Supplementary Materials for the manuscript entitled "Sparse Estimation for Functional Semiparametric Additive Model"

S1 Additional Simulation Results

This section displays the simulation results for Section 3.2, where different methods are fitted for the regression model without scalar covariates.

Model	Counts with the model size												
	1	2	3	4	5	6	7	8	9	10	11	12	13
MARS	0	0	20	71	140	170	197	181	127	69	16	6	3
FSAM-GAMS	0	0	169	261	238	161	98	42	19	9	2	1	0
FSAM-PFLR	5	533	312	98	40	8	4	0	0	0	0	0	0
FSAM-COSSO	0	3	615	202	93	49	27	8	3	0	0	0	0
FSAM-GAM2	0	6	738	168	52	26	3	3	4	0	0	0	0

Table S1: Summary of the number of selected nonparametric components over the 1000 simulations for each model. Model size indicates the number of nonparametric components selected in the model. In FSAM-GAMS we only retain the significant nonparametric components (p-value less than 0.05). Here we implement the function **gam** in the R package **mgcv** to fit FSAM-GAMS. The corresponding p-values of nonparametric components are available from the function **summary.gam**. This selection rule applies to FSAM-PFLR as well, where the p-value is available from the function **lm**.

Model	Frequency of each nonparametric factor									
	$\hat{f}_1$	$\hat{f}_2$	$\hat{f}_3$	$\hat{f}_4$	$\hat{f}_5$	$\hat{f}_6$	$\hat{f}_7$	$\hat{f}_8$	$\hat{f}_9$	$\hat{f}_{10}$
MARS	1000	1000	356	1000	274	226	233	235	211	244
FSAM-GAMS	1000	1000	232	997	155	100	120	106	103	114
FSAM-PFLR	1000	990	232	998	156	101	121	106	98	110
FSAM-COSSO	1000	1000	102	995	52	22	43	31	27	34
FSAM-GAM2	1000	999	79	992	34	13	28	17	16	20
	$\hat{f}_{11}$	$\hat{f}_{12}$	$\hat{f}_{13}$	$\hat{f}_{14}$	$\hat{f}_{15}$	$\hat{f}_{16}$	$\hat{f}_{17}$	$\hat{f}_{18}$	$\hat{f}_{19}$	$\hat{f}_{20}$
MARS	210	208	196	226	222	240	236	228	217	249
FSAM-GAMS	105	108	105	106	96	101	126	111	114	125
FSAM-PFLR	96	105	104	95	91	78	105	104	91	58
FSAM-COSSO	31	38	39	33	34	41	43	39	47	47
FSAM-GAM2	16	18	14	23	18	14	27	16	21	30

Table S2: Summary of frequency of each nonparametric component selected over the 1000 simulations for each model. In FSAM-GAMS we only retain the significant nonparametric components (p-value less than 0.05). This selection rule applies to FSAM-PFLR as well.

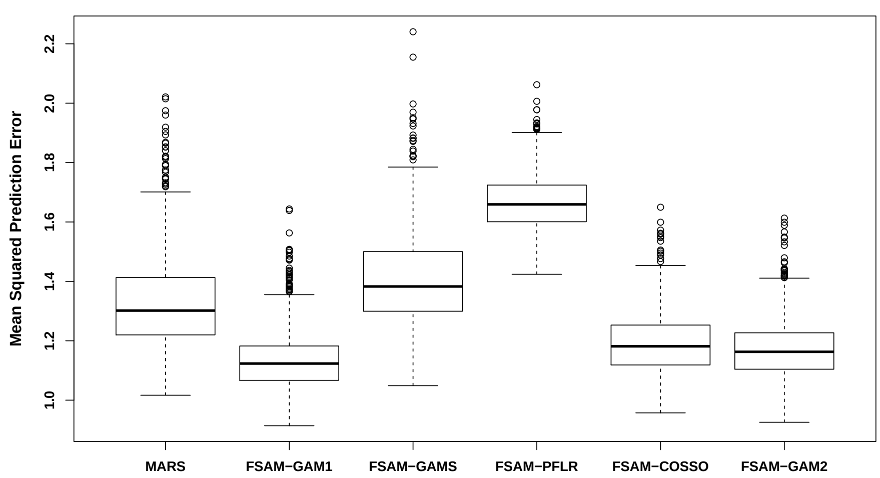

Figure S1: Mean squared prediction errors of each method over 1000 simulations.

S2 Additional Real Application Results

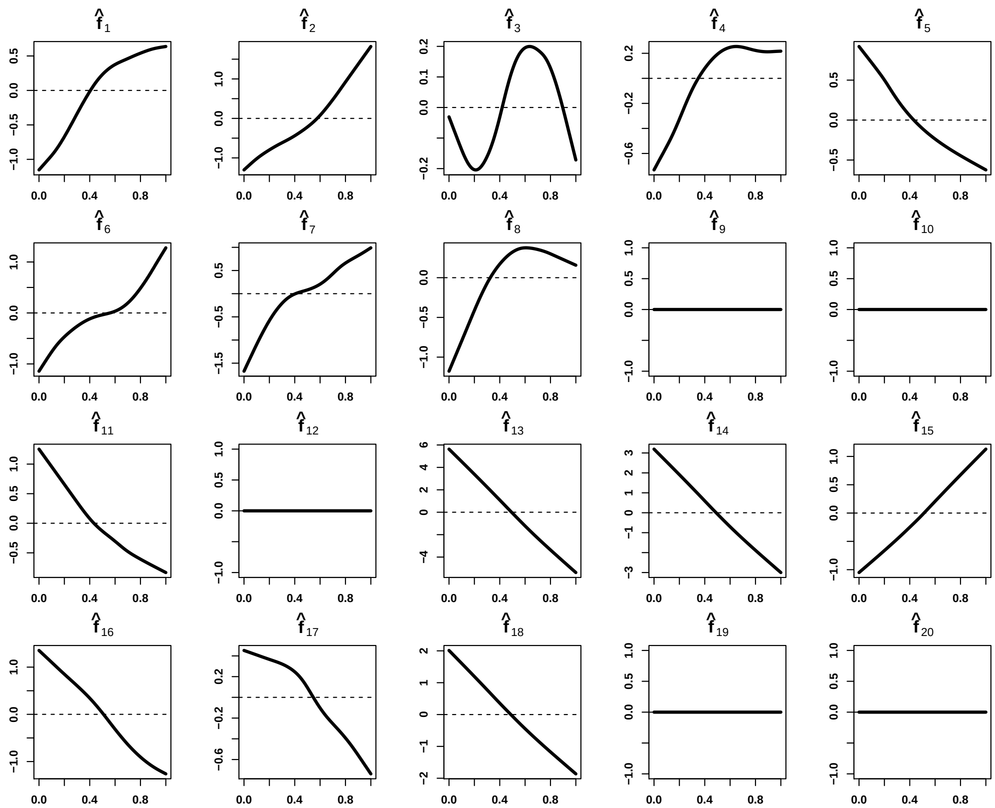

Figure S2: Non-vanishing nonparametric components estimated for the functional semiparametric additive model (4) from the Tecator data. Out of total 20 nonparametric components, 15 nonparametric components are selected.

