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 $m \in \mathbb{N}_+$,

- $D = \mathrm{diag}\{\ ^{1=2}_{1}; \cdots; \ ^{1=2}_{m}\}$, $\ _i = \langle X_i; \ \rangle$, $\ _i = (\ _{i1}; \cdots; \ _{im})^{\mathsf{T}}$ and $\ _i = D^{-1}\ _i$;
- $b_0 = (b_{01}; \cdots; b_{0m})^{\mathsf{T}}$, $\ _{i;m} = \ _i^{\mathsf{T}} b_0$ and $b_{r0} = D b_0$;
- $\hat{\ }_i = (\hat{\ }_{i1}; \cdots; \hat{\ }_{im})^{\mathsf{T}}$, $\hat{\ }_{i;m} = \hat{\ }_i^{\mathsf{T}} b_0$ and $\hat{\ }_i = D^{-1}\hat{\ }_i$.

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 $\|\hat{\ }_i - \ _i\|^2$ and $(\ _i - \hat{\ }_{i;m})^2$.

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

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

and

$$E(\ _i - \hat{\ }_{i;m})^2 \lesssim \frac{1}{N} + \frac{1}{n}\left(1 + \frac{1}{Nh}\right):$$

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 $\Omega_m(n; N)$,*

$$\left\| \frac{1}{n}\sum_{i=1}^{n} \hat{\ }_i \hat{\ }_i^{\mathsf{T}} - \frac{1}{n}\sum_{i=1}^{n} \ _i \ _i^{\mathsf{T}} \right\| = o_p(1):$$

© 2020 Biometrika Trust

Due to the fact that the variance of ξ_{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

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

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* $\Omega_m(n; N)$,

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

where

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

The following lemma is to establish the minimax lower bound for the prediction error. We define that P_θ is a family of probability measures, where θ is the parameter of interest and the corresponding expectation operator is denoted as E_θ. Let $H(\theta; \theta') = \sum_{i=1}^{r} |\theta_i - \theta'_i|$ be the Hamming distance between the binary sequences $\theta = (\theta_1, \ldots, \theta_r)^{\mathsf{T}}$ and $\theta' = (\theta'_1, \ldots, \theta'_r)^{\mathsf{T}}$ on $\{0, 1\}^r = \{(\theta_1, \ldots, \theta_j, \ldots, \theta_r)^{\mathsf{T}} \mid \theta_j = 0 \text{ or } \theta_j = 1\}$. In addition, for the probability measures P_θ and $P_{\theta'}$ with density function p_θ and $p_{\theta'}$ jointly dominated by μ, 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_θ; $\theta \in \{0, 1\}^r$. Let $\tau(\theta)$ be an arbitrary transform of the parameter θ. Consider the pseudo-distance $d(\cdot;\cdot)$ satisfying weak triangle inequality $d(x; z) + d(z; y) \geq A d(x; y)$ with $A \in (0; 1)$ and $d(x; y) = \sum_{j=1}^{r} d_j(x; y)$. If $d_j(\tau(\theta); \tau(\theta')) \geq g_{jr} > 0$ for $H(\theta; \theta') = 1$ such that θ and θ' differ only in the jth coordinate, then for the distance $d(T; \tau(\theta))$,*

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

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 $\tau(\theta)$ 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 $E(\langle \Delta_{(r)k}; \xi_j \rangle^2)$. 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 $\xi_{il_1 l_2} = X_{il_1} X_{il_2}$,

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

We first calculate the bias term,

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

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

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

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

$$
{}_1\left\|\frac{1}{h}\int \mathrm{K}\left(\frac{v-t}{h}\right) \ \ _j(t)\mathrm{d}t- \ \ _j(v)\right\|\left\|\frac{1}{h}\int \mathrm{K}\left(\frac{v-t}{h}\right) \ \ _k(t)\mathrm{d}t- \ \ _k(v)\right\|\lesssim h^4j^ck^c: \tag{S3}
$$

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

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

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 $E(\langle \Delta_{(1)} \ \ _k; \ \ _j\rangle)\lesssim h^2(j^{-a}k^c+k^{-a}j^c)$.

For the variance term, denote

$$
A_i(\ \ _k; \ \ _j)=\sum_{l_1\neq l_2} \ \ _{il_1l_2}\frac{1}{h}\int \mathrm{K}\left(\frac{T_{il_1}-s}{h}\right) \ \ _k(s)\mathrm{d}s\frac{1}{h}\int \mathrm{K}\left(\frac{T_{il_2}-t}{h}\right) \ \ _j(t)\mathrm{d}t
$$

and there is

$$
\mathrm{var}(\langle \Delta_{(1)} \ \ _k; \ \ _j\rangle)\le\frac{1}{n}\frac{1}{N^2(N-1)^2}E\{A_i(\ \ _k; \ \ _j)\}^2:
$$

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

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

with

$$A_{i1}(\xi_j, \xi_k) = E\left[\left\{\int X(u)\,\xi_{k;h}(u)\mathrm{d}u\right\}^2\left\{\int X(u)\,\xi_{j;h}(u)\mathrm{d}u\right\}^2\right]$$

