*Biometrika* (20XX), **XX**, X, *pp.* 1–24
*Printed in Great Britain*

# Supplementary material for "Functional Linear Regression for Discretely Observed Data: from Ideal to Reality"

BY HANG ZHOU, FANG YAO
*Department of Probability and Statistics, School of Mathematical Sciences, Center for Statistical Science, Peking University, Beijing 100871, China*
h_zhou@pku.edu.cn fyao@math.pku.edu.cn

AND HUIMING ZHANG
*Department of Mathematics, University of Macau, Macau 999078, China*
huimingzhang@um.edu.mo

SUMMARY

Section S.1 contains auxiliary lemmas which serve as building blocks for establishing the main theorems, and their proofs are collected together in Section S.3. Section S.2 provides proofs to the main theorems.

## S.1. TECHNICAL LEMMAS

In this section, we present some useful lemmas. It is necessary to introduce the following matrices and vectors for notational convenience. Define, for $m \in \mathbb{N}_+$,

- $D = \text{diag}\{\lambda_1^{1/2}, \ldots, \lambda_m^{1/2}\}$, $\theta_i = \langle X_i, \beta \rangle$, $\xi_i = (\xi_{i1}, \ldots, \xi_{im})^{\mathrm{T}}$ and $\eta_i = D^{-1}\xi_i$;
- $b_0 = (b_{01}, \ldots, b_{0m})^{\mathrm{T}}$, $\theta_{i,m} = \xi_i^{\mathrm{T}} b_0$ and $b_{r0} = Db_0$;
- $\hat{\xi}_i = (\hat{\xi}_{i1}, \ldots, \hat{\xi}_{im})^{\mathrm{T}}$, $\hat{\theta}_{i,m} = \hat{\xi}_i^{\mathrm{T}} b_0$ and $\hat{\eta}_i = D^{-1}\hat{\xi}_i$.

In the sequel, we write $\int pq$ and $\int Apq$ for $\int p(u)q(u)\mathrm{d}u$ and $\int A(u,v)p(u)q(v)\mathrm{d}u\mathrm{d}v$. The following lemma gives the moment bounds of $\|\hat{\eta}_i - \eta_i\|^2$ and $(\theta_i - \hat{\theta}_{i,m})^2$.

LEMMA S1. *Under Conditions 1–3 and 5, for each $1 \le i \le n$, on the high probability set $\Omega_m(n, N)$,*

$$E\|\hat{\eta}_i - \eta_i\|^2 \lesssim \frac{m^{2a+3}}{n}\left(1 + \frac{1}{Nh}\right) + h^4 m^{3a+2c+3} + \frac{m^{a+1}}{N}$$

*and*

$$E(\theta_i - \hat{\theta}_{i,m})^2 \lesssim \frac{1}{N} + \frac{1}{n}\left(1 + \frac{1}{Nh}\right).$$

The following lemma shows that the second order derivative of the likelihood function, also named Hessian matrix, is consistent.

LEMMA S2. *Under Conditions 1–3 and 5, on the high probability set $\Omega_m(n, N)$,*

$$\left\| \frac{1}{n}\sum_{i=1}^{n} \hat{\eta}_i \hat{\eta}_i^{\mathrm{T}} - \frac{1}{n}\sum_{i=1}^{n} \eta_i \eta_i^{\mathrm{T}} \right\| = o_p(1).$$

Due to the fact that the variance of $\xi_{ik}$ tends to zero as $k \to \infty$, we need to do re-parametrization such that the principal scores serve as predictor variables on a common scale of variabilities. Define

$$L_n(b_r) = \frac{1}{n}\left\{\sum_{i=1}^{n}(\hat{\eta}_i^{\mathrm{T}} b_r)Y_i - \frac{(\hat{\eta}_i^{\mathrm{T}} b_r)^2}{2}\right\}, \quad \hat{b}_r = \arg\max_{b_r \in \mathbb{R}^m} L_n(b_r).$$

Then $\hat{b} = D^{-1}\hat{b}_r$ by definition. The following result characterizes the discrepancy between $\hat{b}_r$ and $b_{r0}$, which is a key building block in the proofs of Theorem 3 and 4.

LEMMA S3. *Under conditions 1-5, on the high probability set $\Omega_m(n, N)$,*

$$\|\hat{b}_r - b_{r0}\|^2 = O_p(\alpha_n),$$

*where*

$$\alpha_n = \frac{m}{n} + \left\{\frac{1}{N} + \frac{1}{n}\left(1 + \frac{1}{Nh}\right)\right\}\left\{\frac{m^{2a+3}}{n}\left(1 + \frac{1}{Nh}\right) + h^4 m^{3a+2c+3} + \frac{m^{a+1}}{N}\right\}.$$

The following lemma is to establish the minimax lower bound for the prediction error. We define that $P_\theta$ is a family of probability measures, where $\theta$ is the parameter of interest and the corresponding expectation operator is denoted as $E_\theta$. Let $H(\theta, \theta') = \sum_{i=1}^{r}|\theta_i - \theta_i'|$ be the Hamming distance between the binary sequences $\theta = (\theta_1, \ldots, \theta_r)^{\mathrm{T}}$ and $\theta' = (\theta_1', \ldots, \theta_r')^{\mathrm{T}}$ on $\{0,1\}^r = \{(\theta_1, \ldots, \theta_j, \ldots, \theta_r)^{\mathrm{T}} \mid \theta_j = 0 \text{ or } \theta_j = 1\}$. In addition, for the probability measures $P_\theta$ and $P_{\theta'}$ with density function $p_\theta$ and $p_{\theta'}$ jointly dominated by $\mu$, the integration of their minimal $\int(p_\theta \wedge p_{\theta'})\mathrm{d}\mu$ is denoted as $\|P_\theta \wedge P_{\theta'}\|_a$.

LEMMA S4 (ASSOUAD'S LEMMA). *The estimator $T$ is based on observations of the statistical model $P_\theta$, $\theta \in \{0,1\}^r$. Let $\psi(\theta)$ be an arbitrary transform of the parameter $\theta$. Consider the pseudo-distance $d(\cdot,\cdot)$ satisfying weak triangle inequality $d(x,z) + d(z,y) \geq Ad(x,y)$ with $A \in (0,1)$ and $d(x,y) = \sum_{j=1}^{r} d_j(x,y)$. If $d_j(\psi(\theta), \psi(\theta')) \geq g_{jr} > 0$ for $H(\theta,\theta') = 1$ such that $\theta$ and $\theta'$ differ only in the $j$th coordinate, then for the distance $d(T, \psi(\theta))$,*

$$\max_\theta E_\theta\{d(T, \psi(\theta))\} \geq \frac{g_r A}{2} \min_{H(\theta,\theta')=1} \|P_\theta \wedge P_{\theta'}\|_a, \text{ where } g_r = \sum_{j=1}^{r} g_{jr}.$$

Lemma S4 provides a powerful lower bound for the maximum risk over the discrete parameter set $\{0,1\}^r$, it can be adaptively applied to any parameter $\psi(\theta)$ endowed with the pseudo-distances $d$ ; see Lemma 2 in Yu (1997) and Lemma 2.12 in Tsybakov (2008) for details.

## S.2. PROOFS OF THEOREMS

### *S.2.1. Proof of Theorem 1*

*Proof.* We prove the first assertion of Theorem 1 by evaluating $E(\langle \Delta_{(r)}\phi_k, \phi_j\rangle^2)$. By symmetry, the choice of $r$ does not influence the convergence rate and we assume $r = 1$ in the sequel. By the definition of $\hat{C}_{(1)}$, recall that $\delta_{il_1l_2} = X_{il_1}X_{il_2}$,

$$\langle \Delta_{(1)}\phi_k, \phi_j\rangle = \frac{1}{\lfloor n/2\rfloor}\sum_{i=1}^{\lfloor n/2\rfloor}\frac{1}{N(N-1)}\frac{1}{h^2}\sum_{l_1 \neq l_2}\delta_{il_1l_2} \times \int \mathrm{K}\left(\frac{T_{il_1} - s}{h}\right)\phi_k(s)\mathrm{d}s \int \mathrm{K}\left(\frac{T_{il_2} - t}{h}\right)\phi_j(t)\mathrm{d}t.$$

We first calculate the bias term,

$$
\begin{aligned}
&E(\langle \Delta_{(1)}\phi_k, \phi_j \rangle) \\
=&E\left\{\int X_i(u)\frac{1}{h}\int \mathrm{K}\left(\frac{u-s}{h}\right)\phi_k(s)\mathrm{d}s\mathrm{d}u \int X_i(v)\frac{1}{h}\int \mathrm{K}\left(\frac{v-t}{h}\right)\phi_j(t)\mathrm{d}t\mathrm{d}v\right\} \\
=&\int C(u,v)\frac{1}{h}\int \mathrm{K}\left(\frac{u-s}{h}\right)\phi_k(s)\mathrm{d}s\frac{1}{h}\int \mathrm{K}\left(\frac{v-t}{h}\right)\phi_j(t)\mathrm{d}t\mathrm{d}u\mathrm{d}v \\
=&\int C(u,v)\left\{\frac{1}{h}\int \mathrm{K}\left(\frac{u-s}{h}\right)\phi_k(s)\mathrm{d}s - \phi_k(u)\right\} \\
&\qquad\qquad \times\left\{\frac{1}{h}\int \mathrm{K}\left(\frac{v-t}{h}\right)\phi_j(t)\mathrm{d}t - \phi_j(v)\right\}\mathrm{d}u\mathrm{d}v \\
&+\int C(u,v)\left\{\frac{1}{h}\int \mathrm{K}\left(\frac{u-s}{h}\right)\phi_k(s)\mathrm{d}s - \phi_k(u)\right\}\phi_j(v)\mathrm{d}u\mathrm{d}v \\
&+\int C(u,v)\left\{\frac{1}{h}\int \mathrm{K}\left(\frac{v-t}{h}\right)\phi_j(t)\mathrm{d}t - \phi_j(v)\right\}\phi_k(u)\mathrm{d}u\mathrm{d}v.
\end{aligned}
\tag{S1}
$$

In order to bound each term in the right hand side of (S1), by Taylor expansion and Condition 3,

$$
\begin{aligned}
&\left\|\frac{1}{h}\int \mathrm{K}\left(\frac{v-t}{h}\right)\phi_j(t)\mathrm{d}t - \phi_j(v)\right\|^2 = \int_0^1\left\{\frac{1}{h}\int \mathrm{K}\left(\frac{v-t}{h}\right)\phi_j(t)\mathrm{d}t - \phi_j(v)\right\}^2\mathrm{d}v \\
=&\int_0^1\left[\int_{-1}^1 \mathrm{K}(u)\left\{\phi_j(v) - hu\phi_j^{(1)}(v) + \frac{h^2u^2}{2}\phi_j^{(2)}(v^*)\right\}\mathrm{d}u - \phi_j(v)\right]^2\mathrm{d}v \\
\lesssim& h^4\|\phi_j^{(2)}\|_\infty^2 \lesssim h^4 j^{2c}.
\end{aligned}
\tag{S2}
$$

Then the first term in the right hand side of (S1) is bounded by

$$
\lambda_1\left\|\frac{1}{h}\int \mathrm{K}\left(\frac{v-t}{h}\right)\phi_j(t)\mathrm{d}t - \phi_j(v)\right\|\left\|\frac{1}{h}\int \mathrm{K}\left(\frac{v-t}{h}\right)\phi_k(t)\mathrm{d}t - \phi_k(v)\right\| \lesssim h^4 j^c k^c. \tag{S3}
$$

For the last two terms in the right hand side of (S1),

$$
\begin{aligned}
&\int C(u,v)\left\{\frac{1}{h}\int \mathrm{K}\left(\frac{u-s}{h}\right)\phi_k(s)\mathrm{d}s - \phi_k(u)\right\}\phi_j(v)\mathrm{d}u\mathrm{d}v \\
\le&\lambda_j\left\|\frac{1}{h}\int \mathrm{K}\left(\frac{v-t}{h}\right)\phi_k(t)\mathrm{d}t - \phi_k(v)\right\| \lesssim h^2 j^{-a}k^c.
\end{aligned}
\tag{S4}
$$

Similarly, the last term in (S1) is bounded by $h^2k^{-a}j^c$. Combing equation (S1), (S3) and (S4), under Condition 1–3 and 5, there is $E(\langle \Delta_{(1)}\phi_k, \phi_j\rangle) \lesssim h^2(j^{-a}k^c + k^{-a}j^c)$.

