}} &= \big\|\widehat{\beta}_0(u) - \beta_0(u) + x^{\mathrm{T}}\big(\widehat{\boldsymbol{\beta}}_1(u) - \boldsymbol{\beta}_1(u)\big)\big\|_{\mathcal{Y}} \\ &\leq \big\|\widehat{\beta}_0(u) - \beta_0(u)\big\|_{\mathcal{Y}} + \big\|x^{\mathrm{T}}\big(\widehat{\boldsymbol{\beta}}_1(u) - \boldsymbol{\beta}_1(u)\big)\big\|_{\mathcal{Y}} = O_p((nh)^{-1/2}).\end{aligned}$$

□

S4 Proofs for Non-Euclidean predictor U

Allowing U to come from a general metric space takes a new level of complexity, so we need to introduce some definitions about almost complete convergence.

A sequence of real random variables $\{T_n\}$ is said to converge almost completely to some real random variable T, if and only if, for any $\epsilon > 0$, $\sum_{n=1}^{\infty} \mathbb{P}\left(|T_n - T| > \epsilon\right) < \infty$, and the almost complete convergence of $\{T_n\}$ to T is denoted by $\lim_{n\to\infty} T_n = T, \text{a.co.}$. One says that the rate of almost complete convergence of $\{T_n\}$ to T is of order t_n if and only if, there exists $\epsilon_0 > 0$, $\sum_{n=1}^{\infty} \mathbb{P}\left(|T_n - T| > \epsilon_0 t_n\right) < \infty$, and we write $T_n - T = O_{\text{a.co.}}\left(t_n\right)$.

Proof of Theorem 4. Our first goal is to show that $\widetilde{S}_n\left(y; \boldsymbol{x}, u\right) = S_{\oplus}\left(y; \boldsymbol{x}, u\right) + O(h)$. That is, we need to show that $\widetilde{w}_0 + \widetilde{w}_1 \widetilde{w}_2^{-1} \widetilde{w}_3 = w_0 + w_1 w_2^{-1} w_3 + O(h)$.

We begin by proving that $\widetilde{w}_0(y; u) = w_0(y; u) + O(h)$. By the law of iterated expectations, it follows that

$$\mathbb{E}[\zeta_h(U, u) d^2(Y, y)] = \mathbb{E}\{\mathbb{E}[d^2(Y, y)\zeta_h(U, u)|Y]\} = \mathbb{E}\{d^2(Y, y)\mathbb{E}[\zeta_h(U, u)|Y]\}.$$

We will show that

$$\mathbb{E}[\zeta_h(U, u)|Y = y] = \int \zeta_h(u', u) d\mathbb{P}_{U|Y=y}(u') = \frac{f_{U|Y=y}(u)}{f_U(u)}\left(1 + O(h)\right), \tag{S4.1}$$

where the error term $O(h)$ is uniform over $y \in \mathcal{Y}$. Define $\eta_h(u) = \int K_h(\delta(u', u)) f_{U|Y=y}(u') d\nu(u')$

and $\mu_h(u) = \int K_h(\delta(u',u)) f_U(u') d\nu(u')$. Then,

$$\int \zeta_h(u',u) d\mathbb{P}_{U|Y=y}(u') = \frac{\eta_h(u)}{\mu_h(u)}.$$

By the Lipschitz continuity (A10), we have

$$f_{U|Y=y}(u') = f_{U|Y=y}(u) + O(\delta(u',u)), \quad f_U(u') = f_U(u) + O(\delta(u',u)).$$

Substituting into $\eta_h(u)$, we can obtain

$$\eta_h(u) = f_{U|Y=y}(u) \int K_h(\delta(u',u)) d\nu(u') + \int K_h(\delta(u',u)) O(\delta(u',u)) d\nu(u').$$

Similarly, for $\mu_h(u)$, we have

$$\mu_h(u) = f_U(u) \int K_h(\delta(u',u)) d\nu(u') + \int K_h(\delta(u',u)) O(\delta(u',u)) d\nu(u').$$

Let $A(u) = \int K_h(\delta(u',u)) d\nu(u')$ and $B(u) = \int K_h(\delta(u',u)) \delta(u',u) d\nu(u')$. From (A9), we have $B(u) = O(hA(u))$. Thus,

$$\eta_h(u) = f_{U|Y=y}(u) A(u) + O(B(u)) = f_{U|Y=y}(u) A(u) + O(hA(u)),$$

and

$$\mu_h(u) = f_U(u) A(u) + O(B(u)) = f_U(u) A(u) + O(hA(u)).$$

Therefore,

$$\frac{\eta_h(u)}{\mu_h(u)} = \frac{f_{U|Y=y}(u) A(u) + O(hA(u))}{f_U(u) A(u) + O(hA(u))} = \frac{f_{U|Y=y}(u) + O(h)}{f_U(u) + O(h)} = \frac{f_{U|Y=y}(u)}{f_U(u)} \left(1 + O(h)\right).$$

The error term is uniform due to the uniform Lipschitz constant and the uniform integral conditions in (A2). It then follows from (S4.1) that

$$\begin{aligned}\mathbb{E}[\zeta_h(U,u)d^2(Y,y)] &= \int d^2(y',y)(1+O(h))\frac{f_{U|Y=y}(u)}{f_U(u)}d\mathbb{P}_{Y|y}(y') \\ &= \int d^2(y',y)(1+O(h))\frac{d\mathbb{P}_{Y|U=u}}{d\mathbb{P}_Y}(y')d\mathbb{P}_Y(y') \\ &= \mathbb{E}[d^2(Y,y)(1+O(h)) \mid U=u] = \mathbb{E}[d^2(Y,y) \mid U=u] + O(h).\end{aligned}$$

Thus, we have obtained that $\widetilde{w}_0\left(y;u\right) = w_0\left(y;u\right) + O(h)$.

To get difference between $\widetilde{w}_1$ and w_1, we rewrite $\widetilde{w}_1\left(\boldsymbol{x},u\right) = \left(\widetilde{w}_{1,1}\left(x_1,u\right),\ldots,\widetilde{w}_{1,p}(x_p,u)\right)^{\mathrm{T}}$ and $w_1(\boldsymbol{x},u) = (w_{1,1}(x_1,u),\ldots,w_{1,p}(x_p,u))^{\mathrm{T}}$. Then, we also can get $\mathbb{E}(\zeta_h(U,u)|X_i = x_i) = \frac{f_{U|X_i=x_i}(u)}{f_U(u)}\left(1+O(h)\right)$, yielding that $\widetilde{w}_{1,i}\left(x_i,u\right) = w_{1,i}\left(x_j,u\right) + O(h)$ for each $i=1,\ldots,p$.