S3 Proofs

Consider the regression model:

$$y_i = f_0(\boldsymbol{\zeta}_i) + \boldsymbol{\alpha}_0^\top \boldsymbol{z}_i + \epsilon_i,$$

where $f_0(\boldsymbol{\zeta}) = b_0 + \sum_{j=1}^d f_{0j}(\zeta_j)$ with $f_{0j} \in \bar{H}$ and $\boldsymbol{\alpha}_0 \in \mathbb{R}^p$. Write $g(\boldsymbol{\zeta}, \boldsymbol{z}) = a + \tilde{f}(\boldsymbol{\zeta}) + \boldsymbol{\alpha}^\top \tilde{\boldsymbol{z}} = a + \sum_{j=1}^d \tilde{f}_j(\zeta_j) + \boldsymbol{\alpha}^\top \tilde{\boldsymbol{z}}$ such that $\sum_{i=1}^n \tilde{f}_j(\zeta_{ij}) = 0, j = 1, \ldots, d$ and $\tilde{\boldsymbol{z}} = \boldsymbol{z} - \bar{\boldsymbol{z}}$ which satisfies $\sum_{i=1}^n \tilde{z}_{is} = 0, s = 1, \ldots, p$, where $\bar{\boldsymbol{z}}$ denotes the sample mean of $\boldsymbol{z}_i$'s and $\tilde{\boldsymbol{z}}_i = (\tilde{z}_{i1}, \ldots, \tilde{z}_{ip})^\top$ is the evaluation of $\tilde{\boldsymbol{z}}$ at the data point $\boldsymbol{z}_i, i = 1, \ldots, n$. Similarly, write $g_0(\boldsymbol{\zeta}) = a_0 + \tilde{f}_0(\boldsymbol{\zeta}) + \boldsymbol{\alpha}_0^\top \tilde{\boldsymbol{z}} = a_0 + \sum_{j=1}^d \tilde{f}_{0j}(\zeta_j) + \boldsymbol{\alpha}_0^\top \tilde{\boldsymbol{z}}$ such that $\sum_{i=1}^n \tilde{f}_{0j}(\zeta_{ij}) = 0, j = 1, \ldots, d$, and $\hat{g}(\boldsymbol{\zeta}, \boldsymbol{z}) = \hat{a} + \hat{f}(\boldsymbol{\zeta}) + \hat{\boldsymbol{\alpha}}^\top \tilde{\boldsymbol{z}} = \sum_{j=1}^d \hat{f}_j(\zeta_j) + \hat{\boldsymbol{\alpha}}^\top \tilde{\boldsymbol{z}}$.

Remark 1 The above decomposition for f_0 as sum of a_0 and $\tilde{f}_0$ is different from that as an element of $\{1\} \oplus \sum_{j=1}^d \bar{H}$. This difference applies to the decomposition of $\hat{f}$ as well. The latter representation is given for the sake of identifiability and useful for entropy calculation, which will be illustrated in Lemma 2.

Let $J(g) = J(f)$; that is, the penalty ignores the linear components. Then since $\hat{g}$ minimizes the target function

$$\begin{aligned}
L(g) &= \frac{1}{n}\sum_{i=1}^n (g(\boldsymbol{\zeta}_i, \boldsymbol{z}_i) - y_i)^2 + \tau_n^2 J(g) \\
&= \frac{1}{n}\sum_{i=1}^n \left\{ a + \sum_{j=1}^d \tilde{f}_j(\zeta_{ij}) + \boldsymbol{\alpha}^\top \tilde{\boldsymbol{z}}_i - a_0 - \sum_{j=1}^d \tilde{f}_{0j}(\zeta_{ij}) - \boldsymbol{\alpha}_0^\top \tilde{\boldsymbol{z}}_i - \epsilon_i \right\}^2 + \tau_n^2 J(g) \\
&= (a - a_0)^2 - \left(\frac{2}{n}\sum_{i=1}^n \epsilon_i \right)(a - a_0) + \frac{1}{n}\sum_{i=1}^n \left\{ \sum_{j=1}^d \tilde{f}_j(\zeta_{ij}) + \tilde{\boldsymbol{z}}_i^\top \boldsymbol{\alpha} - \sum_{j=1}^d \tilde{f}_{0j}(\zeta_{ij}) - \tilde{\boldsymbol{z}}_i^\top \boldsymbol{\alpha}_0 - \epsilon_i \right\}^2 \\
&\quad + \tau_n^2 J(g),
\end{aligned}$$

the estimated intercept $\hat{a}$ in $\hat{g}$ must satisfy $\hat{a} = a_0 + \frac{1}{n}\sum_{i=1}^n \epsilon_i$, which implies $\hat{a} - a_0 = O_P(n^{-\frac{1}{2}})$. From now on, we consider the target function

$$\tilde{L}(\tilde{f}, \boldsymbol{\alpha} | \boldsymbol{\zeta}_i, \tilde{\boldsymbol{z}}_i) = \frac{1}{n}\sum_{i=1}^n \left\{ \tilde{f}(\boldsymbol{\zeta}_i) + \boldsymbol{\alpha}^\top \tilde{\boldsymbol{z}}_i - \tilde{f}_0(\boldsymbol{\zeta}_i) - \boldsymbol{\alpha}_0^\top \tilde{\boldsymbol{z}}_i - \epsilon_i \right\}^2 + \tau_n^2 J(\tilde{f}). \tag{1}$$

The solution is denoted as $\hat{g}_n = \hat{f}_n + \hat{\boldsymbol{\alpha}}^\top \tilde{\boldsymbol{z}}$, which is an estimate of $g_0 = \tilde{f}_0 + \boldsymbol{\alpha}_0^\top \tilde{\boldsymbol{z}}$.