For the variance term, denote

$$
A_i(\phi_k, \phi_j) = \sum_{l_1\neq l_2}\delta_{il_1l_2}\frac{1}{h}\int \mathrm{K}\left(\frac{T_{il_1}-s}{h}\right)\phi_k(s)\mathrm{d}s\frac{1}{h}\int \mathrm{K}\left(\frac{T_{il_2}-t}{h}\right)\phi_j(t)\mathrm{d}t
$$

and there is

$$
\mathrm{var}(\langle \Delta_{(1)}\phi_k, \phi_j\rangle) \le \frac{1}{n}\frac{1}{N^2(N-1)^2}E\{A_i(\phi_k,\phi_j)\}^2.
$$

Denote $\phi_{j,h}(s) = h^{-1}\int \mathrm{K}\{(u-s)/h\}\phi_j(u)\mathrm{d}u$ and $\phi_{k,h}(s) = h^{-1}\int \mathrm{K}\{(u-s)/h\}\phi_k(u)\mathrm{d}u$, the second order moment of $A_i(\phi_k,\phi_j)$ can be decomposed as

$$E\{A_i^2(\phi_j,\phi_k)\} = 4!\binom{N}{4}A_{i1}(\phi_j,\phi_k) + 3!\binom{N}{3}A_{i2}(\phi_j,\phi_k) + 2!\binom{N}{2}A_{i3}(\phi_j,\phi_k)$$

with

$$A_{i1}(\phi_j,\phi_k) = E\left[\left\{\int X(u)\phi_{k,h}(u)\mathrm{d}u\right\}^2\left\{\int X(u)\phi_{j,h}(u)\mathrm{d}u\right\}^2\right]$$

$$\begin{aligned}
A_{i2}(\phi_j,\phi_k) =& 2E\left[\left\{\int X(s)\phi_{k,h}(s)\mathrm{d}s\right\}\left\{\int X(s)\phi_{j,h}(s)\mathrm{d}s\right\}\left\{\int X^2(s)\phi_{k,h}(s)\phi_{j,h}(s)\mathrm{d}s\right\}\right]\\
&+E\left(\left\{\int X(s)\phi_{k,h}(s)\mathrm{d}s\right\}^2\left[\int\left\{X^2(s)+\sigma_X^2\right\}\phi_{j,h}^2(s)\mathrm{d}s\right]\right)\\
&+E\left(\left\{\int X(s)\phi_{j,h}(s)\mathrm{d}s\right\}^2\left[\int\left\{X^2(s)+\sigma_X^2\right\}\phi_{k,h}^2(s)\mathrm{d}s\right]\right)\\
=& A_{i21}(\phi_j,\phi_k) + A_{i22}(\phi_j,\phi_k) + A_{i23}(\phi_j,\phi_k),\\
A_{i3}(\phi_j,\phi_k) =& E\left(\left[\int\left\{X^2(u)+\sigma_X^2\right\}\phi_{k,h}^2(u)\mathrm{d}u\right]\left[\int\left\{X^2(u)+\sigma_X^2\right\}\phi_{j,h}^2(u)\mathrm{d}u\right]\right)\\
&+E\left(\left[\int\left\{X^2(u)+\sigma_X^2\right\}\phi_{k,h}(u)\phi_{j,h}(u)\mathrm{d}u\right]^2\right)\\
=& A_{i31}(\phi_j,\phi_k) + A_{i32}(\phi_j,\phi_k)
\end{aligned}$$

It can be checked that $A_{i21} \le A_{i22} + A_{i23}$ and $A_{i32} \le A_{i31}$. In summary,

$$\mathrm{var}(\langle\Delta_{(1)}\phi_k,\phi_j\rangle) \lesssim \frac{1}{n}\left(A_{i1} + \frac{A_{i22}+A_{i23}}{N} + \frac{A_{i31}}{N^2}\right). \tag{S5}$$

Under Condition 1–3 and 5, we can obtain $\|\phi_{k,h}\| = O(1)$ and $E(\langle X,\phi_{k,h}\rangle^4) \lesssim k^{-2a}$ for each $k \le m$. Thus

$$\begin{aligned}
A_{i1}(\phi_j,\phi_k) =& E(\langle X,\phi_{k,h}\rangle^2\langle X,\phi_{j,h}\rangle^2) \le \{E(\langle X,\phi_{k,h}\rangle^4)E(\langle X,\phi_{j,h}\rangle^4)\}^{1/2} \lesssim j^{-a}k^{-a};\\
A_{i2}(\phi_j,\phi_k) \le& 2E\{\langle X,\phi_{j,h}\rangle^2(\|X\phi_{k,h}\|^2 + \sigma_X^2\|\phi_{k,h}\|^2)\}\\
&+2E\{\langle X,\phi_{k,h}\rangle^2(\|X\phi_{j,h}\|^2 + \sigma_X^2\|\phi_{j,h}\|^2)\} \lesssim \lambda_j + \lambda_k;\\
A_{i3}(\phi_j,\phi_k) \le& 2E\{(\|X\phi_{j,h}\|^2 + \sigma_X^2\|\phi_{j,h}\|^2)(\|X\phi_{k,h}\|^2 + \sigma_X^2\|\phi_{k,h}\|^2)\} \lesssim 1.
\end{aligned} \tag{S6}$$

Then the first statement of Theorem 1 comes from combing equation (S4)-(S6) under Codition 5.

For $E(\|\Delta_{(1)}\|_j^2)$, by similar arguments and the definition of $\|\cdot\|_j$,

$$\int \hat{C}_{(1)}(s,t)\phi_j(t)\mathrm{d}t = \frac{1}{\lfloor n/2\rfloor}\sum_{i=1}^{\lfloor n/2\rfloor}\frac{1}{N(N-1)}\frac{1}{h}\sum_{l_1\ne l_2}\delta_{il_1l_2}\mathrm{K}\left(\frac{T_{il_1}-s}{h}\right)\phi_{j,h}(T_{il_2})$$

and $\|\Delta_{(1)}\|_j^2$ can be decomposed to the bias and variance term analogously. For the bias term, by Cauchy–Schwarz inequality and equation (S2),

$$
\begin{aligned}
&\int \left[ E\left\{ \int \hat{C}_{(1)}(s,t)\phi_j(t)\mathrm{d}t \right\} - \int C(s,t)\phi_j(t)\mathrm{d}t \right]^2 \mathrm{d}s \\
=&\int \left\{ \int C_h(s,t)\phi_{j,h}(t)\mathrm{d}t - \int C(s,t)\phi_j(t)\mathrm{d}t \right\}^2 \mathrm{d}s \\
\lesssim&\int \left[ \int C_h(s,t)\{\phi_{j,h}(t)-\phi_j(t)\}\mathrm{d}t \right]^2 \mathrm{d}s + \int \left[ \int \{C_h(s,t)-C(s,t)\}\phi_j(t)\mathrm{d}t \right]^2 \mathrm{d}s \\
\leq&\|\phi_{j,h}-\phi_j\|^2 \int \|C_h(s,\cdot)\|^2 \mathrm{d}s + \lambda_j^2 \|\phi_{j,h}-\phi_j\|^2 \\
\lesssim& h^4 j^{2c},
\end{aligned} \tag{S7}
$$

where $C_h(s,t) = h^{-1}\int \mathrm{K}\{(u-s)/h\}C(u,t)\mathrm{d}u$. By similar arguments of (S5) and (S6), the variance term can be bounded by

$$
\begin{aligned}
&\int \mathrm{var}\left\{ \int \hat{C}_{(1)}(s,t)\phi_j(t)\mathrm{d}t \right\} \mathrm{d}s \\
\leq&\frac{1}{n}\frac{1}{N^2(N-1)^2}\frac{1}{h^2}\int E\left[ \left\{ \sum_{l_1\neq l_2} \delta_{il_1l_2}\mathrm{K}\left(\frac{T_{il_1}-s}{h}\right)\phi_{j,h}(T_{il_2}) \right\}^2 \right] \mathrm{d}s \\
\lesssim&\frac{1}{n}\int E\left\{ X_h^2(s)\langle \phi_{j,h}, X\rangle^2 \right\} \mathrm{d}s + \frac{1}{nN^2} \\
&+\frac{1}{nNh}\int E\left[ \left\{ \int \mathrm{K}\left(\frac{u-s}{h}\right) X_i^2(u)\mathrm{d}u + \sigma_X^2 \right\} \langle \phi_{j,h}, X\rangle^2 \right] \mathrm{d}s \\
\lesssim&\frac{j^{-a}}{n}\left(1+\frac{1}{Nh}\right).
\end{aligned} \tag{S8}
$$

where $X_h(s) = h^{-1}\int \mathrm{K}\{(u-s)/h\}X(u)\mathrm{d}u$. Then the second assertion follows from equation (S7)-(S8), Condition 1–3 and 5. □

### S.2.2. *Proof of Theorem 2*

*Proof.* For the first assertion $\mathrm{pr}\{\Omega_m(n,N)\} \to 1$, by the derivation of Theorem 1 in Hall & Horowitz (2007), it is sufficient to show that $\eta_m^{-2}\|\Delta_{(1)}\|_{\mathrm{HS}}^2 \to 0$ as $n \to \infty$. Theorem 4.2 in Zhang & Wang (2016) implies $\|\Delta_{(1)}\|_{\mathrm{HS}}^2 = O_p\left(n^{-1}\left\{1 + Nh^{-1} + (N^2h^2)^{-1}\right\} + h^4\right)$. Then,

$$
\begin{aligned}
\frac{m^{2a+2}}{n} < \frac{n^{2a+2/2a+4}}{n} &\to 0 \text{ By (i) in Condition 5;} \\
m^{2a+2}h^4 < n^{\frac{2a+2}{2a+4}} n^{-\frac{3a+2c+4}{2a+4}} &\to 0 \text{ By (iii) in Condition 5;} \\
\frac{m^{2a+2}}{nN^2h^2} &\to 0 \text{ By (ii) in Condition 5.}
\end{aligned}
$$

By the proof of Theorem 5.1.8 in Hsing & Eubank (2015), for each $j \leq m$,

$$\begin{aligned}\hat{\phi}_{(1),j} - \phi_j =& \sum_{k \neq j} \frac{\int (\hat{C}_{(1)} - C)\phi_j \phi_k}{(\lambda_j - \lambda_k)} \phi_k + \sum_{k \neq j} \frac{\int (\hat{C}_{(1)} - C)(\hat{\phi}_{(1),j} - \phi_j)\phi_k}{(\lambda_j - \lambda_k)} \phi_k \\ &+ \sum_{k \neq j} \sum_{s=1}^{\infty} \frac{(\lambda_j - \hat{\lambda}_{(1),j})^s}{(\lambda_j - \lambda_k)^{s+1}} \left\{ \int (\hat{C}_{(1)} - C)\hat{\phi}_{(1),j}\phi_k \right\} \phi_k \\ &+ \left\{ \int (\hat{\phi}_{(1),j} - \phi_j)\phi_j \right\} \phi_j,\end{aligned} \tag{S9}$$

such kind of expansions can be found in Li & Hsing (2010) and Hall & Hosseini-Nasab (2006). The bound for $\|\hat{\phi}_{(1),j} - \phi_j\|^2$ can be derived by bounding each terms on the right hand side of (S9).

For the first term in (S9), by Parseval's identity and the definition of $\eta_j$ and $\|\cdot\|_j^2$,

$$\sum_{k \neq j} (\lambda_j - \lambda_k)^{-2} \left\{ \int (\hat{C}_{(1)} - C)\phi_j \phi_k \right\}^2 \leq \eta_j^{-2} \|\Delta_{(1)}\|_j^2. \tag{S10}$$

Combing the second assertion of Theorem 1 and (S10),

$$E \left\{ \left\| \sum_{k \neq j} \frac{\int (\hat{C}_{(1)} - C)\phi_j \phi_k}{(\lambda_j - \lambda_k)} \phi_k \right\|^2 \right\} \lesssim \frac{j^{a+2}}{n} \left( 1 + \frac{1}{Nh} \right) + h^4 j^{2a+2c+2}. \tag{S11}$$

Next, we will show that the remaining terms in (S9) are dominated by (S11). From Bessel's inequality,

$$\begin{aligned}E \left\{ \left\| \sum_{k \neq j} \frac{\int (\hat{C}_{(1)} - C)(\hat{\phi}_{(1),j} - \phi_j)\phi_k}{(\lambda_j - \lambda_k)} \phi_k \right\|^2 \right\} \leq& E \left\{ \frac{\|\hat{C}_{(1)} - C\|^2 \|\hat{\phi}_{(1),j} - \phi_j\|^2}{(2\eta_j)^2} \right\} \\ <& \frac{1}{16} E(\|\hat{\phi}_{(1),j} - \phi_j\|^2),\end{aligned} \tag{S12}$$

where the last inequality comes from the fact $\eta_j^{-1}\|\hat{C}_{(1)} - C\| < 1/2$ on $\Omega_m(n, N)$. Similarly, on the high probability set $\Omega_m(n, N)$,

$$
\begin{aligned}
&E\left[\left\|\sum_{k\neq j}\sum_{s=1}^{\infty}\frac{(\lambda_j-\hat{\lambda}_{(1),j})^s}{(\lambda_j-\lambda_k)^{s+1}}\left\{\int(\hat{C}_{(1)}-C)\hat{\phi}_{(1),j}\phi_k\right\}\phi_k\right\|^2\right]\\
=&E\left[\sum_{k\neq j}\frac{(\lambda_j-\hat{\lambda}_{(1),j})^2}{(\lambda_j-\lambda_k)^2(\hat{\lambda}_{(1),j}-\lambda_k)^2}\left\{\int(\hat{C}_{(1)}-C)\hat{\phi}_{(1),j}\phi_k\right\}^2\right]\\
\leq&2E\left\{\frac{\|\hat{C}_{(1)}-C\|^2}{(2\eta_j-\|\hat{C}_{(1)}-C\|)^2}\left[\sum_{k\neq j}\frac{\{\int(\hat{C}_{(1)}-C)\phi_j\phi_k\}^2}{(\lambda_j-\lambda_k)^2}\right.\right.\\
&\qquad\left.\left.+\sum_{k\neq j}\frac{\{\int(\hat{C}_{(1)}-C)(\hat{\phi}_{(1),j}-\phi_j)\phi_k\}^2}{(\lambda_j-\lambda_k)^2}\right]\right\}\\
\leq&\frac{8}{9}E\left[\frac{\|\hat{C}_{(1)}-C\|^2}{\eta_j^2}\sum_{k\neq j}\frac{\{\int(\hat{C}_{(1)}-C)\phi_j\phi_k\}^2}{(\lambda_j-\lambda_k)^2}+\frac{\|\hat{C}_{(1)}-C\|^4}{\eta_j^4}\|\hat{\phi}_{(1),j}-\phi_j\|^2\right]\\
\leq&\frac{2}{9}E\left[\sum_{k\neq j}\frac{\{\int(\hat{C}_{(1)}-C)\phi_j\phi_k\}^2}{(\lambda_j-\lambda_k)^2}\right]+\frac{1}{18}E(\|\hat{\phi}_{(1),j}-\phi_j\|^2).
\end{aligned}
\tag{S13}
$$