To get the order of $\widetilde{w}_2(u) - w_2(u)$, we have

$$\begin{aligned}\widetilde{w}_2(u) &= \mathbb{E}\{\zeta_h(U,u)[\boldsymbol{X} - \mathbb{E}(\zeta_h(U,u)\boldsymbol{X})][\boldsymbol{X} - \mathbb{E}(\zeta_h(U,u)\boldsymbol{X})]^{\mathrm{T}}\} \\ &= \mathbb{E}(\zeta_h(U,u)\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}) - \mathbb{E}(\zeta_h(U,u)\boldsymbol{X})\mathbb{E}(\zeta_h(U,u)\boldsymbol{X})^{\mathrm{T}}.\end{aligned}$$

A similar argument also yields $\mathbb{E}(\zeta_h(U,u)\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}) = \mathbb{E}(\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} \mid U=u) + \boldsymbol{O}(h)_{p\times p}$. In addition, $\mathbb{E}(\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} \mid U=u) = \mathrm{Cov}(\boldsymbol{X} \mid U=u) + \mathbb{E}(\boldsymbol{X} \mid U=u)\mathbb{E}(\boldsymbol{X} \mid U=u)^{\mathrm{T}}$. This, together with $\mathbb{E}\left(\zeta_h(U,u)\boldsymbol{X}\right) = \mathbb{E}\left(\boldsymbol{X} \mid U=u\right) + \boldsymbol{O}\left(h\right)$ ensures that

$$\widetilde{w}_2(u) = \mathrm{Cov}(\boldsymbol{X} \mid U=u) + \boldsymbol{O}(h)_{p\times p} = w_2(u) + \boldsymbol{O}(h)_{p\times p}.$$

Similarly, we have $\widetilde{w}_3(y;u) = w_3(y;u) + \boldsymbol{O}(h)$. Therefore, $\widetilde{S}_n(y;\boldsymbol{x},u) - S_{\oplus}(y;\boldsymbol{x},u) = O(h)$. With this and the assumption (A1), we get that $d(s_{\oplus}(\boldsymbol{x},u), \widetilde{s}_{\oplus}(\boldsymbol{x},u)) = o(1)$, as $h \to 0$.

Finally, define $r_h = h^{-\frac{\beta_1}{2(\beta_1-1)}}$ and set $D_{j,n} = \left\{y : 2^{j-1} < r_h d(y, s_{\oplus}(u))^{\beta_1/2} \leq 2^j\right\}$. We following the same arguments as before and using assumption (A5), we have that for some $M > 0$, $d(\widetilde{s}_{\oplus}(\boldsymbol{x},u), s_{\oplus}(\boldsymbol{x},u)) \leq 2^{2M/\beta_1} h^{1/(\beta_1-1)}$. Thus, $d(\widetilde{s}_{\oplus}(\boldsymbol{x},u), s_{\oplus}(\boldsymbol{x},u)) = O(h^{1/(\beta_1-1)})$. □

To prove the convergence rate of the stochastic term $d(\widetilde{s}_{\oplus}(\boldsymbol{x},u), \widehat{s}_{\oplus}(\boldsymbol{x},u))$. We consider the case when $\boldsymbol{X}$ is not a random vector, but rather a vector of constants. This reduces the setting to that equation (3.1) becomes $s_{\oplus}(u) = \underset{y\in\mathcal{Y}}{\arg\min}\, \mathbb{E}\left(d^2(Y,y) \mid U = u\right)$, equation (3.2) becomes $\widehat{s}_{\oplus}(u) = \underset{y\in\mathcal{Y}}{\arg\min} \sum_{i=1}^n s_{in}(u)\, d^2(Y_i,y)$, and equation (4.1) becomes $\widetilde{s}_{\oplus}(u) = \underset{y\in\mathcal{Y}}{\arg\min}\, \mathbb{E}\left(\zeta_h(U,u)\, d^2(Y,y)\right)$. We first give some results that will be used in the proof of Theorem 5.

Lemma 4. *We denote that* $\Delta_i = K_h(\delta(U_i,u))/\mathbb{E}\, K_h(\delta(U_i,u))$, $\widehat{r}_1(u) = \frac{1}{n}\sum_{i=1}^n \Delta_i$, *and* $\widehat{r}_2(u) = \frac{1}{n}\sum_{i=1}^n d^2(Y_i,y)\Delta_i$. *Then,*

(i) $\widehat{r}_2(u) - \mathbb{E}\,\widehat{r}_2(u) = O_{a.co.}((\log n/(n\varphi_{\mathcal{U},u}(h)))^{1/2})$,

(ii) $\widehat{r}_1(u) - 1 = O_{a.co.}((\log n/(n\varphi_{\mathcal{U},u}(h)))^{1/2})$.

Proof. We claim there exist positive constants c_1, c_2 (independent of n, h)

such that

$$c_1\,\varphi_{\mathcal{U},u}(h) \;\le\; \mathbb{E}\,K(\delta(U_i,u)/h) \;\le\; c_2\,\varphi_{\mathcal{U},u}(h), \tag{S4.2}$$

and for every integer $m \ge 1$,

$$c_1^m\,\varphi_{\mathcal{U},u}(h) \;\le\; \mathbb{E}\,K^m(\delta(U_i,u)/h) \;\le\; c_2^m\,\varphi_{\mathcal{U},u}(h). \tag{S4.3}$$

Indeed, Assumption (A9) and the fact $\mathbb{1}_{[0,h]}(\delta(U_i,u)) = \mathbb{1}_{[0,1]}(\delta(U_i,u)/h)$ yield $c_1\,\varphi_{\mathcal{U},u}(h) \le \mathbb{E}\,K(\delta(U_i,u)/h) \le c_2\,\varphi_{\mathcal{U},u}(h)$, which is (S4.2). The same comparison applied to $K^m(\cdot)$ gives (S4.3). By (S4.2)–(S4.3), we obtain that

$$c_1c_2^{-m}\varphi_{\mathcal{U},u}(h)^{-(m-1)} \le \mathbb{E}\,\Delta_1^m = \frac{\mathbb{E}\,K(\delta(U_i,u)/h)^m}{(\mathbb{E}\,K(\delta(U_i,u)/h))^m} \le c_2c_1^{-m}\varphi_{\mathcal{U},u}(h)^{-(m-1)}.$$

