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

- $D = \mathrm{diag}\{ {}_1^{1/2}, \ldots, {}_m^{1/2}\}$, ${}_i = \langle X_i, \ \rangle$, ${}_i = ({}_{i1}, \ldots, {}_{im})^{\mathrm{T}}$ and ${}_i = D^{-1} {}_i$;
- $b_0 = (b_{01}, \ldots, b_{0m})^{\mathrm{T}}$, ${}_{i,m} = {}_i^{\mathrm{T}} b_0$ and $b_{r0} = D b_0$;
- $\hat{}_i = (\hat{}_{i1}, \ldots, \hat{}_{im})^{\mathrm{T}}$, $\hat{}_{i,m} = \hat{}_i^{\mathrm{T}} b_0$ 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 $\Omega_m(n, N)$,*

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

Due to the fact that the variance of $_{ik}$ 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

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

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 $\Omega_m(n, N)$,*

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

where

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

The following lemma is to establish the minimax lower bound for the prediction error. We define that P is a family of probability measures, where is the parameter of interest and the corresponding expectation operator is denoted as E. Let $H(\ ,\ ') = \sum_{i=1}^{r} |\ _i - \ '_i|$ be the Hamming distance between the binary sequences $= (\ _1, \dots, \ _r)^{\mathsf{T}}$ and $' = (\ '_1, \dots, \ '_r)^{\mathsf{T}}$ on $\{0,1\}^r = \{(\ _1, \dots, \ _j, \dots, \ _r)^{\mathsf{T}} \mid \ _j = 0$ or $_j = 1\}$. In addition, for the probability measures P and $P_{'}$ with density function p and $p_{'}$ jointly dominated by , the integration of their minimal $\int (p \wedge p_{'})\mathrm{d}$ is denoted as $\|P \wedge P_{'}\|_a$.

LEMMA S4 (ASSOUAD'S LEMMA). *The estimator T is based on observations of the statistical model P; $\in \{0,1\}^r$. Let $(\ )$ be an arbitrary transform of the parameter . 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 $d_j(\ (\ ),\ (\ ')) \geq g_{jr} > 0$ for $H(\ ,\ ') = 1$ such that and $'$ differ only in the jth coordinate, then for the distance $d(T,\ (\ ))$,*

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

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 $(\ )$ 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 $E(\langle \Delta_{(r)}\ _k,\ _j\rangle^2)$. 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 $_{il_1l_2} = X_{il_1}X_{il_2}$,

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

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,

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

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

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

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

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

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

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

$$E\{A_i^2(\ {}_j;\ {}_k)\} = 4!\binom{N}{4}A_{i1}(\ {}_j;\ {}_k) + 3!\binom{N}{3}A_{i2}(\ {}_j;\ {}_k) + 2!\binom{N}{2}A_{i3}(\ {}_j;\ {}_k)$$

with

$$A_{i1}(\ {}_j;\ {}_k) = E\left[\left\{\int X(u)\ {}_{k;h}(u)\mathrm{d}u\right\}^2\left\{\int X(u)\ {}_{j;h}(u)\mathrm{d}u\right\}^2\right]$$

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

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

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

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

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

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 $\|\cdot\|_j$,

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

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 $\Omega_m(n; N)$,

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

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,

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

S134]TJ0 g 0 G [())-55

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

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

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

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

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 $\mathcal{E}(\hat{\ }_n)$,

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

$$= E_{X^*}\left\{\left(\int \ X^* - \int \hat{\ } X^*\right)^2\right\} + \frac{1}{N}E_{X^*}\left\{\int \hat{\ }^2(X^*)^2 - \left(\int \hat{\ } X^*\right)^2\right\} + \frac{{}^2_X}{N}\|\hat{\ }\|^2.$$

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

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

On one hand

$$\left|\sum_{k=1}^{\infty} \ _k\left(\int \hat{\ }^2 \ {}^2_k - \hat{b}_k^2\right)\right| \leq \left|\sum_{k=1}^{\infty} \ _k\int \hat{\ }^2 \ {}^2_k\right| + \left|\sum_{k=1}^{\infty} \ _k\hat{b}_k^2\right|$$

$$\leq \|\hat{\ }\|^2\sum_{k=1}^{\infty} \ _k\| \ _k\|_\infty^2 + \sum_{k=1}^{\infty} \ _k\hat{b}_k^2 < \infty;$$

where the second inequality follows from $\| \ _k\|_\infty = O(1)$ by Condition 3. On the other hand, by Jensen inequality, for any $k \in \mathbb{N}_+$, $\int \hat{\ }^2 \ {}^2_k \geqslant \hat{b}_k^2$ 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 \ {}^2_1 - \hat{b}_1^2 > \ $ and