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

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

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

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

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

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 $\|\cdot\|_j$,

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

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),

$$
\begin{aligned}
&\int \left[E\left\{ \int \hat{C}_{(1)}(s;t)\ \ _j(t)\mathrm{d}t \right\} - \int C(s;t)\ \ _j(t)\mathrm{d}t \right]^2 \mathrm{d}s \\
&= \int \left\{ \int C_h(s;t)\ \ _{j;h}(t)\mathrm{d}t - \int C(s;t)\ \ _j(t)\mathrm{d}t \right\}^2 \mathrm{d}s \\
&\lesssim \int \left[\int C_h(s;t)\{\ \ _{j;h}(t) - \ \ _j(t)\}\mathrm{d}t \right]^2 \mathrm{d}s + \int \left[\int \{C_h(s;t) - C(s;t)\}\ \ _j(t)\mathrm{d}t \right]^2 \mathrm{d}s \\
&\leq \|\ \ _{j;h} - \ \ _j\|^2 \int \|C_h(s;\cdot)\|^2 \mathrm{d}s +
\end{aligned}
$$

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

$$\begin{aligned}\hat{}_{(1);j} - \ _j = &\sum_{k\neq j} \frac{\int(\hat{C}_{(1)} - C)\ _j\ _k}{(\ _j - \ _k)}\ _k + \sum_{k\neq j} \frac{\int(\hat{C}_{(1)} - C)(\hat{}_{(1);j} - \ _j)\ _k}{(\ _j - \ _k)}\ _k \\ &+ \sum_{k\neq j}\sum_{s=1}^{\infty} \frac{(\ _j - \hat{}_{(1);j})^s}{(\ _j - \ _k)^{s+1}} \left\{\int(\hat{C}_{(1)} - C)\hat{}_{(1);j}\ _k\right\}\ _k \\ &+ \left\{\int(\hat{}_{(1);j} - \ _j)\ _j\right\}\ _j;\end{aligned} \tag{S9}$$

such kind of expansions can be found in Li & Hsing (2010) and Hall & Hosseini-Nasab (2006). The bound for $\|\hat{}_{(1);j} - \ _j\|^2$ 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 $\ _j$ and $\|\cdot\|_j^2$,

$$\sum_{k\neq j}(\ _j - \ _k)^{-2}\left\{\int(\hat{C}_{(1)} - C)\ _j\ _k\right\}^2 \leq \ _j^{-2}\|\Delta_{(1)}\|_j^2: \tag{S10}$$

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

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

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

$$\begin{aligned}E\left\{\left\|\sum_{k\neq j}\frac{\int(\hat{C}_{(1)} - C)(\hat{}_{(1);j} - \ _j)\ _k}{(\ _j - \ _k)}\ _k\right\|^2\right\} \leq &E\left\{\frac{\|\hat{C}_{(1)} - C\|^2\|\hat{}_{(1);j} - \ _j\|^2}{(2\ _j)^2}\right\} \\ < &\frac{1}{16}E(\|\hat{}_{(1);j} - \ _j\|^2);\end{aligned} \tag{S12}$$

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

$$
\begin{aligned}
& E\left[\left\|\sum_{k\neq j}\sum_{s=1}^{\infty}\frac{(\ _j - \hat{}_{(1);j})^s}{(\ _j - \ _k)^{s+1}}\left\{\int(\hat{C}_{(1)} - C)\,\hat{}_{(1);j}\ _k\right\}\ _k\right\|^2\right] \\
=& E\left[\sum_{k\neq j}\frac{(\ _j - \hat{}_{(1);j})^2}{(\ _j - \ _k)^2(\hat{}_{(1);j} - \ _k)^2}\left\{\int(\hat{C}_{(1)} - C)\,\hat{}_{(1);j}\ _k\right\}^2\right] \\
\leq& 2E\left\{\frac{\|\hat{C}_{(1)} - C\|^2}{(2\ _j - \|\hat{C}_{(1)} - C\|)^2}\left[\sum_{k\neq j}\frac{\{\int(\hat{C}_{(1)} - C)\ _j\ _k\}^2}{(\ _j - \ _k)^2}\right.\right. \\
& \qquad \left.\left. + \sum_{k\neq j}\frac{\{\int(\hat{C}_{(1)} - C)(\hat{}_{(1);j} - \ _j)\ _k\}^2}{(\ _j - \ _k)^2}\right]\right\} \\
\leq& \frac{8}{9}E\left[\frac{\|\hat{C}_{(1)} - C\|^2}{\ _j^2}\sum_{k\neq j}\frac{\{\int(\hat{C}_{(1)} - C)\ _j\ _k\}^2}{(\ _j - \ _k)^2} + \frac{\|\hat{C}_{(1)} - C\|^4}{\ _j^4}\|\hat{}_{(1);j} - \ _j\|^2\right] \\
\leq& \frac{2}{9}E\left[\sum_{k\neq j}\frac{\{\int(\hat{C}_{(1)} - C)\ _j\ _k\}^2}{(\ _j - \ _k)^2}\right] + \frac{1}{18}E(\|\hat{}_{(1);j} - \ _j\|^2):
\end{aligned}
\tag{S13}
$$

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

S.2.3. Proof of Theorem 3

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

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

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,

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

Theorem 2,

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

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

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

where

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

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

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

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

S.2.4. *Proof of Theorem 4*

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

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

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

$$E_* \left\{ \int \hat{\ }^2 (X^*)^2 - \left(\int \hat{\ } X^* \right)^2 \right\} = \sum_{k=1}^{\infty} \ _k \left(\int \hat{\ }^2 \ \frac{2}{k} - \hat{b}_k^2 \right) = O(1).$$