Further, the law of iterated expectations yields

$$\begin{aligned}\mathbb{E}(|d^2(Y_1,y)|^m\Delta_1^m) &= \frac{\mathbb{E}(K(\delta(U_1,u)/h)^m\mathbb{E}(|d^2(Y_1,y)|^m \mid U_1))}{(\mathbb{E}\,K(\delta(U_1,u)/h))^m}\\ &\le c_2(c_1^{-1}\,\mathrm{diam}(\mathcal{Y}))^m\varphi_{\mathcal{U},u}(h)^{-(m-1)}.\end{aligned}$$

Let $Z_i := d^2(Y_i,y)\Delta_i - \mathbb{E}(d^2(Y_1,y)\Delta_i)$. A binomial expansion together with Jensen's inequality give that

$$\begin{aligned}\mathbb{E}\,|Z_1|^m &= \sum_{k=0}^m \binom{m}{k}\mathbb{E}(|d^2(Y_i,y)\Delta_i|^k)(-\mathbb{E}(|d^2(Y_i,y)\Delta_i|)^{m-k}\\ &\le \sum_{k=0}^m \binom{m}{k}\mathbb{E}(|d^2(Y_i,y)\Delta_i|^k)\mathbb{E}(|d^2(Y_i,y)\Delta_i|^{m-k})\\ &\le c_2(c_1^{-1}\,\mathrm{diam}(\mathcal{Y}))^m\varphi_{\mathcal{U},u}(h)^{-(m-1)}\sum_{k=0}^m\binom{m}{k}.\end{aligned} \tag{S4.4}$$

This, combined with the fact that $\sum_{k=0}^{m}\binom{m}{k}=(1+1)^m$ implies

$$\mathbb{E}\,|Z_1|^m \leq c_2(2c_1^{-1}\operatorname{diam}(\mathcal{Y}))^m\varphi_{\mathcal{U},u}(h)^{-(m-1)}.$$

When $d^2(Y_i,y)\equiv 1$, the same argument yields $\mathbb{E}|\Delta_1-\mathbb{E}\Delta_1|^m \leq c_2^2(2c_1)^m\varphi_{\mathcal{U},u}(h)^{-(m-1)}$. Note that $C^m/(m!)\to 0$ for any fixed C as $m\to\infty$, we conclude that $\mathbb{E}\,|Z_1|^m = O(m!)\varphi_{\mathcal{U},u}(h)^{-(m-1)}$ and $\mathbb{E}|\Delta_1-\mathbb{E}\Delta_1|^m = O(m!)\varphi_{\mathcal{U},u}(h)^{-(m-1)}$. Then, applying a Bernstein-type exponential inequality to $\{Z_i\}_{i=1}^n$, one can show that

$$\sum_{n=1}^{\infty}\mathbb{P}(|\widehat{r}_2(u)-\mathbb{E}\widehat{r}_2(u)|>\varepsilon\sqrt{\log n/n\,\varphi_{\mathcal{U},u}(h)})\;<\;\infty,$$

hence (i) follows by Borel–Cantelli lemma. For (ii), $d^2(Y_i,y)\equiv 1$ (so $Z_i=\Delta_i-\mathbb{E}\Delta_1$) and the rest of the argument is unchanged. This gives $\widehat{r}_1(u)-1 = O_{\text{a.co.}}(\sqrt{\log n/(n\,\varphi_{\mathcal{U},u}(h))})$.

□

We need the following lemma to derive the convergence rate of $d\left(\widetilde{s}_{\oplus}\left(u\right),\widehat{s}_{\oplus}\left(u\right)\right)$.

Lemma 5. *Under Assumptions (A1) and (A7)-(A9), we have*

$$d\left(\widetilde{s}_{\oplus}\left(u\right),\widehat{s}_{\oplus}\left(u\right)\right)=o_{a.s.}\left(1\right).$$

Proof. Recall $\widehat{S}_n\left(y;u\right)=n^{-1}\sum\limits_{i=1}^{n}s_{in}\left(u\right)d^2\left(Y_i,y\right)$, $\Delta_i=\frac{K_h(\delta(U_i,u))}{\mathbb{E}\,K_h(\delta(U_i,u))}$, $\widehat{r}_1(u)=\frac{1}{n}\sum_{i=1}^n\Delta_i$, $\widehat{r}_2(u)=\frac{1}{n}\sum_{i=1}^n d^2(Y_i,y)\,\Delta_i$. We first rewrite $\widehat{S}_n$ and $\widetilde{S}_n$

$$\widehat{S}_n(y;u)\;=\;\frac{n^{-1}\sum_{i=1}^n K_h(\delta(U_i,u))\,d^2(Y_i,y)}{n^{-1}\sum_{i=1}^n K_h(\delta(U_i,u))}=\frac{\widehat{r}_2(u)}{\widehat{r}_1(u)},$$

$$\widetilde{S}_n(y;u) \quad = \quad \frac{\mathbb{E}(K_h(\delta(U_1,u))\, d^2(Y_1,y))}{\mathbb{E}\, K_h(\delta(U_1,u))} = \mathbb{E}(d^2(Y_1,y)\,\Delta_1).$$

Then, we have that

$$\begin{aligned}\widehat{S}_n(y;u) - \widetilde{S}_n(y;u) &= \frac{\widehat{r}_2(u) - \mathbb{E}(d^2(Y_1,y)\,\Delta_1) - \mathbb{E}(d^2(Y_1,y)\,\Delta_1)\,(\widehat{r}_1(u)-1)}{\widehat{r}_1(u)} \\ &= \frac{\widehat{r}_2(u) - \mathbb{E}(d^2(Y_1,y)\,\Delta_1)}{\widehat{r}_1(u)} - \mathbb{E}(d^2(Y_1,y)\,\Delta_1)\cdot\frac{\widehat{r}_1(u)-1}{\widehat{r}_1(u)}.\end{aligned}$$

Let $r_n = (\log n/(n\varphi_{\mathcal{U},u}(h)))^{1/2}$. The Lemma 4 shows that $\widehat{r}_2(u) - \mathbb{E}(d^2(Y_1,y)\,\Delta_1) = O_{\text{a.co.}}(r_n)$, $\widehat{r}_1(u) - 1 = O_{\text{a.co.}}(r_n)$. Since $\widehat{r}_1(u)$ to 1 a.co., $\widehat{r}_1(u)^{-1} = 1 + O_{\text{a.co.}}(r_n)$. Thus,

$$\widehat{S}_n(y;u) - \widetilde{S}_n(y;u) = O_{\text{a.co.}}(r_n).$$

Because convergence almost completely implies convergence in probability and we have assumed that $\lim_{n\to\infty} \frac{\log n}{n\varphi_{\mathcal{U},u}(h)} = 0$ in (A8), we have then shown that $\widehat{S}_n(y;u) - \widetilde{S}_n\,(y;u) = o_p\,(1)$ for any $y \in \mathcal{Y}$. Then, it suffices to to show that for all $\epsilon, \eta > 0$, there exists $\delta > 0$ such that

$$\limsup_{n\to\infty} \mathbb{P}(\sup_{d(y_1,y_2)<\delta} \left| (\widetilde{S}_n - \widehat{S}_n)\,(y_1;u) - (\widetilde{S}_n - \widehat{S}_n)\,(y_2;u) \right| > \epsilon) < \eta.$$

