Journal of Multivariate Analysis 163 (2018) 15–36

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

PCA-based estimation for functional linear regression with functional responses

Masaaki Imaizumi [a], Kengo Kato [b,*]

[a] Institute of Statistical Mathematics, 10–3 Midori-cho, Tachikawa, Tokyo 190–8562, Japan
[b] Graduate School of Economics, University of Tokyo, 7–3–1 Hongo, Bunkyo-ku, Tokyo 113–0033, Japan

ARTICLE INFO

Article history:
Received 22 March 2017
Available online 18 October 2017

AMS subject classifications:
62G08
62G20

Keywords:
Functional data
Functional principal component analysis
Ill-posed inverse problem
Minimax rate

ABSTRACT

This paper studies a regression model where both predictor and response variables are random functions. We consider a functional linear model where the conditional mean of the response variable at each time point is given by a linear functional of the predictor variable. In this paper, we are interested in estimation of the integral kernel $b(s, t)$ of the conditional expectation operator, where s is an output variable while t is a variable that interacts with the predictor variable. This problem is an ill-posed inverse problem, and we consider two estimators based on functional principal component analysis (PCA). We show that under suitable regularity conditions, an estimator based on the single truncation attains the convergence rate for the integrated squared error that is characterized by smoothness of the function $b(s, t)$ in t together with the decay rate of the eigenvalues of the covariance operator, but the rate does not depend on the smoothness of $b(s, t)$ in s. This rate is shown to be minimax optimal, and consequently smoothness of $b(s, t)$ in s does not affect difficulty of estimating b. We also consider an alternative estimator based on the double truncation, and provide conditions under which the alternative estimator attains the optimal rate. We conduct simulations to verify the performance of PCA-based estimators in the finite sample. Finally, we apply our estimators to investigate the relation between the lifetime pattern of working hours and total income, and the relation between the electricity spot price and the wind power in-feed.

1. Introduction

This paper studies a regression model in which both predictor and response variables are random functions. Let X, Y be $L^2(I)$-valued random variables with $I = [0, 1]$, and consider a regression model of the form

$$\mathrm{E}(Y \mid X)(s) = \mathrm{E}\{Y(s)\} + \int_I b(s, t)[X(t) - \mathrm{E}\{X(t)\}]dt. \tag{1}$$

See Section 2 for the precise description of the setup. The focus of this paper is on estimation of the bivariate function $b(s, t)$, which is an ill-posed inverse problem; see Remark 2 in Section 2.

Data collected on dense grids can be typically regarded as realizations of a random function (i.e., a stochastic process), and such data are called functional data. Statistical methodology dealing with functional data is called functional data analysis and has a large number of fruitful applications; see [33]. For example, the functional linear model (1) with functional

* Corresponding author.
E-mail addresses: imaizumi@ism.ac.jp (M. Imaizumi), kkato@e.u-tokyo.ac.jp (K. Kato).

https://doi.org/10.1016/j.jmva.2017.10.001
0047-259X/

predictor and response variables can be used to investigate how a complete daily temperature profile over one year influences daily precipitation on any given day [33, Chapter 16].

In this paper, we consider estimators for the function b based on functional principal component analysis (PCA), which is one of standard techniques used in functional data analysis. Applying basis expansions of X and b using the eigenfunction system $\{\phi_k : k \in \mathbb{N}\}$ with $\mathbb{N} = \{1, 2, \ldots\}$ for the covariance operator of X, we can expand X and b as

$$X(t) = \mathrm{E}\{X(t)\} + \sum_{k=1}^{\infty} \xi_k \phi_k(t) \quad \text{and} \quad b(s,t) = \sum_{k=1}^{\infty} b_{j,k} \phi_j(s) \phi_k(t),$$

where we measure the smoothness of b via how fast $|b_{j,k}|$ decays as $j \to \infty$ or $k \to \infty$. We consider two methods to estimate b based on different characterizations of b. The first method uses the fact that

$$\mathrm{E}\{\xi_k Y(s)\} = \mathrm{E}(\xi_k^2) \sum_{j=1}^{\infty} b_{j,k} \phi_j(s).$$

This method is based on truncation of the series expansion

$$b(s,t) = \sum_{k=1}^{\infty} [\mathrm{E}\{\xi_k Y(s)\} / \mathrm{E}(\xi_k^2)] \phi_k(t)$$

by a finite series $\sum_{k=1}^{m_n}$ with $m_n \to \infty$ as $n \to \infty$ (which we call the single truncation in comparison with the second method below), and replace $\mathrm{E}\{\xi_k Y(\cdot)\}$, $\mathrm{E}(\xi_k^2)$, and ϕ_k by their estimators. This estimator was considered by [14]. The second method uses the expansion of Y as $Y(s) = \sum_j \eta_j \phi_j(s)$. This alternative method is based on truncation of the double series expansion

$$b(s,t) = \sum_{k=1}^{\infty} \{\mathrm{E}(\eta_j \xi_k) / \mathrm{E}(\xi_k^2)\} \phi_j(s) \phi_k(t)$$

by a finite sum $\sum_{j \le m_{n,1}} \sum_{k \le m_{n,2}}$ with $m_{n,1} \to \infty$ and $m_{n,2} \to \infty$ as $n \to \infty$ (which we call the double truncation), and replace $\mathrm{E}(\eta_j \xi_k)$, $\mathrm{E}(\xi_k^2)$, and ϕ_j by their estimators.

Crambes and Mas [14] consider our first estimator, but the focus in [14] is on prediction, and not on estimation of the function b *per se*. These two problems are substantially different, and the authors do not derive sharp rates of convergence for their estimator of b itself. Park and Qian [29] and Hörmann and Kidzínski [21] analyze the estimator from [14] for b with dependent functional data, but they only prove consistency of the estimator. Yao et al. [37] consider a PCA-based estimator similar to (but still different from) our second estimator; however they do not explicitly derive rates of convergence for their estimator.

The object of this paper is to study rates of convergence for estimation of b. First, we show that under suitable regularity conditions, the estimator based on the single truncation (i.e., the estimator of [14]) attains the convergence rate for the integrated squared error that is characterized by smoothness of the function $b(s,t)$ in t together with the decay rate of the eigenvalues of the covariance operator, but the rate does not depend on the smoothness of $b(s,t)$ in s. This rate is shown to be minimax optimal. This means that smoothness of $b(s,t)$ in s does not affect the difficulty of estimating b, which is in sharp contrast with nonparametric estimation of a bivariate regression function. Next, we analyze the second estimator based on the double truncation, and provide conditions under which it attains the optimal rate. We point out that some restrictions on smoothness levels for $b(s,t)$ in s and t are required for the second estimator to achieve the optimal rate. We include the analysis of the second estimator since in applications, the double truncation typically leads to an estimate that is more interpretable than the single truncation, although from a theoretical point of view, the single truncation is enough for the purpose of estimating b; see Remark 1 ahead and the discussion in Chapter 16 of [33]. We also conduct simulations to verify the performance of the estimators in finite samples. In our simulation studies, the second estimator using the double truncation outperforms the first estimator using the single truncation in finite samples. Finally, we apply our estimators to investigate two topics: the relation between the lifetime pattern of working hours and total income using the data from National Longitudinal Survey of Youth conducted by the US Bureau of Labor Statistics [3], and the relation between hourly electricity spot prices and the amount of wind power in-feed using the data from EEX Transparency Platform introduced in [27].

1.1. Literature review

The literature on functional data analysis is now quite broad. We refer to [2,22,33,34] as general references on functional data analysis. Much of the literature on this topic has focused on the functional linear model with a scalar response variable; see [5,6,8–10,12,13,16,19,23,25,28,38]. In particular, Hall and Horowitz [19] consider a PCA-based estimator and an estimator based on Tikhonov regularization for the slope function, and provide conditions under which those estimators attain minimax rates of convergence for the integrated squared error.

The analysis of functional responses was first considered in [32]. Chiou et al. [11] consider a regression model where a predictor variable is finite-dimensional while a response variable is a random function. Functional linear models with

functional predictor and response variables are considered in [1,14,15,20,21,26,37]. Cuevas et al. [15] work with fixed designs, which is a different setting than ours, and prove the consistency of a series estimator of the integral operator with kernel b for the operator norm. We already referred to [14,21,37]. He et al. [20] propose an estimator of b based on functional canonical correlation analysis, but they do not study its asymptotic properties. Lian [26] considers prediction for functional linear regression with functional responses based on a reproducing kernel Hilbert space approach, which is a topic substantially different from ours. The recent paper by Benatia et al. [1] studies a Tikhonov regularization estimation for b and establishes rates of convergence for their estimator; the estimator and the assumptions in [1] are substantially different from ours and so their results are not directly comparable to ours.

It is important to stress here that none of the papers mentioned above derives optimal rates of convergence for estimation of b; the present paper fills this gap and thereby contributes to advancing the understanding of functional data analysis. From a technical point of view, the proofs of the main theorems (Theorems 1 and 2) build upon the techniques developed in [19]. However, given that we are estimating a bivariate function with two different levels of smoothness, rather than a univariate function in the scalar response case, the proofs require a chain of delicate calculations. Furthermore, to establish minimax lower bounds for estimating b, we have to construct a suitable sequence of conditional distributions of Y given X, and because Y takes values in $L^2(I)$, we have to construct a sequence of distributions on $L^2(I)$, which is a significant difference from [19]. To this end, we employ the theory of Gaussian measures on Banach spaces; see Chapter VIII in [35]. Finally, after the present paper was first posted on arXiv in September 2016, manuscript [30] appeared on arXiv in December 2016. In this paper, Pham and Panaretos study rates of convergence for functional time series regression models with functional outputs. However, they study a different estimator (a Fourier–Tikhonov estimator) than ours and do not derive minimax lower bounds. In the present paper, [2] and [22] cover results on functional analysis useful for functional data analysis. For mathematical background on linear inverse problems, we refer to [24].

The rest of the paper is organized as follows. In Section 2, we formally describe the setup and estimators. In Section 3, we present the main results on rates of convergence of the PCA-based estimators for the coefficient function. In Section 4, we present simulation results to verify performance of the PCA-based estimates in the finite sample. In Section 5, we present applications of our estimators to two real data examples. All the proofs are deferred to the Appendix.

1.2. Notation

We use the following notation. For any measurable functions $f : I \to \mathbb{R}$ and $R : I^2 \to \mathbb{R}$, let

$$\|f\| = \left\{\int_I f^2(t)dt\right\}^{1/2} \quad \text{and} \quad \|R\| = \left\{\iint_{I^2} R^2(s,t)dsdt\right\}^{1/2}.$$

For any functions $f, g : I \to \mathbb{R}$, define $f \otimes g : I^2 \to \mathbb{R}$ by $(f \otimes g)(s,t) = f(s)g(t)$ for all $s, t \in I$. Let $\mathcal{L}^2(I) = \{f : I \to \mathbb{R} : f \text{ is measurable}, \ \|f\| < \infty\}$, and define the equivalence relation $\sim$ for real-valued functions f, g defined on I by $f \sim g \Leftrightarrow f = g$ almost everywhere. Define $L^2(I)$ by the quotient space $L^2(I) = \mathcal{L}^2(I)/\sim$ equipped with the inner product $\langle f^\sim, g^\sim\rangle = \int_I f(t)g(t)dt$ for $f, g \in \mathcal{L}^2(I)$, where $f^\sim = \{h \in \mathcal{L}^2(I) : h \sim f\}$; the space $L^2(I)$ is a separable Hilbert space, and as usual, we identify any element in $\mathcal{L}^2(I)$ as an element of $L^2(I)$. Define $L^2(I^2)$ analogously. We also identify any real-valued function f defined almost everywhere on I (or I^2) as a function defined everywhere on I (or I^2) by setting $f(t) = 0$ for any point t at which f is undefined. For any positive sequences a_n, c_n, we write $a_n \sim c_n$ if a_n/c_n is bounded and bounded away from zero. In what follows, let $(\Omega, \mathcal{A}, \mathrm{P})$ denote an underlying probability space.