On one hand

$$\begin{aligned}& \left| \sum_{k=1}^{\infty} \ _k \left(\int \hat{\ }^2 \ \frac{2}{k} - \hat{b}_k^2 \right) \right| \leq \left| \sum_{k=1}^{\infty} \ _k \int \hat{\ }^2 \ \frac{2}{k} \right| + \left| \sum_{k=1}^{\infty} \ _k \hat{b}_k^2 \right| \\ \leq & \| \hat{\ } \|^2 \sum_{k=1}^{\infty} \ _k \| \ _k \|_\infty^2 + \sum_{k=1}^{\infty} \ _k \hat{b}_k^2 < \infty;\end{aligned}$$

where the second inequality follows from $\| \ _k \|_\infty = O(1)$ by Condition 3. On the other hand, by Jensen inequality, for any $k \in \mathbb{N}_+$, $\int \hat{\ }^2 \ \frac{2}{k} \geqslant \hat{b}_k^2$ and the equality holds if and only if $\hat{\ }(s) \hat{\ }_k(s) = \int \hat{\ } \hat{\ }_k$ for all $s \in [0; 1]$, which is the trivial case. Thus, without loss of generality, we assume that there exists a $\ > 0$ such that $\int \hat{\ }^2 \ \frac{2}{1} - \hat{b}_1^2 > \ $ and

$$\left| \sum_{k=1}^{\infty} \ _k \left(\int \hat{\ }^2 \ \frac{2}{k} - \hat{b}_k^2 \right) \right| \geqslant \ _1 \left(\int \hat{\ }^2 \ \frac{2}{1} - \hat{b}_1^2 \right) > \mathrm{C}.$$

Then, the discretely observed prediction error becomes

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

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

$$\begin{aligned}& \mathcal{E}(\tilde{\ }_n) = E_*(\langle \hat{\ } - \ ; X^* \rangle^2) \\ \leq & 2E_* \left(\left\langle \hat{\ } - \ ; \sum_{k=1}^{m} \ _k^* \ _k \right\rangle^2 \right) + 2E_* \left(\left\langle \hat{\ } - \ ; \sum_{k=m+1}^{\infty} \ _k^* \ _k \right\rangle^2 \right) \\ = & 2 \sum_{j=1}^{m} \ _j \left\{ \int (\hat{\ } - \ ) \ _j \right\}^2 + 2E_* \left(\left\langle \hat{\ } - \ ; \sum_{k=m+1}^{\infty} \ _k^* \ _k \right\rangle^2 \right) \\ = & I_1 + I_2.\end{aligned}$$

In this proof, we use $\hat{\ } = \sum_{k=1}^{m} \hat{b}_k$

(S9),

$$\begin{aligned}
&\int \hat{}(s)\ \ _j(s)\mathrm{d}s = \int \sum_{k=1}^{m} \hat{b}_k\, \hat{}_{(1);k}(s)\ \ _j(s)\mathrm{d}s \\
=&\int \sum_{k=1}^{m} \hat{b}_k \left(\ \ _k(s) + \sum_{l\neq k} \frac{\langle \Delta_{(1)}\ \ _l;\ \ _k\rangle}{\ \ _k - \ \ _l}\ \ _l(s) + \sum_{l\neq k} \frac{\langle \Delta_{(1)}\ \ _l;\ \hat{}_{(1);k} - \ \ _k\rangle}{\ \ _k - \ \ _l}\ \ _l(s) \right. \\
&\left. + \sum_{l\neq k}\sum_{s=1}^{\infty} \frac{(\ \ _k - \hat{}_{(1);k})^s}{(\ \ _k - \ \ _l)^{s+1}} \langle \Delta_{(1)}\ \ _l;\ \hat{}_{(1);k}\rangle\ \ _l(s) + \langle \hat{}_{(1);k} - \ \ _k;\ \ _k\rangle\ \ _k(s) \right)\ \ _j(s)\mathrm{d}s \\
=&\hat{b}_j + \sum_{k\neq j}^{m} \frac{\langle \Delta_{(1)}\ \ _j;\ \ _k\rangle}{\ \ _k - \ \ _j} \hat{b}_k + \sum_{k\neq j}^{m} \frac{\langle \Delta_{(1)}\ \ _j;\ \hat{}_{(1);k} - \ \ _k\rangle}{\ \ _k - \ \ _j} \hat{b}_k \\
&+ \sum_{k\neq j}^{m}\sum_{s=1}^{\infty} \frac{(\ \ _k - \hat{}_{(1);k})^s}{(\ \ _k - \ \ _j)^{s+1}} \langle \Delta_{(1)}\ \ _j;\ \hat{}_{(1);k}\rangle \hat{b}_k + \langle \hat{}_{(1);j} - \ \ _j;\ \ _j\rangle \hat{b}_j :
\end{aligned}$$