Since the kernel function is assumed to be non-negative, we have that $n^{-1}\sum_{i=1}^{n} |s_{in}\,(u)| = 1$. Then, $\left|\widehat{S}_n\,(y_1;u) - \widehat{S}_n\,(y_2;u)\right| \leq 2\,\mathrm{diam}\,(\mathcal{Y})\,d\,(y_1,y_2)\,n^{-1}\sum_{i=1}^{n}|s_{in}\,(u)| = O\,(d\,(y_1,y_2))$. Similarly, $\left|\widetilde{S}_n\,(y_1;u) - \widetilde{S}_n\,(y_2;u)\right| = O\,(d\,(y_1,y_2))$. Hence, $d\,(\widetilde{s}_\oplus\,(u)\,, \widehat{s}_\oplus\,(u)) = o_p(1)$.

Further, since $\widehat{S}_n(y;u)$ convergences to $\widetilde{S}_n(y;u)$ almost surely, the probability of the event $\{\lim_{n\to\infty}\widehat{S}_n(y;u) = \widetilde{S}_n(y;u)\}$ equals to 1. When the even $\{\lim_{n\to\infty} \widehat{S}_n(y;u) = \widetilde{S}_n(y;u)\}$ occurs, we have $d(\widehat{s}_\oplus(u), \widetilde{s}_\oplus(u)) \to 0$ by Assumption (A1). Consequently, we can get that $d\left(\widehat{s}_\oplus(u), \widetilde{s}_\oplus(u)\right) \to 0$ almost surely. □

Lemma 6. *Under Assumptions (A1), (A2) and (A6)-(A9), we have*

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

$$\textit{and} \qquad d(\widetilde{s}_\oplus(u), \widehat{s}_\oplus(u)) = O_{a.s.}((n\varphi_{\mathcal{U},u}(h)/\log n)^{\frac{-1}{2(\beta_2-1)}}),$$

Proof. We use similar methods as before. Let $T_{n,h}\left(y\right) = \widehat{S}_n\left(y;u\right) - \widetilde{S}_n\left(y;u\right)$, $D_i(y) = d^2(Y_i, y) - d^2(Y_i, \widetilde{s}_\oplus\left(u\right))$. Then, we have

$$|T_{n,h}\left(y\right) - T_{n,h}(\widetilde{s}_\oplus\left(u\right))| \le |n^{-1}\sum_{i=1}^{n}(s_{in} - \zeta_{hi})D_i(y)| + |n^{-1}\sum_{i=1}^{n}[\zeta_{hi}D_i(y) - \mathbb{E}(\zeta_{hi}D_i\left(y\right))]|, \tag{S4.5}$$

Based on the proof of Lemma 5 and $|D_i\left(y\right)| \le 2\,\mathrm{diam}\left(\mathcal{Y}\right) d\left(y, \widetilde{s}_\oplus\left(u\right)\right)$, we can get that $|n^{-1}\sum_{i=1}^{n}\left(s_{in} - \zeta_{hi}\right) D_i(y)| = O_p[d\left(y, \widetilde{s}_\oplus(u)\right)\left((n\varphi_{\mathcal{U},u}(h))\right)^{-1/2}]$, where c is independent of y and $\widetilde{s}_\oplus\left(u\right)$. Then, for a given $R > 0$, we can define event

$$C(R) = \Big\{\sup_{d(y,\widetilde{s}_\oplus(u))<h'} \left|n^{-1}\sum_{i=1}^{n}(s_{in} - \zeta_{hi})D_i(y)\right| \le Rh'((n\varphi_{\mathcal{U},u}(h)))^{-1/2}\Big\},$$

so that $\mathbb{P}(C^c(R))$ can be made arbitrarily small by choosing R and n sufficiently large.

Now look at the second term on the right hand side of Equation (S4.5). To manage this term, define $m_{(u,y)}(U, Y) : \mathcal{U} \times \mathcal{Y} \to \mathbb{R}$ as $m_{(u,y)}(U, Y) =$

$K_h(\delta(U,u))d^2(Y,y)/\mathbb{E}(K_h(\delta(U,u)))$ and the corresponding function class $\mathcal{M}_{n,h'} = \left\{m_{(u,y)} - m_{(u,\widetilde{s}_\oplus(u))} : d\left(y, \widetilde{s}_\oplus\left(u\right)\right) < h'\right\}$. An envelope function for $\mathcal{M}_{nh'}$ is

$$M_{nh'}\left(U,Y\right) = \frac{2\,\mathrm{diam}\left(\mathcal{Y}\right)h' K_h\left(\delta\left(U,u\right)\right)}{\mathbb{E}\left(K_h\left(\delta\left(U,u\right)\right)\right)}.$$

It follows from (S4.3) that $\mathbb{E}(M^2_{nh'}(u)) = O(h'^2\varphi^{-1}_{\mathcal{U},u}(h))$. Combining this fact with Theorems 2.7.11 and 2.14.2 of Van Der Vaart & Wellner (1996), for small h', under assumption (A2), it follows that