2. Setup and estimators

Suppose that we observe a pair of random functions (X, Y) indexed by $I = [0,1]$, where $X = \{X(t) : t \in I\}$ and $Y = \{Y(t) : t \in I\}$ are predictor and response variables, respectively. We assume that X and Y are $L^2(I)$-valued random variables such that $\mathrm{E}(\|X\|^2) < \infty$ and $\mathrm{E}(\|Y\|^2) < \infty$—recall that a measurable stochastic process with paths in $L^2(I)$ almost surely induces an $L^2(I)$-valued random variable, and vice versa; see [31] or [4]. We consider a functional linear regression model

$$\mathrm{E}(Y \mid X)(s) = \mathrm{E}\{Y(s)\} + \int_I b(s,t)[X(t) - \mathrm{E}\{X(t)\}]dt, \tag{2}$$

where $\mathrm{E}(Y \mid X)$ is the conditional expectation of Y as an $L^2(I)$-valued random variable conditionally on the σ-field generated by X – which is well-defined since $\mathrm{E}(\|Y\|) < \infty$, and $\mathrm{E}(Y \mid X)$ itself is an $L^2(I)$-valued random variable; see Chapter 5 in [35] – and $(s,t) \mapsto b(s,t)$ is the coefficient function assumed to be in $L^2(I^2)$, i.e., $\|b\|^2 = \iint_{I^2} b^2(s,t)dsdt < \infty$. The equality in (2) should be understood as an equality between $L^2(I)$-valued random variables. Alternatively, the model (2) can be described in a more conventional way as

$$Y(s) = \mathrm{E}\{Y(s)\} + \int_I b(s,t)[X(t) - \mathrm{E}\{X(t)\}]dt + \mathcal{E}(s), \quad \mathrm{E}(\mathcal{E} \mid X)(s) = 0,$$

where $\mathcal{E} = Y - \mathrm{E}(Y \mid X)$.

The goal of this paper is estimation of the function $(s,t) \mapsto b(s,t)$, and to this end we shall employ the functional principal component analysis (PCA). Consider the covariance function, defined, for all $s, t \in I$, by $K(s,t) = \mathrm{cov}\{X(s), X(t)\}$. The assumption that $\mathrm{E}(\|X\|^2) < \infty$ ensures that $K \in L^2(I^2)$. In addition, we assume that the integral operator from $L^2(I)$ into itself with kernel K, namely the covariance operator of X, is injective (which is equivalent to the condition that $\mathrm{var}(\langle f, X\rangle) > 0$ for all $f \in L^2(I)$ with $\|f\| = 1$). The covariance operator is self-adjoint and positive definite. The Hilbert–Schmidt theorem (see [34, Theorem VI.16]) then ensures that K admits the spectral expansion

$$K(s,t) = \sum_{k=1}^{\infty} \kappa_k \phi_k(s)\phi_k(t)$$

in $L^2(I^2)$, where $\kappa_1 \geq \kappa_2 \geq \cdots > 0$ are a non-increasing sequence of eigenvalues tending to zero and $\{\phi_k : k \in \mathbb{N}\}$ is an orthonormal basis of $L^2(I)$ consisting of eigenfunctions of the integral operator, namely, $\int_I K(s,t)\phi_k(t)dt = \kappa_k\phi_k(s)$ for all $k \in \mathbb{N}$. We will later assume that there are no ties in the κ_js, i.e., $\kappa_1 > \kappa_2 > \cdots > 0$. Given that $\{\phi_k : k \in \mathbb{N}\}$ is an orthonormal basis of $L^2(I)$, we have the following expansion in $L^2(I)$:

$$X(t) = \mathrm{E}\{X(t)\} + \sum_{k=1}^{\infty} \xi_k \phi_k(t),$$

where each ξ_k is defined by

$$\xi_k = \int_I [X(t) - \mathrm{E}\{X(t)\}]\phi_k(t)dt.$$

By Parseval's identity and Fubini's theorem, $\sum_{k=1}^{\infty} \mathrm{E}(\xi_k^2) = \int_I \mathrm{var}\{X(t)\}dt < \infty$ and

$$\mathrm{E}(\xi_k\xi_\ell) = \iint_{I^2} K(s,t)\phi_k(s)\phi_\ell(t)dsdt = \begin{cases} \kappa_k & \text{if } k = \ell, \\ 0 & \text{if } k \neq \ell. \end{cases} \tag{3}$$

Furthermore, since $\{\phi_j \otimes \phi_k : j, k \in \mathbb{N}\}$ is an orthonormal basis of $L^2(I^2)$, we have

$$b(s,t) = \sum_{k=1}^{\infty} b_{j,k}\phi_j(s)\phi_k(t)$$

in $L^2(I^2)$ with $b_{j,k} = \iint_{I^2} b(s,t)\phi_j(s)\phi_k(t)dsdt$. This yields

$$\int_I b(s,t)[X(t) - \mathrm{E}\{X(t)\}]dt = \sum_{j=1}^{\infty} \left(\sum_{k=1}^{\infty} b_{j,k}\xi_k \right) \phi_j(s).$$

Now, because of (3) and given that the expansion of X holds in $L^2(I \times \Omega, dt \otimes d\mathrm{P})$ too (i.e., $\mathrm{E}[\|X - \mathrm{E}\{X(\cdot)\} - \sum_{k=1}^{N} \xi_k\phi_k\|^2] = \sum_{k=N+1}^{\infty} \mathrm{E}(\xi_k^2) \to 0$ as $N \to \infty$), we have $\mathrm{E}\{\xi_k Y(s)\} = \kappa_k \sum_{j=1}^{\infty} b_{j,k}\phi_j(s)$, where the equality holds in $L^2(I)$, and therefore we obtain the following characterization of b:

$$b(s,t) = \sum_{k=1}^{\infty} \frac{\mathrm{E}\{\xi_k Y(s)\}}{\kappa_k} \phi_k(t). \tag{4}$$

This characterization leads to a method for estimating b.

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent copies of (X, Y) as $(L^2(I) \times L^2(I))$-valued random variables. We estimate K by the empirical covariance function $\widehat{K}$ defined, for all $s, t, \in I$, by

$$\widehat{K}(s,t) = \frac{1}{n} \sum_{i=1}^{n} \{X_i(s) - \overline{X}(s)\}\{X_i(t) - \overline{X}(t)\},$$

where $\overline{X} = (X_1 + \cdots + X_n)/n$. Let

$$\widehat{K}(s,t) = \sum_{k=1}^{\infty} \widehat{\kappa}_k \widehat{\phi}_k(s)\widehat{\phi}_k(t) \tag{5}$$

be the spectral expansion of $\widehat{K}$ in $L^2(I^2)$, where $\widehat{\kappa}_1 \geq \widehat{\kappa}_2 \geq \cdots \geq 0$ are a non-increasing sequence of eigenvalues tending to zero and $\{\widehat{\phi}_k : k \in \mathbb{N}\}$ is an orthonormal basis of $L^2(I)$ consisting of eigenfunctions of the integral operator with kernel $\widehat{K}$, namely, for all $k \in \mathbb{N}$

$$\int_I \widehat{K}(s,t)\widehat{\phi}_k(t)dt = \widehat{\kappa}_k\widehat{\phi}_k(s).$$

The spectral expansion in (5) is possible because the integral operator with kernel $\widehat{K}$ is of finite rank (at most $n-1$), and so in addition to an orthonormal system of $L^2(I)$ consisting of eigenfunctions corresponding to the positive eigenvalues, we can add functions so that the augmented system of functions $\{\widehat{\phi}_k : k \in \mathbb{N}\}$ becomes an orthonormal basis of $L^2(I)$. Furthermore, let

$$\widehat{\xi}_{i,k} = \int_I \{X_i(t) - \overline{X}(t)\}\widehat{\phi}_k(t)dt.$$

Using the characterization in (4), we consider the following estimator based on the single truncation:

$$\widehat{b}(s,t) = \sum_{k=1}^{m_n} \frac{1}{n\widehat{\kappa}_k} \sum_{i=1}^{n} \widehat{\xi}_{i,k} Y_i(s) \widehat{\phi}_k(t), \tag{6}$$

where $m_n \to \infty$ as $n \to \infty$. This estimator was considered in [14].

We also consider an alternative estimator based on truncating the double series, namely, the double truncation. Let $\mathcal{E} = Y - \mathrm{E}(Y \mid X)$, and consider the expansions $Y(s) = \sum_{j=1}^{\infty} \eta_j \phi_j(s)$ and $\mathcal{E}(s) = \sum_{j=1}^{\infty} \varepsilon_j \phi_j(s)$ in $L^2(I)$. Now, given that

$$\int_I \left[\int_I b(s,t)[X(t) - \mathrm{E}\{X(t)\}]dt \right] \phi_j(s)ds = \sum_{k=1}^{\infty} b_{j,k}\xi_k$$

for each $j \in \mathbb{N}$, we have

$$\eta_j = a_j + \sum_{k=1}^{\infty} b_{j,k}\xi_k + \varepsilon_j \quad \forall j \in \mathbb{N},$$

where $a_j = \mathrm{E}(\eta_j) = \int_I \mathrm{E}\{Y(s)\}\phi_j(s)ds$ for all $j \in \mathbb{N}$. Therefore, we have $\mathrm{E}(\eta_j\xi_k) = b_{j,k}\mathrm{E}(\xi_k^2) = \kappa_k b_{j,k}$, namely,

$$b_{j,k} = \mathrm{E}(\eta_j\xi_k)/\kappa_k.$$

Based on this characterization, we consider the following alternative estimator:

$$\widetilde{b}(s,t) = \sum_{j=1}^{m_{n,1}} \sum_{k=1}^{m_{n,2}} \widetilde{b}_{j,k} \widehat{\phi}_j(s) \widehat{\phi}_k(t), \tag{7}$$

where $m_{n,1} \to \infty$ and $m_{n,2} \to \infty$ as $n \to \infty$, and each $\widetilde{b}_{j,k}$ is defined by

$$\widetilde{b}_{j,k} = \frac{1}{n} \sum_{i=1}^{n} \widehat{\eta}_{i,j} \widehat{\xi}_{i,k} / \widehat{\kappa}_k$$

with $\widehat{\eta}_{i,j} = \int_I Y_i(s)\widehat{\phi}_j(s)ds$.

In the next section, we will derive rates of convergence of the estimators $\widehat{b}$ and $\widetilde{b}$ for the integrated squared error.

Remark 1 (*Motivation of the Double Truncation*)**.** It will turn out in the next section that $\widehat{b}$ with properly chosen m_n is rate optimal, and from a theoretical point of view, the single truncation is enough for the purpose of estimating b. However, in practice, the double truncation would be a preferred option since, in general using a small number of basis functions works as a smoothing operator regardless of choice of basis functions, and hence, compared with the single truncation, the double truncation typically results in an estimate of $b(s,t)$ more regular in s and thereby yielding a more interpretable estimate. See the discussion in Chapter 16 of [33] and the real data analysis in Section 5. Hence the analysis of our second estimator is of some importance. Further, in our simulation studies, the second estimator $\widetilde{b}$ using the double truncation outperforms the first estimator $\widehat{b}$ using the single truncation in the finite sample.

Remark 2 (*Ill-posedness of Estimation of b*)**.** The problem of estimating b can be regarded as a problem of estimating an unknown operator in the operator equation, and therefore is an ill-posed inverse problem. For any $R \in L^2(I^2)$, let $T_R : L^2(I) \to L^2(I)$ denote the integral operator with kernel R defined, for all $h \in L^2(I)$, by

$$(T_R h)(s) = \int_I R(s,t)h(t)dt.$$

The adjoint operator T_R^* of T_R is also an integral operator and of the form

$$(T_R^* h)(t) = \int_I R(s,t)h(s)ds.$$

Now, let $C_{XY}(s,t) = \mathrm{cov}\{X(s), Y(t)\}$ for all $s, t \in I$. Then, using the symmetry of K, we have that for any $h \in L^2(I)$,

$$(T_{C_{XY}} h)(t) = \iint_{I^2} K(t,u)b(s,u)h(s)dsdu = (T_K T_b^* h)(t),$$

i.e., $T_{C_{XY}} = T_K T_b^*$. Since we are assuming that T_K is injective, we have that $T_b^* = T_K^{-1} T_{C_{XY}}$. Both $\mathrm{cov}\{X(s), Y(t)\}$ and K can be directly estimated from the data. However, since T_K is a compact operator [34, Theorems VI.22 and VI.23], T_K^{-1} is necessarily unbounded [24, p. 23], and therefore the problem of recovering T_b^* is ill-posed [24, Section 15.1]. In fact, consider $C_{XY}^N = C_{XY} + \kappa_N \phi_N \otimes \phi_1$, which converges to C_{XY} in $L^2(I^2)$ as $N \to \infty$, i.e., $T_{C_{XY}^N}$ converges to $T_{C_{XY}}$ in the Hilbert–Schmidt norm. It is seen that $b^N = b + \phi_1 \otimes \phi_N$ satisfies $T_{b^N}^* = T_K^{-1} T_{C_{XY}^N}$, but $\|b^N - b\| = \|\phi_1 \otimes \phi_N\| = 1$.