By Cauchy–Schwarz inequality,

$$I_1 \lesssim \sum_{j=1}^{m}\ \ _j(\hat{b}_j - b_{0j})^2 + \sum_{j=1}^{m}\ \ _j \left(\sum_{k\neq j}^{m} \frac{\langle \Delta_{(1)}\ \ _j;\ \ _k\rangle}{\ \ _k - \ \ _j} \hat{b}_k \right)^2 + E; \tag{S18}$$

where

$$\begin{aligned}
E =& \sum_{j=1}^{m}\ \ _j \left(\sum_{k\neq j}^{m} \frac{\langle \Delta_{(1)}\ \ _j;\ \hat{}_{(1);k} - \ \ _k\rangle}{\ \ _k - \ \ _j} \hat{b}_k \right)^2 \\
&+ \sum_{j=1}^{m}\ \ _j \left\{ \sum_{k\neq j}^{m}\sum_{s=1}^{\infty} \frac{(\ \ _k - \hat{}_{(1);k})^s}{(\ \ _k - \ \ _j)^{s+1}} \langle \Delta_{(1)}\ \ _j;\ \hat{}_{(1);k}\rangle \hat{b}_k \right\}^2 + \sum_{j=1}^{m}\ \ _j \langle \hat{}_{(1);j} - \ \ _j;\ \ _j\rangle^2 \hat{b}_j^2
\end{aligned}$$

95 is the remaining term.

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

$$\sum_{j=1}^{m}\ \ _j(\hat{b}_j - b_j)^2 = \|\hat{b}_r - b_{r0}\|^2 = O_p(\ \ _n); \tag{S19}$$

and

$$\begin{aligned}
&\sum_{j=1}^{m}\ \ _j \left(\sum_{k\neq j}^{m} \frac{\langle \Delta_{(1)}\ \ _j;\ \ _k\rangle}{\ \ _k - \ \ _j} \hat{b}_k \right)^2 \\
\leq& 2\sum_{j=1}^{m}\ \ _j \left(\sum_{k\neq j}^{m} \frac{\langle \Delta_{(1)}\ \ _j;\ \ _k\rangle}{\ \ _k - \ \ _j} (\hat{b}_k - b_{0k}) \right)^2 + 2\sum_{j=1}^{m}\ \ _j \left(\sum_{k\neq j}^{m} \frac{\langle \Delta_{(1)}\ \ _j;\ \ _k\rangle}{\ \ _k - \ \ _j} b_{0k} \right)^2 :
\end{aligned} \tag{S20}$$

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

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

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

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

Thus, Under Condition 5

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

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

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

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

$$\begin{aligned} E_4 =& \sum_{j=1}^{m} \ \ _j \left\{ \sum_{k\neq j}^{m} \frac{\langle \Delta_{(1)} \ \ _j ; \hat{\ }_{(1);k} - \ \ _k \rangle}{\ \ _k - \ \ _j} (\hat{b}_k - b_{0k}) \right\}^2 \\ \leq& \sum_{j=1}^{m} \ \ _j \sum_{k\neq j}^{m} \frac{\langle \Delta_{(1)} \ \ _j ; \hat{\ }_{(1);k} - \ \ _k \rangle^2}{(\ \ _k - \ \ _j)^2} \sum_{k=1}^{m} (\hat{b}_k - b_{0k})^2 \\ \leq& m^a \|\hat{b}_r - b_{r0}\|^2 \sum_{j=1}^{m} \ \ _j \frac{\|\Delta_{(1)}\|_{\mathrm{HS}}^2}{\ \ _j^2} \sum_{k=1}^{m} \| \hat{\ }_{(1);k} - \ \ _k \|^2 \\ =& o_p(\ \ _n); \end{aligned} \tag{S21}$$

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

$$E_6 = \sum_{j=1}^{m} \ \ _j \langle \hat{\ }_{(1);j} - \ \ _j ; \ \ _j \rangle^2 (\hat{b}_j - b_{0j})^2 \leq \frac{\|\Delta_{(1)}\|_{\mathrm{HS}}^2}{(2 \ \ _m)^2} \sum_{j=1}^{m} \ \ _j (\hat{b}_j - b_{0j})^2 = o_p(\ \ _n): \tag{S23}$$

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

As for I_2,

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

where

$$I_{22} = E_* \left(\left\langle \sum_{k=m+1}^{\infty} b_{0k} \quad _k; \sum_{k=m+1}^{\infty} \quad ^*_k \quad _k \right\rangle^2 \right) = \sum_{k=m+1}^{\infty} \quad _k b_{0k}^2 = O(m^{1-a-2b}):$$

As for I_{21}, by the orthogonality of the series $\{ \quad _k \}_{k=1}^{\infty}$, k $\quad \left(\quad _{k=m+1} \quad ^* \quad _k \right\rangle^2 \right) \quad \left(\quad _{k=m+1} \quad _k \right)^2 \Big),$

Xfor 0 19

where $e_i \sim N(0, \sigma^2)$ and $E(\epsilon_{ik}^2) = \sigma_k$. Denote P_θ the conditional probability measure of $\{Y_i(\cdot, \cdot)\}_{i=1}^n$ given X, and its corresponding density is denoted q_θ as

It is sufficient to find g_r and a feasible lower bounds for $\min_{H(\theta, \theta_0)=1} E(\|P_{\hat{\theta}} - P_{\theta_0}\|_a)$ by applying Lemma S4 with $\Phi(\theta) = \sum_{k=1}^r \sigma_k b_k \theta_k$. The prediction error

$$d(\hat{\theta}, \theta) = E(\tilde{\epsilon}_n) = \sum_{j=1}^{\infty} \int_{\mathcal{T}_j} (\hat{\theta} - \theta)_j^2 = \sum_{j=1}^{\infty} \lambda_j (\hat{b}_j - b_j)^2 \tag{S24}$$