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

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 $\mathcal{E}(\tilde{\ }_n)$,

$$\mathcal{E}(\tilde{\ }_n) = E_*(\langle\hat{\ } - \ ; X^*\rangle^2)$$

$$\leq 2E_*\left(\left\langle\hat{\ } - \ ; \sum_{k=1}^{m} \ {}^*_k \ _k\right\rangle^2\right) + 2E_*\left(\left\langle\hat{\ } - \ ; \sum_{k=m+1}^{\infty} \ {}^*_k \ _k\right\rangle^2\right)$$

$$= 2\sum_{j=1}^{m} \ _j\left\{\int(\hat{\ } - \ ) \ _j\right\}^2 + 2E_*\left(\left\langle\hat{\ } - \ ; \sum_{k=m+1}^{\infty} \ {}^*_k \ _k\right\rangle^2\right)$$

$$= I_1 + I_2.$$

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

(S9),

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

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

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

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

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

$$\leq m^a \|\hat{b}_r - b_{r0}\|^2 \sum_{j=1}^{m} \ {}_j \sum_{k\neq j}^{m} \frac{\langle \Delta_{(1)} \ {}_j; \ {}_k \rangle^2}{(\ {}_k - \ {}_j)^2}:$$

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

$$E\left\{ \sum_{j=1}^{m} \ {}_j \sum_{k\neq j}^{m} \frac{\langle \Delta_{(1)} \ {}_j; \ {}_k \rangle^2}{(\ {}_k - \ {}_j)^2} \right\} \lesssim \frac{1}{n}(m^{3-a} + 1) + h^4(m^{3-a+2c} + 1):$$

Thus, Under Condition 5

$$\sum_{j=1}^{m} \ {}_j \left\{ \sum_{k\neq j}^{m} \frac{\langle \Delta_{(1)} \ {}_j; \ {}_k \rangle}{{}_k - \ {}_j} (\hat{b}_k - b_{0k}) \right\}^2 = O_p\left(\left\{ \frac{m^3 + m^a}{n} + h^4(m^{3+2c} + m^a) \right\} \ {}_n \right)$$

$$= o_p(\ {}_n):$$

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

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

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

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

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

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

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

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

As for I_{21}, by the orthogonality of the series $\{\quad _k\}_{k=1}^{\infty}$, k where $e_i \sim N(0;\sigma^2)$ and $E(\xi^2_{ik}) = \lambda_k$. Denote P_θ the conditional probability measure of $\{Y_i(\cdot;\theta)\}_{i=1}^n$ given X, and its corresponding density is denoted q_θ as

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

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

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

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

where $d_k(\beta(\theta);\beta(\theta')) = \lambda_k\delta^2(\theta'_k - \theta_k)^2 b_k^2$; $k = 1;2;\dots;r$: Assume θ differs from θ' only in the jth coordinate, thus $H(\theta;\theta') = 1$, and $d_j(\beta(\theta);\beta(\theta')) = \delta^2\lambda_j b_j^2$; which implies $g_r = \delta^2\sum_{k=1}^r \lambda_k b_k^2 = \delta^2 O(r^{1-a-2b})$. For any estimator $\hat\beta = \sum_{j\in N_+} \hat b_j \phi_j$ based on $\{X_i;Y_i\}_{i=1}^n$, apply Lemma S4 with $A = 1/2$

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

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

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

To guarantee the positiveness of the right hand side of (S26), we need to show that $\{K(P_\theta | P_{\theta'})/2\}^{1/2}$ is sufficient small for a suitable $\delta > 0$. For a fixed $a_j \in \{1;2;\dots;r\}$, the log-likelihood ratio for normal noise in Condition 7 with $\kappa_2 = (2\sigma^2)^{-1}$ is