Remark 3 (*Rationales for Using $\{\widehat{\phi}_k : k \in \mathbb{N}\}$ for Decomposing Y and Other Estimators*)**.** There are in fact alternative choices to decompose Y instead of the empirical eigenfunctions $\{\widehat{\phi}_k : k \in \mathbb{N}\}$ of the covariance operator of X. For instance, one could use a neutral (i.e., non-stochastic) basis of $L^2(I)$, such as splines or wavelet bases. The analysis of the estimator with a neutral basis is basically in lines with (or simpler than) that of $\widetilde{b}$. In contrast, [37] consider the empirical eigenfunctions of the covariance operator of Y to decompose Y. We would like to state here rationales to use the empirical eigenfunctions $\{\widehat{\phi}_k : k \in \mathbb{N}\}$ of the covariance operator of X to decompose Y. First, there is no general guidance on how to choose a neutral basis to decompose Y; however, the X-related eigenfunctions are already at hand, and we can bypass the problem of selecting neutral bases for Y. Second, although the empirical eigenfunctions of the covariance operator of Y would be natural to decompose Y, for the resulting estimator to achieve sharp rates of convergence, we would have to make additional non-trivial assumptions on the eigenvalues of the covariance operator of Y (e.g., an assumption such that the eigenvalues are "well-separated" from one another) to insure sufficient estimation accuracy for the empirical eigenfunctions. Finally, in applications, there are situations where the periodic behaviors of X and Y tend to be similar (e.g., the Canadian weather data example of [33]), and in such cases, the X-related eigenfunctions appear to be suitable for decomposing Y.

3. Main results

3.1. Rates of convergence

In this subsection, we derive rates of convergence of the estimators $\widehat{b}$ and $\widetilde{b}$ defined in (6) and (7), respectively. To this end, we make the following assumption. Recall that $\mathcal{E} = Y - \mathrm{E}(Y \mid X)$.

Assumption 1. There exist constants $\alpha > 1$, $\beta > \alpha/2 + 1$, $\gamma > 1/2$, and $C_1 > 1$ such that

$$\mathrm{E}(\|Y\|^2) < \infty, \quad \mathrm{E}(\|X\|^2) < \infty, \quad \mathrm{E}(\|\mathcal{E}\|^2 \mid X) \le C_1 \text{ almost surely}, \tag{8}$$

$$\mathrm{E}(\xi_k^4) \le C_1 \kappa_k^2 \ \forall k \in \mathbb{N}, \tag{9}$$

$$\kappa_k \le C_1 k^{-\alpha}, \ \kappa_k - \kappa_{k+1} \ge C_1^{-1} k^{-\alpha-1} \ \forall k \in \mathbb{N}, \tag{10}$$

$$|b_{j,k}| \le C_1 j^{-\gamma} k^{-\beta} \ \forall j, k \in \mathbb{N}. \tag{11}$$

Some comments on Assumption 1 are in order. The first row (8) is a standard moment condition. The second (9) and third rows (10) are adapted from [19]. Condition (9) is standard in the literature on functional linear models. Concretely, Condition (9) is automatically satisfied if X is Gaussian, since in that case ξ_k are Gaussian. In Condition (10), as in [5] and [19], we require that the eigenvalues $\{\kappa_k : k \in \mathbb{N}\}$ are "well-separated", namely, $\kappa_k - \kappa_{k+1} \ge C_1^{-1} k^{-\alpha-1}$ for all $k \in \mathbb{N}$. This condition is used to ensure sufficient estimation accuracy of the empirical eigenfunctions $\widehat{\phi}_k$. This condition also ensures that, since $\kappa_k \to 0$ as $k \to \infty$,

$$\kappa_k = \sum_{j=k}^{\infty} (\kappa_j - \kappa_{j+1}) \ge \frac{1}{C_1} \sum_{j=k}^{\infty} j^{-\alpha-1} \ge \frac{k^{-\alpha}}{C_1 \alpha}.$$

So $\kappa_k \sim k^{-\alpha}$ as $k \to \infty$. The value of α measures "ill-posedness" of the estimation problem, so that the larger α is, the more difficult estimation of b will be. For given constants $\alpha > 1$ and $C_1 > 1$, the class of distributions of X verifying (8)–(10) is rich enough, and described as

$$\Big\{ \mathrm{P} \circ X^{-1} : X = \sum_k \sqrt{\kappa_k}\, U_k \phi_k, \ \{\phi_k : k \in \mathbb{N}\} \text{ is an orthonormal basis of } L^2(I),$$
$$\{U_k\} \sim \mathcal{WN}(0, 1), \ \mathrm{E}(U_k^4) \le C_1 \kappa_k \le C_1\, k^{-\alpha} \forall k \in \mathbb{N}, \ \kappa_k - \kappa_{k+1} \ge C_1^{-1} k^{-\alpha-1} \ \forall k \in \mathbb{N} \Big\},$$

where $\{U_k\} \sim \mathcal{WN}(0, 1)$ means that $\{U_k\}$ is a white noise process (i.e., an uncorrelated sequence of random variables) with mean zero and unit variance.

The last condition (11) is a smoothness condition on b, where the smoothness is measured through the eigenfunction system $\{\phi_k : k \in \mathbb{N}\}$, which is natural in our setting. Since $b(s, t)$ is a bivariate function, however, there are potentially a number of variations on how $b_{j,k}$ decays as $j \to \infty$ or $k \to \infty$. We focus on a simple case where $|b_{j,k}|$ decays like $j^{-\gamma} k^{-\beta}$ as $j \to \infty$ or $k \to \infty$, and γ measures smoothness of $b(s, t)$ in s while β measures smoothness of $b(s, t)$ in t. We also require that $\beta > \alpha/2 + 1$ for a technical reason; see the discussion after Theorem 1.

The following theorem establishes rates of convergence for $\widehat{b}$.

Theorem 1. *Consider the estimator $\widehat{b}$ defined in (6). Suppose that Assumption 1 is satisfied. Choose m_n in such a way that $m_n \to \infty$ and $m_n = o\{n^{1/(2\alpha+2)}\}$. Then*

$$\|\widehat{b} - b\|^2 = O_{\mathrm{P}}\{n^{-1}m_n^{\alpha+1} + m_n^{-2\beta+1}\}.$$

Therefore, by choosing $m_n \sim n^{1/(\alpha+2\beta)}$, we have $\|\widehat{b} - b\|^2 = O_{\mathrm{P}}\{n^{-(2\beta-1)/(\alpha+2\beta)}\}$.

Remark 4. It is not difficult to verify from the proof of Theorem 1 that the results of the theorem hold uniformly over a class of distributions $\mathcal{F}(\alpha, \beta, \gamma, C_1)$ of (X, Y) that verify (2) and (8)–(11) for given constants $\alpha > 1$, $\beta > \alpha/2 + 1$, $\gamma > 1/2$, and $C_1 > 1$. In particular, by choosing $m_n \sim n^{1/(\alpha+2\beta)}$, we have

$$\lim_{D\to\infty} \limsup_{n\to\infty} \sup_{F\in\mathcal{F}(\alpha,\beta,\gamma,C_1)} \mathrm{P}_F\left\{\|\widehat{b} - b\|^2 > Dn^{-(2\beta-1)/(\alpha+2\beta)}\right\} = 0,$$

where P_F denotes the probability under F. We will show in Theorem 3 that the rate $n^{-(2\beta-1)/(\alpha+2\beta)}$ is minimax optimal.

The requirement that $m_n = o\{n^{1/(2\alpha+2)}\}$ comes from the following reason. In the proof of Theorem 1, we require that there exists a sufficiently small constant $c > 0$ such that, with probability approaching 1, $|\widehat{\kappa}_k - \kappa_\ell| \geq c|\kappa_k - \kappa_\ell|$ for all $1 \leq k \leq m_n$ and $\ell \neq k$. Since $|\widehat{\kappa}_k - \kappa_\ell| \geq |\kappa_k - \kappa_\ell| - |\widehat{\kappa}_k - \kappa_k|$ and $\sup_{k\geq 1}|\widehat{\kappa}_k - \kappa_k| \leq \|\widehat{K} - K\| = O_{\mathrm{P}}(n^{-1/2})$ by Lemma 4.2 in [2], it suffices to have that $n^{1/2}\inf_{1\leq k\leq m_n, \ell\neq k}|\kappa_k - \kappa_\ell| \to \infty$. Now, for any $1 \leq k \leq m_n$ and $\ell \neq k$, $|\kappa_k - \kappa_\ell| \geq \min(\kappa_k - \kappa_{k+1}, \kappa_{k-1} - \kappa_k) \geq C_1^{-1}k^{-\alpha-1} \geq C_1^{-1}m_n^{-\alpha-1}$, and to ensure that $n^{1/2}m_n^{-\alpha-1} \to \infty$, we need that $m_n = o\{n^{1/(2\alpha+2)}\}$. In addition, in order that $m_n \sim n^{1/(\alpha+2\beta)}$ satisfies $m_n = o\{n^{1/(2\alpha+2)}\}$, we need that $\beta > \alpha/2 + 1$.

The theorem shows that the value of γ does not affect rates of convergence of $\widehat{b}$, which is perhaps not surprising in view of the definition of $\widehat{b}$. What is interesting is the fact that $\widehat{b}$ with m_n properly chosen is rate optimal, which means that smoothness of $b(s, t)$ in s does not affect the difficulty of estimating b. This is in sharp contrast with nonparametric estimation of a bivariate regression function. It should be noted that the results of Theorem 1 continue to hold even if the condition that $|b_{j,k}| \leq C_1 j^{-\gamma}k^{-\beta}$ for all $j, k \in \mathbb{N}$ is replaced by a weaker condition that $|b_{j,k}| \leq C_1\ell_j k^{-\beta}$ for all $j, k \in \mathbb{N}$ for some (given) positive sequence $\{\ell_j : j \in \mathbb{N}\}$ such that $\sum_{j=1}^{\infty}\ell_j^2 < \infty$. However, the value of γ does matter for the analysis of the second estimator $\widetilde{b}$.

Crambes and Mas [14] study prediction based on the estimator $\widehat{b}$. They prove that, assuming $\mathrm{E}\{Y(t)\} = \mathrm{E}\{X(t)\} = 0$ for all $t \in I$, the estimator $\widehat{Y}_{n+1}(s) = \int_I \widehat{b}(s, t)X_{n+1}(t)dt$ with an appropriate choice of the cut-off level m_n attains the minimax rate for estimation of $\mathrm{E}(Y_{n+1} \mid X_{n+1})$ under the mean integrated squared error (MISE). Importantly, the prediction problem considered in [14] is related to, but substantially different from, the problem of estimating b considered in the present paper; the former is not an ill-posed inverse problem (is not a type of problems formulated as solving an integral equation; see Remark 2), and Crambes and Mas [14] do not derive sharp rates of convergence for $\widehat{b}$ itself and hence do not cover Theorem 1—the proof of Theorem 2 in [14] does not lead to the results of our Theorem 1 since from the beginning their proof is bounding $\mathrm{E}[\|\int_I\{\widehat{b}(\cdot, t) - b(\cdot, t)\}X_{n+1}(t)dt\|^2]$; furthermore, Crambes and Mas [14] assume a stronger moment condition on ξ_k; see (6) in their paper. The estimator $\widehat{b}$ with dependent functional data is analyzed in [21,29], but the authors only prove consistency of $\widehat{b}$ and thus do not cover Theorem 1; to be precise, they prove consistency of the integral operator with kernel $\widehat{b}$ for the operator norm.

Next, we derive rates of convergence for our second estimator.

Theorem 2. *Consider the estimator $\widetilde{b}$ defined in (7). Suppose that Assumption 1 is satisfied. Furthermore, suppose that $\gamma > \beta/2 + 1$. Then provided that $\max(m_{n,1}, m_{n,2}) = o\{n^{1/(2\alpha+2)}\}$, we have*

$$\|\widetilde{b} - b\|^2 = O_{\mathrm{P}}\left\{n^{-1}(m_{n,1} + m_{n,2}^{\alpha+1}) + m_{n,1}^{-2\gamma+1} + m_{n,2}^{-2\beta+1}\right\}. \tag{12}$$

Therefore, by choosing $m_{n,1} \sim \min\{n^{1/(2\gamma)}, (n/\ln n)^{1/(2\alpha+2)}\}$ and $m_{n,2} \sim n^{1/(\alpha+2\beta)}$, we have

$$\|\widetilde{b} - b\|^2 = O_{\mathrm{P}}\left[\max\{(n/\ln n)^{-(2\gamma-1)/(2\alpha+2)}, n^{-(2\gamma-1)/(2\gamma)}, n^{-(2\beta-1)/(\alpha+2\beta)}\}\right]. \tag{13}$$

Since the estimator $\widetilde{b}(s, t)$ depends on $\widehat{\phi}_1(s), \ldots, \widehat{\phi}_{m_{n,1}}(s)$, the accumulation of these estimation errors contributes to the term $m_{n,1}/n$ in the bound (12), while the term $m_{n,1}^{-2\gamma+1}$ comes from the bias. Because of these terms, γ appears in the bound (13), and in contrast to $\widehat{b}$, the second estimator $\widetilde{b}$ has suboptimal rates in some cases; of course there could be a room to improve upon the bound (12). Still, the estimator $\widetilde{b}$ is able to attain the optimal rate $n^{-(2\beta-1)/(\alpha+2\beta)}$ provided that

$$\begin{aligned} &\beta \leq \frac{(2\gamma-1)\alpha + 2\gamma}{2} \qquad \text{if } \gamma > \alpha + 1, \\ &\beta < \frac{(2\gamma-1)\alpha + 2\alpha + 2}{2(2\alpha - 2\gamma + 3)} \quad \text{if } \gamma \leq \alpha + 1, \end{aligned} \tag{14}$$

which actually covers wide regions of (α, β, γ). Fig. 1 depicts regions of (β, γ) where $\widetilde{b}$ attains the rate $n^{-(2\beta-1)/(\alpha+2\beta)}$ for different values of α. We plot two regions $(A) = \{(\beta, \gamma) : \alpha/2 + 1 < \beta \leq \{(2\gamma - 1)\alpha + 2\gamma\}/2, \gamma > \alpha + 1\}$ and