Let $\mathscr{F}^d = \{f : f \in 1 \oplus \left(\bigoplus_{j=1}^d \bar{H}\right),\ J(f) < \infty\}$, where $J(f) = \sum_{j=1}^d ||P^j f||$ with P^j denoting the orthogonal projection from $\mathscr{F}$ onto $\bar{H}$. Therefore, the conditional expectation, g_0 is an element of

$$\mathscr{G} = \left\{ g : g(\boldsymbol{\zeta}, \boldsymbol{z}) = \sum_{j=1}^d f_j(\zeta_j) + \boldsymbol{\alpha}^\top \tilde{\boldsymbol{z}}, \boldsymbol{\alpha} \in \mathbb{R}^p, \sum_{j=1}^d f_j \in \mathscr{F}^d, \sum_{i=1}^n f_j(\zeta_{ij}) = 0 \right\},$$

under the assumption that $J(f_0) < \infty$. Following Mammen and van de Geer (1997), for $g(\boldsymbol{\zeta}, \boldsymbol{z}) = f(\boldsymbol{\zeta}) + \boldsymbol{\alpha}^\top \tilde{\boldsymbol{z}} \in \mathscr{G}$, $J(g)$ is set to be $J(f)$; thus $J(g_0) < \infty$. Now consider two subsets of $\mathscr{G}$, $\mathscr{G}_1 = \{g_1 : g_1(\boldsymbol{\zeta}, \boldsymbol{z}) = \sum_{j=1}^d f_j(\zeta_j), f_j'\text{s satisfy } \sum_{i=1}^n f_j(\zeta_{ij}) = 0, g_1 \in \mathscr{F}^d\}$ and $\mathscr{G}_2 = \{g_2 : g_2(\boldsymbol{\zeta}, \boldsymbol{z}) = \boldsymbol{\alpha}^\top \tilde{\boldsymbol{z}}, \boldsymbol{\alpha} \in \mathbb{R}^p\}$. Every element of $\mathscr{G}$ can be written as sum of two elements, one from each of $\mathscr{G}_1$ and $\mathscr{G}_2$.

Before stating a proposition that will be employed later, we first introduce some notation and the concept of entropy. Let Q be the joint distribution of $\boldsymbol{\zeta}$ and $\boldsymbol{z}$ and Q_n the corresponding empirical distribution. Obviously the support of Q is $\mathscr{X} = [0,1]^d \times \mathbb{R}^p$. For any function g supported on $\mathscr{X}$, if $\int |g|^2 \,\mathrm{d}Q < \infty$, then define

$$||g||_{2,Q} = \left(\int |g|^2 \,\mathrm{d}Q \right)^{\frac{1}{2}}.$$

We refer to $||\cdot||_{2,Q}$ as the $L_2(Q)$ metric; similarly we can define the $L_2(Q_n)$ metric or the $||\cdot||_n$ metric by replacing Q with Q_n. We can now define the entropy of $\mathscr{G}$ with respect to the $||\cdot||_n$ metric. For any $\delta > 0$, we can find a collection of functions $g_1, \ldots, g_N$, such that for each $g \in \mathscr{G}$, there is a $j = j(g) \in \{1, \ldots, N\}$ such that $||g - g_j||_n \le \delta$. Let $N(\delta, \mathscr{G}, ||\cdot||_n)$ be the smallest value of N for which such a covering by balls with radius δ exists. Then $H(\delta, \mathscr{G}, ||\cdot||_n) = \log\{N(\delta, \mathscr{G}, ||\cdot||_n)\}$ is called the δ-entropy of $\mathscr{G}$ (for the $||\cdot||_n$ metric). Similarly, we can define the entropy of $\mathscr{G}$ for other metrics like the $||\cdot||_\infty$ metric. It is trivial that $H(\delta, \mathscr{G}, ||\cdot||_n) \le H(\delta, \mathscr{G}, ||\cdot||_\infty)$. For distinction, we write $||\cdot||_\infty$ to denote the supremum norm of a function, $||\cdot||_E$ to denote the Euclidean norm of a vector, $||\cdot||$ to denote the Sobolev norm defined in the RKHS, $||\cdot||_{2,Q}$ to denote the $L_2(Q)$ metric and $||\cdot||_n$ to denote the $L_2(Q_n)$ metric. Following the notation on page 167 from van de Geer (2000), for any $g \in \mathscr{G}$,

$$||g||_n^2 = \frac{1}{n}\sum_{i=1}^n g^2(\boldsymbol{\zeta}_i, \boldsymbol{z}_i), \quad (\epsilon, g)_n = \frac{1}{n}\sum_{i=1}^n \epsilon_i g(\boldsymbol{\zeta}_i, \boldsymbol{z}_i), \quad ||y - g||_n^2 = \frac{1}{n}\sum_{i=1}^n (y_i - g(\boldsymbol{\zeta}_i, \boldsymbol{z}_i))^2.$$

For readability, we refer to the assumptions (A.1) - (A.4) as Assumption 1, Assumption 2, etc.

Assumption 1 *Both* $\boldsymbol{\zeta}$ *and* $\boldsymbol{z}$ *are statistically independent of* ϵ*. Furthermore,* $\mathrm{E}(\epsilon) = 0$ *and* $\max_{1\le j\le p} \mathrm{E}(|\boldsymbol{z}_{(j)}|) < \infty$*, where* $\boldsymbol{z}_{(j)}$ *denotes the* j*th component of* $\boldsymbol{z}$*.*

Assumption 2 $\Lambda_{\max}[\mathrm{var}\{h(\boldsymbol{\zeta})\}] < \infty$ *and* $0 < \Lambda_{\min}\{\mathrm{var}(\boldsymbol{z}^\star)\} \le \Lambda_{\max}\{\mathrm{var}(\boldsymbol{z}^\star)\} < \infty$*.*

Assumption 3 ϵ_i*'s are (uniformly) sub-Gaussian, i.e., there exist some constants* K *and* σ_0^2*, such that*

$$K^2(\mathrm{E}\, e^{\epsilon_i^2/K^2} - 1) \le \sigma_0^2.$$

Assumption 2 implies that $0 < \Lambda_{\min}\{\mathrm{var}(\boldsymbol{z})\} \le \Lambda_{\max}\{\mathrm{var}(\boldsymbol{z})\} < \infty$. The sample variance-covariance matrix of $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$ is denoted as $\boldsymbol{S}_{\boldsymbol{z}}^2$, i.e., $\boldsymbol{S}_{\boldsymbol{z}}^2 = \frac{1}{n}\sum_{i=1}^n \tilde{\boldsymbol{z}}_i \tilde{\boldsymbol{z}}_i^\top$.

Lemma 1 If $\hat{g}_n$ is the minimizer of $L(g)$ and Assumptions 3 and that $\tau_n = o(1)$ are met, then there exists a constant σ not depending on n, such that

$$||\hat{g}_n - g_0||_n \le \sigma$$

almost surely for sufficiently large n. Furthermore, if Assumptions 1 and 2 are satisfied as well, then almost surely

$$||\hat{g}_n||_n \le R$$

for some positive constant R (independent of n) as long as n is sufficiently large.