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

$$d(\theta, \theta^0) = d(\{\theta_k\}, \{\theta_k^0\}) = \sum_{k=1}^r \int_{\mathcal{T}_k} (\theta - \theta^0)_k^2 = \sum_{k=1}^r d_k(\theta, \theta^0),$$

where $d_k(\theta, \theta^0) = \sigma_k^2 (\theta_k - \theta_k^0)^2 b_k^2$, $k = 1, 2, \ldots, r$. Assume θ differs from θ^0 only in the jth coordinate, thus $H(\theta, \theta^0) = 1$, and $d_j(\theta, \theta^0) = \sigma_j^2 b_j^2$, which implies $g_r = \sigma_r^2 \sum_{k=1}^r \sigma_k b_k^2 = \sigma_r^2 O(r^{1-a-2b})$. For any estimator $\hat{\theta} = \sum_{j \in \mathbb{N}_+} \hat{\theta}_j \phi_j$ based on $\{X_i, Y_i\}_{i=1}^n$, apply Lemma S4 with $A = 1/2$

$$\sup_{X, \theta \in \mathcal{G}} E\{E(\tilde{\epsilon}_n)\} \geq \max_{\theta \in \{0,1\}^r} E\{d(\hat{\theta}, \theta)\} \geq \frac{g_r}{4} \min_{H(\theta, \theta^0)=1} E(\|P_{\hat{\theta}} - P_{\theta_0}\|_a). \tag{S25}$$

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

$$E(\|P_{\hat{\theta}} - P_{\theta_0}\|_a) \leq 1 - E[\{K(P_\theta \mid P_{\theta_0})/2\}^{1/2}]. \tag{S26}$$

To guarantee the positiveness of the right hand side of (S26), we need to show that $\{K(P_\theta \mid P_{\theta_0})/2\}^{1/2}$ is sufficient small for a suitable $\theta > 0$. For a fixed a, $j \in \{1, 2, \ldots, r\}$, the log-likelihood ratio for normal noise in Condition 7 with $\sigma_k^{-2} = (2\sigma^2)^{-1}$ is

$$\log \frac{q_\theta}{q} = \frac{1}{2} \sum_{i=1}^n \left\{ \int_{\mathcal{T}} Y_i \, d(\theta - \theta^0) - \int_{\mathcal{T}} X_i (\theta - \theta^0) \int_{\mathcal{T}} X_i (\theta - \theta^0) \right\}^2.$$

Note that $K(P_\theta \mid P_{\theta_0}) = E\{\log(q_{\theta_0}/q_\theta) \mid X\}$,

$$\begin{aligned} E[\{K(P_\theta \mid P_{\theta_0})/2\}^{1/2}] &= E\left[\frac{n}{4\sigma^2} \sum_{k=1}^r (\theta_k - \theta_k^0)^2 b_k^2 \lambda_{1k}^2 \right]^{1/2} \\ &\geq \left(\frac{n}{4\sigma^2} \sum_{k=1}^r (\theta_k - \theta_k^0)^2 b_k^2 E(\epsilon_{1k}^2) \right)^{1/2} = \frac{n}{4\sigma^2} \sigma^2 b_j^2 \lambda_j^{1/2}; \end{aligned} \tag{S27}$$

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

$$E(\|P_{\hat{\theta}} - P_{\theta_0}\|_a) \geq 1 - \min_{1 \leq j \leq r} \frac{C}{2} R j^{-(a+2b)} \}^{1/2} = 1 - \frac{C}{2} R r^{-(a+2b)} \}^{1/2}.$$

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

$$\min_{H(\theta, \theta^0)=1} E(\|P_{\hat{\theta}} - P_{\theta_0}\|_a) \geq 1 - \frac{C}{2} R \frac{n}{r^{a+2b}} \}^{1/2} = 1 - \frac{C}{2} \frac{R}{L^{a+2b}} \}^{1/2} > 0.$$

where $e_i \sim N(0, \sigma^2)$ and $E(\epsilon_{ik}^2) = \sigma_k$. Denote P_θ the conditional probability measure of $\{Y_i(\cdot, \cdot)\}_{i=1}^n$ given X, and its corresponding density is denoted q_θ as

It is sufficient to find g_r and a feasible lower bounds for $\min_{H(\theta, \theta_0)=1} E(\|P_{\hat{\theta}} - P_{\theta_0}\|_a)$ by applying Lemma S4 with $\Phi(\theta) = \sum_{k=1}^r \sigma_k b_k \theta_k$. The prediction error

$$d(\hat{\theta}, \theta) = E(\tilde{\epsilon}_n) = \sum_{j=1}^{\infty} \int_{\mathcal{T}_j} (\hat{\theta} - \theta)_j^2 = \sum_{j=1}^{\infty} \lambda_j (\hat{b}_j - b_j)^2 \tag{S24}$$