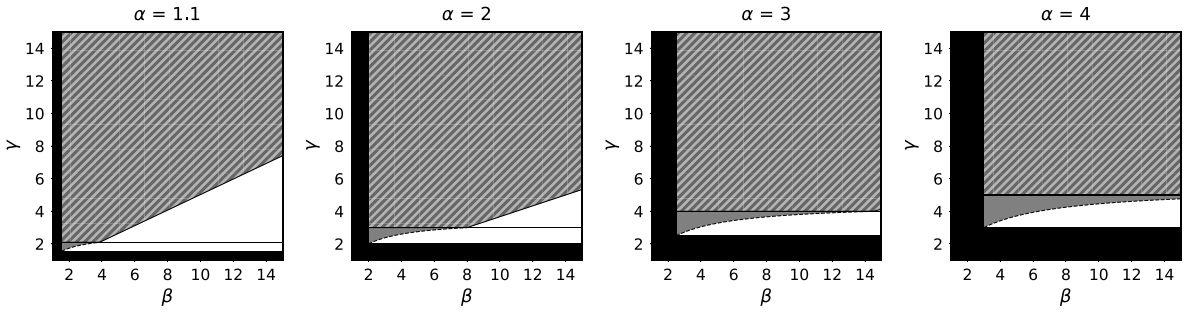

Fig. 1. Regions of (β, γ) for different values of α. When the parameters (β, γ) are contained in the striped gray region (A) or the dark gray region (B), the estimator $\widetilde{b}$ attains the rate $n^{-(2\beta-1)/(\alpha+2\beta)}$. The black region corresponds to the region where $\beta \leq \alpha/2+1$ or $\gamma \leq \alpha/2+1$.

$(B) = \{(\beta,\gamma) : \alpha/2+1 < \beta < \{(2\gamma-1)\alpha+2\alpha+2\}/\{2(2\alpha-2\gamma+3)\}, \alpha/2+1 < \gamma \leq \alpha+1\}$ in Fig. 1. The figure shows that the estimator $\widetilde{\beta}$ is able to attain the minimax optimal rate when the slope function $b(s,t)$ is sufficiently smooth in s relative to smoothness in t; further, γ need not be larger than β in order that $\widetilde{b}$ attains the minimax optimal rate, and in fact for large α, γ can be much smaller than β.

Yao et al. [37] consider an estimator for b that is related to, but still different from, our second estimator $\widetilde{b}$. Their estimator is based on applying the functional PCA to both X and Y. Let $L(s,t) = \mathrm{cov}\{Y(s), Y(t)\}$ be the covariance function of Y, and let $L(s,t) = \sum_{j=1}^{\infty} \rho_j \psi_j(s)\psi_j(t)$ be the spectral expansion of L, where $\rho_1 \geq \rho_2 \geq \cdots > 0$ and $\{\psi_j : j \in \mathbb{N}\}$ is an orthonormal basis of $L^2(I)$ (we assume here that this expansion is possible). Then $Y(s) = \mathrm{E}\{Y(s)\} + \sum_{j=1}^{\infty} \zeta_j \psi_j(s)$ where $\zeta_j = \int_I [Y(s) - \mathrm{E}\{Y(s)\}]\psi_j(s)ds$, and observe that b can be expanded in $L^2(I^2)$ as $b(s,t) = \sum_{k=1}^{\infty}\{\mathrm{cov}(\zeta_j, \xi_k)/\kappa_k\}\psi_j(s)\phi_k(t)$. The method of estimation of b in [37] is to approximate the infinite series by a finite sum, and replace $\mathrm{cov}(\zeta_j, \xi_k)$, κ_k, ψ_j, and ϕ_k by their estimators. However, Yao et al. [37] do not explicitly derive rates of convergence of this estimator, although it should be noted that they assume that only discrete measurements with measurement errors for X and Y are available. The analysis of the estimator of Yao et al. [37] requires a substantially different set of assumptions than ours and thus is not pursued in the present paper.

3.2. Minimax lower bounds

In this subsection, we derive minimax lower bounds for estimation of b. Since the minimax rate of convergence is bounded from below by that in a restricted class of distributions (recall that "max" is taken with respect to the underlying distribution), in the case of deriving minimax lower bounds, we may focus the analysis to a restricted class of distributions, as long as the resulting lower bound is matched with the upper bound. Hence, in this subsection, we consider the following setting.

Let $\alpha > 1, \beta > 1/2, \gamma > 1/2$, and $C_1 > 1$ be given constants. Let $\mathcal{E}$ be an $L^2(I)$-valued Gaussian random variable such that $\mathrm{E}(\langle f, \mathcal{E}\rangle) = 0$ and $\mathrm{E}(\langle f, \mathcal{E}\rangle^2) > 0$ for all $f \in L^2(I)$ with $\|f\| = 1$ (recall that an $L^2(I)$-valued random variable Z is said to be Gaussian if $\langle f, Z\rangle$ is normally distributed for each $f \in L^2(I)$). Let $R(s,t) = \mathrm{E}\{\mathcal{E}(s)\mathcal{E}(t)\}$ be the covariance function of $\mathcal{E}$, and let $R(s,t) = \sum_{j=1}^{\infty} \lambda_j \phi_j(s)\phi_j(t)$ be the spectral expansion of R, where $\lambda_1 \geq \lambda_2 \geq \cdots > 0$ and $\{\phi_j : j \in \mathbb{N}\}$ is an orthonormal basis of $L^2(I)$. Now, let $X = \sum_{k=1}^{\infty} k^{-\alpha/2} U_k \phi_k$ for $U_1, U_2, \ldots$ being mutually independent $\mathcal{U}[-\sqrt{3}, \sqrt{3}]$ random variables independent from $\mathcal{E}$, and generate, as an $L^2(I)$-valued random variable, $Y(\cdot) = \int_I b(\cdot, t)X(t)dt + \mathcal{E}(\cdot)$, where $b \in L^2(I^2)$. Since U_k has mean zero and unit variance, we have $\kappa_k = k^{-\alpha}$, and so

$$\kappa_k - \kappa_{k+1} = \alpha \int_k^{k+1} u^{-\alpha-1} du \geq \alpha(k+1)^{-\alpha-1} \geq \alpha 2^{-\alpha-1} k^{-\alpha-1}.$$

In addition, $\xi_k = k^{-\alpha/2}U_k$, and so $\mathrm{E}(\xi_k^4) = 9k^{-2\alpha}/5$. Define a class of functions for b as

$$\mathcal{B}(\beta, \gamma, C_1) = \left\{ b = \sum_{k=1}^{\infty} b_{j,k}\phi_j \otimes \phi_k : |b_{j,k}| \leq C_1 j^{-\gamma} k^{-\beta} \ \forall j, k \in \mathbb{N} \right\}.$$

Theorem 3. *Work with the setting described as above. Then there exists a constant $c > 0$ such that*

$$\liminf_{n\to\infty} \inf_{\overline{b}^n} \sup_{b \in \mathcal{B}(\beta,\gamma,C_1)} \mathrm{P}_b\{\|\overline{b}^n - b\|^2 \geq c n^{-(2\beta-1)/(\alpha+2\beta)}\} > 0,$$

where P_b denotes the probability under b, and $\sup_{\overline{b}^n}$ is taken over all estimators $\overline{b}^n$ of b based on independent copies $(X_1, Y_1), \ldots, (X_n, Y_n)$ of (X, Y).

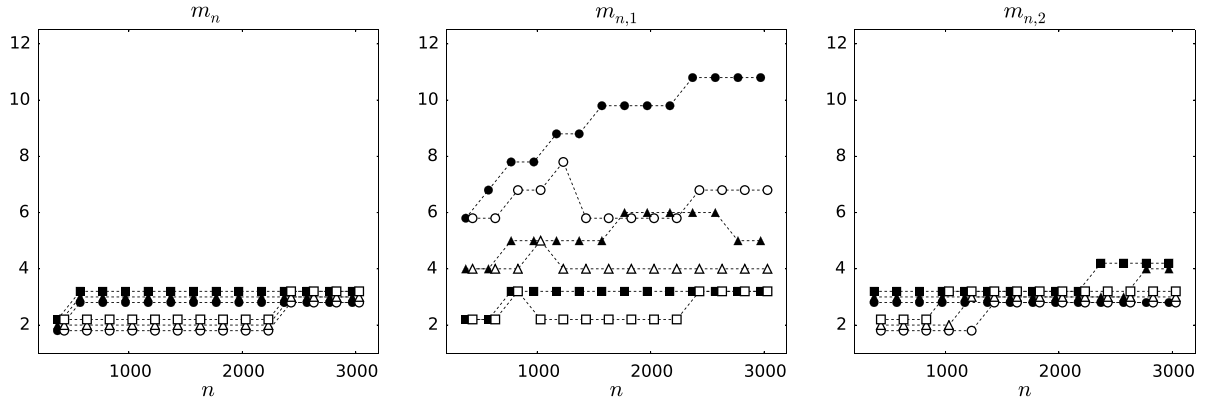

Fig. 2. The values of m_n (left panel) and $(m_{n,1}, m_{n,2})$ (middle and right panels) minimizing MISE against n for each parameter configuration. $(\alpha, \beta) = (1.2, 3.0)$ (black marker) and $(\alpha, \beta) = (2.4, 3.0)$ (white marker), and $\gamma = 2.5$ (circle), $\gamma = 3.0$ (triangle) and $\gamma = 4.0$ (square).

This theorem shows that, under Assumption 1, the first PCA-based estimator $\widehat{b}$ with m_n properly chosen is minimax rate optimal, while the second PCA-based estimator $\widetilde{b}$ with $(m_{n,1}, m_{n,2})$ properly chosen is minimax rate optimal provided that the additional restriction (14) is satisfied.

Remark 5. One might be tempted to argue that the conclusion of Theorem 3 would be derived from the following observation: taking integration of $Y(s) = \int_I b(s,t)X(t)dt + \mathcal{E}(s)$, we arrive at the functional linear model

$$Y^\sharp = \int_I b_\sharp(t)X(t)dt + \varepsilon^\sharp,$$

where $Y^\sharp = \int_I Y(s)ds$, $b_\sharp(t) = \int_I b(s,t)ds$, and $\varepsilon^\sharp = \int_I \mathcal{E}(s)ds$. Since for any estimator $\overline{b}^n$ of b, $\|\overline{b}^n_\sharp - b_\sharp\| \le \|\overline{b}^n - b\|$ where $\overline{b}^n_\sharp(t) = \int_I \overline{b}^n(s,t)ds$, the conclusion of Theorem 3 would follow from Theorem 1 in [19].

However, this argument contains a gap. The reason is that, when applying Theorem 1 in [19], we implicitly restrict estimators of $b_\sharp$ to those based on $(Y_1^\sharp, X_1), \ldots, (Y_n^\sharp, X_n)$, thereby discarding the information that the entire paths of $Y_1, \ldots, Y_n$ are fully observed, which results in restricting a class of estimators of $b_\sharp$. Therefore, formally, the conclusion of Theorem 3 does not directly follow from Theorem 1 in [19]. The proof of Theorem 3 builds on constructing a suitable sequence of conditional distributions of Y given X, and since Y takes values in $L^2(I)$, we have to construct a sequence of distributions on $L^2(I)$, which is a significant difference from [19]. To this end, we employ the theory of Gaussian measures on Banach spaces; see Chapter VIII in [35].

4. Simulation results

In this section, we present simulation results to verify the performance of the estimators in finite samples. We consider the following data generating process. Let $\phi_1 \equiv 1$, $\phi_{j+1}(t) = 2^{1/2}\cos(j\pi t)$ for $j \in \mathbb{N}$, and generate (X, Y) as follows:

$$Y(\cdot) = \int_I b(\cdot,t)X(t)dt + \mathcal{E}(\cdot), \quad X = \sum_{k=1}^{50} k^{-\alpha/2}U_k\phi_k, \quad \mathcal{E} = \sum_{j=1}^{50} j^{-1.1/2}Z_j\phi_j,$$

$$b = \sum_{j,k=1}^{50} b_{j,k}\phi_j \otimes \phi_k, \quad b_{1,1} = 0.3,\ b_{j,k} = 4(-1)^{j+k}j^{-\gamma}k^{-\beta} \text{ for } (j,k) \neq (1,1),$$

where $U_k \sim \mathcal{U}[-3^{1/2}, 3^{1/2}]$ and $Z_j \sim \mathcal{N}(0,1)$ are all independent, and the following sample sizes for n are examined: 400, 600, ..., 2800, 3000. We consider the following configurations for (α, β, γ):

$$(1.2, 3, 2.5), \quad (1.2, 3, 3), \quad (1.2, 3, 4), \quad (2.4, 3, 2.5), \quad (2.4, 3, 3), \quad (2.4, 3, 4),$$

which satisfy restriction (14). The number of repetitions for each simulation is 1000. The numerical results obtained in this section were carried out by using the matrix language Ox [17].