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

Note that $K(P_\theta | P_{\theta'}) = -E\{\log(q_{\theta'}/q_\theta) | X\}$,

$$\begin{aligned} E[\{K(P_\theta | P_{\theta'})/2\}^{1/2}] &= E\left[\left\{\frac{n}{4\sigma^2}\sum_{k=1}^r (\theta'_k - \theta_k)^2 b_k^2 \xi^2_{1k}\right\}^{1/2}\right] \\ &\leq \left\{\frac{n}{4\sigma^2}\sum_{k=1}^r (\theta'_k - \theta_k)^2 b_k^2 E(\xi^2_{1k})\right\}^{1/2} = \left(\frac{n}{4\sigma^2}\delta^2 b_j^2\lambda_j\right)^{1/2}; \end{aligned} \tag{S27}$$

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

$$E(\|P_\theta \wedge P_{\theta'}\|_a) \geq 1 - \min_{1\leq j\leq r}\left(\frac{C}{2}Rnj^{-(a+2b)}\right)^{1/2} = 1 - \left(\frac{C}{2}Rnr^{-(a+2b)}\right)^{1/2}:$$

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

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

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

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

$$d(\hat{\beta};\beta) = E(\tilde{\epsilon}_n) = \sum_{j=1}^\infty \lambda_j \left(\int (\hat{\beta}-\beta)\phi_j\right)^2 = \sum_{j=1}^\infty \lambda_j(\hat{b}_j - b_j)^2 \tag{S24}$$

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

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

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

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

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

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

To guarantee the positiveness of the right hand side of (S26), we need to show that $\{K(P_\theta|P_{\theta'})/2\}^{1/2}$ is sufficient small for a suitable $\delta > 0$. For a fixed $a_j \in \{1;2;\ldots;r\}$, the log-likelihood ratio for normal noise in Condition 7 with $\kappa_2 = (2\sigma^2)^{-1}$ is

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

Note that $K(P_\theta|P_{\theta'}) = -E\{\log(q_{\theta'}/q_\theta)|X\}$,

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

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

$$E(\|P_\theta \wedge P_{\theta'}\|_a) \geq 1 - \min_{1\leq j\leq r}\left(\frac{C\delta}{2}Rnj^{-(a+2b)}\right)^{1/2} = 1 - \left(\frac{C\delta}{2}Rnr^{-(a+2b)}\right)^{1/2}:$$

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

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

where $e_i \sim N(0; \ ^2)$ and $E(\ ^2_{ik}) = \ _k$. Denote P the conditional probability measure of $\{Y_i(\ ;\ )\}_{i=1}^n$ given X, and its corresponding density is denoted as q.

It is sufficient to find g_r and a feasible lower bounds for $\min_{H(\ ;\ ^0)=1} E(\|P \wedge P_0\|_a)$ by applying Lemma S4 with $(\ ) = \sum_{k=1}^{r} \ _k b_k \ _k$. The prediction error

$$d(\hat{\ };\ ) = E(\tilde{\ }_n) = \sum_{j=1}^{\infty} \ _j \int (\hat{\ } \ ) \ _j \ ^2 = \sum_{j=1}^{\infty} \ _j (\hat{b}_j \ b_j)^2 \qquad \text{(S24)}$$

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

$$d \ (\ ); \ ^0 = d(\ _{;}\ ;\ _{0;}) = \sum_{k=1}^{r} \ _k \int (\ _{;} \ _{0;}) \ _k \ ^2 = \sum_{k=1}^{r} d_k \ (\ ); \ ^0 ;$$

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

$$\sup_{X\ ;\ \in G} E\{E(\tilde{\ }_n)\} \ \max_{\in\{0;1\}^r} E\{d(\hat{\ };\ _{;}\ )\} \ \frac{g_r}{4} \min_{H(\ ;\ ^0)=1} E(\|P \wedge P_0\|_a): \qquad \text{(S25)}$$

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

$$E(\|P \wedge P_0\|_a) \ 1 \ E[\{K(P \mid P_0)/2\}^{1/2}]: \qquad \text{(S26)}$$

To guarantee the positiveness of the right hand side of (S26), we need to show that $\{K(P \mid P_0)/2\}^{1/2}$ is sufficient small for a suitable $\ > 0$. For a fixed $a_j \in \{1;2;\ldots;r\}$, the log-likelihood ratio for normal noise in Condition 7 with $\ _2 = (2\ ^2)^{1}$ is

$$\log \frac{q_0}{q} = \frac{1}{2} \sum_{i=1}^{n} \left[Y_i \int \ _{;}\ X_i \int X_i(\ _{;} \ _{0;}) \int X_i(\ _{;} \ _{0;}) \right]^2$$