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

Combining this with Equation (S4.5) and the definition of the $C(R)$,

$$\mathbb{E}\big(\mathbb{1}_{C(R)}\sup_{d(y,\widetilde{s}_\oplus(u))<h'}\left|T_{n,h}\left(y\right) - T_{n,h}\left(\widetilde{s}_\oplus\left(u\right)\right)\right|\big) \leq ah'(n\varphi_{\mathcal{U},u}(h)))^{-1/2},$$

where $\mathbb{1}_{C(R)}$ is indicator function for $C(R)$ and a is a constant depending on R. To complete the proof, set $r_{n,h} = (n\varphi_{\mathcal{U},u}(h))^{\frac{\beta_2}{4(\beta_2-1)}}$ and $Q_{j,n} = \left\{y : 2^{j-1} < r_{n,h}d(y, \widetilde{s}_\oplus\left(u\right))^{\beta_2/2} \leq 2^j\right\}$. Choose η_2 satisfying Assumption (A6) and set $\widetilde{\eta} := (\eta_2/2)^{\beta_2/2}$. For any integer M,

$$\begin{aligned}&\mathbb{P}\big(r_{n,h}d(\widetilde{s}_\oplus(u),\widehat{s}_\oplus(u))^{\beta_2/2} > 2^M\big) \leq \mathbb{P}(C^c(R)) + \mathbb{P}\left(2d\left(\widetilde{s}_\oplus\left(u\right),\widehat{s}_\oplus\left(u\right)\right) > \eta\right)\\ &+ \sum_{\substack{j\geq M\\ 2^j\leq r_{n,h}\tilde{\eta}}}\mathbb{P}[(\mathbb{1}_{C(R)}\sup_{y\in Q_{j,n}}\left|T_{n,h}\left(y;u\right) - T_{n,h}\left(\widetilde{s}_\oplus\left(u\right);u\right)\right|) \geq C\frac{2^{2(j-1)}}{r^2_{n,h}}],\end{aligned}\tag{S4.6}$$

By Lemma 5, we have $\mathbb{P}\left(2d(\widetilde{s}_{\oplus}(u),\widehat{s}_{\oplus}(u))>\eta\right)\to 0$. The fact that $d(y,\widetilde{s}_{\oplus}(u))<\left(2^j/r_{n,h}\right)^{2/\beta_2}$ on $Q_{j,n}(u)$ implies that the sum on the right hand side of Equation (S4.6) is bounded above by

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

which converges since $\beta_2>1$. Thus $d\left(\widetilde{s}_{\oplus}\left(u\right),\widehat{s}_{\oplus}\left(u\right)\right)=O_p(r_{n,h}^{-2/\beta_2})=O_p((n\varphi_{\mathcal{U},u}(h))^{\frac{-1}{2(\beta_2-1)}})$.

Then we prove the almost sure convergence rate. Using the same technique as in the proof of Theorem 1 and $\widehat{S}_n(y;u)-\widetilde{S}_n(y;u)=O_{\text{a.co.}}((\log n/(n\varphi_{\mathcal{U},u}(h)))^{1/2})$, let $t_{n,h}=(n\varphi_{\mathcal{U},u}(h)/\log n)^{\frac{\beta_2}{4(\beta_2-1)}}$ and $C_{j,n}=\left\{y:2^{j-1}<t_{n,h}d(y,\widetilde{s}_{\oplus}\left(u\right))^{\beta_2/2}\leq 2^j\right\}$ we can get that there exists a constant $C>0$, then

$$\sup_{d(y,\widetilde{s}_{\oplus}(u))<(2^j/t_{n,h})^{2/\beta_2}}\left|T_{n,h}\left(y\right)-T_{n,h}\left(\widetilde{s}_{\oplus}\left(u\right)\right)\right|\leq C(2^j/t_{n,h})^{2/\beta_2}(\log n/(n\varphi_{\mathcal{U},u}(h)))^{1/2}.$$

For a given $M>0$, we define event $A=\{\lim_{n\to\infty}|\widehat{S}_n(y;u)-\widetilde{S}_n(y;u)|t_{n,h}^{2/\beta_2}\leq M\}$ to satisfy $\mathbb{P}(A)=1$. Thus, $\mathbb{P}(\sum_{j>1}\mathbb{1}\sup_{y\in C_{j,n}}\left|T_{n,h}\left(y\right)-T_{n,h}\left(\widetilde{s}_{\oplus}\left(u\right)\right)\right|>c_j\frac{2^{2(j-1)}}{t_{n,h}^2}|A)=0$. By the same arguments as before, we have $\mathbb{P}(\lim_{n\to\infty}d(\widetilde{s}_{\oplus}(u),\widehat{s}_{\oplus}(u))t_{n,h}^{2/\beta_2}\leq M|A)=1$. Since $\mathbb{P}(\lim_{n\to\infty}d(\widetilde{s}_{\oplus}(u),\widehat{s}_{\oplus}(u))t_{n,h}^{2/\beta_2}\leq M)=\mathbb{P}(\lim_{n\to\infty}d(\widetilde{s}_{\oplus}(u),\widehat{s}_{\oplus}(u))t_{n,h}^{2/\beta_2}\leq M|A)\mathbb{P}(A)+\mathbb{P}(\lim_{n\to\infty}d(\widetilde{s}_{\oplus}(u),\widehat{s}_{\oplus}(u))t_{n,h}^{2/\beta_2}\leq M|A^c)\mathbb{P}(A^c)=1$, we conclude that

$$d(\widetilde{s}_{\oplus}(u),\widehat{s}_{\oplus}(u))=O_{a.s.}((n\varphi_{\mathcal{U},u}(h)/\log n)^{\frac{-1}{2(\beta_2-1)}}).$$

□

Lemma 7. *Under Assumptions (A1) and (A7)-(A9), we have*

$$d\left(\widetilde{s}_{\oplus}\left(\boldsymbol{x},u\right),\widehat{s}_{\oplus}\left(\boldsymbol{x},u\right)\right)=o_{a.s.}\left(1\right).$$

Proof. We can write $\widehat{S}_n\left(y;\boldsymbol{x},u\right)-\widetilde{S}_n\left(y;\boldsymbol{x},u\right)=\widehat{w}_0\left(y;u\right)-\widetilde{w}_0\left(y;u\right)+\widehat{w}_1\left(\boldsymbol{x},u\right)\widehat{w}_2^{-1}\left(u\right)\widehat{w}_3\left(y;u\right)-\widetilde{w}_1\left(\boldsymbol{x},u\right)\widetilde{w}_2^{-1}\left(u\right)\widetilde{w}_3\left(y;u\right)$. In Lemma 5, we have $\widehat{w}_0\left(y;u\right)=\widetilde{w}_0\left(y;u\right)+O_{\text{a.co.}}(\sqrt{\log n/n\varphi_{\mathcal{U},u}\left(h\right)})$. Let $\widehat{w}_1\left(\boldsymbol{x},u\right)=\left(\widehat{w}_{1,1}\left(x_1,u\right),\ldots,\widehat{w}_{1,p}\left(x_p,u\right)\right)^{\mathrm{T}}$ and $\widetilde{w}_1\left(\boldsymbol{x},u\right)=\left(\widetilde{w}_{1,1}\left(x_1,u\right),\ldots,\widetilde{w}_{1,p}\left(x_p,u\right)\right)^{\mathrm{T}}$. Similarly, it can be shown that $\widehat{w}_{1,j}\left(x_j,u\right)=\widetilde{w}_{1,j}\left(x_j,u\right)+O_{\text{a.co.}}(\sqrt{\log n/n\varphi_{\mathcal{U},u}\left(h\right)})$ for $j=1,..,p$.