In this experiment, we simulate values of the MISE of $\widehat{b}$ for $m_n \in \{1, \ldots, 20\}$ and $\widetilde{b}$ for $(m_{n,1}, m_{n,2}) \in \{1, \ldots, 20\}^2$ in each case, and report the optimal MISE. The selected values of m_n and $(m_{n,1}, m_{n,2})$ in each configuration are reported in Fig. 2. It is observed that (i) the values of $m_{n,1}$ selected become smaller as γ increases; (ii) the values of $m_{n,2}$ are less sensitive to n than those of $m_{n,1}$; and (iii) the values of m_n are close to those of $m_{n,2}$.

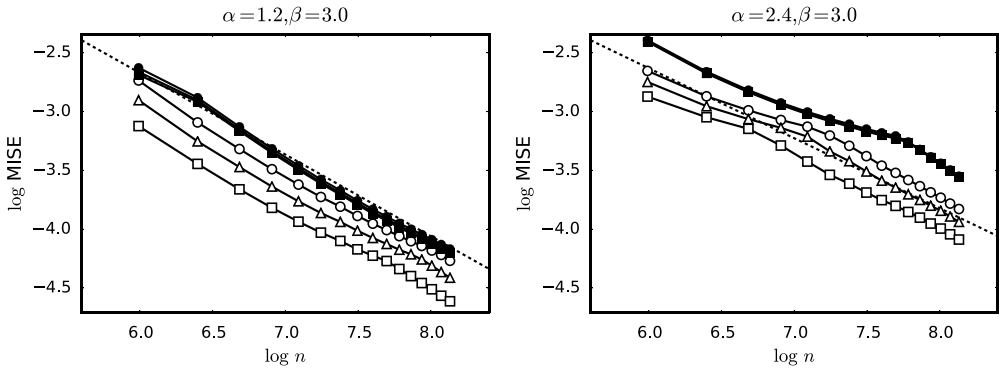

Fig. 3. Plots of log MISE against $\ln n$ for each parameter configuration : $(\alpha, \beta) = (1.2, 3.0)$ (left panel) and $(\alpha, \beta) = (2.4, 3.0)$ (right panel), and $\gamma = 2.5$ (circle), $\gamma = 3.0$ (triangle) and $\gamma = 4.0$ (square). The black and white markers correspond to values of log MISE of $\widehat{b}$ and $\widetilde{b}$, respectively. The dashed line has slope $-(2\beta - 1)/(\alpha + 2\beta)$.

Next, Fig. 3 plots the values of the log MISE against $\ln n$. It is observed that (i) the values of the log MISE of $\widehat{b}$ are almost identical for different values of γ; (ii) in contrast, the log MISE of $\widetilde{b}$ decreases as γ increases, but the slope is not sensitive to the value of γ, which indicates that the rate at which the MISE of $\widetilde{b}$ decreases is independent of γ, but the constant depends on γ and decreases as γ increases; (iii) all the slopes are close to $-(2\beta - 1)/(\alpha + 2\beta)$, at least for large n. These observations are consistent with our theoretical results. Finally, in this limited experiment, the second estimator $\widetilde{b}$ performs better than the first estimator $\widehat{b}$, especially when $\gamma = 4$; the difference in the log MISE is roughly 0.5 in that case, which means that the MISE of $\widehat{b}$ is $e^{0.5} \approx 1.65$ times that of $\widetilde{b}$. So, the simulation results would encourage using the double truncation rather than the single truncation.

5. Real data analysis

5.1. Working hours and income data

We investigate the relation between the lifetime pattern of working hours and total income using data from the National Longitudinal Survey of Youth [3]. This is a major data set in the field of human resources. It consists of a sample of 12,686 American youth born between 1957 and 1964. We use data of yearly working time (in hours) and total net family income in a year from Round 1 (1979 survey year) to Round 25 (2012 survey year).

We include cohorts who answered the 25 rounds and omit outliers who were above the 95% quantile in terms of income. Thus we have working hours and income data for 353 observations, which are plotted in Fig. 4. In the latter, the black dashed lines show the working hour data $X_i(t)$ and the income data $Y_i(t)$ for each respondent $i \in \{1, \ldots, 353\}$. The mean of working hours increases at early ages (between 20 and 30) and the mean of income monotonically increases through all the rounds of the survey. Since the income data are slightly discretized, some observations with high income take similar values. We note that both working hours and income data vary a lot; hence we pay attention to how smoothing effects of the estimators make the statistical analysis interpretable.

We use the income as a response variable and the working hours as a predictor variable. The values of m_n and $(m_{n,1}, m_{n,2})$ are selected by minimizing the cross-validation criteria as in [37], viz.

$$\min_{m_n} \sum_{i=1}^{n} \int \left[Y_i(s) - \overline{Y}(s) - \int \widehat{b}_{(-i,m_n)}(s,t)\{X_i(t) - \overline{X}(t)\}dt \right]^2 ds,$$

$$\min_{m_{n,1},m_{n,2}} \sum_{i=1}^{n} \int \left[Y_i(s) - \overline{Y}(s) - \int \widetilde{b}_{(-i,m_{n,1},m_{n,2})}(s,t)\{X_i(t) - \overline{X}(t)\}dt \right]^2 ds,$$

where $\widehat{b}_{(-i,m)}(s,t)$ and $\widetilde{b}_{(-i,m_1,m_2)}(s,t)$ are the estimates without the ith observation and with the truncation levels m_n and $(m_{n,1}, m_{n,2})$, respectively. Using these criteria, we chose $m_n = 4$ for $\widehat{b}$ and $(m_{n,1}, m_{n,2}) = (4, 4)$ for $\widetilde{b}$.

Figs. 5 and 6 plot graphs of the estimates $\widehat{b}$ and $\widetilde{b}$, respectively. Figs. 7 and 8 plot the slices of the estimates with $s \in \{1990, 2000\}$ and $t \in \{1990, 2000\}$. The overall shapes of the estimates as functions of s or t are roughly similar, but $\widetilde{b}$ is smoother in s than $\widehat{b}$ because of the double truncation. Our functional regression analysis reveals that working hours, not only in advanced ages but also in middle ones, can have positive effects on the income in advanced ages, and the positive effects get larger as the cohorts get older. In contrast, high working hours at a young age have negative effects on the income in later years.

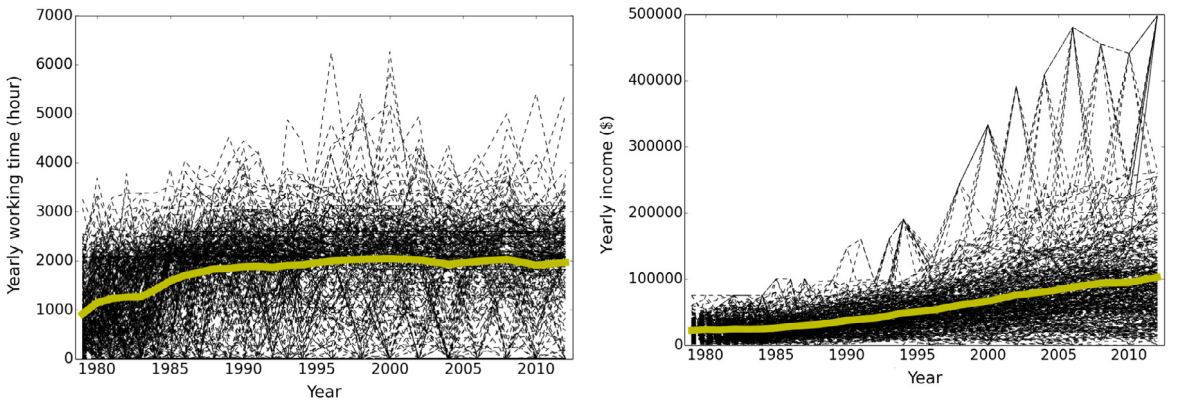

Fig. 4. Plots of labor and income data. Yearly working hour (left panel) and yearly net income (right panel) against a survey year. Each black line is data for each cohort i and the yellow line is a mean function. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

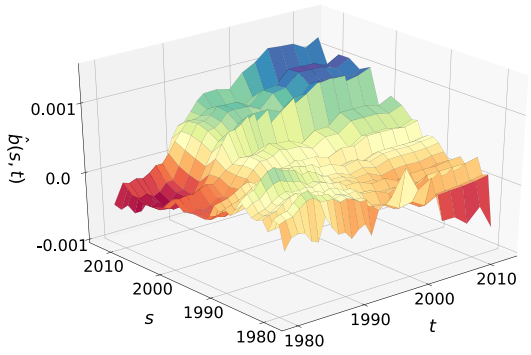

Fig. 5. $\widehat{b}(s,t)$ with the labor and income data. $m_n = 4$.

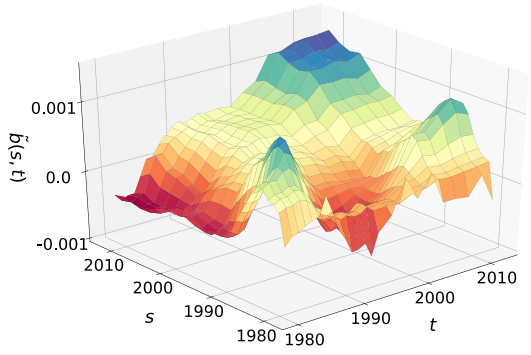

Fig. 6. $\widetilde{b}(s,t)$ with the labor and income data. $m_{n,1} = 4$ and $m_{n,2} = 4$.

The negative effects may be interpreted as follows: individuals who work a lot in youth possibly had less education, and their income does not increase much as they get old. This result is consistent with economic theory of human capital [7]: educational investment at a young age increases lifetime revenue via high earnings later in life.

5.2. Electricity prices

We investigate the mechanism of electricity spot prices of the German power market traded at the European Energy Exchange (EEX). In the German electricity market, the amount of renewable energy sources has a certain effect on the demand for the electricity because of the purchase guarantee, and the wind power in-feed has the largest influence; a detailed

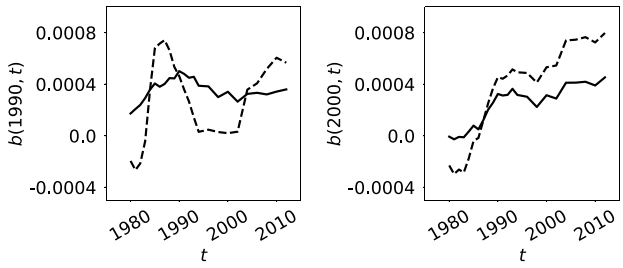

Fig. 7. Sliced $\widehat{b}$ (solid) and $\widetilde{b}$ (dashed) against t.

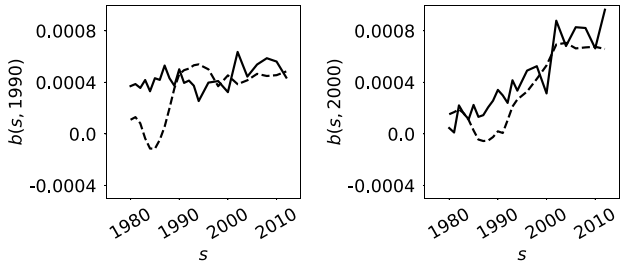

Fig. 8. Sliced $\widehat{b}$ (solid) and $\widetilde{b}$ (dashed) against s.