Proof: Since $\hat{g}_n$ minimizes $L(g)$, then it must satisfy

$$||y - \hat{g}_n||_n^2 + \tau_n^2 J(\hat{g}_n) \le ||y - g_0||_n^2 + \tau_n^2 J(g_0);$$

thus

$$||\hat{g}_n - g_0||_n^2 + \tau_n^2 J(\hat{g}_n) \le 2(\epsilon, \hat{g}_n - g_0)_n + \tau_n^2 J(g_0). \tag{2}$$

From Assumption 3, we have $\mathrm{E}(\epsilon_1^2) < \infty$. Then $\frac{1}{n}\sum_{i=1}^n \epsilon_i^2 = O(1)$ almost surely. By the Cauchy-Schwarz inequality, it follows

$$||\hat{g}_n - g_0||_n^2 \le ||\hat{g}_n - g_0||_n O(1) + o(1).$$

Therefore, there exist positive constants σ and R such that, almost surely, for all large n,

$$||\hat{g}_n - g_0||_n \le \sigma.$$

Additionally, since f_0 is a continuous function defined on $[0,1]^d$ and $\Lambda_{\max}\{\text{var}(\boldsymbol{z})\}$ is finite, we see that $\mathrm{E}\{g_0(\boldsymbol{\zeta},\boldsymbol{z})\}^2 < \infty$. By the strong law of large numbers, we have almost surely, for all large n,

$$||\hat{g}_n||_n \leq ||\hat{g}_n - g_0||_n + ||g_0||_n \leq R.$$

□

Note that we incorporate the estimated intercept $\hat{a}$ in $\hat{g}$. Actually this will not make a difference if we remove $\hat{a}$ from $\hat{g}$ given the fact $|\hat{a}-a_0| = O_P(n^{-\frac{1}{2}})$ as shown above. Denote $B_n(g_0,\sigma) = \{g \in \mathscr{G} : ||g-g_0||_n \leq \sigma\}$. Due to Lemma 1, we restrict our attention to $B_n(g_0,\sigma)$ from now on. It follows that $\sup_{g\in B_n(g_0,\sigma)} ||g||_n \leq R$, with a similar argument to that used in showing Lemma 1. Let $\mathscr{G}^\top$ denote $B_n(g_0,\sigma) \cap \{g \in \mathscr{G} : J(g) \leq C\}$, where C is a positive constant. Correspondingly, let $\mathscr{G}_1^\top = \mathscr{G}^\top \cap \mathscr{G}_1$ and $\mathscr{G}_2^\top = \mathscr{G}^\top \cap \mathscr{G}_2$.

Proposition 1 Under Assumptions 1, 2 and 3, there exist constants T_0 and C_0, both of which are independent of n, such that

$$\mathbf{P}\left\{\sup_{g\in\mathscr{G}^\top} \frac{|\frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_i g(\boldsymbol{\zeta}_i,\boldsymbol{z}_i)|}{||g||_n^{1-\frac{1}{2l}}} \geq T\right\} \leq 2\exp\left(-\frac{T^2}{C_0^2}\right), \tag{3}$$

for all $T \geq T_0$.

To prove Proposition 1, we need the following lemmas.

Lemma 2 Assuming that Assumption 1 and 2 are met, then there exists a positive constant A, which does not depend on n, such that the entropy of $\mathscr{G}^\top$ satisfies

$$H(\delta, \mathscr{G}^\top, ||\cdot||_n) \leq A\delta^{-\frac{1}{l}}, \quad \forall \delta > 0$$

for sufficiently large n.

Proof: First we study the entropy of $\mathscr{G}_1^\top$. As shown in Lemma A.1 in Lin and Zhang (2006),

$$H(\delta, \{g_1 : g_1(\boldsymbol{\zeta}) = \sum_{j=1}^d f_j(\zeta_j), f_j'\text{s satisfy} \sum_{i=1}^n f_j(\zeta_{ij}) = 0, J(g_1) \leq 1\}, ||\cdot||_\infty) \leq A_0 d^{(l+1)/l}\delta^{-\frac{1}{l}},$$

for all $\delta > 0$, $n \geq 1$ and some $A_0 > 0$ not depending on δ, n or d. Therefore it can be claimed that

$$H(\delta, \mathscr{G}_1^\top, ||\cdot||_\infty) \leq A_1\delta^{-\frac{1}{l}} \quad \forall \delta > 0, \tag{4}$$

where A_1 is a positive constant not depending on n or δ.

For any g writing as $g(\boldsymbol{\zeta}, \boldsymbol{z}) = \sum_{j=1}^{d} f_j(\zeta_j) + \boldsymbol{\alpha}^\top \tilde{\boldsymbol{z}} \in \mathscr{G}^\top$, where $g_1(\boldsymbol{\zeta}, \boldsymbol{z}) = \sum_{j=1}^{d} f_j(\zeta_j)$ satisfies $J(g_1) \leq C$, then we have $||g - g_0||_n \leq \sigma$ and $\sum_{j=1}^{d} ||f_j - \tilde{f}_{0j}||_\infty \leq 2dC$, based on Lemma A.1 in Lin and Zhang (2006). $\boldsymbol{\alpha}^\top \tilde{\boldsymbol{z}}$ therefore satisfies, for all large n,

$$\begin{aligned} ||\boldsymbol{\alpha}^\top \tilde{\boldsymbol{z}} - \boldsymbol{\alpha}_0^\top \tilde{\boldsymbol{z}}||_n &= \left\| \left\{ g(\boldsymbol{\zeta}, \boldsymbol{z}) - \sum_{j=1}^{d} f_j \right\} - \left\{ g_0(\boldsymbol{\zeta}, \boldsymbol{z}) - \sum_{j=1}^{d} \tilde{f}_{0j} \right\} \right\|_n \\ &\leq ||g - g_0||_n + \sum_{j=1}^{d} ||f_j - \tilde{f}_{0j}||_n \\ &\leq 2\sigma + 2dC. \end{aligned} \tag{5}$$

As a result, for any $q(\boldsymbol{\zeta}, \boldsymbol{z}) = \boldsymbol{\alpha}^\top \tilde{\boldsymbol{z}} \in \mathscr{G}_2^\top$, $||q||_n \leq M$ holds for some constant M and sufficiently large n, based on the triangular inequality and the fact that $||\boldsymbol{\alpha}_0 \tilde{\boldsymbol{z}}||_n$ is finite for sufficiently large n. It is from the fact that for sufficiently large n, $||\boldsymbol{\alpha}_0^\top \tilde{\boldsymbol{z}}||_n^2 \leq (\Lambda_{\max}(\text{var}\,(\boldsymbol{z})) + \epsilon)||\boldsymbol{\alpha}_0||_E^2$ holds almost surely for any given $\epsilon > 0$ and $\Lambda_{\max}(\text{var}\,(\boldsymbol{z}))$ is finite if Assumption 2 is met. Additionally, $||\boldsymbol{\alpha}^\top \tilde{\boldsymbol{z}} - \boldsymbol{\alpha}_0^\top \tilde{\boldsymbol{z}}||_n^2 > (\Lambda_{\min}(\text{var}\,(\boldsymbol{z})) - \epsilon)||\boldsymbol{\alpha} - \boldsymbol{\alpha}_0||_E^2$ holds almost surely for any given $\epsilon > 0$. Therefore, $||\boldsymbol{\alpha} - \boldsymbol{\alpha}_0||_E \leq C_a$ for some constant C_a and any $g_2(\boldsymbol{\zeta}, \boldsymbol{z}) = \boldsymbol{\alpha}^\top \tilde{\boldsymbol{z}} \in \mathscr{G}_2^\top$. It follows that $H(\delta, \mathscr{G}_2^\top, ||\cdot||_n) \leq A_2 \log\left(\frac{1}{\delta}\right)$ almost surely for some constant A_2 not dependent on n, when n is sufficiently large.