The proof is complete by combing (S9) to (S13) and the fact $\|\{\int(\hat{\phi}_{(1),j}-\phi_j)\phi_j\}\phi_j\| = \|\hat{\phi}_{(1),j}-\phi_j\|^2/2$ (Hsing & Eubank, 2015, Theorem 5.1.7). □

### S.2.3. *Proof of Theorem 3*

*Proof.* The $\mathcal{L}^2$ discrepancy between $\beta$ and $\hat{\beta}$ can be decomposed as

$$
\begin{aligned}
\|\hat{\beta}-\beta\|^2 =& \left\|\sum_{k=1}^m \hat{b}_k\hat{\phi}_k - \sum_{k=1}^{\infty} b_{0k}\phi_k\right\|^2 = \left\|\sum_{k=1}^m \frac{1}{2}\hat{b}_k(\hat{\phi}_{(1),k}+\hat{\phi}_{(2),k}) - \sum_{k=1}^{\infty} b_{0k}\phi_k\right\|^2\\
\leq&\frac{1}{2}\left\|\sum_{k=1}^m \hat{b}_k\hat{\phi}_{(1),k} - \sum_{k=1}^{\infty} b_{0k}\phi_k\right\|^2 + \frac{1}{2}\left\|\sum_{k=1}^m \hat{b}_k\hat{\phi}_{(2),k} - \sum_{k=1}^{\infty} b_{0k}\phi_k\right\|^2.
\end{aligned}
$$

These two terms on the right hand side of last equation admit the same asymptotic behavior, we only need to calculate the first term. By Cauchy–Schwarz inequality,

$$
\begin{aligned}
&\left\|\sum_{k=1}^m \hat{b}_k\hat{\phi}_{(1),k} - \sum_{k=1}^{\infty} b_{0k}\phi_k\right\|^2\\
\leq&3\left\|\sum_{k=1}^m (\hat{b}_k-b_{0k})\hat{\phi}_{(1),k}\right\|^2 + 3\left\|\sum_{k=1}^m b_{0k}(\hat{\phi}_{(1),k}-\phi_k)\right\|^2 + 3\left\|\sum_{k=m+1}^{\infty} b_{0k}\phi_k\right\|^2.
\end{aligned}
\tag{S14}
$$

The first term in the right hand side of (S14) is bounded by $\|D^{-1}\|^2\|\hat{b}_r - b_{r0}\|^2 = O_p(m^a\alpha_n)$, which is due to the compatibility of the matrix norm and the vector norm. The last term in the right hand side of (S14) is $O(m^{1-2b})$. For the second term in the right hand side of (S14), by

Theorem 2,

$$
\begin{aligned}
& E\left\{\left\|\sum_{k=1}^{m} b_{0k}(\hat{\phi}_{(1),k}-\phi_k)\right\|^2\right\} \\
\leq & \sum_{k_1,k_2}^{m} b_{0k_1}b_{0k_2}E(\|\hat{\phi}_{(1),k_1}-\phi_{k_1}\|\|\hat{\phi}_{(1),k_2}-\phi_{k_2}\|) \\
\leq & \sum_{k_1,k_2}^{m} b_{0k_1}b_{0k_2}\{E(\|\hat{\phi}_{(1),k_1}-\phi_{k_1}\|^2)E(\|\hat{\phi}_{(1),k_2}-\phi_{k_2}\|^2)\}^{1/2} \\
= & \left[\sum_{k=1}^{m} b_{0k}\{E(\|\hat{\phi}_{(1)}-\phi_k\|^2)\}^{1/2}\right]^2 \\
\lesssim & \left[\frac{1+m^{\frac{a}{2}+2-b}}{n^{\frac{1}{2}}}\left\{1+\frac{1}{(Nh)^{\frac{1}{2}}}\right\}+h^2(1+m^{a+c+2-b})\right]^2 \\
= & O\left(\frac{1}{nNh}\right)+o(\alpha_n),
\end{aligned}
\tag{S15}
$$

where the last equality holds under Condition 5. Combing (S14) and (S15),

$$
\left\|\sum_{k=1}^{m}\hat{b}_k\hat{\phi}_{(1),k}-\sum_{k=1}^{\infty}b_{0k}\phi_k\right\|^2 = O_p\left(\frac{m^{a+1}}{n}+m^{1-2b}+\delta_n\right),
$$

where

$$
\delta_n = m^a\left\{\frac{1}{N}+\frac{1}{n}\left(1+\frac{1}{Nh}\right)\right\}\left\{\frac{m^{2a+3}}{n}\left(1+\frac{1}{Nh}\right)+h^4m^{3a+2c+3}+\frac{m^{a+1}}{N}\right\}+\frac{1}{nNh}.
$$

Under condition 6, there is $Nh > \mathrm{C}$ and

$$
\begin{aligned}
\frac{1}{N}\frac{m^{2a+3}}{n} &\leq n^{-\frac{2a+2}{a+2b}}n^{\frac{2a+3}{a+2b}}n^{-1} = O\left(\frac{m}{n}\right); \\
\frac{1}{N}\times h^4m^{3a+2c+3} &\leq \frac{1}{N}\left(\underline{N}^{1/4}n^{-\frac{2a+b+c+1}{2(a+2b)}}\right)^4 m^{3a+2c+3} = O\left(\frac{m}{n}\right); \\
\frac{1}{N}\times\frac{m^{a+1}}{N} &\leq \frac{m^{a+1}}{n^{\frac{2a+2b}{a+2b}}} = O\left(\frac{m}{n}\right); \\
\frac{1}{n}\frac{m^{2a+3}}{n} &= \frac{m^{2a+2}}{n}\frac{m}{n} = o\left(\frac{m}{n}\right); \\
\frac{1}{n}h^4m^{3a+2c+3} &\leq \frac{1}{n}n^{-\frac{3a+2c+4}{2a+4}}n^{\frac{3a+2c+3}{a+2b}} \leq \frac{1}{n} = o\left(\frac{m}{n}\right); \\
\frac{1}{n}\frac{m^{a+1}}{N} &\leq \frac{1}{n}\frac{m^{a+1}}{m^{2a+2}} = o\left(\frac{m}{n}\right); \\
\frac{1}{nNh} &\leq \frac{1}{n} = o\left(\frac{m}{n}\right).
\end{aligned}
\tag{S16}
$$

Then we obtain $\delta_n = O_p(m^{a+1}/n) = O_p(n^{(1-2b)/(a+2b)})$. □

### *S.2.4. Proof of Theorem 4*

*Proof.* By the definition of $\mathcal{E}(\hat\theta_n)$,

$$\begin{aligned}\mathcal{E}(\hat\theta_n) =&E_*\left[\left\{\int \beta X^* - \frac{1}{N}\sum_{j=1}^{N}\hat\beta(T_j^*)X_j^*\right\}^2\right]\\ =&E_{X^*}\left\{\left(\int \beta X^* - \int\hat\beta X^*\right)^2\right\} + \frac{1}{N}E_{X^*}\left\{\int\hat\beta^2(X^*)^2 - \left(\int\hat\beta X^*\right)^2\right\} + \frac{\sigma_X^2}{N}\|\hat\beta\|^2.\end{aligned}$$

We can show that for any $\hat\beta = \sum_{k\geqslant 1}\hat b_k^2\phi_k$ with $\|\hat\beta\|_2 < \infty$,

$$E_*\left\{\int\hat\beta^2(X^*)^2 - \left(\int\hat\beta X^*\right)^2\right\} = \sum_{k=1}^{\infty}\lambda_k\left(\int\hat\beta^2\phi_k^2 - \hat b_k^2\right) = O(1).$$

On one hand

$$\begin{aligned}\left|\sum_{k=1}^{\infty}\lambda_k\left(\int\hat\beta^2\phi_k^2 - \hat b_k^2\right)\right| \leq& \left|\sum_{k=1}^{\infty}\lambda_k\int\hat\beta^2\phi_k^2\right| + \left|\sum_{k=1}^{\infty}\lambda_k\hat b_k^2\right|\\ \leq&\|\hat\beta\|^2\sum_{k=1}^{\infty}\lambda_k\|\phi_k\|_\infty^2 + \sum_{k=1}^{\infty}\lambda_k\hat b_k^2 < \infty,\end{aligned}$$

where the second inequality follows from $\|\phi_k\|_\infty = O(1)$ by Condition 3. On the other hand, by Jensen inequality, for any $k \in \mathbb{N}_+$, $\int\hat\beta^2\phi_k^2 \geqslant \hat b_k^2$ and the equality holds if and only if $\hat\beta(s)\hat\phi_k(s) = \int\hat\beta\hat\phi_k$ for all $s\in[0,1]$, which is the trivial case. Thus, without loss of generality, we assume that there exists a $\delta > 0$ such that $\int\hat\beta^2\phi_1^2 - \hat b_1^2 > \delta$ and

$$\left|\sum_{k=1}^{\infty}\lambda_k\left(\int\hat\beta^2\phi_k^2 - \hat b_k^2\right)\right| \geqslant \lambda_1\left(\int\hat\beta^2\phi_1^2 - \hat b_1^2\right) > \mathrm{C}.$$

Then, the discretely observed prediction error becomes

$$\mathcal{E}(\hat\theta_n) = \mathcal{E}(\tilde\theta_n) + O_p\left(\frac{1}{N}\right). \tag{S17}$$

Next we focus on the asymptotic behavior of $\mathcal{E}(\tilde\theta_n)$. By the definition of $\mathcal{E}(\tilde\theta_n)$,

$$\begin{aligned}\mathcal{E}(\tilde\theta_n) =& E_*(\langle\hat\beta - \beta, X^*\rangle^2)\\ \leq&2E_*\left(\left\langle\hat\beta-\beta, \sum_{k=1}^{m}\xi_k^*\phi_k\right\rangle^2\right) + 2E_*\left(\left\langle\hat\beta-\beta, \sum_{k=m+1}^{\infty}\xi_k^*\phi_k\right\rangle^2\right)\\ =&2\sum_{j=1}^{m}\lambda_j\left\{\int(\hat\beta-\beta)\phi_j\right\}^2 + 2E_*\left(\left\langle\hat\beta-\beta, \sum_{k=m+1}^{\infty}\xi_k^*\phi_k\right\rangle^2\right)\\ =&I_1 + I_2.\end{aligned}$$

In this proof, we use $\hat\beta = \sum_{k=1}^{m}\hat b_k\phi_{(1),k}$ instead of $\hat\beta = 0.5\sum_{k=1}^{m}\hat b_k(\hat\phi_{(1),k} + \hat\phi_{(2),k})$ to reduce the notation burden since this does not affect the asymptotic behavior. For $I_1$, by expansions

(S9),

$$
\begin{aligned}
&\int \hat\beta(s)\phi_j(s)\mathrm{d}s = \int \sum_{k=1}^{m} \hat b_k \hat\phi_{(1),k}(s)\phi_j(s)\mathrm{d}s \\
=&\int \sum_{k=1}^{m} \hat b_k \Bigg( \phi_k(s) + \sum_{l\neq k} \frac{\langle \Delta_{(1)}\phi_l, \phi_k\rangle}{\lambda_k-\lambda_l}\phi_l(s) + \sum_{l\neq k} \frac{\langle \Delta_{(1)}\phi_l, \hat\phi_{(1),k}-\phi_k\rangle}{\lambda_k-\lambda_l}\phi_l(s) \\
&+ \sum_{l\neq k}\sum_{s=1}^{\infty} \frac{(\lambda_k-\hat\lambda_{(1),k})^s}{(\lambda_k-\lambda_l)^{s+1}} \langle \Delta_{(1)}\phi_l, \hat\phi_{(1),k}\rangle \phi_l(s) + \langle \hat\phi_{(1),k}-\phi_k, \phi_k\rangle\phi_k(s) \Bigg) \phi_j(s)\mathrm{d}s \\
=&\hat b_j + \sum_{k\neq j}^{m} \frac{\langle \Delta_{(1)}\phi_j, \phi_k\rangle}{\lambda_k-\lambda_j}\hat b_k + \sum_{k\neq j}^{m} \frac{\langle \Delta_{(1)}\phi_j, \hat\phi_{(1),k}-\phi_k\rangle}{\lambda_k-\lambda_j}\hat b_k \\
&+ \sum_{k\neq j}^{m}\sum_{s=1}^{\infty} \frac{(\lambda_k-\hat\lambda_{(1),k})^s}{(\lambda_k-\lambda_j)^{s+1}} \langle \Delta_{(1)}\phi_j, \hat\phi_{(1),k}\rangle \hat b_k + \langle \hat\phi_{(1),j}-\phi_j, \phi_j\rangle \hat b_j.
\end{aligned}
$$

By Cauchy–Schwarz inequality,

$$
I_1 \lesssim \sum_{j=1}^{m} \lambda_j(\hat b_j - b_{0j})^2 + \sum_{j=1}^{m} \lambda_j \left( \sum_{k\neq j}^{m} \frac{\langle \Delta_{(1)}\phi_j, \phi_k\rangle}{\lambda_k-\lambda_j}\hat b_k \right)^2 + E, \tag{S18}
$$

where

$$
\begin{aligned}
E =& \sum_{j=1}^{m} \lambda_j \left( \sum_{k\neq j}^{m} \frac{\langle \Delta_{(1)}\phi_j, \hat\phi_{(1),k}-\phi_k\rangle}{\lambda_k-\lambda_j}\hat b_k \right)^2 \\
&+ \sum_{j=1}^{m} \lambda_j \left\{ \sum_{k\neq j}^{m}\sum_{s=1}^{\infty} \frac{(\lambda_k-\hat\lambda_{(1),k})^s}{(\lambda_k-\lambda_j)^{s+1}} \langle \Delta_{(1)}\phi_j, \hat\phi_{(1),k}\rangle \hat b_k \right\}^2 + \sum_{j=1}^{m} \lambda_j \langle \hat\phi_{(1),j}-\phi_j, \phi_j\rangle^2 \hat b_j^2
\end{aligned}
$$

is the remaining term.