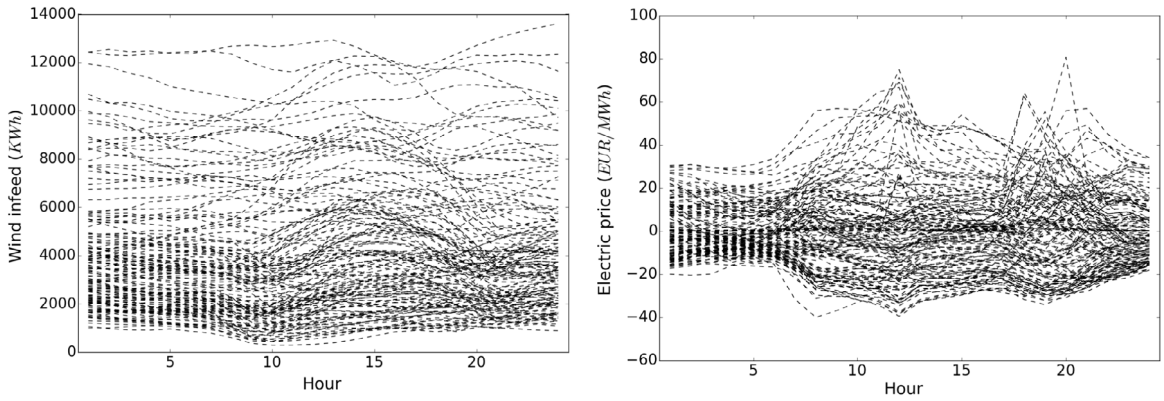

Fig. 9. Plots of wind infeed and electricity price data. Wind power infeed (left panel) and centered electric price (right panel).

discussion is found in [27]. With this background, we analyze how the wind power in-feed affects the electricity price in the German power market.

The data on prices of the German electricity market are taken from the European Energy Exchange, and the data on wind power in Germany are taken from the EEX Transparency Platform as in [27]. These data sets contain hourly electricity prices and wind power in-feed from January 2006 to September 2008, and we take $Y_i(t)$ and $X_i(t)$ to be the electricity price and wind power in-feed at time $t \in \{1, \ldots, 24\}$ and week $i \in \{1, \ldots, 143\}$; each $Y_i(t)$ is centered around its sample mean.

Fig. 9 shows a plot of the data. The functional data in this example are likely to be dependent across i, but we expect that the convergence results in this paper could be extended to weakly dependent functional data. A formal analysis with dependent functional data is beyond the scope of the paper.

For this data set, we chose $m_n = 2$ for $\widehat{b}$ and $(m_{n,1}, m_{n,2}) = (2, 1)$ for $\widetilde{b}$ by the cross-validation. Figs. 10 and 11 display graphs of the estimates $\widehat{b}$ and $\widetilde{b}$, respectively. Figs. 12 and 13 show slices of the estimates in the morning and evening (9 am and 5 pm), which show that, as before, $\widetilde{b}$ is smoother in s than $\widehat{b}$ because of the double truncation. Fig. 10 shows high fluctuations of the estimate $\widehat{b}$, which makes it difficult to interpret the estimate. In contrast, from Fig. 11, it is observed that $\widetilde{b}$ is negative in the region $(s, t) \in [8, 17] \times [8, 17]$, but is close to zero elsewhere.

This result shows that the wind power in-feed has negative effects on the electricity price in the daytime, but except for the daytime, the effect of the wind power in-feed is small. It reflects two economic phenomena of markets: price sensitivity and supply–demand balance. For the sensitivity, the market is active during the daytime and hence the correlation negatively

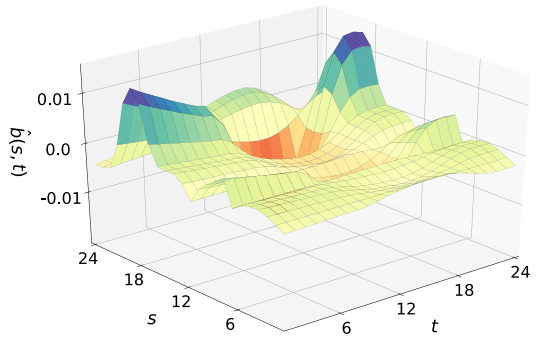

Fig. 10. $\widehat{b}(s,t)$ with the electricity price and wind power infeed data. $m_n = 2$.

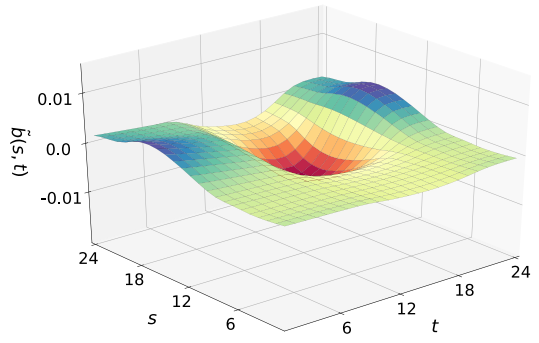

Fig. 11. $\widetilde{b}(s,t)$ with the electricity price and wind power infeed data. $m_{n,1} = 2$ and $m_{n,2} = 1$.

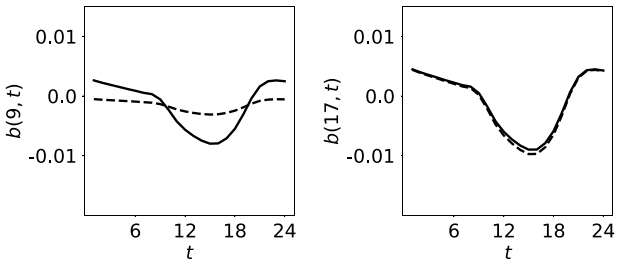

Fig. 12. Sliced $\widehat{b}$ (solid) and $\widetilde{b}$ (dashed) against t.

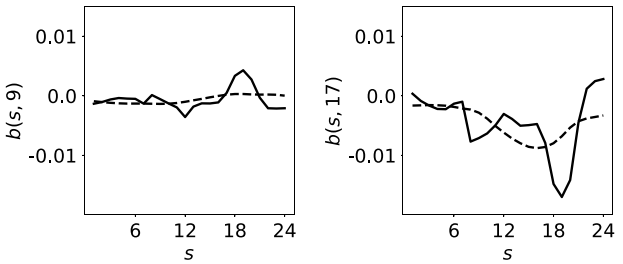

Fig. 13. Sliced $\widehat{b}$ (solid) and $\widetilde{b}$ (dashed) against s.

varies in the daytime. For the balance, more wind in-feed yields the excess supply of electricity and therefore decreases the price.

Acknowledgments

The authors would like to thank the Editor-in-Chief, Christian Genest, the Associate Editor and two anonymous referees for their careful reading of the manuscript and constructive comments that helped improve on the quality of the manuscript. M. Imaizumi is supported by Grant-in-Aid for JSPS Research Fellow (15J10206) from the JSPS, and K. Kato is supported by Grant-in-Aid for Scientific Research (C) (15K03392) from the JSPS.

Appendix. Proof of Theorem 1

In what follows, the notation $\lesssim$ signifies that the left-hand side is bounded by the right-hand side up to a constant that depends only on $\alpha, \beta, \gamma, C_1$. We first note that $\widehat{b}$ is invariant with respect to choices of signs of the $\widehat{\phi}_k$s, and so without loss of generality, we may assume that

$$\int_I \widehat{\phi}_k(t)\phi_k(t)dt \geq 0 \quad \forall k \in \mathbb{N}. \tag{A.1}$$

Recall that $m_n = o\{n^{1/(2\alpha+2)}\}$. Lemma 4.2 in [2] yields

$$\sup_{k\in\mathbb{N}} |\widehat{\kappa}_k - \kappa_k| \leq \|\widehat{K} - K\| \equiv \widehat{\Delta}. \tag{A.2}$$

Since

$$\mathrm{E}(\|X - \mathrm{E}\{X(\cdot)\}\|^4) = \mathrm{E}\left\{\left(\sum_{k=1}^{\infty}\xi_k^2\right)^2\right\} = \sum_{k=1}^{\infty}\mathrm{E}(\xi_k^2\xi_\ell^2) \leq \sum_{k,\ell\in\mathbb{N}}\{\mathrm{E}(\xi_k^4)\}^{1/2}\{\mathrm{E}(\xi_\ell^4)\}^{1/2} \lesssim \left(\sum_{k=1}^{\infty}\kappa_k\right)^2 \lesssim 1,$$

we have $\widehat{\Delta} = O_{\mathrm{P}}(n^{-1/2})$. Define the event

$$A_n = \{|\widehat{\kappa}_k - \kappa_\ell| \geq |\kappa_k - \kappa_\ell|/\sqrt{2} \text{ for all } 1 \leq k \leq m_n \text{ and for all } \ell \neq k\}.$$

It is seen that, since $|\kappa_k - \kappa_\ell| \geq \min(\kappa_{k-1} - \kappa_k, \kappa_k - \kappa_{k+1}) \geq C_1^{-1}k^{-\alpha-1} \geq C_1^{-1}m_n^{-\alpha-1}$ whenever $1 \leq k \leq m_n$ and $\ell \neq k$, and since $n^{-1/2} = o(m_n^{-\alpha-1})$, we have $\mathrm{P}(A_n) \to 1$. Furthermore, arguing as in Hall and Horowitz [19, pp. 83–84], we have

$$(1 - Cm_n^{2\alpha+2}\widehat{\Delta}^2)\,\|\widehat{\phi}_k - \phi_k\|^2 \leq 8\underbrace{\sum_{\ell:\ell\neq k}(\kappa_k - \kappa_\ell)^{-2}\left[\int\{\widehat{K}(s,t) - K(s,t)\}\phi_k(s)\phi_\ell(t)dsdt\right]^2}_{=\widehat{u}_k^2} \quad \forall k \in \{1,\ldots,m_n\},$$

where C is a constant that depends only on C_1 and $\mathrm{E}(\widehat{u}_k^2) \lesssim k^2/n$. Since $m_n^{2\alpha+2}\widehat{\Delta}^2 = o_{\mathrm{P}}(1)$, we conclude that

$$\|\widehat{\phi}_k - \phi_k\|^2 \leq 8\{1 + o_{\mathrm{P}}(1)\}\widehat{u}_k^2 \quad \text{and} \quad \mathrm{E}(\widehat{u}_k^2) \lesssim k^2/n, \tag{A.3}$$

where $o_{\mathrm{P}}(1)$ is uniform in $j \in \{1,\ldots,m_n\}$.

In what follows, we will freely use the estimates in (A.2) and (A.3). In particular, since $\beta > 3/2$, we have

$$\sum_{k=1}^{m_n} k^{-2\beta}\|\widehat{\phi}_k - \phi_k\|^2 = O_{\mathrm{P}}\left(n^{-1}\sum_{j=1}^{\infty}k^{-2\beta+2}\right) = O_{\mathrm{P}}(n^{-1}).$$

In what follows, integrations such as $\int_I f(t)dt$ and $\iint_{I^2} R(s,t)dsdt$ are abbreviated as $\int f$ and $\iint R$.

Let $\mathcal{E}_i = Y_i - \mathrm{E}(Y_i \mid X_i)$, and expand Y_i and $\mathcal{E}_i$ as $Y_i = \sum_{j=1}^{\infty}\eta_{i,j}\phi_j$ and $\mathcal{E}_i = \sum_{j=1}^{\infty}\varepsilon_{i,j}\phi_j$, where $\eta_{i,j} = \int Y_i\phi_j$ and $\varepsilon_{i,j} = \int \mathcal{E}_i\phi_j$. Observe that $\widehat{b}$ admits the following alternative expansion in $L^2(I)$:

$$\widehat{b} = \sum_{j=1}^{\infty}\sum_{k=1}^{m_n}\widehat{b}_{j,k}(\phi_j \otimes \widehat{\phi}_k),$$

where $\widehat{b}_{j,k} = n^{-1}\sum_{i=1}^{n}\eta_{i,j}\widehat{\xi}_{i,k}/\widehat{\kappa}_k$. Because $\{\phi_j \otimes \widehat{\phi}_k : j,k \in \mathbb{N}\}$ is an orthonormal basis of $L^2(I^2)$, expand b as $b = \sum_{k=1}^{\infty}\breve{b}_{j,k}(\phi_j \otimes \widehat{\phi}_k)$ with $\breve{b}_{j,k} = \iint b(\phi_j \otimes \widehat{\phi}_k)$.

Now, setting $\eta_{i,j}^c = \eta_{i,j} - n^{-1}\sum_{i'=1}^{n}\eta_{i',j}$ and $\varepsilon_{i,j}^c = \varepsilon_{i,j} - n^{-1}\sum_{i'=1}^{n}\varepsilon_{i',j}$, observe that

$$\eta_{i,j}^c = \sum_{\ell=1}^{\infty}\breve{b}_{j,\ell}\widehat{\xi}_{i,\ell} + \varepsilon_{i,j}^c.$$

Plugging this expression into $\widehat{b}_{j,k}$ together with the facts that

$$n^{-1}\sum_{i=1}^{n}\widehat{\xi}_{i,k}=0,\ n^{-1}\sum_{i=1}^{n}\widehat{\xi}_{i,\ell}\widehat{\xi}_{i,k}=\iint\widehat{K}(\widehat{\phi}_\ell\otimes\widehat{\phi}_k)=\begin{cases}\widehat{\kappa}_k & \text{if } \ell=k,\\ 0 & \text{if } \ell\neq k,\end{cases} \tag{A.4}$$

we have $\widehat{b}_{j,k}=\check{b}_{j,k}+n^{-1}\sum_{i=1}^{n}\varepsilon_{i,j}\widehat{\xi}_{i,k}/\widehat{\kappa}_k$. Therefore,

$$(\widehat{b}_{j,k}-b_{j,k})^2\lesssim(\check{b}_{j,k}-b_{j,k})^2+\widehat{\kappa}_k^{-2}\left(\frac{1}{n}\sum_{i=1}^{n}\varepsilon_{i,j}\widehat{\xi}_{i,k}\right)^2.$$

We divide the rest of the proof into three steps.