As pointed out earlier, every element in $\mathscr{G}^\top$ can be written as sum of two elements from $\mathscr{G}_1^\top$ and $\mathscr{G}_2^\top$, respectively. Consequently, $H(\delta, \mathscr{G}^\top, ||\cdot||_n) \leq H(\delta/2, \mathscr{G}_1^\top, ||\cdot||_n) + H(\delta/2, \mathscr{G}_2^\top, ||\cdot||_n) \leq A\delta^{-\frac{1}{l}}$, for sufficiently large n and some positive A, which is independent of n. □

Lemma 3 Assume that Assumption 3 is met. Then for all $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_n)^\top \in \mathbb{R}^n$ and $a > 0$,

$$\mathbf{P}\left(\left| \sum_{i=1}^{n} \epsilon_i \gamma_i \right| \geq a \right) \leq 2 \exp\left\{ -\frac{a^2}{8(K^2 + \sigma_0^2) \sum_{i=1}^{n} \gamma_i^2} \right\}.$$

Proof: See the proof of Lemma 8.2 of van de Geer (2000). □

Lemma 4 Assuming that Assumptions 1, 2 and 3 are met, then for some constant B depending only on K and σ_0, and for any $\delta > 0$, we have

$$\mathbf{P}\left\{ \sup_{g \in \mathscr{G}^\top} \left| \frac{1}{n} \sum_{i=1}^{n} \epsilon_i g(\boldsymbol{\zeta}_i, \boldsymbol{z}_i) \right| \geq \delta \right\} \leq 2 \exp\left(-\frac{n\delta^2}{B^2 R^2} \right),$$

where $\sup_{g \in B_n(g_0, \sigma)} ||g||_n \leq R$, as long as n is sufficiently large.

Proof: Let for each $i = 0, 1, \ldots, T_i$ be a $2^{-i}R$-covering set of $\mathscr{G}^\top$, i.e., for each $g \in \mathscr{G}^\top$ there is a $g^i \in T_i$ such that $||g - g^i||_n \leq 2^{-i}R,\ i = 0, 1, \ldots$. Without loss of generality, we assume that $T_i \subset \mathscr{G}^\top,\ i = 0, 1, \ldots$. Note that

$$\left|\frac{1}{n}\sum_{i=1}^{n}\epsilon_i\left\{g(\boldsymbol{\zeta}_i, \boldsymbol{z}_i) - g^S(\boldsymbol{\zeta}_i, \boldsymbol{z}_i)\right\}\right| \leq \sqrt{\mathrm{E}\left(\epsilon_i^2\right)}||g - g^S||_n \text{ almost surely}$$

for sufficiently large n, applying the strong law of large numbers. The inequality above implies that, as long as a sufficiently large S is chosen, then almost surely, we have $\left|\frac{1}{n}\sum_{i=1}^{n}\epsilon_i\left\{g(\boldsymbol{\zeta}_i, \boldsymbol{z}_i) - g^S(\boldsymbol{\zeta}_i, \boldsymbol{z}_i)\right\}\right| \leq \delta/2$ for sufficiently large n. Therefore, it suffices to prove an exponential inequality for

$$\mathbf{P}\left\{\sup_{g\in T}\left|\frac{1}{n}\sum_{i=1}^{n}\epsilon_i g(\boldsymbol{\zeta}_i, \boldsymbol{z}_i)\right| \geq \delta/2\right\},$$

where $T = \cup_{i=1}^{\infty}T_i$.

Since $\sup_{g\in B_n(g_0,\sigma)}||g||_n \leq R$, T_0 can be chosen as $\{0\}$. For any $j \in N^+$, $g^j = \sum_{i=1}^{j}(g^i - g^{(i-1)})$. Let $C_2^2 = 8(K^2 + \sigma_0^2)$. Since for any $g \in T$, $\left|\frac{1}{n}\sum_{i=1}^{n}\epsilon_i g(\boldsymbol{\zeta}_i, \boldsymbol{z}_i)\right| \leq \sum_{j=1}^{\infty}\left|\frac{1}{n}\sum_{i=1}^{n}\epsilon_i(g^j - g^{(j-1)})\right|$, we have that for any nonnegative sequence $\{\eta_j\}$ satisfying $\sum_{j=1}^{\infty}\eta_j \leq 1$,

$$\begin{aligned}\mathbf{P} = \mathbf{P}&\left\{\sup_{g\in T}\left|\frac{1}{n}\sum_{i=1}^{n}\epsilon_i g(\boldsymbol{\zeta}_i, \boldsymbol{z}_i)\right| \geq \delta/2\right\}\\
&\leq \sum_{j=1}^{\infty}\mathbf{P}\left\{\sup_{g\in T}\left|\frac{1}{n}\sum_{i=1}^{n}\epsilon_i(g^j - g^{(j-1)})\right| \geq \eta_j\delta/2\right\}\\
&\leq 2\sum_{j=1}^{\infty}\exp\left\{2H(2^{-j}R, \mathscr{G}^\top, ||\cdot||_n) - \frac{n\delta^2\eta_j^2}{36C_2^2 2^{-2j}R^2}\right\}.\end{aligned}$$

The last expression comes from the fact that

$$||g^j - g^{(j-1)}||_n \leq ||g^j - g||_n + ||g - g^{(j-1)}||_n \leq 2^{-j}R + 2^{-j+1}R = 3(2^{-j}R)$$

and Lemma 3.

As shown in Lemma 2, $H(\delta, \mathscr{G}^\top, ||\cdot||_n) \leq A\delta^{-\frac{1}{l}}\ \forall\delta > 0$, for some $A > 0$ independent of n. We have

$$\sqrt{n}\delta \geq 24C_2\left(\int_0^R H^{\frac{1}{2}}(x, \mathscr{G}^\top, ||\cdot||_n)dx \vee R\right),$$

for sufficiently large n. We choose

$$\eta_j = \frac{12C_2 2^{-j}RH^{\frac{1}{2}}(2^{-j}R, \mathscr{G}^\top, ||\cdot||_n)}{\sqrt{n}\delta} \vee \frac{2^{-j}\sqrt{j}}{2E},$$

where $E=\sum_{j=1}^{\infty}2^{-j}\sqrt{j}$. Then

$$\sum_{j=1}^{\infty}\eta_j \le \sum_{j=1}^{\infty}\frac{12C_2 2^{-j}RH^{\frac{1}{2}}(2^{-j}R,\mathscr{G}^\top,||\cdot||_n)}{\sqrt{n}\delta}+\sum_{j=1}^{\infty}\frac{2^{-j}\sqrt{j}}{2E}\le\frac{1}{2}+\frac{1}{2}=1.$$