For the first two terms in the right hand side of (S18), by Lemma S3

$$
\sum_{j=1}^{m} \lambda_j(\hat b_j - b_j)^2 = \|\hat b_r - b_{r0}\|^2 = O_p(\alpha_n), \tag{S19}
$$

and

$$
\begin{aligned}
&\sum_{j=1}^{m} \lambda_j \left( \sum_{k\neq j}^{m} \frac{\langle \Delta_{(1)}\phi_j, \phi_k\rangle}{\lambda_k-\lambda_j}\hat b_k \right)^2 \\
\leq& 2\sum_{j=1}^{m} \lambda_j \left( \sum_{k\neq j}^{m} \frac{\langle \Delta_{(1)}\phi_j, \phi_k\rangle}{\lambda_k-\lambda_j}(\hat b_k - b_{0k}) \right)^2 + 2\sum_{j=1}^{m} \lambda_j \left( \sum_{k\neq j}^{m} \frac{\langle \Delta_{(1)}\phi_j, \phi_k\rangle}{\lambda_k-\lambda_j} b_{0k} \right)^2 .
\end{aligned} \tag{S20}
$$

For the first term in the right hand side of (S20), by Cauchy–Schwarz inequality and Lemma S3

$$\sum_{j=1}^{m}\lambda_j\left\{\sum_{k\neq j}^{m}\frac{\langle\Delta_{(1)}\phi_j,\phi_k\rangle}{\lambda_k-\lambda_j}(\hat{b}_k-b_{0k})\right\}^2\leq\sum_{j=1}^{m}\lambda_j\sum_{k\neq j}^{m}(\hat{b}_k-b_{0k})^2\sum_{k\neq j}^{m}\frac{\langle\Delta_{(1)}\phi_j,\phi_k\rangle^2}{(\lambda_k-\lambda_j)^2}$$
$$\leq m^a\|\hat{b}_r-b_{r0}\|^2\sum_{j=1}^{m}\lambda_j\sum_{k\neq j}^{m}\frac{\langle\Delta_{(1)}\phi_j,\phi_k\rangle^2}{(\lambda_k-\lambda_j)^2}.$$

By Theorem 1 and Lemma 7 in Dou et al. (2012),

$$E\left\{\sum_{j=1}^{m}\lambda_j\sum_{k\neq j}^{m}\frac{\langle\Delta_{(1)}\phi_j,\phi_k\rangle^2}{(\lambda_k-\lambda_j)^2}\right\}\lesssim\frac{1}{n}(m^{3-a}+1)+h^4(m^{3-a+2c}+1).$$

Thus, Under Condition 5

$$\sum_{j=1}^{m}\lambda_j\left\{\sum_{k\neq j}^{m}\frac{\langle\Delta_{(1)}\phi_j,\phi_k\rangle}{\lambda_k-\lambda_j}(\hat{b}_k-b_{0k})\right\}^2=O_p\left(\left\{\frac{m^3+m^a}{n}+h^4(m^{3+2c}+m^a)\right\}\alpha_n\right)$$
$$=o_p\left(\alpha_n\right).$$

By (S41) in the proof of Lemma S3, the second term of (S20) is $o_p\left(\alpha_n\right)$. For the remaining part, we divide $E$ into several parts,

$$\begin{aligned}E\lesssim&\sum_{j=1}^{m}\lambda_j\left(\sum_{k\neq j}^{m}\frac{\langle\Delta_{(1)}\phi_j,\hat{\phi}_{(1),k}-\phi_k\rangle}{\lambda_k-\lambda_j}b_{0k}\right)^2\\
&+\sum_{j=1}^{m}\lambda_j\left\{\sum_{k\neq j}^{m}\sum_{s=1}^{\infty}\frac{(\lambda_k-\hat{\lambda}_{(1),k})^s}{(\lambda_k-\lambda_j)^{s+1}}\langle\Delta_{(1)}\phi_j,\hat{\phi}_{(1),k}\rangle b_{0k}\right\}^2\\
&+\sum_{j=1}^{m}\lambda_j\langle\hat{\phi}_{(1),j}-\phi_j,\phi_j\rangle^2b_{0j}^2+\sum_{j=1}^{m}\lambda_j\left\{\sum_{k\neq j}^{m}\frac{\langle\Delta_{(1)}\phi_j,\hat{\phi}_{(1),k}-\phi_k\rangle}{\lambda_k-\lambda_j}(\hat{b}_k-b_{0k})\right\}^2\\
&+\sum_{j=1}^{m}\lambda_j\left\{\sum_{k\neq j}^{m}\sum_{s=1}^{\infty}\frac{(\lambda_k-\hat{\lambda}_{(1),k})^s}{(\lambda_k-\lambda_j)^{s+1}}\langle\Delta_{(1)}\phi_j,\hat{\phi}_{(1),k}\rangle(\hat{b}_k-b_{0k})\right\}^2\\
&+\sum_{j=1}^{m}\lambda_j\langle\hat{\phi}_{(1),j}-\phi_j,\phi_j\rangle^2(\hat{b}_j-b_{0j})^2\\
=&E_1+E_2+E_3+E_4+E_5+E_6.\end{aligned}$$

By equation (S42) to (S46) in the proof of Lemma S3, $E_1+E_2+E_3=o_p(\alpha_n)$. The following equations show that $E_4, E_5$ and $E_6$ are also $o_p(\alpha_n)$.

$$
\begin{aligned}
E_4 =& \sum_{j=1}^{m} \lambda_j \left\{ \sum_{k\neq j}^{m} \frac{\langle \Delta_{(1)}\phi_j, \hat{\phi}_{(1),k}-\phi_k\rangle}{\lambda_k-\lambda_j} (\hat{b}_k - b_{0k}) \right\}^2 \\
\leq & \sum_{j=1}^{m} \lambda_j \sum_{k\neq j}^{m} \frac{\langle \Delta_{(1)}\phi_j, \hat{\phi}_{(1),k}-\phi_k\rangle^2}{(\lambda_k-\lambda_j)^2} \sum_{k=1}^{m} (\hat{b}_k - b_{0k})^2 \\
\leq & m^a \|\hat{b}_r - b_{r0}\|^2 \sum_{j=1}^{m} \lambda_j \frac{\|\Delta_{(1)}\|_{\text{HS}}^2}{\eta_j^2} \sum_{k=1}^{m} \|\hat{\phi}_{(1),k}-\phi_k\|^2 \\
=& o_p(\alpha_n);
\end{aligned}
\tag{S21}
$$

$$
\begin{aligned}
E_5 =& \sum_{j=1}^{m} \lambda_j \left\{ \sum_{k\neq j}^{m} \sum_{s=1}^{\infty} \frac{(\lambda_k - \hat{\lambda}_{(1),k})^s}{(\lambda_k-\lambda_j)^{s+1}} \langle \Delta_{(1)}\phi_j, \hat{\phi}_{(1),k}\rangle (\hat{b}_k - b_{0k}) \right\}^2 \\
\leq & \sum_{j=1}^{m} \lambda_j \sum_{k\neq j}^{m} \left\{ \sum_{s=1}^{\infty} \frac{(\lambda_k - \hat{\lambda}_{(1),k})^s}{(\lambda_k-\lambda_j)^{s+1}} \right\}^2 \langle \Delta_{(1)}\phi_j, \hat{\phi}_{(1),k}\rangle^2 \sum_{k=1}^{m} (\hat{b}_k - b_{0k})^2 \\
\lesssim & m^a \|\hat{b}_r - b_{r0}\|^2 \sum_{j=1}^{m} \lambda_j \sum_{k\neq j}^{m} \frac{\langle \Delta_{(1)}\phi_j, \hat{\phi}_{(1),k}\rangle^2}{(\lambda_k-\lambda_j)^2} \\
\lesssim & m^a \|\hat{b}_r - b_{r0}\|^2 \sum_{j=1}^{m} \lambda_j \left\{ \sum_{k\neq j}^{m} \frac{\langle \Delta_{(1)}\phi_j, \phi_k\rangle^2}{(\lambda_k-\lambda_j)^2} + \sum_{k\neq j}^{m} \frac{\langle \Delta_{(1)}\phi_j, \hat{\phi}_{(1),k}-\phi_k\rangle^2}{(\lambda_k-\lambda_j)^2} \right\} \\
\lesssim & m^a \|\hat{b}_r - b_{r0}\|^2 \sum_{j=1}^{m} \lambda_j \left( \|\hat{\phi}_{(1),j}-\phi_j\|^2 + \sum_{k\neq j}^{m} \|\hat{\phi}_{(1),k}-\phi_k\|^2 \right) \\
=& o_p(\alpha_n);
\end{aligned}
\tag{S22}
$$

$$
E_6 = \sum_{j=1}^{m} \lambda_j \langle \hat{\phi}_{(1),j}-\phi_j, \phi_j\rangle^2 (\hat{b}_j - b_{0j})^2 \leq \frac{\|\Delta_{(1)}\|_{\text{HS}}^2}{(2\eta_m)^2} \sum_{j=1}^{m} \lambda_j (\hat{b}_j - b_{0j})^2 = o_p(\alpha_n). \tag{S23}
$$

Thus, under Condition 5, combining (S20) to (S23) we have $I_1 = O_p(\alpha_n)$.

As for $I_2$,

$$
\begin{aligned}
I_2 =& E_* \left( \left\langle \hat{\beta} - \beta, \sum_{k=m+1}^{\infty} \xi_k^* \phi_k \right\rangle^2 \right) \\
\lesssim & E_* \left( \left\langle \sum_{k=1}^{m} \hat{b}_k \hat{\phi}_{(1),k}, \sum_{k=m+1}^{\infty} \xi_k^* \phi_k \right\rangle^2 \right) + E_* \left( \left\langle \sum_{k=1}^{\infty} b_{0k} \phi_k, \sum_{k=m+1}^{\infty} \xi_k^* \phi_k \right\rangle^2 \right) \\
=& I_{21} + I_{22},
\end{aligned}
$$

where

$$I_{22} = E_* \left( \left\langle \sum_{k=m+1}^{\infty} b_{0k}\phi_k, \sum_{k=m+1}^{\infty} \xi_k^* \phi_k \right\rangle^2 \right) = \sum_{k=m+1}^{\infty} \lambda_k b_{0k}^2 = O(m^{1-a-2b}).$$

As for $I_{21}$, by the orthogonality of the series $\{\phi_k\}_{k=1}^{\infty}$,

$$\begin{aligned}
&E_* \left( \left\langle \sum_{k=1}^{m} \hat{b}_k \hat{\phi}_{(1),k}, \sum_{k=m+1}^{\infty} \xi_k^* \phi_k \right\rangle^2 \right) \\
=&E_* \left( \left\langle \sum_{k=1}^{m} \hat{b}_k (\hat{\phi}_{(1),k} - \phi_k), \sum_{k=m+1}^{\infty} \xi_k^* \phi_k \right\rangle^2 \right) \\
\le&E_* \left( \left\| \sum_{k=m+1}^{\infty} \xi_k^* \phi_k \right\|^2 \right) \left\| \sum_{k=1}^{m} \hat{b}_k (\hat{\phi}_{(1),k} - \phi_k) \right\|^2 \\
\lesssim&m^{1-a} \left( \left\| \sum_{k=1}^{m} (\hat{b}_k - b_{0k})(\hat{\phi}_{(1),k} - \phi_k) \right\|^2 + \left\| \sum_{k=1}^{m} b_{0k} (\hat{\phi}_{(1),k} - \phi_k) \right\|^2 \right) \\
\le&m^{1-a} \sum_{k=1}^{m} (\hat{b}_k - b_{0k})^2 \sum_{k=1}^{m} \|\hat{\phi}_{(1),k} - \phi_k\|^2 + m^{1-a} \times \left\| \sum_{k=1}^{m} b_{0k} (\hat{\phi}_{(1),k} - \phi_k) \right\|^2 \\
=&O_p \left( \frac{m^{1-a}}{nNh} \right) + o_p(\alpha_n),
\end{aligned}$$

where the last equality comes from (S15), Lemma S3 and Theorem 2. Thus $I_2 = O_p(m^{1-a}/(nNh)) + o_p(\alpha_n)$. Combining the rate of $I_1$ and $I_2$, under Condition 5,

$$\sum_{j=1}^{\infty} \lambda_j \left\{ \int (\hat{\beta} - \beta)\phi_j \right\}^2 = O_p \left( \frac{m}{n} + m^{1-a-2b} + \delta_n' \right),$$

where

$$\delta_n' = \left\{ \frac{1}{N} + \frac{1}{n} \left( 1 + \frac{1}{nNh} \right) \right\} \left\{ \frac{m^{2a+3}}{n} \left( 1 + \frac{1}{Nh} \right) + h^4 m^{3a+2c+3} + \frac{m^{a+1}}{N} \right\} + \frac{m^{1-a}}{nNh}.$$