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

$$d(\theta, \theta^0) = d(\{\theta_k\}, \{\theta_k^0\}) = \sum_{k=1}^r \int_{\mathcal{T}_k} (\theta - \theta^0)_k^2 = \sum_{k=1}^r d_k(\theta, \theta^0),$$

where $d_k(\theta, \theta^0) = \sigma_k^2 (\theta_k - \theta_k^0)^2 b_k^2$, $k = 1, 2, \ldots, r$. Assume θ differs from θ^0 only in the jth coordinate, thus $H(\theta, \theta^0) = 1$, and $d_j(\theta, \theta^0) = \sigma_j^2 b_j^2$, which implies $g_r = \sigma_r^2 \sum_{k=1}^r \sigma_k b_k^2 = \sigma_r^2 O(r^{1-a-2b})$. For any estimator $\hat{\theta} = \sum_{j \in \mathbb{N}_+} \hat{\theta}_j \phi_j$ based on $\{X_i, Y_i\}_{i=1}^n$, apply Lemma S4 with $A = 1/2$

$$\sup_{X, \theta \in \mathcal{G}} E\{E(\tilde{\epsilon}_n)\} \geq \max_{\theta \in \{0,1\}^r} E\{d(\hat{\theta}, \theta)\} \geq \frac{g_r}{4} \min_{H(\theta, \theta^0)=1} E(\|P_{\hat{\theta}} - P_{\theta_0}\|_a). \tag{S25}$$

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

$$E(\|P_{\hat{\theta}} - P_{\theta_0}\|_a) \leq 1 - E[\{K(P_\theta \mid P_{\theta_0})/2\}^{1/2}]. \tag{S26}$$

To guarantee the positiveness of the right hand side of (S26), we need to show that $\{K(P_\theta \mid P_{\theta_0})/2\}^{1/2}$ is sufficient small for a suitable $\theta > 0$. For a fixed a, $j \in \{1, 2, \ldots, r\}$, the log-likelihood ratio for normal noise in Condition 7 with $\sigma_k^{-2} = (2\sigma^2)^{-1}$ is

$$\log \frac{q_\theta}{q} = \frac{1}{2} \sum_{i=1}^n \left\{ \int_{\mathcal{T}} Y_i \, d(\theta - \theta^0) - \int_{\mathcal{T}} X_i (\theta - \theta^0) \int_{\mathcal{T}} X_i (\theta - \theta^0) \right\}^2.$$

Note that $K(P_\theta \mid P_{\theta_0}) = E\{\log(q_{\theta_0}/q_\theta) \mid X\}$,

$$\begin{aligned} E[\{K(P_\theta \mid P_{\theta_0})/2\}^{1/2}] &= E\left[\frac{n}{4\sigma^2} \sum_{k=1}^r (\theta_k - \theta_k^0)^2 b_k^2 \lambda_{1k}^2 \right]^{1/2} \\ &\geq \left(\frac{n}{4\sigma^2} \sum_{k=1}^r (\theta_k - \theta_k^0)^2 b_k^2 E(\epsilon_{1k}^2) \right)^{1/2} = \frac{n}{4\sigma^2} \sigma^2 b_j^2 \lambda_j^{1/2}; \end{aligned} \tag{S27}$$

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

$$E(\|P_{\hat{\theta}} - P_{\theta_0}\|_a) \geq 1 - \min_{1 \leq j \leq r} \frac{C}{2} R j^{-(a+2b)} \}^{1/2} = 1 - \frac{C}{2} R r^{-(a+2b)} \}^{1/2}.$$

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

$$\min_{H(\theta, \theta^0)=1} E(\|P_{\hat{\theta}} - P_{\theta_0}\|_a) \geq 1 - \frac{C}{2} R \frac{n}{r^{a+2b}} \}^{1/2} = 1 - \frac{C}{2} \frac{R}{L^{a+2b}} \}^{1/2} > 0.$$

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

$$
\begin{aligned}
E\left(\sum_{j=1}^{m} G_{1;j}^2\right) &= \sum_{j=1}^{m} {}_jE\left(\sum_{k\neq j}^{m} \frac{\langle \Delta_{(2)}\ {}_k;\ {}_j\rangle}{{}_k-{}_j} b_{0k}\right)^2 \\
&= \sum_{j=1}^{m} {}_jE\left\{\sum_{k_1;k_2\neq j}^{m} \frac{\langle \Delta_{(2)}\ {}_{k_1};\ {}_j\rangle\langle \Delta_{(2)}\ {}_{k_2};\ {}_j\rangle}{({}_{k_1}-{}_j)({}_{k_2}-{}_j)} b_{0k_1}b_{0k_2}\right\}: \\
&\leq \sum_{j=1}^{m} {}_j \sum_{k_1;k_2\neq j}^{m} \frac{\left\{E(\langle \Delta_{(2)}\ {}_{k_1};\ {}_j\rangle^2)E(\langle \Delta_{(2)}\ {}_{k_1};\ {}_j\rangle^2)\right\}^{1=2} b_{0k_1}b_{0k_2}}{({}_{k_1}-{}_j)({}_{k_2}-{}_j)} \\
&= \sum_{j=1}^{m} {}_j \left\{\sum_{k\neq j}^{m} \frac{\left\{E(\langle \Delta_{(2)}\ {}_{k_1};\ {}_j\rangle^2)\right\}^{1=2}}{{}_k-{}_j} b_{0k}\right\}^2 \\
&\lesssim \sum_{j=1}^{m} {}_j \left[\sum_{k\neq j}^{m} \frac{b_{0k}}{{}_k-{}_j}\left\{\frac{j^{-\frac{a}{2}}k^{-\frac{a}{2}}}{n^{\frac{1}{2}}} + h^2\left(k^c j^{-a} + k^{-a} j^c\right)\right\}\right]^2 \\
&\lesssim \sum_{j=1}^{m} \left(\frac{1}{n} + h^4 j^{2c}\right) j^{-a} = O\left(\frac{1}{n} + h^4 m^{2c-a+1}\right) = o({}_n):
\end{aligned}
\tag{S41}
$$