Note that $\eta_j \ge \frac{12C_2 2^{-j}RH^{1/2}(2^{-j}R,\mathscr{G}^\top,||\cdot||_n)}{\sqrt{n}\delta}$. Plugging this into the expression of $\mathbf{P}$, it follows

$$\begin{aligned}
\mathbf{P} &\le 2\sum_{j=1}^{\infty}\exp\left\{2H(2^{-j}R,\mathscr{G}^\top,||\cdot||_n)-\frac{n\delta^2\eta_j^2}{36C_2^2 2^{-2j}R^2}\right\}\\
&\le 2\sum_{j=1}^{\infty}\exp\left(\frac{n\delta^2\eta_j^2}{72C_2^2 2^{-2j}R^2}-\frac{n\delta^2\eta_j^2}{36C_2^2 2^{-2j}R^2}\right)\\
&= 2\sum_{j=1}^{\infty}\exp\left(-\frac{n\delta^2\eta_j^2}{72C_2^2 2^{-2j}R^2}\right)\\
&\le 2\sum_{j=1}^{\infty}\exp\left(-\frac{n\delta^2}{72C_2^2 2^{-2j}R^2}\frac{2^{-2j}j}{4E^2}\right)\ \text{(since } \eta_j \ge \frac{2^{-j}\sqrt{j}}{2E}\text{)}\\
&= 2\sum_{j=1}^{\infty}\exp\left(-\frac{n\delta^2 j}{288C_2^2E^2R^2}\right)\\
&\le 2\exp\left(-\frac{n\delta^2}{B^2R^2}\right)\ \text{ for some } B>0.
\end{aligned}$$

□

Proof of Proposition 1: T_0 is defined as $\sup\left\{(2^{-j}R)^{\frac{1}{2l}-1}\times 24C_2\left(\frac{2l}{2l-1}A(2^{-j+1}R)^{\frac{2l-1}{2l}}\vee 2^{-j+1}R\right),\ j=1,2\ldots,\right\}$. Then for $T\ge T_0$,

$$\begin{aligned}
&\mathbf{P}\left\{\sup_{g\in\mathscr{G}^\top}\frac{|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\epsilon_i g(\boldsymbol{\zeta}_i,\boldsymbol{z}_i)|}{||g||_n^{1-\frac{1}{2l}}}\ge T\right\}\\
&\le \sum_{j=1}^{\infty}\mathbf{P}\left\{\sup_{g\in\mathscr{G}^\top,2^{-j}R<||g||_n\le 2^{-j+1}R}|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\epsilon_i g(\boldsymbol{\zeta}_i,\boldsymbol{z}_i)|\ge T(2^{-j}R)^{1-\frac{1}{2l}}\right\}\\
&\le 2\sum_{j=1}^{\infty}\exp\left\{-\frac{T^2(2^{-j}R)^{2-\frac{1}{l}}}{B^2R^2}\right\}\ \text{(using Lemma 4)}\\
&\le 2\exp\left(-\frac{T^2}{C_0^2}\right),
\end{aligned}$$

for some constant $C_0>0$. □

Proof of Theorem 1: By Lemma 2, for sufficiently large n, we have

$$H\left(\delta,\left\{\frac{g-g_0}{J(g_0)+J(g)}:g\in B_n(g_0,\sigma)\right\},||\cdot||_n\right)<A^\top\delta^{-\frac{1}{l}},\quad \forall\delta>0$$

for some constant $A^\top$ not depending on n. Now we can apply Proposition 1 to the class $\left\{\frac{g-g_0}{J(g_0)+J(g)} : g \in B_n(g_0,\sigma)\right\}$. Consequently,

$$\frac{(\epsilon, \hat{g}_n - g_0)_n}{||\hat{g}_n - g_0||_n^{1-\frac{1}{2l}}(J(g_0)+J(\hat{g}_n))^{\frac{1}{2l}}} = O_P(n^{-\frac{1}{2}}). \tag{6}$$

Incorporating (6) in (2), we have

$$||\hat{g}_n - g_0||_n^2 + \tau_n^2 J(\hat{g}_n) \le O_P(n^{-\frac{1}{2}})||\hat{g}_n - g_0||_n^{1-\frac{1}{2l}}\left\{J(g_0)+J(\hat{g}_n)\right\}^{\frac{1}{2l}} + \tau_n^2 J(g_0).$$

If $O_P(n^{-\frac{1}{2}})||\hat{g}_n - g_0||_n^{1-\frac{1}{2l}}\left\{J(g_0)+J(\hat{g}_n)\right\}^{\frac{1}{2l}} < \tau_n^2 J(g_0)$, it follows that

$$||\hat{g}_n - g_0||_n^2 + \tau_n^2 J(\hat{g}_n) \le 2\tau_n^2 J(g_0); \tag{7}$$

otherwise,

$$||\hat{g}_n - g_0||_n^2 + \tau_n^2 J(\hat{g}_n) \le O_P(n^{-\frac{1}{2}})||\hat{g}_n - g_0||_n^{1-\frac{1}{2l}}\left\{J(g_0)+J(\hat{g}_n)\right\}^{\frac{1}{2l}}. \tag{8}$$

Next we will verify the result for separated cases. For the case of inequality (7), it is trivial that

$$||\hat{g}_n - g_0||_n = O_P(\tau_n), \qquad J(\hat{g}_n) = O_P(1)J(g_0). \tag{9}$$

For the case of inequality (8), there are two possibilities.

(i) If $J(\hat{g}_n) \ge J(g_0)$, it follows that $||\hat{g}_n - g_0||_n^2 + \tau_n^2 J(\hat{g}_n) \le O_P(n^{-\frac{1}{2}})||\hat{g}_n - g_0||_n^{1-\frac{1}{2l}}\left\{J(\hat{g}_n)\right\}^{\frac{1}{2l}}$. Then $\left\{J(\hat{g}_n)\right\}^{\frac{1}{2l}} \le O_P(n^{-\frac{1}{4l-2}})||\hat{g}_n - g_0||_n^{\frac{1}{2l}}\tau_n^{-\frac{2}{2l-1}}$. Thus

$$||\hat{g}_n - g_0||_n^2 \le O_P(n^{-\frac{1}{2}})||\hat{g}_n - g_0||_n^{1-\frac{1}{2l}}\left\{J(\hat{g}_n)\right\}^{\frac{1}{2l}} \le O_P(n^{-\frac{l}{2l-1}})||\hat{g}_n - g_0||_n \tau_n^{-\frac{2}{2l-1}}.$$

In other words,

$$||\hat{g}_n - g_0||_n = O_P(n^{-\frac{l}{2l-1}})\tau_n^{-\frac{2}{2l-1}}, \quad J(\hat{g}_n) = O_P(n^{-\frac{2l}{2l-1}})\tau_n^{-\frac{4l+2}{2l-1}}. \tag{10}$$