Similarly, under Condition 6 and (S16), $O_p(\delta_n') = O_p(m/n) = O_p(n^{1-a-2b/(a+2b)})$ and the proof is comlete. □

### S.2.5. *Proof of Theorem 5*

*Proof.* To obtain the lower bound, we will construct $\beta$ as a transformation of a new parameter in some discrete space. Define $\mathcal{X} := \{X_i(t), t \in [0,1]\}_{i=1}^{n}$ and assume $X_i(t)$ are random copies of $X(t)$ with eigenfunctions $\{\phi_k\}_{k=1}^{\infty}$ and eigenvalues $\{\lambda_k\}_{k=1}^{\infty}$. Note that $\mathcal{G}$ is the class of $X$ and $\beta$ satisfying properties 1 and 2 in Theorem 5. For each $\theta = (\theta_1, \dots \theta_r) \in \{0,1\}^r$ and a small $\epsilon > 0$, define $\beta_{\theta,\epsilon} = \epsilon \sum_{k=1}^{r} \theta_k b_k \phi_k$ with $b_k \le Ck^{-b}$. The responses $\{Y_i(\theta,\epsilon)\}_{i=1}^{n}$ are driven by

$$Y_i(\theta,\epsilon) = \int \beta_{\theta,\epsilon} X_i + e_i = \eta_i(\theta,\epsilon) + e_i \quad \text{with} \quad \eta_i(\theta,\epsilon) = \epsilon \sum_{k=1}^{r} \theta_k b_k \xi_{ik},$$

where $e_i \sim N(0,\sigma^2)$ and $E(\xi_{ik}^2) = \lambda_k$. Denote $P_\theta$ the conditional probability measure of $\{Y_i(\theta,\epsilon)\}_{i=1}^n$ given $\mathcal{X}$, and its corresponding density is denoted as $q_\theta$.

It is sufficient to find $g_r$ and a feasible lower bounds for $\min_{H(\theta,\theta')=1} E(\|P_\theta \wedge P_{\theta'}\|_a)$ by applying Lemma S4 with $\psi(\theta) = \epsilon \sum_{k=1}^r \theta_k b_k \phi_k$. The prediction error

$$d(\hat\beta,\beta) = \mathcal{E}(\tilde\theta_n) = \sum_{j=1}^\infty \lambda_j \left\{\int (\hat\beta - \beta)\phi_j\right\}^2 = \sum_{j=1}^\infty \lambda_j (\hat b_j - b_j)^2 \tag{S24}$$

can be viewed as a semi-distance in $\mathcal{L}^2[0,1]$ and $d(x,z) + d(z,y) \geq d(x,y)/2$. By definition,

$$d\left(\psi(\theta), \psi\left(\theta'\right)\right) = d(\beta_{\theta,\epsilon}, \beta_{\theta',\epsilon}) = \sum_{k=1}^r \lambda_k \left\{\int (\beta_{\theta,\epsilon} - \beta_{\theta',\epsilon})\phi_k\right\}^2 = \sum_{k=1}^r d_k\left(\psi(\theta), \psi\left(\theta'\right)\right),$$

where $d_k\left(\psi(\theta), \psi\left(\theta'\right)\right) = \lambda_k \epsilon^2 (\theta_k' - \theta_k)^2 b_k^2,\ k = 1,2,\ldots,r$. Assume $\theta$ differs from $\theta'$ only in the $j$th coordinate, thus $H\left(\theta,\theta'\right) = 1$, and $d_j\left(\psi(\theta), \psi\left(\theta'\right)\right) = \epsilon^2 \lambda_j b_j^2$, which implies $g_r = \epsilon^2 \sum_{k=1}^r \lambda_k b_k^2 = \epsilon^2 O(r^{1-a-2b})$. For any estimator $\hat\beta = \sum_{j\in\mathbb{N}_+} \hat b_j \phi_j$ based on $\{X_i, Y_i\}_{i=1}^n$, apply Lemma S4 with $A = 1/2$

$$\sup_{X^*,\beta\in\mathcal{G}} E\{\mathcal{E}(\tilde\theta_n)\} \geq \max_{\theta\in\{0,1\}^r} E\{d(\hat\beta, \beta_{\theta,\epsilon})\} \geq \frac{g_r}{4} \min_{H(\theta,\theta')=1} E(\|P_\theta \wedge P_{\theta'}\|_a). \tag{S25}$$

By the property of the total variation distance and Pinsker's inequality (Tsybakov, 2008, Lemma 2.1 and 2.5),

$$E(\|P_\theta \wedge P_{\theta'}\|_a) \geq 1 - E[\{K(P_\theta \mid P_{\theta'})/2\}^{1/2}]. \tag{S26}$$

To guarantee the positiveness of the right hand side of (S26), we need to show that $\{K(P_\theta \mid P_{\theta'})/2\}^{1/2}$ is sufficient small for a suitable $\epsilon > 0$. For a fixed a $j \in \{1,2,\ldots,r\}$, the log-likelihood ratio for normal noise in Condition 7 with $K_{\sigma^2} = (2\sigma^2)^{-1}$ is

$$\log\left(\frac{q_{\theta'}}{q_\theta}\right) = \frac{1}{\sigma^2}\sum_{i=1}^n \left[\left(Y_i - \int \beta_{\theta,\epsilon} X_i\right) \int X_i(\beta_{\theta,\epsilon} - \beta_{\theta',\epsilon}) - \left\{\int X_i(\beta_{\theta,\epsilon} - \beta_{\theta',\epsilon})\right\}^2\right]$$

Note that $K(P_\theta \mid P_{\theta'}) = -E_\theta\{\log\left(q_{\theta'}/q_\theta\right) \mid \mathcal{X}\}$,

$$\begin{aligned} E[\{K(P_\theta \mid P_{\theta'})/2\}^{1/2}] &= E\left[\left\{\frac{n}{4\sigma^2}\sum_{k=1}^r (\theta_k' - \theta_k)^2 b_k^2 \xi_{1k}^2\right\}^{1/2}\right] \\ &\leq \left\{\frac{n}{4\sigma^2}\sum_{k=1}^r (\theta_k' - \theta_k)^2 b_k^2 E(\xi_{1k}^2)\right\}^{1/2} = \left(\frac{n}{4\sigma^2}\epsilon^2 b_j^2 \lambda_j\right)^{1/2}, \end{aligned} \tag{S27}$$

where the last inequality is by Jensen's inequality. By $|b_j| \leq \mathrm{C} j^{-b},\ \lambda_j \leq \mathrm{R} j^{-a}$ for $j \geq 1$ and constant $\mathrm{C}, \mathrm{R} > 0$,

$$E(\|P_\theta \wedge P_{\theta'}\|_a) \geq 1 - \min_{1\leq j\leq r} \frac{\epsilon \mathrm{C}}{2\sigma}\left(\mathrm{R} n j^{-(a+2b)}\right)^{1/2} = 1 - \frac{\epsilon \mathrm{C}}{2\sigma}\left(\mathrm{R} n r^{-(a+2b)}\right)^{1/2}.$$

If we put $r = Ln^{1/(a+2b)}$, for sufficiently small $\epsilon > 0$,

$$\min_{H(\theta,\theta')=1} E(\|P_\theta \wedge P_{\theta'}\|_a) \geq 1 - \tfrac{\epsilon \mathrm{C}}{2\sigma}\left(\mathrm{R}\tfrac{n}{r^{a+2b}}\right)^{1/2} = 1 - \tfrac{\epsilon \mathrm{C}}{2\sigma}\left(\tfrac{\mathrm{R}}{L^{a+2b}}\right)^{1/2} > 0.$$

Then the proof of Lemma S4 for the Gaussian noise is completed by (S25), (S26), (S27) and $g_r = \epsilon^2 \sum_{j=1}^{r} \lambda_j b_j^2 = O\left(n^{(1-a-2b)/(a+2b)}\right)$. For general noises satisfying Condition 7, the bound (S27) still holds up to a constant and the proof is analogous. □

## S.3. Proofs of Lemmas

### *S.3.1. Proof of Lemma S1*

*Proof.* We focus on $E\{(\hat{\xi}_{ik} - \xi_{ik})^2\}$ first, for each $i \le n/2$

$$
\begin{aligned}
E\{(\hat{\xi}_{ik} - \xi_{ik})^2\} \le & 2E\left[\left\{\frac{1}{N}\sum_{j=1}^{N} X_{ij}\phi_k(T_{ij}) - \langle X_i, \phi_k\rangle\right\}^2\right] \\
& + 2E\left(\left[\frac{1}{N}\sum_{j=1}^{N} X_{ij}\{\hat{\phi}_{(2),k}(T_{ij}) - \phi_k(T_{ij})\}\right]^2\right).
\end{aligned}
$$

By the central limit theorem, the first term in the right hand side of last equation is bounded by $CN^{-1}$. For the second term,

$$
\begin{aligned}
& E\left(\left[\frac{1}{N}\sum_{j=1}^{N} X_{ij}\{\hat{\phi}_{(2),k}(T_{ij}) - \phi_k(T_{ij})\}\right]^2\right) \\
= & \frac{1}{N^2}\sum_{j_1 \ne j_2} E[\delta_{ij_1j_2}\{\hat{\phi}_{(2),k}(T_{ij_1}) - \phi_k(T_{ij_1})\}\{\hat{\phi}_{(2),k}(T_{ij_2}) - \phi_k(T_{ij_2})\}] \\
& + \frac{1}{N^2}\sum_{j=1}^{N} E[X_{ij}^2\{\hat{\phi}_{(2),k}(T_{ij}) - \phi_k(T_{ij})\}^2].
\end{aligned}
\tag{S28}
$$

We first show that the second term in the right hand side of (S28) is $o(N^{-1})$,

$$
\begin{aligned}
& \frac{1}{N^2}\sum_{j=1}^{N} E[X_{ij}^2\{\hat{\phi}_{(2),k}(T_{ij}) - \phi_k(T_{ij})\}^2] \\
= & \frac{1}{N}E\left(\int E[\{X_i^2(u) + \sigma_X^2\}\{\hat{\phi}_{(2),k}(u) - \phi_k(u)\}^2 \mid \hat{\phi}_{(2),k}]\mathrm{d}u\right) \\
\lesssim & \frac{1}{N}E(\|\hat{\phi}_{(2),k} - \phi_k\|^2) = o\left(\frac{1}{N}\right),
\end{aligned}
\tag{S29}
$$

where the last inequality comes from Condition 3 and Theorem 2.

As for the first term in the right hand side of (S28), notice that $\hat{\phi}_{(2),k}$ is independent of $X_i$ and $T_{ij}$ for each $i \le n/2$. Therefore, for each $j_1 \ne j_2$, $X_{ij_1}\hat{\phi}_{(2),k}(T_{ij_1})$ and $X_{ij_2}\hat{\phi}_{(2),k}(T_{ij_2})$ are independent conditional on $X_i$ and $\hat{\phi}_{(2),k}$,

$$\begin{aligned}
&E[\delta_{ij_1j_2}\{\hat{\phi}_{(2),k}(T_{ij_1})-\phi_k(T_{ij_1})\}\{\hat{\phi}_{(2),k}(T_{ij_2})-\phi_k(T_{ij_2})\}]\\
=&E(E[\delta_{ij_1j_2}\{\hat{\phi}_{(2),k}(T_{ij_1})-\phi_k(T_{ij_1})\}\{\hat{\phi}_{(2),k}(T_{ij_2})-\phi_k(T_{ij_2})\}\mid X_i,\hat{\phi}_{(2),k}])\\
=&E\left[\int X_i(u)\{\hat{\phi}_{(2),k}(u)-\phi(u)\}\mathrm{d}u\right]^2\\
\leq&E(\|X_i\|^2)E(\|\hat{\phi}_{(2),k}-\phi_k\|^2)\\
\lesssim&\frac{k^{a+2}}{n}\left(1+\frac{1}{Nh}\right)+h^4k^{2a+2c+2}.
\end{aligned}$$

Thus, the first assertion of Lemma S1 has been proofed for $i\leq n/2$ by

$$E(\|\hat{\eta}_i-\eta_i\|^2)=\sum_{k=1}^{m}k^aE\{(\hat{\xi}_{ik}-\xi_{ik})^2\}\lesssim\frac{m^{2a+3}}{n}\left(1+\frac{1}{Nh}\right)+h^4m^{3a+2c+3}+\frac{m^{a+1}}{N}.$$

For the second assertion,

$$(\theta_i-\hat{\theta}_{i,m})^2\leq 3(\theta_i-\theta_{i,m})^2+3(\theta_{i,m}-\tilde{\theta}_{i,m})^2+3(\tilde{\theta}_{i,m}-\hat{\theta}_{i,m})^2 \tag{S30}$$

where $\tilde{\theta}_{i,m}=\tilde{\xi}_i^{\mathrm{T}}b_0$ and $\tilde{\xi}_i=(\tilde{\xi}_{i1},\cdots,\tilde{\xi}_m)^{\mathrm{T}}$ with $\tilde{\xi}_{ik}=N^{-1}\sum_{j=1}^{N}X_{ij}\phi_k(T_{ij})$.