For $G_{2;j}$,

$$
\begin{aligned}
\sum_{j=1}^{m} G_{2;j}^2 &= \sum_{j=1}^{m} {}_j \left\{\sum_{k\neq j}^{m} \frac{\langle \Delta_{(2)}(\hat{}_{(2);k}-{}_k);\ {}_j\rangle}{{}_k-{}_j} b_{0k}\right\}^2 \\
&= \sum_{j=1}^{m} {}_j \sum_{k_1;k_2\neq j}^{m} \frac{\langle \Delta_{(2)}(\hat{}_{(2);k_1}-{}_{k_1});\ {}_j\rangle\langle \Delta_{(2)}(\hat{}_{(2);k_2}-{}_{k_2});\ {}_j\rangle}{({}_{k_1}-{}_j)({}_{k_2}-{}_j)} b_{0k_1}b_{0k_2} \\
&\leq \|\Delta_{(2)}\|_{\mathrm{HS}}^2 \sum_{j=1}^{m} {}_j \sum_{k_1;k_2\neq j}^{m} \frac{\|\hat{}_{(2);k_1}-{}_{k_1}\|\|\hat{}_{(2);k_2}-{}_{k_2}\|}{|{}_{k_1}-{}_j||{}_{k_2}-{}_j|} b_{0k_1}b_{0k_2}:
\end{aligned}
\tag{S42}
$$

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

$$
\begin{aligned}
&E\left(\sum_{j=1}^{m} {}_j \sum_{k_1;k_2\neq j}^{m} \frac{\|\hat{}_{(2);k_1}-{}_{k_1}\|\|\hat{}_{(2);k_2}-{}_{k_2}\|}{|{}_{k_1}-{}_j||{}_{k_2}-{}_j|} b_{0k_1}b_{0k_2}\right) \\
&\leq \sum_{j=1}^{m} {}_j \sum_{k_1;k_2\neq j}^{m} \frac{\{E(\|\hat{}_{(2);k_1}-{}_{k_1}\|^2)E(\|\hat{}_{(2);k_2}-{}_{k_2}\|^2)\}^{1=2}}{|{}_{k_1}-{}_j||{}_{k_2}-{}_j|} b_{0k_1}b_{0k_2} \\
&= \sum_{j=1}^{m} {}_j \left(\sum_{k\neq j} \frac{\{E(\|\hat{}_{(2);k}-{}_k\|^2)\}^{1=2}}{|{}_k-{}_j|} b_k\right)^2 \\
&\lesssim \frac{1}{n}\left\{1 + \frac{1}{Nh} + m^{2a+5-2b}\log m\left(1+\frac{1}{Nh}\right)\right\} + h^4 m^{3a+5-2b+2c}\log m:
\end{aligned}
\tag{S43}
$$

Thus,

$$\sum_{j=1}^{m} G_{2;j}^2 = O_p(\|\Delta_{(2)}\|_{\mathrm{HS}}^2) O_p\left(\sum_{j=1}^{m} \sum_{k_1;k_2\neq j}^{m} \frac{\|\hat{\ }_{(2);k_1} - \ _{k_1}\|\|\hat{\ }_{(2);k_2} - \ _{k_2}\|}{|\ _{k_1} - \ _j||\ _{k_2} - \ _j|} b_{0k_1} b_{0k_2}\right)$$
$$= o_p(\ _n):$$

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

$$\sum_{j=1}^{m} G_{3;j}^2 = \sum_{j=1}^{m} \ _j \left\{\sum_{k\neq j}^{m} \sum_{s=1}^{\infty} \frac{(\ _k - \hat{\ }_{(2);k})^s}{(\ _k - \ _j)^{s+1}} \langle \Delta_{(2)} \hat{\ }_{(2);k}; \ _j\rangle b_{0k}\right\}^2$$
$$\leq 2\sum_{j=1}^{m} \ _j \left\{\sum_{k\neq j}^{m} \sum_{s=1}^{\infty} \frac{(\ _k - \hat{\ }_{(2);k})^s}{(\ _k - \ _j)^{s+1}} \langle \Delta_{(2)} \ _{(2);k}; \ _j\rangle b_{0k}\right\}^2$$
$$+ 2\sum_{j=1}^{m} \ _j \left\{\sum_{k\neq j}^{m} \sum_{s=1}^{\infty} \frac{(\ _k - \hat{\ }_{(2);k})^s}{(\ _k - \ _j)^{s+1}} \langle \Delta_{(2)}(\hat{\ }_{(2);k} - \ _k); \ _j\rangle b_{0k}\right\}^2$$
$$= G_{3;j;1} + G_{3;j;2}:$$