Note that $K(P \mid P_0) = \ E\{\log(q_0/q) \mid X\}$,

$$E[\{K(P \mid P_0)/2\}^{1/2}] = E\left[\left\{\frac{n}{4\ ^2} \sum_{k=1}^{r} (\ ^0_k \ _k)^2 b_k^2 \ ^2_{1k}\right\}^{1/2}\right]$$
$$\left\{\frac{n}{4\ ^2} \sum_{k=1}^{r} (\ ^0_k \ _k)^2 b_k^2 E(\ ^2_{1k})\right\}^{1/2} = \frac{n}{4\ ^2} \ ^2 b_j^2 \ _j^{1/2}; \qquad \text{(S27)}$$

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

$$E(\|P \wedge P_0\|_a) \ 1 \ \min_{1 \ j \ r} \frac{C}{2} Rn j^{(a+2b)} \ ^{1/2} = 1 \ \frac{C}{2} Rn r^{(a+2b)} \ ^{1/2}:$$

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

$$\min_{H(\ ;\ ^0)=1} E(\|P \wedge P_0\|_a) \ 1 \ \frac{C}{2} R\frac{n}{r^{a+2b}} \ ^{1/2} = 1 \ \frac{C}{2} \frac{R}{L^{a+2b}} \ ^{1/2} > 0.$$

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

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 $\beta(\theta) = \sum_{k=1}^r \theta_k b_k \phi_k$. The prediction error

$$d(\hat{\beta};\beta) = E(\tilde{\epsilon}_n) = \sum_{j=1}^{\infty} \lambda_j \int (\hat{\beta}-\beta)\phi_j^2 = \sum_{j=1}^{\infty} \lambda_j (\hat{b}_j - b_j)^2 \tag{S24}$$

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,

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

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

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

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 $\gamma > 0$. For a fixed $a_j \in \{1;2;\dots;r\}$, the log-likelihood ratio for normal noise in Condition 7 with $\kappa_2 = (2\sigma^2)^{-1}$ is

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

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

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

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

$$E(\|P_\theta \wedge P_{\theta'}\|_a) \geq 1 - \min_{1 \leq j \leq r}\left(\frac{C\gamma}{2}\sqrt{Rnj^{-(a+2b)}}\right)^{1/2} = 1 - \left(\frac{C\gamma}{2}\sqrt{Rnr^{-(a+2b)}}\right)^{1/2}:$$

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

$$\min_{H(\theta;\theta')=1} E(\|P_\theta \wedge P_{\theta'}\|_a) \geq 1 - \left(\frac{C\gamma}{2}\sqrt{\frac{Rn}{r^{a+2b}}}\right)^{1/2} = 1 - \left(\frac{C\gamma}{2}\sqrt{\frac{R}{L^{a+2b}}}\right)^{1/2} > 0.$$

where $e_i \quad N(0; \; ^2)$ and $E(\;^2_{ik}) = \;_k$. Denote P the conditional probability measure of $f Y_i(\;;\;)g^n_{i=1}$ given X, and its corresponding density is denoted q as

It is suf cient to nd g_r and a feasible lower bounds for $\min_{H(\;;\;^0)=1} E(kP \;\wedge\; P_{\;0}k_a)$ by applying Lemma S4 with $(\;) = \sum_{k=1}^{r} \;_k b_k \;_k$. The prediction error

$$d(\hat{\;}; \;) = E(\tilde{\;}_n) = \sum_{j=1}^{1} \;_j \int (\hat{\;} \quad \;)\;_j \quad ^2 = \sum_{j=1}^{1} \;_j (\hat{b}_j \quad b_j)^2 \qquad \text{(S24)}$$

can be viewed as a semi-distance in $L^2[0;1]$ and $d(x;z) + d(z;y) \quad d(x;y)=2$. By de nition,

$$d \quad (\;); \quad ^0 = d(\;_{;} \;; \;_{0;}) = \sum_{k=1}^{r} \;_k \int (\;_{;} \quad \;_{0;})\;_k \quad ^2 = \sum_{k=1}^{r} d_k \quad (\;); \quad ^0 \;;$$