To get difference between $\widehat{w}_2(u)$ and $\widetilde{w}_2(u)$, let $\mu_{\boldsymbol{X}_h}=\mathbb{E}[\zeta_{hi}\boldsymbol{X}]$, $\bar{\boldsymbol{X}}_h=n^{-1}\sum_{j=1}^{n}s_{jn}\boldsymbol{X}_j$,

$$\begin{aligned}\widehat{w}_2(u)-\widetilde{w}_2(u)&=\frac{1}{n}\sum_{i=1}^{n}s_{in}(\boldsymbol{X}_i-\bar{\boldsymbol{X}}_h)(\boldsymbol{X}_i-\bar{\boldsymbol{X}}_h)^{\mathrm{T}}-\mathbb{E}[\zeta_{hi}(\boldsymbol{X}-\mu_{\boldsymbol{X}_h})(\boldsymbol{X}-\mu_{\boldsymbol{X}_h})^{\mathrm{T}}]\\&=\underbrace{\frac{1}{n}\sum_{i=1}^{n}s_{in}\left[(\boldsymbol{X}_i-\bar{\boldsymbol{X}}_h)(\boldsymbol{X}_i-\bar{\boldsymbol{X}}_h)^{\mathrm{T}}-(\boldsymbol{X}_i-\mu_{\boldsymbol{X}_h})(\boldsymbol{X}_i-\mu_{\boldsymbol{X}_h})^{\mathrm{T}}\right]}_{A}\\&\quad+\underbrace{(\frac{1}{n}\sum_{i=1}^{n}s_{in}(\boldsymbol{X}_i-\mu_{\boldsymbol{X}_h})(\boldsymbol{X}_i-\mu_{\boldsymbol{X}_h})^{\mathrm{T}}-\mathbb{E}[\zeta_{hi}(\boldsymbol{X}-\mu_{\boldsymbol{X}_h})(\boldsymbol{X}-\mu_{\boldsymbol{X}_h})^{\mathrm{T}}])}_{B}.\end{aligned}$$

By analogy with the proof of Lemma 5, for every element (k,l) in B, we have $B_{kl}=O_{\text{a.co.}}(\sqrt{\log n/n\varphi_{\mathcal{U},u}(h)})$. For Part A, we decompose A as follows:

$$\begin{aligned}A&=\frac{1}{n}\sum_{i=1}^{n}s_{in}((\boldsymbol{X}_i-\mu_{\boldsymbol{X}_h}+\mu_{\boldsymbol{X}_h}-\bar{\boldsymbol{X}}_h)(\boldsymbol{X}_i-\mu_{\boldsymbol{X}_h}+\mu_{\boldsymbol{X}_h}-\bar{\boldsymbol{X}}_h)^{\mathrm{T}}-(\boldsymbol{X}_i-\mu_{\boldsymbol{X}_h})(\boldsymbol{X}_i-\mu_{\boldsymbol{X}_h})^{\mathrm{T}})\\&=(\mu_{\boldsymbol{X}_h}-\bar{\boldsymbol{X}}_h)(n^{-1}\sum_{i=1}^{n}s_{in}(\boldsymbol{X}_i-\bar{\boldsymbol{X}}_h)^{\mathrm{T}})+(n^{-1}\sum_{i=1}^{n}s_{in}(\boldsymbol{X}_i-\mu_{\boldsymbol{X}}))(\mu_{\boldsymbol{X}_h}-\bar{\boldsymbol{X}}_h)^{\mathrm{T}}.\end{aligned}$$

Thus, any element (k,l) of A is $O_{\text{a.co.}}(\sqrt{\log n/n\varphi_{\mathcal{U},u}(h)})O(1)+O(1)O_{\text{a.co.}}(\sqrt{\log n/n\varphi_{\mathcal{U},u}(h)})$. Consequently, the every element (k,l) of $\widehat{w}_2(u)-\widetilde{w}_2(u)$ is $O_{\text{a.co.}}(\sqrt{\log n/n\varphi_{\mathcal{U},u}(h)})$. Similarity, we can find the every element (k,l) in $\widehat{w}_3(y;u)-\widetilde{w}_3(y;u)$ is $O_{\text{a.co.}}(\sqrt{\log n/n\varphi_{\mathcal{U},u}(h)})$.

Since $\widehat{w}_2^{-1}(u)-\widetilde{w}_2^{-1}(u)=\widehat{w}_2^{-1}(u)(\widetilde{w}_2(u)-\widehat{w}_2(u))\widetilde{w}_2^{-1}(u)$, the bound on $\widehat{w}_2(u)-\widetilde{w}_2(u)$ implies that the every element (k,l) of $\widehat{w}_2^{-1}(u)-\widetilde{w}_2^{-1}(u)$ is $O_{\text{a.co.}}(\sqrt{\log n/n\varphi_{\mathcal{U},u}(h)})$. Next, consider the $\widehat{w}_1(\boldsymbol{x},u)\widehat{w}_2^{-1}(u)\widehat{w}_3(y;u)-\widetilde{w}_1(\boldsymbol{x},u)\widetilde{w}_2^{-1}(u)\widetilde{w}_3(y;u)$. Decompose this difference as $(\widehat{w}_1-\widetilde{w}_1)\widehat{w}_2^{-1}\widehat{w}_3+\widetilde{w}_1(\widehat{w}_2^{-1}-\widetilde{w}_2^{-1})\widehat{w}_3+\widetilde{w}_1\widetilde{w}_2^{-1}(\widehat{w}_3-\widetilde{w}_3)$. Note that $\widetilde{w}_1(\boldsymbol{x},u)$, $\widetilde{w}_2^{-1}(u)$, and $\widetilde{w}_3(y;u)$ are bounded, and we have shown that each element in $\widehat{w}_1-\widetilde{w}_1$, $\widehat{w}_2^{-1}-\widetilde{w}_2^{-1}$, and $\widehat{w}_3-\widetilde{w}_3$ is $O_{\text{a.co.}}(\sqrt{\log n/n\varphi_{\mathcal{U},u}(h)})$. Combining this with $\widehat{w}_0\left(y;u\right)-\widetilde{w}_0\left(y;u\right)=O_{\text{a.co.}}(\sqrt{\log n/n\varphi_{\mathcal{U},u}\left(h\right)})$, we conclude that $\widehat{S}_n\left(y;\boldsymbol{x},u\right)-\widetilde{S}_n\left(y;\boldsymbol{x},u\right)=O_{a.s.}(\sqrt{\log n/n\varphi_{\mathcal{U},u}(h)})$.