The first part in the right hand side of (S30) is bounded by $E(|\theta_i-\theta_{i,m}|^2)=\sum_{k>m}\lambda_kb_{0k}^2=O(m^{1-a-2b})\lesssim N^{-1}$ under Condition 5. For the second part in the right hand side of (S30),

$$E(|\theta_{i,m}-\tilde{\theta}_{i,m}|^2)=E\left(\left[\sum_{k=1}^{m}\left\{\xi_{ik}-\frac{1}{N}\sum_{j=1}^{N}X_{ij}\phi_k(T_{ij})\right\}b_{0k}\right]^2\right)\lesssim N^{-1}.$$

For the last part in the right hand side of (S30),

$$\begin{aligned}
E\{(\tilde{\theta}_{i,m}-\hat{\theta}_{i,m})^2\}=&E\left(\sum_{k_1,k_2}^{m}b_{0k_1}b_{0k_2}\left[\frac{1}{N}\sum_{j=1}^{N}X_{ij}\{\hat{\phi}_{(2),k_1}(T_{ij})-\phi_{k_1}(T_{ij})\}\right]\right.\\
&\left.\times\left[\frac{1}{N}\sum_{j=1}^{N}X_{ij}\{\hat{\phi}_{(2),k_2}(T_{ij})-\phi_{k_2}(T_{ij})\}\right]\right).
\end{aligned} \tag{S31}$$

We first bound the expectation of each term in the right hand side of (S31),

$$\begin{aligned}
&\frac{1}{N^2}E\left(\left[\sum_{j=1}^{N}X_{ij}\{\hat{\phi}_{(2),k_1}(T_{ij})-\phi_{k_1}(T_{ij})\}\right]\left[\sum_{j=1}^{N}X_{ij}\{\hat{\phi}_{(2),k_2}(T_{ij})-\phi_{k_2}(T_{ij})\}\right]\right)\\
=&\frac{1}{N^2}\sum_{j_1\neq j_2}^{N}E[\delta_{ij_1j_2}\{\hat{\phi}_{(2),k_1}(T_{ij_1})-\phi_{k_1}(T_{ij_1})\}\{\hat{\phi}_{(2),k_2}(T_{ij_2})-\phi_{k_2}(T_{ij_2})\}]\\
&+\frac{1}{N^2}\sum_{j=1}^{N}E[X_{ij}^2\{\hat{\phi}_{(2),k_1}(T_{ij})-\phi_{k_1}(T_{ij})\}\{\hat{\phi}_{(2),k_2}(T_{ij})-\phi_{k_2}(T_{ij})\}]\\
=&\mathrm{I}_{k_1,k_2}+\mathrm{II}_{k_1,k_2}.
\end{aligned}$$

For $\mathrm{I}_{k_1,k_2}$, due to the fact that $X_i$ and $\hat{\phi}_{(2),k}$ are independent for each $i \le n/2$,

$$\begin{aligned}\mathrm{I}_{k_1,k_2} =& \frac{1}{N^2}\sum_{j_1\neq j_2}^{N} E[\delta_{ij_1j_2}\{\hat{\phi}_{(2),k_1}(T_{ij_1}) - \phi_{k_1}(T_{ij_1})\}\{\hat{\phi}_{(2),k_1}(T_{ij_2}) - \phi_{k_1}(T_{ij_2})\}] \\ =& \frac{N-1}{N} E\left[E\left\{\int X_i(\hat{\phi}_{(2),k_1} - \phi_{k_1})\int X_i(\hat{\phi}_{(2),k_2} - \phi_{k_2}) \mid \hat{\phi}_{(2),k_1}, \hat{\phi}_{(2),k_2}\right\}\right] \\ \le& \frac{N-1}{N} E\{E(\|X_i\|^2\|\hat{\phi}_{(2),k_1} - \phi_{k_1}\|\|\hat{\phi}_{(2),k_2} - \phi_{k_2}\| \mid \hat{\phi}_{(2),k_1}, \hat{\phi}_{(2),k_2})\} \\ \lesssim& E(\|\hat{\phi}_{(2),k_1} - \phi_{k_1}\|\|\hat{\phi}_{(2),k_2} - \phi_{k_2}\|) \\ \le& \{E(\|\hat{\phi}_{(2),k_1} - \phi_{k_1}\|^2)E(\|\hat{\phi}_{(2),k_2} - \phi_{k_2}\|^2)\}^{1/2},\end{aligned}$$

where the second last inequality is valid by the sample splitting and the last inequality is from Cauchy–Schwarz inequality.

By Theorem 2,

$$\left|\sum_{k_1,k_2}^{m} b_{0k_1}b_{0k_2}\mathrm{I}_{k_1,k_2}\right| \lesssim \sum_{k_1=1}^{m} b_{0k_1}\{E(\|\hat{\phi}_{(2),k_1} - \phi_{k_1}\|^2)\}^{1/2}\sum_{k_2=1}^{m} b_{0k_2}\{E(\|\hat{\phi}_{(2),k_2} - \phi_{k_2}\|^2)\}^{1/2}$$

$$\lesssim \frac{1}{n}\left(1 + \frac{1}{Nh}\right) + h^4(1 + m^{2a+2c-2b+4}) \lesssim \frac{1}{n}\left(1 + \frac{1}{Nh}\right) + o\left(\frac{1}{N}\right),$$

where the last equality holds under Condition 5.

After similar calculation as (S29), we can show that $\left|\sum_{k_1,k_2}^{m} b_{k_1}b_{k_2}\mathrm{II}_{k_1,k_2}\right| = o(N^{-1})$. Thus the second assertion of Lemma S1 holds for $i \le n/2$ and the proof is complete. □

### S.3.2. *Proof of Lemma S2*

*Proof.* By triangle inequality,

$$\begin{aligned}&\left\|\frac{1}{n}\sum_{i=1}^{n}\hat{\eta}_i\hat{\eta}_i^{\mathrm{T}} - \frac{1}{n}\sum_{i=1}^{n}\eta_i\eta_i^{\mathrm{T}}\right\| \le \left\|\frac{1}{n}\sum_{i=1}^{n}\hat{\eta}_i(\hat{\eta}_i - \eta_i)^{\mathrm{T}}\right\| + \left\|\frac{1}{n}\sum_{i=1}^{n}(\hat{\eta}_i - \eta_i)\eta_i^{\mathrm{T}}\right\| \\ \le& \left\|\frac{1}{n}\sum_{i=1}^{n}(\hat{\eta}_i - \eta_i)(\hat{\eta}_i - \eta_i)^{\mathrm{T}}\right\| + 2\left\|\frac{1}{n}\sum_{i=1}^{n}\eta_i(\hat{\eta}_i - \eta_i)^{\mathrm{T}}\right\|.\end{aligned} \tag{S32}$$

For the first term in the right hand side of (S32), under Condition 5,

$$\begin{aligned}E\left\{\left\|\frac{1}{n}\sum_{i=1}^{n}(\hat{\eta}_i - \eta_i)(\hat{\eta}_i - \eta_i)^{\mathrm{T}}\right\|\right\} \le& \frac{1}{n}\sum_{i=1}^{n}E\{\|(\hat{\eta}_i - \eta_i)(\hat{\eta}_i - \eta_i)^{\mathrm{T}}\|\} \\ =& \frac{1}{n}\sum_{i=1}^{n}E\left(\|\hat{\eta}_i - \eta_i\|^2\right) = o(1).\end{aligned}$$

Then it is sufficient to show the second term in the right hand side of (S32) is $o_p(1)$. By triangle and Cauchy–Schwarz inequality,

$$E\left\{\left\|\frac{1}{n}\sum_{i=1}^{n}(\hat{\eta}_i - \eta_i)\eta_i^{\mathrm{T}}\right\|\right\} \le \frac{1}{n}\sum_{i=1}^{n}\left\{E\left(\|\hat{\eta}_i - \eta_i\|^2\right)E\left(\|\eta_i\|^2\right)\right\}^{1/2} = o(1),$$

where the last equality holds under Condition 5 and Lemma S1.

*S.3.3. Proof of Lemma S3*

*Proof.* It is sufficient to show that for any given $\varepsilon > 0$, there exist a large constant C such that

$$\mathrm{pr}\left\{ \sup_{\|u\|=\mathrm{C}, u\in\mathbb{R}^m} L_n(b_{r0} + {\alpha_n}^{1/2}u) < L_n(b_{r0}) \right\} \geq 1 - \varepsilon. \tag{S33}$$

Equation (S33) implies that there exists a local maximizer $\hat{b}_r$ such that $\|\hat{b}_r - b_{r0}\|^2 = O_p(\alpha_n)$. The true likelihood function $l(\beta)$ can also be regarded as a function of $b_r$. Define $l(b_r)$ as

$$l(b_r) = \frac{1}{n}\sum_{i=1}^{n}\left\{ \left(\eta_i^{\mathrm{T}} b_r + \sum_{k=m+1}^{\infty} \xi_{ik} b_k\right) Y_i - \frac{1}{2}\left(\eta_i^{\mathrm{T}} b_r + \sum_{k=m+1}^{\infty} \xi_{ik} b_k\right)^2 \right\}.$$

It follows from Taylor's expansion that

$$\begin{aligned} & L_n(b_{r0} + {\alpha_n}^{1/2}u) - L_n(b_{r0}) \\ =& {\alpha_n}^{1/2} u^{\mathrm{T}} \left.\frac{\partial L_n(b_r)}{\partial b_r}\right|_{b_r=b_{r0}} + \frac{\alpha_n}{2} u^{\mathrm{T}} \left.\frac{\partial^2 L_n(b_r)}{\partial b_r \partial {b_r}^{\mathrm{T}}}\right|_{b_r=b^*} u \\ =& {\alpha_n}^{1/2} u^{\mathrm{T}} \left.\frac{\partial l(b_r)}{\partial b_r}\right|_{b_r=b_{r0}} + {\alpha_n}^{1/2} u^{\mathrm{T}} \left( \left.\frac{\partial L_n(b_r)}{\partial b_r}\right|_{b_r=b_{r0}} - \left.\frac{\partial l(b_r)}{\partial b_r}\right|_{b_r=b_{r0}} \right) \\ & + \frac{\alpha_n}{2} u^{\mathrm{T}} \left.\frac{\partial^2 L_n(b_r)}{\partial b_r \partial {b_r}^{\mathrm{T}}}\right|_{b_r=b^*} u \\ =& J_1 + J_2 + J_3, \end{aligned}$$

where each element of $b^*$ lies between the corresponding element of $b_{r0}$ and $b_{r0} + {\alpha_n}^{1/2}u$.

By the fact that $E\{(Y_i - \theta_i)\eta_i | X_i\} = 0$. By the multivariate central limit theory,

$$\left\| \left.\frac{\partial l(b_r)}{\partial b_r}\right|_{b_r=b_{r0}} \right\| = \left\| \frac{1}{n}\sum_{i=1}^{n} (Y_i - \theta_i)\eta_i \right\| = O_p\left( \left(\frac{m}{n}\right)^{1/2} \right).$$

Thus

$$J_1 = {\alpha_n}^{1/2} u^{\mathrm{T}} \left.\frac{\partial l(b_r)}{\partial b_r}\right|_{b_r=b_{r0}} \leq {\alpha_n}^{1/2} \left\| \left.\frac{\partial l(b_r)}{\partial b_r}\right|_{b_r=b_{r0}} \right\| \times \|u\| = O_p\left(\alpha_n \|u\|\right).$$

By the law of large numbers,

$$\frac{1}{n}\sum_{i=1}^{n} \eta_i \eta_i^{\mathrm{T}} \xrightarrow{p} I_m,$$

where $I_m$ denotes the identity matrix in $\mathbb{R}^{m\times m}$. Combining this with Lemma S2, there is

$$\left\| \frac{1}{n} \left.\frac{\partial^2 L_n(b_r)}{\partial b_r \partial {b_r}^{\mathrm{T}}}\right|_{b_r=b^*} + I_m \right\| = o_p(1).$$

Therefore, $J_3 = -\alpha_n \|u\|^2 \{1 + o_p(1)\}$.