where $d_k(\;(\;);\;(\;^0)) = \;_k \;^2(\;^0_k \quad \;_k)^2 b_k^2$; $k = 1;2;:::;r$: Assume differs from ^{0} only in the j th coordinate, thus $H(\;;\;^0) = 1$, and $d_j(\;(\;);\;(\;^0)) = \;^2 \;_j b_j^2$; which implies $g_r = \;^2 \sum_{k=1}^{r} \;_k b_k^2 = \;^2 O(r^{1 \;\; a \;\; 2b})$. For any estimator $\hat{\;} = \sum_{j \in N_+} \hat{b}_j \;_j$ based on $f X_i; Y_i g^n_{i=1}$, apply Lemma S4 with $A = 1=2$

$$\sup_{X \;;\; 2G} E f E(\tilde{\;}_n) g \quad \max_{2f\,0;1g^r} E f d(\hat{\;}; \;_{;})g \quad \frac{g_r}{4} \min_{H(\;;\;^0)=1} E(kP \;\wedge\; P_{\;0}k_a): \qquad \text{(S25)}$$

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

$$E(kP \;\wedge\; P_{\;0}k_a) \quad 1 \quad E[f K(P \;\; j \;P_{\;0})=2g^{1=2}]: \qquad \text{(S26)}$$

To guarantee the positiveness of the right hand side of (S26), we need to show that $f K(P \;\; j\; P_{\;0})=2g^{1=2}$ is suf cient small for a suitable > 0. For a xed $aj \; 2 \; f\, 1;2;:::;r\, g$, the log-likelihood ratio for normal noise in Condition 7 with $\;_2 = (2 \quad ^2) \quad ^1$ is

$$\log \frac{q_{\;0}}{q} = \frac{1}{2} \sum_{i=1}^{n} \left[Y_i \int \;_{;} X_i \quad \int X_i(\;_{;} \quad \;_{0;}) \quad \int X_i(\;_{;} \quad \;_{0;}) \right]^2$$

Note that $K(P \;\; j \; P_{\;0}) = \quad E \;\; f \log(q_{\;0}=q)\; j\; X g$,

$$E[f K(P \;\; j\; P_{\;0})=2g^{1=2}] = E\left[\left\{ \frac{n}{4\;^2} \sum_{k=1}^{r} (\;^0_k \quad \;_k)^2 b_k^2 \;^2_{1k} \right\}^{1=2} \right]$$

$$\left\{ \frac{n}{4\;^2} \sum_{k=1}^{r} (\;^0_k \quad \;_k)^2 b_k^2 E(\;^2_{1k}) \right\}^{1=2} = \frac{n}{4\;^2} \;^2 b_j^2 \;_j \quad ^{1=2}; \qquad \text{(S27)}$$

where the last inequality is by Jensen's inequality. By $jb_j j \quad Cj \quad ^b$; $\;_j \quad Rj \quad ^a$ for $j \quad 1$ and constant $C; R > 0$,

$$E(kP \;\wedge\; P_{\;0}k_a) \quad 1 \quad \min_{1 \;\; j \;\; r} \frac{C}{2} \quad Rnj \quad ^{(a+2b)} \quad ^{1=2} = 1 \quad \frac{C}{2} \quad Rnr \quad ^{(a+2b)} \quad ^{1=2}:$$

If we put $r = Ln^{1=(a+2b)}$, for suf ciently small > 0,

$$\min_{H(\;;\;^0)=1} E(kP \;\wedge\; P_{\;0}k_a) \quad 1 \quad \frac{C}{2} \quad R\frac{n}{r^{a+2b}} \quad ^{1=2} = 1 \quad \frac{C}{2} \quad \frac{R}{L^{a+2b}} \quad ^{1=2} > 0.$$

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

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

$$d(\hat\beta;\beta) = E(\tilde\epsilon_n) = \sum_{j=1}^{\infty} \lambda_j \left\{\int (\hat\beta - \beta)\phi_j\right\}^2 = \sum_{j=1}^{\infty} \lambda_j (\hat b_j - b_j)^2 \qquad \text{(S24)}$$

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

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

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

$$\sup_{X;\beta \in G} E\{E(\tilde\epsilon_n)\} \ge \max_{\theta \in \{0;1\}^r} E\{d(\hat\beta;\beta_{\cdot;\theta})\} \ge \frac{g_r}{4} \min_{H(\theta;\theta')=1} E(\|P_\theta \wedge P_{\theta'}\|_a). \qquad \text{(S25)}$$

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

$$E(\|P_\theta \wedge P_{\theta'}\|_a) \ge 1 - E[\{K(P_\theta | P_{\theta'})/2\}^{1/2}]. \qquad \text{(S26)}$$