Thus, $\widehat{S}_n\left(y;\boldsymbol{x},u\right)-\widetilde{S}_n\left(y;\boldsymbol{x},u\right)=o_p\left(1\right)$ for any $y\in\mathcal{Y}$. Then, we just need to show that for any $\eta>0$, $\limsup_{n\to\infty}\mathbb{P}\big(\sup_{d(y_1,y_2)<h'}\big|(\widetilde{S}_n-\widehat{S}_n)\left(y_1;\boldsymbol{x},u\right)-(\widetilde{S}_n-\widehat{S}_n)\left(y_2;\boldsymbol{x},u\right)\big|>\eta\big)\to 0$ as $h'\to 0$. We can write

$$\begin{aligned}|\widehat{S}_n\left(y_1\right)-\widehat{S}_n\left(y_2\right)|&=\left|\widehat{w}_0\left(y_1\right)-\widehat{w}_0\left(y_2\right)+\widehat{w}_1\left(\boldsymbol{x}\right)\widehat{w}_2^{-1}\left(\widehat{w}_3\left(y_1\right)-\widehat{w}_3\left(y_2\right)\right)\right|\\&\le|\widehat{w}_0\left(y_1\right)-\widehat{w}_0\left(y_2\right)|+\|\widehat{w}_1\left(\boldsymbol{x}\right)\widehat{w}_2^{-1}\|\|\left(\widehat{w}_3\left(y_1\right)-\widehat{w}_3\left(y_2\right)\right)\|\end{aligned}$$

From Lemma 5, the first term is $O_p\left(d\left(y_1,y_2\right)\right)$. Further, $\|\widehat{w}_1\left(\boldsymbol{x}\right)\widehat{w}_2^{-1}(u)\|\|\left(\widehat{w}_3\left(y_1\right)-\widehat{w}_3(y_2)\right)\|=O_p\left(d\left(y_1,y_2\right)\right)$. Thus, $|\widehat{S}_n\left(y_1;\boldsymbol{x},u\right)-\widehat{S}_n\left(y_2;\boldsymbol{x},u\right)|=O_p\left(d\left(y_1,y_2\right)\right)$. Similarly, we have $|\widetilde{S}_n(y_1;\boldsymbol{x},u)-\widetilde{S}_n(y_2;\boldsymbol{x},u)|=O\left(d\left(y_1,y_2\right)\right)$. Hence, $d\left(\widetilde{s}_{\oplus}(\boldsymbol{x},u),\widehat{s}_{\oplus}(\boldsymbol{x},u)\right)=o_p(1)$.

Further, since $\widehat{S}_n(y;\boldsymbol{x},u)$ convergences to $\widetilde{S}_n(y;\boldsymbol{x},u)$ almost surely, the probability of the event $\{\lim_{n\to\infty}\widehat{S}_n(y;\boldsymbol{x},u)=\widetilde{S}_n(y;\boldsymbol{x},u)\}$ equals to 1. When the even $\{\lim_{n\to\infty}\widehat{S}_n(y;\boldsymbol{x},u)=\widetilde{S}_n(y;\boldsymbol{x},u)\}$ occurs, we have $d(\widehat{s}_{\oplus}(\boldsymbol{x},u),\widetilde{s}_{\oplus}(\boldsymbol{x},u))\to 0$ by Assumption (A1). Consequently, we can get that $d\left(\widehat{s}_{\oplus}(\boldsymbol{x},u),\widetilde{s}_{\oplus}(\boldsymbol{x},u)\right)\to 0$ almost surely.

□

Proof of Theorem 5. We will follow similar arguments as Theorem 2 and Theorem 6. Set $T_{n,h}(y;\boldsymbol{x},u)=\widehat{S}_n(y;\boldsymbol{x},u)-\widetilde{S}_n(y;\boldsymbol{x},u)$, $D_i=d^2(Y_i,y)-d^2(Y_i,\widetilde{s}_{\oplus}(\boldsymbol{x},u))$, $\zeta_{hi}(u)=\zeta_h(U_i,u)$, $\bar{\boldsymbol{X}}_h=n^{-1}\sum_{i=1}^n s_{in}\boldsymbol{X}_i$, $\bar{D}_h=n^{-1}\sum_{i=1}^n s_{in}D_i$, $\mu_{\boldsymbol{X}_h}=\mathbb{E}(\zeta_{hi}\boldsymbol{X}_i)$, and $\mu_{D_h}=\mathbb{E}(\zeta_{hi}D_i)$. Then we consider the follow inequality,

$$|T_{n,h}(y;\boldsymbol{x},u)-T_{n,h}(\widetilde{s}_{\oplus}(\boldsymbol{x},u);\boldsymbol{x},u)|\le|\frac{1}{n}\sum_{i=1}^n(s_{in}-\zeta_{hi})D_i|+|\frac{1}{n}\sum_{i=1}^n(\zeta_{hi}D_i-\mathbb{E}(\zeta_{hi}D_i))|+$$

$$|\widehat{w}_1\widehat{w}_2^{-1}\frac{1}{n}\sum_{i=1}^n s_{in}(\boldsymbol{X}_i-\bar{\boldsymbol{X}}_h)(D_i-\bar{D}_h)-\widetilde{w}_1\widetilde{w}_2^{-1}\mathbb{E}\{\zeta_{hi}(\boldsymbol{X}_i-\mu_{\boldsymbol{X}_h})(D_i-\mu_{D_h})\}|$$

$$=:T_1(y)+T_2(y)+T_3(y).$$

We have shown that the first term is $O_p[d(y,\widetilde{s}_{\oplus}(u))(n\varphi_{\mathcal{U},u}(h))^{-1/2}]$ and the O_p term is independent of y and $\widetilde{s}_{\oplus}(\boldsymbol{x},u)$.