Step 1. We wish to bound $\sum_{j=1}^{\infty}\sum_{k=1}^{m_n}(\check{b}_{j,k}-b_{j,k})^2$. We will use the following expansion: if $\inf_{\ell:\ell\neq k}|\widehat{\kappa}_k-\kappa_\ell|>0$, then

$$\widehat{\phi}_k-\phi_k=\sum_{\ell:\ell\neq k}(\widehat{\kappa}_k-\kappa_\ell)^{-1}\phi_\ell\iint(\widehat{K}-K)(\widehat{\phi}_k\otimes\phi_\ell)+\phi_k\int(\widehat{\phi}_k-\phi_k)\phi_k. \tag{A.5}$$

See Lemma 5.1 in [19]. Observe that

$$\begin{aligned}\check{b}_{j,k}-b_{j,k}&=\sum_{\ell:\ell\neq k}b_{j,\ell}(\widehat{\kappa}_k-\kappa_\ell)^{-1}\iint(\widehat{K}-K)(\widehat{\phi}_k\otimes\phi_\ell)+b_{j,k}\int(\widehat{\phi}_k-\phi_k)\phi_k\\&=\sum_{\ell:\ell\neq k}b_{j,\ell}(\kappa_k-\kappa_\ell)^{-1}\iint(\widehat{K}-K)(\phi_k\otimes\phi_\ell)\\&\quad+\sum_{\ell:\ell\neq k}b_{j,\ell}\{(\widehat{\kappa}_k-\kappa_\ell)^{-1}-(\kappa_k-\kappa_\ell)^{-1}\}\iint(\widehat{K}-K)(\phi_k\otimes\phi_\ell)\\&\quad+\sum_{\ell:\ell\neq k}b_{j,\ell}(\widehat{\kappa}_k-\kappa_\ell)^{-1}\iint(\widehat{K}-K)\{(\widehat{\phi}_k-\phi_k)\otimes\phi_\ell\}+b_{j,k}\int(\widehat{\phi}_k-\phi_k)\phi_k\\&\equiv T_{j,k,1}+T_{j,k,2}+T_{j,k,3}+T_{j,k,4}.\end{aligned}$$

It is seen that $|T_{j,k,4}|\lesssim j^{-\gamma}k^{-\beta}\|\widehat{\phi}_k-\phi_k\|$. Next, since

$$\left|\iint(\widehat{K}-K)\{(\widehat{\phi}_k-\phi_k)\otimes\phi_\ell\}\right|\leq\|\widehat{K}-K\|\cdot\|\widehat{\phi}_k-\phi_k\|,$$

we have, on the event A_n,

$$|T_{j,k,3}|\lesssim j^{-\gamma}\|\widehat{K}-K\|\cdot\|\widehat{\phi}_k-\phi_k\|\sum_{\ell:\ell\neq k}\frac{\ell^{-\beta}}{|\kappa_k-\kappa_\ell|}.$$

In view of the assumption that $k^{-\alpha}\lesssim\kappa_k\lesssim k^{-\alpha}$, choose $k_0\geq 1$ and $C>1$ large enough so that $\kappa_k/\kappa_{[k/C]}\leq 1/2$ and $\kappa_{[Ck]+1}/\kappa_k\leq 1/2$ for all $k\geq k_0$, where $[a]$ denotes the largest integer not exceeding a. We may choose k_0 and C in such a way that they depend only on α and C_1. Now, partition the sum $\sum_{\ell:\ell\neq k}$ into $\sum_{\ell=1}^{[k/C]}$, $\sum_{\ell=[k/C]+1,\neq k}^{[Ck]}$ and $\sum_{\ell=[Ck]+1}^{\infty}$. Observe that

$$\sum_{\ell=1}^{[k/C]}\frac{\ell^{-\beta}}{(\kappa_\ell-\kappa_k)}\leq\sum_{\ell=1}^{[k/C]}\frac{\ell^{-\beta}}{\kappa_\ell(1-\kappa_k/\kappa_{[k/C]})}\lesssim\sum_{\ell=1}^{[k/C]}\ell^{-\beta+\alpha}\lesssim\begin{cases}1 & \text{if } \beta>\alpha+1,\\ \ln k & \text{if } \beta=\alpha+1,\\ k^{\alpha-\beta+1} & \text{if } \beta<\alpha+1,\end{cases}$$

and

$$\sum_{\ell=[Ck]+1}^{\infty}\frac{\ell^{-\beta}}{(\kappa_k-\kappa_\ell)}\leq\sum_{\ell=[Ck]+1}^{\infty}\frac{\ell^{-\beta}}{\kappa_k(1-\kappa_{[Ck]+1}/\kappa_k)}\lesssim k^{\alpha}\sum_{\ell=[Ck]+1}^{\infty}\ell^{-\beta}\lesssim k^{\alpha-\beta+1}.$$

For $[k/C]<\ell<k$, observe that

$$\kappa_\ell-\kappa_k\geq k^{-\alpha-1}\left\{C_1^{-1}+k^{\alpha+1}\sum_{p=\ell}^{k-2}(\kappa_p-\kappa_{p+1})\right\}\geq k^{-\alpha-1}C_1^{-1}\left\{1+\sum_{p=\ell}^{k-2}(k/p)^{\alpha+1}\right\}\geq k^{-\alpha-1}C_1^{-1}(k-\ell).$$

Likewise, for $k<\ell\leq[Ck]$, $\kappa_k-\kappa_\ell\gtrsim k^{-\alpha-1}|k-\ell|$. Hence

$$\sum_{\ell=[k/C]+1,\neq k}^{[Ck]}\frac{\ell^{-\beta}}{|\kappa_\ell-\kappa_k|}\lesssim k^{\alpha+1}\sum_{\ell=[k/C]+1,\neq k}^{[Ck]}\frac{\ell^{-\beta}}{|k-\ell|}\lesssim k^{\alpha-\beta+1}\ln k.$$

This yields

$$\sum_{\ell:\ell\neq k}\frac{\ell^{-\beta}}{|\kappa_k-\kappa_\ell|}\lesssim\begin{cases}1 & \text{if } \beta>\alpha+1\\ k^{\alpha-\beta+1}\ln k & \text{if } \beta\leq\alpha+1,\end{cases} \tag{A.6}$$

and so, on the event A_n,

$$|T_{j,k,3}|\lesssim j^{-\gamma}(1+k^{\alpha-\beta+1}\ln k)\cdot\|\widehat{K}-K\|\cdot\|\widehat{\phi}_k-\phi_k\|.$$

Turning to $T_{j,k,1}$, observe that for each $\ell\neq k$,

$$\iint(\widehat{K}-K)(\phi_k\otimes\phi_\ell)=\frac{1}{n}\sum_{i=1}^n\xi_{i,k}\xi_{i,\ell}-\overline{\xi}_k\overline{\xi}_\ell,$$

which implies that $\mathrm{E}(T_{j,k,1}^2)$ is bounded, viz.

$$\begin{aligned}\mathrm{E}(T_{j,k,1}^2)&=\mathrm{E}\left[\left\{\sum_{\ell:\ell\neq k}\frac{b_{j,\ell}}{\kappa_k-\kappa_\ell}\left(\frac{1}{n}\sum_{i=1}^n\xi_{i,k}\xi_{i,\ell}-\overline{\xi}_k\overline{\xi}_\ell\right)\right\}^2\right]\\&\lesssim n^{-1}\mathrm{E}\left\{\xi_k^2\left(\sum_{\ell:\ell\neq k}\frac{b_{j,\ell}}{\kappa_k-\kappa_\ell}\xi_\ell\right)^2\right\}+\mathrm{E}\left\{\overline{\xi}_k^2\left(\sum_{\ell:\ell\neq k}\frac{b_{j,\ell}}{\kappa_k-\kappa_\ell}\overline{\xi}_\ell\right)^2\right\}.\end{aligned} \tag{A.7}$$

The first term on the right-hand side is bounded via the Cauchy–Schwarz inequality as

$$n^{-1}\{\mathrm{E}(\xi_k^4)\}^{1/2}\left[\mathrm{E}\left\{\left(\sum_{\ell:\ell\neq k}\frac{b_{j,\ell}\xi_\ell}{\kappa_k-\kappa_\ell}\right)^4\right\}\right]^{1/2},$$

where $\{\mathrm{E}(\xi_k^4)\}^{1/2}\lesssim k^{-\alpha}$. Now, observe that

$$\begin{aligned}\mathrm{E}\left\{\left(\sum_{\ell:\ell\neq k}\frac{b_{j,k}\xi_\ell}{\kappa_k-\kappa_\ell}\right)^4\right\}&\leq\sum_{\ell_1:\ell_1\neq k}\cdots\sum_{\ell_4:\ell_4\neq k}\left|\frac{b_{j,\ell_1}}{\kappa_k-\kappa_{\ell_1}}\right|\cdots\left|\frac{b_{j,\ell_4}}{\kappa_k-\kappa_{\ell_4}}\right|\mathrm{E}(|\xi_{\ell_1}\cdots\xi_{\ell_4}|)\\&\lesssim j^{-4\gamma}\sum_{\ell_1:\ell_1\neq k}\cdots\sum_{\ell_4:\ell_4\neq k}\frac{\ell_1^{-\beta}}{|\kappa_k-\kappa_{\ell_1}|}\cdots\frac{\ell_4^{-\beta}}{|\kappa_k-\kappa_{\ell_4}|}\mathrm{E}(|\xi_{\ell_1}\cdots\xi_{\ell_4}|)\end{aligned}$$

and a repeated application of Hölder's inequality yields

$$\mathrm{E}(|\xi_{\ell_1}\cdots\xi_{\ell_4}|)\leq\{\mathrm{E}(\xi_{\ell_1}^4)\}^{1/4}\cdots\{\mathrm{E}(\xi_{\ell_4}^4)\}^{1/4}\lesssim\ell_1^{-\alpha/2}\cdots\ell_4^{-\alpha/2}.$$

Hence the first term on the right-hand side of (A.7) satisfies

$$n^{-1}\mathrm{E}\left\{\xi_k^2\left(\sum_{\ell:\ell\neq k}\frac{b_{j,\ell}}{\kappa_k-\kappa_\ell}\xi_\ell\right)^2\right\}\lesssim n^{-1}j^{-2\gamma}k^{-\alpha}\left(\sum_{\ell:\ell\neq k}\frac{\ell^{-\beta-\alpha/2}}{|\kappa_k-\kappa_\ell|}\right)^2\lesssim n^{-1}j^{-2\gamma}k^{-\alpha},$$

where the last inequality follows from a similar estimate to (A.6) together with the assumption that $\beta>\alpha/2+1$. Similarly, we may bound the second term on the right-hand side of (A.7) by

$$\{\mathrm{E}(\overline{\xi}_k^4)\}^{1/2}\left[\mathrm{E}\left\{\left(\sum_{\ell:\ell\neq k}\frac{b_{j,k}\overline{\xi}_\ell}{\kappa_k-\kappa_\ell}\right)^4\right\}\right]^{1/2}\lesssim n^{-2}j^{-2\gamma}k^{-\alpha}.$$

To see this, observe that $\mathrm{E}(\overline{\xi}_k^4)\lesssim n^{-3}\mathrm{E}(\xi_k^4)+n^{-2}\{\mathrm{E}(\xi_k^2)\}^2\lesssim n^{-2}k^{-2\alpha}$, and likewise,

$$\mathrm{E}\left\{\left(\sum_{\ell:\ell\neq k}\frac{b_{j,k}\overline{\xi}_\ell}{\kappa_k-\kappa_\ell}\right)^4\right\}\lesssim n^{-2}j^{-4\gamma}.$$

Hence, we conclude that $\mathrm{E}(T_{j,k,1}^2)\lesssim n^{-1}j^{-2\gamma}k^{-\alpha}$.