To guarantee the positiveness of the right hand side of (S26), we need to show that $\{K(P_\theta | P_{\theta'})/2\}^{1/2}$ is sufficient small for a suitable $\gamma > 0$. For a fixed $j \in \{1;2;\ldots;r\}$, the log-likelihood ratio for normal noise in Condition 7 with $\kappa_2 = (2\sigma^2)^{-1}$ is

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

Note that $K(P_\theta | P_{\theta'}) = -E\{\log(q_{\theta'}/q_\theta) | X\}$,

$$\begin{aligned} E[\{K(P_\theta | P_{\theta'})/2\}^{1/2}] &= E\left[\left\{\frac{n}{4\sigma^2} \sum_{k=1}^{r} \gamma^2 (\theta'_k - \theta_k)^2 b_k^2 \xi_{1k}^2\right\}^{1/2}\right] \\ &\le \left\{\frac{n}{4\sigma^2} \sum_{k=1}^{r} \gamma^2 (\theta'_k - \theta_k)^2 b_k^2 E(\xi_{1k}^2)\right\}^{1/2} = \left(\frac{n}{4\sigma^2} \gamma^2 b_j^2 \lambda_j\right)^{1/2}; \end{aligned} \qquad \text{(S27)}$$

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

$$E(\|P_\theta \wedge P_{\theta'}\|_a) \ge 1 - \min_{1 \le j \le r} \left(\frac{C}{2}\gamma Rnj^{-(a+2b)}\right)^{1/2} = 1 - \left(\frac{C}{2}\gamma Rnr^{-(a+2b)}\right)^{1/2}.$$

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

$$\min_{H(\theta;\theta')=1} E(\|P_\theta \wedge P_{\theta'}\|_a) \ge 1 - \left(\frac{C}{2}\gamma R\frac{n}{r^{a+2b}}\right)^{1/2} = 1 - \left(\frac{C}{2}\gamma\frac{R}{L^{a+2b}}\right)^{1/2} > 0.$$

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

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 $\beta(\theta) = \sum_{k=1}^r \theta_k b_k \phi_k$. The prediction error

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

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,

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

where $d_k(\beta(\theta);\beta(\theta')) = \lambda_k\tau^2(\theta'_k - \theta_k)^2 b_k^2$; $k = 1;2;\ldots;r$. Assume θ differs from θ' only in the jth coordinate, thus $H(\theta;\theta') = 1$, and $d_j(\beta(\theta);\beta(\theta')) = \tau^2\lambda_j b_j^2$; which implies $g_r = \tau^2 \sum_{k=1}^r \lambda_k b_k^2 = \tau^2 O(r^{1-a-2b})$. 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 $A = 1/2$

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

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 $\tau > 0$. For a fixed $a_j \in \{1;2;\ldots;r\}$, the log-likelihood ratio for normal noise in Condition 7 with $\kappa_2 = (2\sigma^2)^{-1}$ is

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

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

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

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

$$E(\|P_\theta \wedge P_{\theta'}\|_a) \geq 1 - \min_{1\leq j\leq r}\left(\frac{C}{2}\tau Rnj^{-(a+2b)}\right)^{1/2} = 1 - \left(\frac{C}{2}\tau Rnr^{-(a+2b)}\right)^{1/2}:$$

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

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

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

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

For $G_{2;j}$,

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

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

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

Thus,

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

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

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

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

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

Similarly,

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

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),

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

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

$$\max_{\ \in\{0,1\}^r} E\ \{d(T;\ (\ ))\} \geq \frac{1}{2^r} \sum_{\ \in\{0,1\}^r} \sum_{j=1}^{r} E\ \{d_j(T;\ (\ ))\}$$
$$= \frac{1}{2^r} \sum_{j=1}^{r} \sum_{\ \in\{0,1\}^r} E\ \{d_j(T;\ (\ ))\} = \frac{1}{2^r} \sum_{j=1}^{r} \sum_{\ \in\{0,1\}^r} \frac{E\ \{d_j(T;\ (\ ))\} + E_{\ ^j}\{d_j(T;\ (\ ^j))\}}{2}$$
$$= \frac{1}{2^{r+1}} \sum_{j=1}^{r} \sum_{\ \in\{0,1\}^r} \Bigg(E\ \Bigg[\{d_j(T;\ (\ )) + d_j + 1 \sum_{j^J}^{r}))\}$$

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.