If we can show that

$$\left\| \left.\frac{\partial L_n(b_r)}{\partial b_r}\right|_{b_r=b_{r0}} - \left.\frac{\partial l(b_r)}{\partial b_r}\right|_{b_r=b_{r0}} \right\| = O_p({\alpha_n}^{1/2}), \tag{S34}$$

then $J_2 = O_p(\alpha_n \|u\|)$ and $J_3$ uniformly dominates both $J_1$ and $J_2$, which leads to (S33). We decompose the left hand side of (S34) into several parts,

$$\begin{aligned}
&\left\| \left.\frac{\partial L_n(b_r)}{\partial b_r}\right|_{b_r=b_{r0}} - \left.\frac{\partial l(b_r)}{\partial b_r}\right|_{b_r=b_{r0}} \right\| \\
=&\left\| \frac{1}{n}\sum_{i=1}^{n}(\theta_i - \hat{\theta}_{i,m})\hat{\eta}_i + \frac{1}{n}\sum_{i=1}^{n}(Y_i - \theta_i)(\hat{\eta}_i - \eta_i) \right\| \\
\leq&\left\| \frac{1}{n}\sum_{i=1}^{n}(\theta_i - \hat{\theta}_{i,m})\hat{\eta}_i \right\| + \left\| \frac{1}{n}\sum_{i=1}^{n}(Y_i - \theta_i)(\hat{\eta}_i - \eta_i) \right\| \\
=&H_1 + H_2.
\end{aligned}$$

Consider $H_2$ first,

$$H_2^2 = \frac{1}{n^2}\sum_{i_1 \neq i_2}(Y_{i_1} - \theta_{i_1})(Y_{i_2} - \theta_{i_2})(\hat{\eta}_{i_1} - \eta_{i_1})^{\mathrm{T}}(\hat{\eta}_{i_2} - \eta_{i_2}) + \frac{1}{n^2}\sum_{i=1}^{n}(Y_i - \theta_i)^2\|\hat{\eta}_i - \eta_i\|^2.$$

Notice that $E\{(Y_{i_1} - \theta_{i_1})(\hat{\eta}_{i_1} - \eta_{i_1}) \mid X_i, T_{ij}, \varepsilon_{ij}\} = 0$,

$$\begin{aligned}
E(H_2^2) =& \frac{1}{n^2}\sum_{i=1}^{n} E\left\{(Y_i - \theta_i)^2\|\hat{\eta}_i - \eta_i\|^2\right\} \\
=&\frac{1}{n^2}\sum_{i=1}^{n} E\left[E\left\{(Y_i - \theta_i)^2\|\hat{\eta}_i - \eta_i\|^2 \mid X_i, T_{ij}, \varepsilon_{ij}\right\}\right] \\
=&\frac{\sigma_Y^2}{n^2}\sum_{i=1}^{n} E(\|\hat{\eta}_i - \eta_i\|)^2 = o(\alpha_n),
\end{aligned} \tag{S35}$$

where the last equality holds under Lemma S1 and Condition 5.

As for $H_1$,

$$H_1 \leq \left\| \frac{1}{n}\sum_{i=1}^{n}(\theta_i - \hat{\theta}_{i,m})\eta_i \right\| + \left\| \frac{1}{n}\sum_{i=1}^{n}(\theta_i - \hat{\theta}_{i,m})(\hat{\eta}_i - \eta_i) \right\| = H_{11} + H_{12}.$$

For $H_{12}$,

$$\begin{aligned}
E\left(H_{12}\right) \leq& \frac{1}{n}\sum_{i=1}^{n} E\{\|(\theta_i - \hat{\theta}_{i,m})(\hat{\eta}_i - \eta_i)\|\} \\
\leq&\frac{1}{n}\sum_{i=1}^{n}[E\{(\theta_i - \hat{\theta}_{i,m})^2\}E\left(\|\hat{\eta}_i - \eta_i\|^2\right)]^{1/2} = O({\alpha_n}^{1/2}).
\end{aligned} \tag{S36}$$

As for $H_{11}$,

$$H_{11}^2 \leq 2\left\| \frac{1}{\lfloor n/2 \rfloor}\sum_{i=1}^{\lfloor n/2 \rfloor}(\theta_i - \hat{\theta}_{i,m})\eta_i \right\|^2 + 2\left\| \frac{1}{\lfloor n/2 \rfloor}\sum_{i=\lfloor n/2 \rfloor+1}^{n}(\theta_i - \hat{\theta}_{i,m})\eta_i \right\|^2. \tag{S37}$$

For the fact that both terms on the right hands of (S37) admit the same convergence rates, we only need to calculate one of them.

$$
\begin{aligned}
&E\left\{\left\|\frac{1}{\lfloor n/2\rfloor}\sum_{i=1}^{\lfloor n/2\rfloor}(\theta_i-\hat{\theta}_{i,m})\eta_i\right\|^2\right\}\\
=&\frac{1}{\lfloor n/2\rfloor^2}\left[\sum_{i_1\neq i_2}^{\lfloor n/2\rfloor}E\{(\theta_{i_1}-\hat{\theta}_{i_1,m})(\theta_{i_2}-\hat{\theta}_{i_2,m})\eta_{i_1}^{\mathrm{T}}\eta_{i_2}\}+\sum_{i=1}^{\lfloor n/2\rfloor}E\{(\theta_i-\hat{\theta}_{i,m})^2\|\eta_i\|^2\}\right].
\end{aligned}
\tag{S38}
$$

Notice that $\|\eta_i\|^2=\sum_{k=1}^m\lambda_k^{-1}\xi_{ik}^2$ and by similar arguments in proofing the second statement of Lemma S1, we obtain

$$
\frac{1}{\lfloor n/2\rfloor^2}\sum_{i=1}^{\lfloor n/2\rfloor}E\{(\theta_i-\hat{\theta}_{i,m})^2\|\eta_i\|^2\}=O\left(\frac{m}{n}\left\{\frac{1}{n}\left(1+\frac{1}{Nh}\right)+\frac{1}{N}\right\}\right)=o(\alpha_n).
$$

As for the first term in the right hand side of (S38), given $\hat{\phi}_{(2),k}$, $(\theta_{i_1}-\hat{\theta}_{i_1,m})\eta_{i_1}$ and $(\theta_{i_2}-\hat{\theta}_{i_2,m})\eta_{i_2}$ are independent for $1\le i_1\neq i_2\le\lfloor n/2\rfloor$. Thus

$$
\frac{1}{\lfloor n/2\rfloor^2}\sum_{i_1\neq i_2}^{\lfloor n/2\rfloor}E\{(\theta_{i_1}-\hat{\theta}_{i_1,m})(\theta_{i_2}-\hat{\theta}_{i_2,m})\eta_{i_1}^{\mathrm{T}}\eta_{i_2}\}\asymp E[\|E\{(\theta_{i_1}-\hat{\theta}_{i_1,m})\eta_{i_1}\mid\hat{\phi}_{(2)}\}\|^2].
\tag{S39}
$$

By expansion (S9), the $j$th element of $E\{(\theta_{i_1}-\hat{\theta}_{i_1,m})\eta_{i_1}\mid\hat{\phi}_{(2)}\}$ is

$$
\begin{aligned}
&E\left(\sum_{k=1}^m\langle X_i,\hat{\phi}_{(2),k}-\phi_k\rangle b_{0k}\frac{\xi_{ij}}{\lambda_j^{1/2}}\mid\hat{\phi}_{(2)}\right)\\
=&E\left\{\sum_{k=1}^m\int X_i(s)\sum_{l\neq k}\frac{\langle\Delta_{(2)}\phi_k,\phi_l\rangle}{\lambda_k-\lambda_l}\phi_l(s)\mathrm{d}s b_{0k}\frac{\xi_{ij}}{\lambda_j{}^{1/2}}\mid\hat{\phi}_{(2)}\right\}\\
&+E\left\{\sum_{k=1}^m\int X_i(s)\sum_{l\neq k}\frac{\langle\Delta_{(2)}(\hat{\phi}_{(2),k}-\phi_k),\phi_l\rangle}{\lambda_k-\lambda_l}\phi_l(s)\mathrm{d}s b_{0k}\frac{\xi_{ij}}{\lambda_j^{1/2}}\mid\hat{\phi}_{(2)}\right\}\\
&+E\left\{\sum_{k=1}^m\int X_i(s)\sum_{l\neq k}\sum_{s=1}^{\infty}\frac{(\lambda_k-\hat{\lambda}_{(2),k})^s}{(\lambda_k-\lambda_l)^{s+1}}\langle\Delta_2\hat{\phi}_{(2),k},\phi_l\rangle\phi_l(s)\mathrm{d}s b_{0k}\frac{\xi_{ij}}{\lambda_j^{1/2}}\mid\hat{\phi}_{(2)}\right\}\\
&+E\left\{\sum_{k=1}^m\int X_i(s)\langle\hat{\phi}_{(2),k}-\phi_k,\phi_k\rangle\phi_k(s)\mathrm{d}s b_{0k}\frac{\xi_{ij}}{\lambda_j^{1/2}}\mid\hat{\phi}_{(2)}\right\}\\
=&G_{1,j}+G_{2,j}+G_{3,j}+G_{4,j},
\end{aligned}
\tag{S40}
$$

where

$$
\begin{aligned}
G_{1,j}=&\lambda_j^{1/2}\sum_{k\neq j}^m\frac{\langle\Delta_{(2)}\phi_k,\phi_j\rangle}{\lambda_k-\lambda_j}b_{0k},\quad G_{2,j}=\lambda_j^{1/2}\sum_{k\neq j}^m\frac{\langle\Delta_{(2)}(\hat{\phi}_{(2),k}-\phi_k),\phi_j\rangle}{\lambda_k-\lambda_j}b_{0k},\\
G_{3,j}=&\lambda_j^{1/2}\sum_{k\neq j}\sum_{s=1}^{\infty}\frac{(\lambda_k-\hat{\lambda}_{(2),k})^s}{(\lambda_k-\lambda_j)^{s+1}}\langle\Delta_{(2)}\hat{\phi}_{(2),k},\phi_j\rangle b_{0k},\quad G_{4,j}=\lambda_j^{1/2}\langle\hat{\phi}_{(2),j}-\phi_j,\phi_j\rangle b_{0j}.
\end{aligned}
$$

Start with $G_{1,j}$, by Cauchy–Schwarz inequality, Theorem 1 and Lemma 7 in Dou et al. (2012),

$$\begin{aligned}
E\left(\sum_{j=1}^{m} G_{1,j}^2\right) &= \sum_{j=1}^{m} \lambda_j E\left(\sum_{k\neq j}^{m} \frac{\langle \Delta_{(2)}\phi_k, \phi_j\rangle}{\lambda_k-\lambda_j} b_{0k}\right)^2 \\
&= \sum_{j=1}^{m} \lambda_j E\left\{\sum_{k_1,k_2\neq j}^{m} \frac{\langle \Delta_{(2)}\phi_{k_1}, \phi_j\rangle \langle \Delta_{(2)}\phi_{k_2}, \phi_j\rangle}{(\lambda_{k_1}-\lambda_j)(\lambda_{k_2}-\lambda_j)} b_{0k_1} b_{0k_2}\right\}. \\
&\le \sum_{j=1}^{m} \lambda_j \sum_{k_1,k_2\neq j}^{m} \frac{\left\{E(\langle \Delta_{(2)}\phi_{k_1}, \phi_j\rangle^2) E(\langle \Delta_{(2)}\phi_{k_1}, \phi_j\rangle^2)\right\}^{1/2} b_{0k_1} b_{0k_2}}{(\lambda_{k_1}-\lambda_j)(\lambda_{k_2}-\lambda_j)} \\
&= \sum_{j=1}^{m} \lambda_j \left\{\sum_{k\neq j}^{m} \frac{\left\{E(\langle \Delta_{(2)}\phi_{k_1}, \phi_j\rangle^2)\right\}^{1/2}}{\lambda_k-\lambda_j} b_{0k}\right\}^2 \\
&\lesssim \sum_{j=1}^{m} \lambda_j \left[\sum_{k\neq j}^{m} \frac{b_{0k}}{\lambda_k-\lambda_j} \left\{\frac{j^{-\frac{a}{2}} k^{-\frac{a}{2}}}{n^{\frac{1}{2}}} + h^2\left(k^c j^{-a} + k^{-a} j^c\right)\right\}\right]^2 \\
&\lesssim \sum_{j=1}^{m} \left(\frac{1}{n} + h^4 j^{2c}\right) j^{-a} = O\left(\frac{1}{n} + h^4 m^{2c-a+1}\right) = o(\alpha_n).
\end{aligned} \tag{S41}$$

For $G_{2,j}$,

$$\begin{aligned}
\sum_{j=1}^{m} G_{2,j}^2 &= \sum_{j=1}^{m} \lambda_j \left\{\sum_{k\neq j}^{m} \frac{\langle \Delta_{(2)}(\hat{\phi}_{(2),k}-\phi_k), \phi_j\rangle}{\lambda_k-\lambda_j} b_{0k}\right\}^2 \\
&= \sum_{j=1}^{m} \lambda_j \sum_{k_1,k_2\neq j}^{m} \frac{\langle \Delta_{(2)}(\hat{\phi}_{(2),k_1}-\phi_{k_1}), \phi_j\rangle \langle \Delta_{(2)}(\hat{\phi}_{(2),k_2}-\phi_{k_2}), \phi_j\rangle}{(\lambda_{k_1}-\lambda_j)(\lambda_{k_2}-\lambda_j)} b_{0k_1} b_{0k_2} \\
&\le \|\Delta_{(2)}\|_{\text{HS}}^2 \sum_{j=1}^{m} \lambda_j \sum_{k_1,k_2\neq j}^{m} \frac{\|\hat{\phi}_{(2),k_1}-\phi_{k_1}\| \|\hat{\phi}_{(2),k_2}-\phi_{k_2}\|}{|\lambda_{k_1}-\lambda_j||\lambda_{k_2}-\lambda_j|} b_{0k_1} b_{0k_2}.
\end{aligned} \tag{S42}$$