(ii) If $J(\hat{g}_n) < J(g_0)$, it follows that $||\hat{g}_n - g_0||_n^2 + \tau_n^2 J(\hat{g}_n) \le O_P(n^{-\frac{1}{2}})||\hat{g}_n - g_0||_n^{1-\frac{1}{2l}}\left\{J(g_0)\right\}^{\frac{1}{2l}}$. After some simple algebra, we have

$$||\hat{g}_n - g_0||_n = O_P(n^{-\frac{l}{2l+1}})\left\{J(g_0)\right\}^{\frac{1}{2l+1}}, \quad J(\hat{g}_n) = J(g_0)O_P(1). \tag{11}$$

When $J(g_0) = J(f_0) > 0$ and $\tau_n^{-1} = n^{\frac{l}{2l+1}}\left\{J(f_0)\right\}^{\frac{2l-1}{4l+2}}$, then we obtain the same result from (9), (10) and (11). To be more specific, $||\hat{g}_n - g_0||_n = O_P(n^{-\frac{l}{2l+1}})\left\{J(f_0)\right\}^{\frac{1}{2l+1}}$ and $J(\hat{f}_n) = J(f_0)O_P(1)$. When $J(g_0) = J(f_0) = 0$, then both

$O_P(n^{-\frac{1}{2}})||\hat{g}_n - g_0||_n^{1-\frac{1}{2l}} \{J(g_0) + J(\hat{g}_n)\}^{\frac{1}{2l}} < \tau_n^2 J(g_0)$ and $J(\hat{g}_n) < J(g_0)$ are impossible, which indicate that we only need to consider (10) under this circumstance. When $\tau_n^{-1} = n^{1/4}$, $||\hat{g}_n - g_0||_n = O_P(n^{-\frac{1}{2}})$ and $J(\hat{f}_n) = O_P(n^{-\frac{1}{2}})$. □

In Corollary 1, we only need to show that $\hat{f}_n$ and $\hat{\boldsymbol{\alpha}}$ defined above satisfy Corollary 1 as well since the estimated intercept $\hat{a}$ converges to a_0 with a rate of $O_P(n^{-\frac{1}{2}})$, as indicated at the very beginning. To prove Corollary 1, we need to quantify the ratio of $||\cdot||_n$ and $||\cdot||_{2,Q}$ norm for both $\hat{f}_n$ and $\hat{g}_n$. Entropy with bracketing is an important tool in studying magnitude of the ratio. Let $N_B(\delta, \mathscr{G}, ||\cdot||_{2,Q})$ be the smallest value of N for which there exist pairs of functions $\{[g_j^L, g_j^U]\}_{j=1}^N$ such that $||g_j^U - g_j^L||_{2,Q} \le \delta$ for all $j = 1, \ldots, N$, and such that for each $g \in \mathscr{G}$, there exists $j = j(g) \in \{1, \ldots, N\}$ such that $g_j^L \le g \le g_j^U$. Then $H_B(\delta, \mathscr{G}, ||\cdot||_{2,Q}) = \log N_B(\delta, \mathscr{G}, ||\cdot||_{2,Q})$ is called the δ-entropy with bracketing of $\mathscr{G}$ (for the $L_2(Q)$ metric). Following lemmas are needed to compute the ratio of $||g||_n$ and $||g||_{2,Q}$ for any $g \in \mathscr{G}$.

Lemma 5 For all $\delta > 0$, $H_B(\delta, \mathscr{G}, ||\cdot||_{2,Q}) \le H(\delta/2, \mathscr{G}, ||\cdot||_\infty)$

Proof: See Lemma 2.1 of van de Geer (2000). □

Lemma 6 Let $\mathscr{A}$ denote a collection of functions defined on $\mathscr{X}$. Suppose that $\mathscr{A}$ is uniformly bounded, i.e., $\sup_{a \in \mathscr{A}} ||a||_\infty \le M$ for some constant M, and that for some $0 < \nu < 2$, $\sup_{\delta > 0} \delta^\nu H_B(\delta, \mathscr{A}, ||\cdot||_{2,Q}) < \infty$. Then for all $\eta > 0$ there exists a constant C such that

$$\limsup_{n \to \infty} \mathbf{P}\left(\sup_{a \in \mathscr{A}, ||a||_{2,Q} > Cn^{-1/(2+\nu)}} \left| \frac{||a||_n}{||a||_{2,Q}} - 1 \right| > \eta \right) = 0.$$

See Theorem 2.3 of Mammen and van de Geer (1997) or van de Geer (1988), Lemma 6.3.4.

Before proving the corollary, we restate the extra assumption.

Assumption 4 *The support of $\boldsymbol{z}$ is compact in $\mathbb{R}^p$.*

Proof of Corollary 1 We first need to show that $\hat{g}_n$ is bounded. As shown above, $||\frac{\hat{f}_n}{1+J(\hat{f}_n)}||_\infty \le C$ for some constant C and $J(\hat{f}_n) = O_P(1)$ with an suitable τ_n. Therefore, $||\hat{f}_n||_\infty = O_p(1)$. Additionally, provided that Assumption (4) holds, then $||\hat{\boldsymbol{\alpha}}^\top \tilde{\boldsymbol{z}}||_\infty$ is bounded in probability as well, since it has been shown that $\hat{\boldsymbol{\alpha}} = O_P(1)$. Thus $||\hat{g}_n||_\infty \le$

$||\hat{f}_n||_\infty + ||\hat{\boldsymbol{\alpha}}^\top \tilde{\boldsymbol{z}}||_\infty = O_P(1)$ in Lemma 2. We henceforth consider a subset of $\mathscr{G}$, $\{g : g \in \mathscr{G}, ||g||_\infty \leq C, J(g) \leq C\}$, which is still denoted as $\mathscr{G}^\top$. Similarly, let $\mathscr{G}_1^\top$ denote $\{g_1 : g_1(\boldsymbol{\zeta}, \boldsymbol{z}) = \sum_{j=1}^d f_j(\zeta_j), \sum_{i=1}^n f_j(\zeta_{ij}) = 0, j = 1, \ldots, d, ||g_1||_\infty \leq C_f, J(g_1) \leq C\}$, where C_f is a positive constant, and $\mathscr{G}_2^\top$ for $\{g_2 : g_2(\boldsymbol{\zeta}, \boldsymbol{z}) = \boldsymbol{\alpha}^\top \tilde{\boldsymbol{z}}, ||\boldsymbol{\alpha}||_E \leq C_{\boldsymbol{\alpha}}\}$ with $C_{\boldsymbol{\alpha}}$ being a positive constant that does not depend on $\boldsymbol{\alpha}$.