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

$$G_{3;j;1} = 2\sum_{j=1}^{m} \ _j \sum_{k_1\neq k_2\neq j}^{m} \left\{\sum_{s=1}^{\infty} \frac{(\ _{k_1} - \hat{\ }_{(2);k_1})^s}{(\ _{k_1} - \ _j)^{s+1}}\right\} \left\{\sum_{s=1}^{\infty} \frac{(\ _{k_2} - \hat{\ }_{(2);k_2})^s}{(\ _{k_2} - \ _j)^{s+1}}\right\}$$
$$\times \langle \Delta_{(2)} \ _{(2);k_1}; \ _j\rangle \langle \Delta_{(2)} \ _{(2);k_2}; \ _j\rangle b_{0k_1} b_{0k_2}$$
$$= 2\sum_{j=1}^{m} \ _j \sum_{k_1\neq k_2\neq j}^{m} \frac{\ _{k_1} - \hat{\ }_{(2);k_1}}{(\ _{k_1} - \ _j)(\hat{\ }_{(2);k_1} - \ _j)} \frac{\ _{k_2} - \hat{\ }_{(2);k_2}}{(\ _{k_2} - \ _j)(\hat{\ }_{(2);k_2} - \ _j)}$$
$$\times \langle \Delta_{(2)} \ _{(2);k_1}; \ _j\rangle \langle \Delta_{(2)} \ _{(2);k_2}; \ _j\rangle b_{0k_1} b_{0k_2} \tag{S44}$$
$$\lesssim \frac{\|\Delta_{(2)}\|_{\mathrm{HS}}^2}{(2\ _m - \|\Delta_{(2)}\|_{\mathrm{HS}})^2} \sum_{j=1}^{m} \ _j E\left\{\sum_{k_1;k_2\neq j}^{m} \frac{\langle \Delta_{(2)} \ _{k_1}; \ _j\rangle \langle \Delta_{(2)} \ _{k_2}; \ _j\rangle}{(\ _{k_1} - \ _j)(\ _{k_2} - \ _j)} b_{0k_1} b_{0k_2}\right\}$$
$$\lesssim \sum_{j=1}^{m} G_{1;j}^2 = o_p(\ _n):$$

190 Similarly,

$$G_{3;j;2} = 2\sum_{j=1}^{m} \ _j \left\{\sum_{k\neq j}^{m} \sum_{s=1}^{\infty} \frac{(\ _k - \hat{\ }_{(2);k})^s}{(\ _k - \ _j)^{s+1}} \langle \Delta_{(2)}(\hat{\ }_{(2);k} - \ _k); \ _j\rangle b_{0k}\right\}^2$$
$$\lesssim \frac{\|\Delta_{(2)}\|_{\mathrm{HS}}^2}{(2\ _m - \|\Delta_{(2)}\|_{\mathrm{HS}})^2} \sum_{j=1}^{m} \ _j \left\{\sum_{k\neq j}^{m} \frac{\langle \Delta_{(2)}(\hat{\ }_{(2);k} - \ _k); \ _j\rangle}{\ _k - \ _j} b_{0k}\right\}^2 \tag{S45}$$
$$\lesssim \sum_{j=1}^{m} G_{2;j}^2 = o_p(\ _n):$$

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

$$\begin{aligned}\sum_{j=1}^{m} G_{4;j}^2 &\lesssim \sum_{j=1}^{m} \ _j b_j^2 \| \hat{}_{(2);j} - \ _j\|^4 \le \sum_{j=1}^{m} \ _j b_j^2 \| \hat{}_{(2);j} - \ _j\|^2 \frac{\|\Delta_{(2)}\|_{\mathrm{HS}}^2}{\ _j^2} \\ &\lesssim \|\Delta_{(2)}\|_{\mathrm{HS}}^2 \sum_{j=1}^{m} j^{a+2-2b} \| \hat{}_{(2);j} - \ _j\|^2 = o_p(\ _n);\end{aligned} \tag{S46}$$

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

$$E[\|E\{(\ _{i_1} - \hat{}_{i_1})\ _{i_1}| \hat{}_{(2)}\}\|^2] = o(\ _n):$$

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 $\ = (\ _1;\dots;\ _j;\dots;\ _r)^{\mathsf{T}} \in \{0;1\}^r$, define $\ ^j = (\ _1;\dots;1-\ _j;\dots;\ _r)^{\mathsf{T}}$ as the perturbation of $\$ such that $\ ^j$ differs from $\$ only in the jth position. By assumptions,

$$\begin{aligned}&\max_{\in\{0;1\}^r} E\ \{d(T;\ (\ ))\} \ge \frac{1}{2^r} \sum_{\in\{0;1\}^r} \sum_{j=1}^{r} E\ \{d_j(T;\ (\ ))\} \\ =&\frac{1}{2^r}\sum_{j=1}^{r}\sum_{\in\{0;1\}^r} E\ \{d_j(T;\ (\ ))\} = \frac{1}{2^r}\sum_{j=1}^{r}\sum_{\in\{0;1\}^r} \frac{E\ \{d_j(T;\ (\ ))\} + E\ _j\{d_j(T;\ (\ ^j))\}}{2} \\ =&\frac{1}{2^{r+1}}\sum_{j=1}^{r}\sum_{\in\{0;1\}^r}\Big(E\ \Big[\{d_j(T;\ (\ )) + d_j\end{aligned}$$

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.