By Theorem 2, Lemma 7 in Dou et al. (2012) and Cauchy–Schwarz inequality,

$$\begin{aligned}
E&\left(\sum_{j=1}^{m} \lambda_j \sum_{k_1,k_2\neq j}^{m} \frac{\|\hat{\phi}_{(2),k_1}-\phi_{k_1}\| \|\hat{\phi}_{(2),k_2}-\phi_{k_2}\|}{|\lambda_{k_1}-\lambda_j||\lambda_{k_2}-\lambda_j|} b_{0k_1} b_{0k_2}\right) \\
&\le \sum_{j=1}^{m} \lambda_j \sum_{k_1,k_2\neq j}^{m} \frac{\{E(\|\hat{\phi}_{(2),k_1}-\phi_{k_1}\|^2) E(\|\hat{\phi}_{(2),k_2}-\phi_{k_2}\|^2)\}^{1/2}}{|\lambda_{k_1}-\lambda_j||\lambda_{k_2}-\lambda_j|} b_{0k_1} b_{0k_2} \\
&= \sum_{j=1}^{m} \lambda_j \left(\sum_{k\neq j} \frac{\{E(\|\hat{\phi}_{(2),k}-\phi_k\|^2)\}^{1/2}}{|\lambda_k-\lambda_j|} b_k\right)^2 \\
&\lesssim \frac{1}{n}\left\{1 + \frac{1}{Nh} + m^{2a+5-2b} \log m \left(1 + \frac{1}{Nh}\right)\right\} + h^4 m^{3a+5-2b+2c} \log m.
\end{aligned} \tag{S43}$$

Thus,

$$\sum_{j=1}^{m} G_{2,j}^2 = O_p(\|\Delta_{(2)}\|_{\mathrm{HS}}^2) O_p\left(\sum_{j=1}^{m} \lambda_j \sum_{k_1,k_2\neq j}^{m} \frac{\|\hat{\phi}_{(2),k_1}-\phi_{k_1}\|\|\hat{\phi}_{(2),k_2}-\phi_{k_2}\|}{|\lambda_{k_1}-\lambda_j||\lambda_{k_2}-\lambda_j|} b_{0k_1} b_{0k_2}\right)$$
$$= o_p(\alpha_n).$$

For $G_{3,j,1}$,

$$\begin{aligned}\sum_{j=1}^{m} G_{3,j}^2 =& \sum_{j=1}^{m} \lambda_j \left\{\sum_{k\neq j}^{m}\sum_{s=1}^{\infty} \frac{(\lambda_k-\hat{\lambda}_{(2),k})^s}{(\lambda_k-\lambda_j)^{s+1}} \langle \Delta_{(2)}\hat{\phi}_{(2),k}, \phi_j\rangle b_{0k}\right\}^2 \\ \leq& 2\sum_{j=1}^{m} \lambda_j \left\{\sum_{k\neq j}^{m}\sum_{s=1}^{\infty} \frac{(\lambda_k-\hat{\lambda}_{(2),k})^s}{(\lambda_k-\lambda_j)^{s+1}} \langle \Delta_{(2)}\phi_{(2),k}, \phi_j\rangle b_{0k}\right\}^2 \\ &+ 2\sum_{j=1}^{m} \lambda_j \left\{\sum_{k\neq j}^{m}\sum_{s=1}^{\infty} \frac{(\lambda_k-\hat{\lambda}_{(2),k})^s}{(\lambda_k-\lambda_j)^{s+1}} \langle \Delta_{(2)}(\hat{\phi}_{(2),k}-\phi_k), \phi_j\rangle b_{0k}\right\}^2 \\ =& G_{3,j,1} + G_{3,j,2}.\end{aligned}$$

For $G_{3,j,1}$,

$$\begin{aligned}G_{3,j,1} =& 2\sum_{j=1}^{m} \lambda_j \sum_{k_1\neq k_2\neq j}^{m} \left\{\sum_{s=1}^{\infty} \frac{(\lambda_{k_1}-\hat{\lambda}_{(2),k_1})^s}{(\lambda_{k_1}-\lambda_j)^{s+1}}\right\}\left\{\sum_{s=1}^{\infty} \frac{(\lambda_{k_2}-\hat{\lambda}_{(2),k_2})^s}{(\lambda_{k_2}-\lambda_j)^{s+1}}\right\} \\ & \times \langle \Delta_{(2)}\phi_{(2),k_1}, \phi_j\rangle\langle \Delta_{(2)}\phi_{(2),k_2}, \phi_j\rangle b_{0k_1}b_{0k_2} \\ =& 2\sum_{j=1}^{m} \lambda_j \sum_{k_1\neq k_2\neq j}^{m} \frac{\lambda_{k_1}-\hat{\lambda}_{(2),k_1}}{(\lambda_{k_1}-\lambda_j)(\hat{\lambda}_{(2),k_1}-\lambda_j)} \frac{\lambda_{k_2}-\hat{\lambda}_{(2),k_2}}{(\lambda_{k_2}-\lambda_j)(\hat{\lambda}_{(2),k_2}-\lambda_j)} \\ & \times \langle \Delta_{(2)}\phi_{(2),k_1}, \phi_j\rangle\langle \Delta_{(2)}\phi_{(2),k_2}, \phi_j\rangle b_{0k_1}b_{0k_2} \\ \lesssim& \frac{\|\Delta_{(2)}\|_{\mathrm{HS}}^2}{(2\eta_m-\|\Delta_{(2)}\|_{\mathrm{HS}})^2} \sum_{j=1}^{m}\lambda_j E\left\{\sum_{k_1,k_2\neq j}^{m} \frac{\langle \Delta_{(2)}\phi_{k_1}, \phi_j\rangle\langle \Delta_{(2)}\phi_{k_2}, \phi_j\rangle}{(\lambda_{k_1}-\lambda_j)(\lambda_{k_2}-\lambda_j)} b_{0k_1}b_{0k_2}\right\} \\ \lesssim& \sum_{j=1}^{m} G_{1,j}^2 = o_p(\alpha_n).\end{aligned} \tag{S44}$$

Similarly,

$$\begin{aligned}G_{3,j,2} =& 2\sum_{j=1}^{m} \lambda_j \left\{\sum_{k\neq j}^{m}\sum_{s=1}^{\infty} \frac{(\lambda_k-\hat{\lambda}_{(2),k})^s}{(\lambda_k-\lambda_j)^{s+1}} \langle \Delta_{(2)}(\hat{\phi}_{(2),k}-\phi_k), \phi_j\rangle b_{0k}\right\}^2 \\ \lesssim& \frac{\|\Delta_{(2)}\|_{\mathrm{HS}}^2}{(2\eta_m-\|\Delta_{(2)}\|_{\mathrm{HS}})^2} \sum_{j=1}^{m}\lambda_j \left\{\sum_{k\neq j}^{m} \frac{\langle \Delta_{(2)}(\hat{\phi}_{(2),k}-\phi_k), \phi_j\rangle}{\lambda_k-\lambda_j} b_{0k}\right\}^2 \\ \lesssim& \sum_{j=1}^{m} G_{2,j}^2 = o_p(\alpha_n).\end{aligned} \tag{S45}$$

For the last term $G_{4,j}$, by the fact that $|\langle \hat{\phi}_{(2),j} - \phi_j, \phi_j \rangle| = \|\hat{\phi}_{(2),j} - \phi_j\|^2/2$ and the perturbation results in Bosq (2000),

$$
\begin{aligned}
&\sum_{j=1}^{m} G_{4,j}^2 \lesssim \sum_{j=1}^{m} \lambda_j b_j^2 \|\hat{\phi}_{(2),j} - \phi_j\|^4 \leq \sum_{j=1}^{m} \lambda_j b_j^2 \|\hat{\phi}_{(2),j} - \phi_j\|^2 \frac{\|\Delta_{(2)}\|_{\text{HS}}^2}{\eta_j^2} \\
\lesssim & \|\Delta_{(2)}\|_{\text{HS}}^2 \sum_{j=1}^{m} j^{a+2-2b} \|\hat{\phi}_{(2),j} - \phi_j\|^2 = o_p(\alpha_n),
\end{aligned} \tag{S46}
$$

where the calculation of last equality is analogous to (S43). Combing equation (S40) to (S46), we get

$$
E[\|E\{(\theta_{i_1} - \hat{\theta}_{i_1})\eta_{i_1} | \hat{\phi}_{(2)}\}\|^2] = o(\alpha_n).
$$

Then equation (S34) holds by combing of equation (S35)– (S39). The proof of Lemma S3 is complete.

### S.3.4. *Proof of Lemma S4*

*Proof.* For any $\theta = (\theta_1, \ldots, \theta_j, \ldots, \theta_r)^{\mathrm{T}} \in \{0,1\}^r$, define $\theta^j = (\theta_1, \ldots, 1 - \theta_j, \ldots, \theta_r)^{\mathrm{T}}$ as the perturbation of $\theta$ such that $\theta^j$ differs from $\theta$ only in the $j$th position. By assumptions,

$$
\begin{aligned}
&\max_{\theta \in \{0,1\}^r} E_\theta\{d(T, \psi(\theta))\} \geq \frac{1}{2^r} \sum_{\theta \in \{0,1\}^r} \sum_{j=1}^{r} E_\theta\{d_j(T, \psi(\theta))\} \\
=& \frac{1}{2^r} \sum_{j=1}^{r} \sum_{\theta \in \{0,1\}^r} E_\theta\{d_j(T, \psi(\theta))\} = \frac{1}{2^r} \sum_{j=1}^{r} \sum_{\theta \in \{0,1\}^r} \frac{E_\theta\{d_j(T, \psi(\theta))\} + E_{\theta^j}\{d_j(T, \psi(\theta^j))\}}{2} \\
=& \frac{1}{2^{r+1}} \sum_{j=1}^{r} \sum_{\theta \in \{0,1\}^r} \left( E_\theta \left[ \{d_j(T, \psi(\theta)) + d_j(T, \psi(\theta^j))\} \frac{d_j(T, \psi(\theta))}{d_j(T, \psi(\theta)) + d_j(T, \psi(\theta^j))} \right] \right. \\
&\left. + E_{\theta^j} \left[ \{d_j(T, \psi(\theta)) + d_j(T, \psi(\theta^j))\} \frac{d_j(T, \psi(\theta^j))}{d_j(T, \psi(\theta)) + d_j(T, \psi(\theta^j))} \right] \right) \\
\geq& \frac{A}{2^{r+1}} \sum_{j=1}^{r} \sum_{\theta \in \{0,1\}^r} \left( \left[ E_\theta \left\{ d_j(\psi(\theta), \psi(\theta^j)) \frac{d_j(T, \psi(\theta))}{d_j(T, \psi(\theta)) + d_j(T, \psi(\theta^j))} \right\} \right. \right. \\
&\left. \left. + E_{\theta^j} \left\{ d_j(\psi(\theta), \psi(\theta^j)) \frac{d_j(T, \psi(\theta^j))}{d_j(T, \psi(\theta)) + d_j(T, \psi(\theta^j))} \right\} \right] \right) \\
\geq& \sum_{j=1}^{r} \frac{A g_{jr}}{2^{r+1}} \sum_{\theta \in \{0,1\}^r} \inf_{f_i \geq 0; f_1 + f_2 = 1} \{E_\theta(f_1) + E_{\theta^j}(f_2)\} \\
=& \sum_{j=1}^{r} \frac{A g_{jr}}{2^{r+1}} \sum_{\theta \in \{0,1\}^r} \|P_\theta \wedge P_{\theta^j}\|_a \geq \frac{A g_r}{2} \min_{H(\theta, \theta')=1} \|P_\theta \wedge P_{\theta'}\|_a,
\end{aligned}
$$

where the second equality is due to the identity $\sum_{j=1}^{r} \sum_{\theta \in \{0,1\}^r} E_\theta\{d_j(T, \psi(\theta))\} = \sum_{j=1}^{r} \sum_{\theta \in \{0,1\}^r} E_{\theta^j}\{d_j(T, \psi(\theta^j))\}$, the third last inequality is by $d(x, z) + d(z, y) \geq Ad(x, y)$, the second last inequality is due to $d_j\left(\psi(\theta), \psi(\theta^j)\right) \geq g_{jr}$, the last equality is from the definition of $\|P_\theta \wedge P_{\theta^j}\|_a$ (Tsybakov, 2008, Lemma 2.1), and the last inequality is by $g_r = \sum_{j=1}^{r} g_{jr}$, and $\|P_\theta \wedge P_{\theta^j}\|_a \geq \min_{H(\theta, \theta')=1} \|P_\theta \wedge P_{\theta'}\|_a$. □

## REFERENCES


BOSQ, D. (2000). *Linear Processes in Function Spaces*. New York: Springer.

DOU, W. W., POLLARD, D. & ZHOU, H. H. (2012). Estimation in functional regression for general exponential families. *Ann. Statist.* **40**, 2421–2451.

HALL, P. & HOROWITZ, J. L. (2007). Methodology and convergence rates for functional linear regression. *Ann. Statist.* **35**, 70–91.

HALL, P. & HOSSEINI-NASAB, M. (2006). On properties of functional principal components. *J. R. Statist. Soc.* B **68**, 109–126.

HSING, T. & EUBANK, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Chichester: Wiley.

LI, Y. & HSING, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *Ann. Statist.* **38**, 3321–3351.

TSYBAKOV, A. B. (2008). *Introduction to nonparametric estimation*. Springer Science & Business Media.

YU, B. (1997). Assouad, fano, and le cam. In *Festschrift for Lucien Le Cam*. Springer, pp. 423–435.

ZHANG, X. & WANG, J.-L. (2016). From sparse to dense functional data and beyond. *Ann. Statist.* **44**, 2281–2321.