Next, we control the second term uniformly as before. Under Assumption (A2), we obtain

$$\mathbb{E}\Big(\sup_{d(y,\widetilde{s}_\oplus(\boldsymbol{x},u))<h'}\Big|\frac{1}{n}\sum_{i=1}^{n}\zeta_{hi}D_i-\mathbb{E}(\zeta_{hi}D_i)\Big|\Big)=O(h'(n\varphi_{\mathcal{U},u}\left(h\right))^{-1/2}).$$

Also, we can obtain that $\sup_{d(y,\widetilde{s}_\oplus(\boldsymbol{x},u))<\delta'} T_3(y)=O_p(h'(n\varphi_{\mathcal{U},u}\left(h\right))^{-1/2}d(y,\widetilde{s}_\oplus))$. Define $D(R')=\{\sup_{d(y,\widetilde{s}_\oplus(\boldsymbol{x},u))<\delta'}(T_1(y)+T_3(y))\leq\frac{R'h'}{(n\varphi_{\mathcal{U},u}(h))^{1/2}}\}$, where $R'>0, h'>0$. Then, $\mathbb{P}(D^c(R'))$ can be made arbitrarily small by choosing R' and n sufficiently large. Therefore, we have

$$\mathbb{E}\big(\mathbb{1}_{D(R')}\sup_{d(y,\widetilde{s}_\oplus(\boldsymbol{x},u))<h'}\left|T_{n,h}\left(y;\boldsymbol{x},u\right)-T_{n,h}\left(\widetilde{s}_\oplus\left(\boldsymbol{x},u\right);\boldsymbol{x},u\right)\right|\big)\leq Ch'(n\varphi_{\mathcal{U},u}\left(h\right))^{-1/2},$$

where $\mathbb{1}_{D(R')}$ is the indicator function for the set $D(R')$ and C is a constant depending on R' and the assumption (A6).

To finish the proof, let $t_n=(n\varphi_{\mathcal{U},u}(h))^{\frac{\beta_2}{4(\beta_2-1)}}$ and define $S_{j,n}=\{y: 2^{j-1}<t_nd(y,\widetilde{s}_\oplus(\boldsymbol{x},u)\leq 2^j\}$. Choose η_2 satisfying (A6) and let $\tilde{\eta}:=(\eta_2/2)^{\beta_2/2}$. For any integer M,

$$\begin{aligned}&\mathbb{P}\left(t_nd(\widetilde{s}_\oplus(\boldsymbol{x},u),\widehat{s}_\oplus(\boldsymbol{x},u))^{\beta_2/2}>2^M\right)\leq\mathbb{P}\left(D^c(R')\right)+\mathbb{P}\left(2d(\widetilde{s}_\oplus(\boldsymbol{x},u),\widehat{s}_\oplus(\boldsymbol{x},u))>\tilde{\eta}\right)\\&+\sum_{\substack{j\geq M\\2^j\leq t_n\tilde{\eta}}}\mathbb{P}[(\mathbb{1}_{D(R')}\sup_{y\in S_{j,n}}\left|T_{n,h}\left(y;\boldsymbol{x},u\right)-T_{n,h}\left(\widetilde{s}_\oplus\left(\boldsymbol{x},u\right);\boldsymbol{x},u\right)\right|)\geq C\frac{2^{2(j-1)}}{t_n^2}],\end{aligned}\tag{S4.7}$$

where the second term tends to 0 for any $\tilde{\eta}>0$ by Lemma 7. Applying Markov inequality shows that the third term of (S4.7) is bounded by

$$4aC^{-1}\sum_{\substack{j\geq M\\2^j\leq t_n\tilde{\eta}}}\frac{2^{2j(1-\beta_2)/\beta_2}}{t_n^{2(1-\beta_2)/\beta_2}\sqrt{n\varphi_{\mathcal{U},u}(h)}}\leq 4aC^{-1}\sum_{j\geq M}\left(\frac{1}{4^{(\beta_2-1)/\beta_2}}\right)^j,$$

which converges since $\beta_2 > 1$. Thus, $d\left(\widetilde{s}_{\oplus}\left(\boldsymbol{x}, u\right), \widehat{s}_{\oplus}\left(\boldsymbol{x}, u\right)\right) = O_p((n\varphi_{\mathcal{U},u}(h))^{\frac{-1}{2(\beta_2-1)}})$.

Next, we derive the almost sure convergence rate. Recall we have established the fact that $\widehat{S}_n(y; \boldsymbol{x}, u) - \widetilde{S}_n(y; \boldsymbol{x}, u) = O_{\text{a.co.}}((\log n/(n\varphi_{\mathcal{U},u}(h)))^{1/2})$. Define $t_{n,h} = (n\varphi_{\mathcal{U},u}(h)/\log n)^{\frac{\beta_2}{4(\beta_2-1)}}$ and define the sets

$$C_{j,n} = \left\{y : 2^{j-1} < t_{n,h} d(y, \widetilde{s}_{\oplus}\left(\boldsymbol{x}, u\right))^{\beta_2/2} \leq 2^j\right\}$$

. Let B be the event on which there exists a constant $C > 0$ such that

$$\sup_{d(y,\widetilde{s}_{\oplus}(\boldsymbol{x},u))<(2^j/t_{n,h})^{2/\beta_2}} \left|T_{n,h}\left(y\right) - T_{n,h}\left(\widetilde{s}_{\oplus}\left(\boldsymbol{x}, u\right)\right)\right| \leq C(2^j/t_{n,h})^{2/\beta_2}(\log n/(n\varphi_{\mathcal{U},u}(h)))^{1/2}.$$

Then, using the same technique as before, we have $\mathbb{P}(B) = 1$. Thus,

$$\mathbb{P}\Big(\sum_{j>M} \mathbb{1}\Big\{\sup_{y\in C_{j,n}} \left|T_{n,h}\left(y\right) - T_{n,h}\left(\widetilde{s}_{\oplus}\left(\boldsymbol{x}, u\right)\right)\right| > C\frac{2^{2(j-1)}}{t_{n,h}^2}\Big\} \mid B\Big) = 0$$

for sufficiently large M and C. Using an essentially identical argument, we have

$$\mathbb{P}(\lim_{n\to\infty} d(\widetilde{s}_{\oplus}(\boldsymbol{x}, u), \widehat{s}_{\oplus}(\boldsymbol{x}, u))t_{n,h}^{2/\beta_2} \leq M \mid B) = 1.$$

By the law of total probability, we conclude that

$$d(\widetilde{s}_{\oplus}(\boldsymbol{x}, u), \widehat{s}_{\oplus}(\boldsymbol{x}, u)) = O_{a.s.}((n\varphi_{\mathcal{U},u}(h)/\log n)^{\frac{-1}{2(\beta_2-1)}}).$$

□

Bibliography

Scott, D. W. (1992), *Multivariate Density Estimation: Theory, Practice, and Visualization*, Wiley.

Stellato, B., Banjac, G., Goulart, P., Bemporad, A. & Boyd, S. (2020), 'Osqp: An operator splitting solver for quadratic programs', *Mathematical Programming Computation* **12**(4), 637–672.

Van Der Vaart, A. W. & Wellner, J. A. (1996), *Weak convergence and empirical processes*, Springer Series in Statistics, Springer, New York, NY.