Next, we shall provide a uniform bound for both $||g_1||_n/||g_1||_{2,Q}$, $g_1 \in \mathscr{G}_1^\top$ and $||g||_n/||g||_{2,Q}$, $g \in \mathscr{G}^\top$. For the former one, Lemma 5.6 of van de Geer (2000) is employed. As shown in Lemma 2, $H(\delta, \mathscr{G}_1^\top, ||\cdot||_n) \leq A\delta^{-\frac{1}{l}}$ for some constant A. Take $\delta_n = (2A)^{l/(2l+1)} n^{-l/(2l+1)}$ and $H(\delta) = \delta^{-\frac{1}{l}}$. Then $n\delta_n^2 \to \infty$, and $n\delta_n^2 = 2A\delta_n^{-\frac{1}{l}} = 2AH(\delta_n)$ for all n. Thus we have

$$\limsup_{n\to\infty} \mathbf{P}\left(\sup_{g_1 \in \mathscr{G}_1^\top} \frac{||g_1||_n}{||g_1||_{2,Q} \vee \delta_n} > 14\right) \leq \limsup_{n\to\infty} 4\mathbf{P}\left\{\sup_{u>0} \frac{H(u, \mathscr{G}_1^\top, ||\cdot||_n)}{H(u)} > A\right\} = 0 \quad (12)$$

Inequality (12) implies that with probability arbitrarily close to 1,

$$||\hat{f}_n - \tilde{f}_0||_n^2 \leq \max\left\{196||\hat{f}_n - \tilde{f}_0||_{2,Q}^2, O(n^{-2l/(2l+1)})\right\}, \quad (13)$$

for sufficiently large n. We use Lemma 6 to derive a uniform bound on $||g||_n/||g||_{2,Q}$ for $g \in \mathscr{G}^\top$. Based on Lemma 5 and combining (4), $H_B(\delta, \mathscr{G}_1^\top, ||\cdot||_{2,Q}) \leq H(\delta/2, \mathscr{G}_1^\top, ||\cdot||_\infty)$ $\leq A_1\delta^{-\frac{1}{l}}$ for some constant A_1. Since the support of $\boldsymbol{z}$ is compact, it is straightforward that $H_B(\delta, \mathscr{G}_2^\top, ||\cdot||_{2,Q}) \leq H(\delta/2, \mathscr{G}_2^\top, ||\cdot||_\infty) \leq A_2 \log(1/\delta)$ for some constant A_2. Therefore, $H_B(\delta, \mathscr{G}^\top, ||\cdot||_{2,Q}) \leq A\delta^{-\frac{1}{l}}$ for some constant A. Taking $\mathscr{A} = \mathscr{G}^\top$ and $\nu = \frac{1}{l}$, then the condition $\sup_{\delta>0} \delta^\nu H_B(\delta, \mathscr{A}, ||\cdot||_{2,Q}) < \infty$ is satisfied in Lemma 6. We can derive from Lemma 6, that with probability arbitrarily close to 1,

$$||\hat{g}_n - g_0||_{2,Q}^2 \leq \max(\eta_1||\hat{g}_n - g_0||_n^2, O(n^{-2l/(2l+1)})) = O_P(n^{-2l/(2l+1)}), \quad (14)$$

for some constant η_1 and sufficiently large n.

Note that

$$\begin{aligned}
||\hat{g}_n - g_0||_{2,Q}^2 &= ||\hat{f}_n(\boldsymbol{\zeta}) - \tilde{f}_0(\boldsymbol{\zeta}) + (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)^\top(\boldsymbol{z} - \bar{\boldsymbol{z}})||_{2,Q}^2 \\
&= ||\hat{f}_n(\boldsymbol{\zeta}) - \tilde{f}_0(\boldsymbol{\zeta}) + (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)^\top\{\boldsymbol{z}^\star + h(\boldsymbol{\zeta}) - \bar{\boldsymbol{z}}\}||_{2,Q}^2 \\
&= ||\hat{f}_n(\boldsymbol{\zeta}) - \tilde{f}_0(\boldsymbol{\zeta}) + (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)^\top\{h(\boldsymbol{\zeta}) - \bar{\boldsymbol{z}}\}||_{2,Q}^2 + ||(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)^\top \boldsymbol{z}^\star||_{2,Q}^2 \\
&= O_P(n^{-2l/(2l+1)}). \qquad (15)
\end{aligned}$$

The last equation holds according to (14). Since $||(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)^\top \boldsymbol{z}^\star||_{2,Q}^2 \geq \Lambda_{\min}\{\mathrm{var}(\boldsymbol{z}^\star)\}||\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0||_E^2$, $||\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0||_E = O_P(n^{-l/(2l+1)})$ based on (15) when Assumption (2) is met.

Now we can verify the consistency of $\hat{f}_n$. Take $C^\star = \max_{1 \le j \le p} |z_j|$. Then $C^\star < \infty$ when Assumption (4) is satisfied. Given that $||\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0||_E = O_P(n^{-l/(2l+1)})$ and

$$\begin{aligned} ||\hat{g}_n - g_0||_n^2 &= ||\hat{f}_n - \tilde{f}_0 + (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)^\top \tilde{\boldsymbol{z}}||_n^2 \\ &\ge ||\hat{f}_n - \tilde{f}_0||_n^2 + \frac{2}{n}\sum_{i=1}^{n} \left\{ \hat{f}_n(\boldsymbol{\zeta}_i) - \tilde{f}_0(\boldsymbol{\zeta}_i) \right\} (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)^\top (\boldsymbol{z}_i - \bar{\boldsymbol{z}}) \\ &\ge ||\hat{f}_n - \tilde{f}_0||_n^2 - 4C^\star ||\hat{f}_n - \tilde{f}_0||_n ||\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0||_E \end{aligned}$$

we have

$$||\hat{f}_n - \tilde{f}_0||_n^2 \le 4C^\star ||\hat{f}_n - \tilde{f}_0||_n O_P(n^{-l/(2l+1)}) + ||\hat{g}_n - g_0||_n^2$$

Therefore, in either Case (i), $0 < J(f_0) < \infty$ and $\tau_n^{-1} = n^{\frac{l}{2l+1}} \{J(f_0)\}^{\frac{2l-1}{4l+2}}$, or Case (ii), $J(f_0) = 0$, and $\tau_n^{-1} = n^{1/4}$,

$$||\hat{f}_n - \tilde{f}_0||_n = O_P(n^{-l/(2l+1)})$$

The proof is completed. □

References

Lin, Y. and Zhang, H. H. (2006). Component selection and smoothing in multivariate nonparametric regression. *The Annals of Statistics*, 34(5):2272–2297.

Mammen, E. and van de Geer, S. (1997). Penalized quasi-likelihood estimation in partial linear models. *The Annals of Statistics*, 25(3):1014–1035.

van de Geer, S. A. (1988). Regression analysis and empirical processes. *CWI Tracts*, 45:1–161.

van de Geer, S. A. (2000). *Empirical Processes in M-estimation.* Cambridge Univ. Press, Cambridge.