Finally, we shall bound $|T_{j,k,2}|$. To this end, observe that, on the event A_n,

$$|T_{j,k,2}| \lesssim j^{-\gamma}\|\widehat{K}-K\| \sum_{\ell:\ell\neq k} \frac{\ell^{-\beta}\widehat{v}_{k,\ell}}{|\kappa_k-\kappa_\ell|^2},$$

where

$$\widehat{v}_{k,\ell} = \left| \frac{1}{n}\sum_{i=1}^n \xi_{i,k}\xi_{i,\ell} - \overline{\xi}_k\overline{\xi}_\ell \right|.$$

Then we have

$$\mathrm{E}\left\{ \left(\sum_{\ell:\ell\neq k} \frac{\ell^{-\beta}}{|\kappa_k-\kappa_\ell|^2}\widehat{v}_{k,\ell} \right)^2 \right\} \leq \left[\sum_{\ell:\ell\neq k} \frac{\ell^{-\beta}}{|\kappa_k-\kappa_\ell|^2} \{\mathrm{E}(\widehat{v}_{k,\ell}^2)\}^{1/2} \right]^2 \lesssim n^{-1}k^{-\alpha} \left(\sum_{\ell:\ell\neq k} \frac{\ell^{-\beta-\alpha/2}}{|\kappa_k-\kappa_\ell|^2} \right)^2,$$

and the far right-hand side is $\lesssim n^{-1}(k^{-\alpha}+k^{2\alpha-2\beta+4})$, because

$$\begin{aligned}
\sum_{\ell:\ell\neq k} \frac{\ell^{-\beta-\alpha/2}}{|\kappa_k-\kappa_\ell|^2} &= \left(\sum_{\ell=1}^{[k/C]} + \sum_{\ell=[k/C]+1,\neq k}^{[Ck]} + \sum_{\ell=[Ck]+1}^{\infty} \right) \frac{\ell^{-\beta-\alpha/2}}{|\kappa_k-\kappa_\ell|^2} \\
&\lesssim \sum_{\ell=1}^{[k/C]} \ell^{3\alpha/2-\beta} + k^{2\alpha+2} \sum_{\ell=[k/C]+1,\neq k}^{[Ck]} \frac{\ell^{-\beta-\alpha/2}}{|k-\ell|^2} + k^{2\alpha} \sum_{\ell=[Ck]+1}^{\infty} \ell^{-\beta-\alpha/2} \\
&\lesssim 1 + k^{3\alpha/2-\beta+1}\ln k + k^{3\alpha/2-\beta+2} + k^{3\alpha/2-\beta+1} \lesssim 1 + k^{3\alpha/2-\beta+2}.
\end{aligned}$$

Summarizing, using (A.2) and (A.3), we have

$$\sum_{j=1}^{\infty}\sum_{k=1}^{m_n}(T_{j,k,1}^2+\cdots+T_{j,k,4}^2) = O_{\mathrm{P}}[n^{-1}+n^{-2}\{m_n^3+m_n^{2\alpha-2\beta+5}(\ln m_n)^2\}].$$

Given that $m_n = o\{n^{1/(2\alpha+2)}\}$, $m_n^3 = o(n)$, so that the last expression is $O_{\mathrm{P}}\{n^{-1}+n^{-2}m_n^{2\alpha-2\beta+5}(\ln m_n)^2\}$. Furthermore, $m_n^{2\alpha-2\beta+5}(\ln m_n)^2 = o(m_n^{\alpha+3}) = o(n)$ because $\beta > \alpha/2+1$. Hence we conclude that

$$\sum_{j=1}^{\infty}\sum_{k=1}^{m_n}(T_{j,k,1}^2+\cdots+T_{j,k,4}^2) = O_{\mathrm{P}}(n^{-1}).$$

Step 2. We wish to bound $\sum_{j=1}^{\infty}\sum_{k=1}^{m_n}\widehat{\kappa}_k^{-2}(n^{-1}\sum_{i=1}^n\varepsilon_{i,j}\widehat{\xi}_{i,k})^2$. Observe that for $k\in\{1,\ldots,m_n\}$,

$$|\widehat{\kappa}_k/\kappa_k - 1| \lesssim k^{\alpha}|\widehat{\kappa}_k-\kappa_k| \leq m_n^{\alpha}\|\widehat{K}-K\| = o_{\mathrm{P}}(1),$$

from which we have $\max_{1\leq k\leq m_n}|\kappa_k/\widehat{\kappa}_k - 1| = o_{\mathrm{P}}(1)$ and hence $\sum_{k=1}^{m_n}\widehat{\kappa}_k^{-1} \leq \{1+o_{\mathrm{P}}(1)\}\sum_{k=1}^{m_n}\kappa_k^{-1}$. Given that conditionally on $X_1^n = \{X_1,\ldots,X_n\}$, $\varepsilon_{1,j},\ldots,\varepsilon_{n,j}$ are mutually independent with mean zero, we have

$$\mathrm{E}\left\{ \left(n^{-1}\sum_{i=1}^n \varepsilon_{i,j}\widehat{\xi}_{i,k} \right)^2 \mid X_1^n \right\} = n^{-2}\sum_{i=1}^n \mathrm{E}(\varepsilon_{i,j}^2 \mid X_1^n)\widehat{\xi}_{i,k}^2.$$

Further, because by the Monotone Convergence Theorem for conditional expectation and Bessel's inequality,

$$\sum_{j=1}^{\infty}\mathrm{E}(\varepsilon_{i,j}^2 \mid X_1^n) = \mathrm{E}\left(\sum_{j=1}^{\infty}\varepsilon_{i,j}^2 \mid X_1^n \right) \leq \mathrm{E}(\|\mathcal{E}_i\|^2 \mid X_1^n) = \mathrm{E}(\|\mathcal{E}_i\|^2 \mid X_i) \leq C_1,$$

we have

$$\mathrm{E}\left\{ \sum_{j=1}^{\infty}\sum_{k=1}^{m_n}\widehat{\kappa}_k^{-2}\left(\frac{1}{n}\sum_{i=1}^n \varepsilon_{i,j}\widehat{\xi}_{i,k} \right)^2 \,\middle|\, X_1^n \right\} \lesssim n^{-1}\sum_{k=1}^{m_n}\widehat{\kappa}_k^{-1} \leq n^{-1}\{1+o_{\mathrm{P}}(1)\}\sum_{k=1}^{m_n}\kappa_k^{-1} = O_{\mathrm{P}}(n^{-1}m_n^{\alpha+1}).$$

This yields

$$\sum_{j=1}^{\infty}\sum_{k=1}^{m_n}\widehat{\kappa}_k^{-2}\left(\frac{1}{n}\sum_{i=1}^n \varepsilon_{i,j}\widehat{\xi}_{i,k} \right)^2 = O_{\mathrm{P}}(n^{-1}m_n^{\alpha+1}).$$

Summarizing, we conclude that

$$\sum_{j=1}^{\infty}\sum_{k=1}^{m_n}(\widehat{b}_{j,k}-b_{j,k})^2 = O_{\mathrm{P}}(n^{-1}m_n^{\alpha+1}).$$

Step 3. Conclusion. Recall that $\widehat{b}=\sum_{j=1}^{\infty}\sum_{k=1}^{m_n}\widehat{b}_{j,k}(\phi_j\otimes\widehat{\phi}_k)$, and observe that

$$\widehat{b}-b=\sum_{j=1}^{\infty}\sum_{k=1}^{m_{n,2}}(\widehat{b}_{j,k}-b_{j,k})(\phi_j\otimes\widehat{\phi}_k)+\sum_{j=1}^{\infty}\sum_{k=1}^{m_n}b_{j,k}\{\phi_j\otimes(\widehat{\phi}_k-\phi_k)\}+B_n,$$

where $B_n=b-\sum_{j=1}^{\infty}\sum_{k=1}^{m_n}b_{j,k}(\phi_j\otimes\phi_k)$. Given that $\|B_n\|^2=\sum_{j=1}^{\infty}\sum_{k>m_n}b_{j,k}^2=O(m_n^{-2\beta+1})$, we have

$$\|\widehat{b}-b\|^2=O_{\mathrm{P}}(n^{-1}m_n^{\alpha+1}+m_n^{-2\beta+1})+\iint\left[\sum_{j=1}^{\infty}\sum_{k=1}^{m_n}b_{j,k}\{\phi_j\otimes(\widehat{\phi}_k-\phi_k)\}\right]^2.$$

Now, observe that, using Parseval's identity, the second term on the right-hand side is

$$\sum_{j=1}^{\infty}\int\left\{\sum_{k=1}^{m_n}b_{j,k}(\widehat{\phi}_k-\phi_k)\right\}^2\le m_n\sum_{j=1}^{\infty}\sum_{k=1}^{m_n}b_{j,k}^2\|\widehat{\phi}_k-\phi_k\|^2\lesssim m_n\sum_{k=1}^{m_n}k^{-2\beta}\|\widehat{\phi}_k-\phi_k\|^2,$$

which is $O_{\mathrm{P}}(n^{-1}m_n)$. This completes the proof for the first assertion. The second assertion follows directly from the first assertion. □

A.1. Proof of Theorem 2

The proof is parallel to that of Theorem 1. We freely use the results in the proof of Theorem 1. Since $\widetilde{b}$ is invariant with respect to choices of signs of $\widehat{\phi}_k$'s, one can assume (A.1) without loss of generality. Let $\overline{m}_n=\max(m_{n,1},m_{n,2})$, and define the event

$$A_n'=\{|\widehat{\kappa}_k-\kappa_\ell|\ge|\kappa_k-\kappa_\ell|/\sqrt{2}\text{ for all }1\le k\le\overline{m}_n\text{ and for all }\ell\ne k\},$$

for which we have $\mathrm{P}(A_n')\to 0$ since $\overline{m}_n=o\{n^{1/(2\alpha+2)}\}$.

Expand $\mathcal{E}_i=Y_i-\mathrm{E}(Y_i\mid X_i)$ as $\mathcal{E}_i=\sum_j\widehat{\varepsilon}_{i,j}\widehat{\phi}_j$ with $\widehat{\varepsilon}_{i,j}=\int\mathcal{E}_i\widehat{\phi}_j$. Let $\widehat{\eta}_{i,j}^c=\widehat{\eta}_{i,j}-n^{-1}\sum_{i'=1}^n\widehat{\eta}_{i',j}$ and $\widehat{\varepsilon}_{i,\ell}^c=\widehat{\varepsilon}_{i,\ell}-n^{-1}\sum_{i'=1}^n\widehat{\varepsilon}_{i',\ell}$. Next, observe that

$$\widehat{\eta}_{i,j}^c=\sum_{k=1}^{\infty}\check{b}_{j,k}^*\widehat{\xi}_{i,k}+\widehat{\varepsilon}_{i,j}^c,$$

where $\check{b}_{j,k}^*=\iint b(\widehat{\phi}_j\otimes\widehat{\phi}_k)$. Hence, using the relation in (A.4), we have $\widehat{\kappa}_k\widetilde{b}_{j,k}=\widehat{\kappa}_k\check{b}_{j,k}^*+n^{-1}\sum_{i=1}^n\widehat{\varepsilon}_{i,j}\widehat{\xi}_{i,k}$, which yields

$$(\widetilde{b}_{j,k}-b_{j,k})^2\lesssim(\check{b}_{j,k}^*-b_{j,k})^2+\widehat{\kappa}_k^{-2}\left(n^{-1}\sum_{i=1}^n\widehat{\varepsilon}_{i,j}\widehat{\xi}_{i,k}\right)^2.$$

Further observe that

$$\begin{aligned}\check{b}_{j,k}^*-b_{j,k}&=\iint b(\widehat{\phi}_j\otimes\widehat{\phi}_k-\phi_j\otimes\phi_k)\\&=\iint b\{(\widehat{\phi}_j-\phi_j)\otimes\phi_k\}+\iint b\{\phi_j\otimes(\widehat{\phi}_k-\phi_k)\}+\iint b\{(\widehat{\phi}_j-\phi_j)\otimes(\widehat{\phi}_k-\phi_k)\}\\&\equiv I_{j,k}+II_{j,k}+III_{j,k}.\end{aligned}$$

Step 1 in the proof of Theorem 1 shows that $\sum_{j=1}^{m_{n,1}}\sum_{k=1}^{m_{n,2}}II_{j,k}^2=O_{\mathrm{P}}(n^{-1})$, and likewise we have $\sum_{j=1}^{m_{n,1}}\sum_{k=1}^{m_{n,2}}I_{j,k}^2=O_{\mathrm{P}}(n^{-1})$. Furthermore, using (A.5), observe that $\iint b\{(\widehat{\phi}_j-\phi_j)\otimes(\widehat{\phi}_k-\phi_k)\}=\sum_{p,q}b_{p,q}\widehat{w}_{j,p}\widehat{w}_{k,q}$, where

$$\widehat{w}_{j,p}=\begin{cases}(\widehat{\kappa}_j-\kappa_p)^{-1}\iint(\widehat{K}-K)(\widehat{\phi}_j\otimes\phi_p) & \text{if } p\ne j,\\ \int(\widehat{\phi}_j-\phi_j)\phi_j & \text{if } p=j.\end{cases}$$

For each $p \neq j$, on the event A_n', $|\widehat{w}_{j,p}| \lesssim |\kappa_j - \kappa_p|^{-1}\|\widehat{K} - K\|$, which yields that on the event A_n',

$$\left|\sum_{p,q} b_{p,q}\widehat{w}_{j,p}\widehat{w}_{k,q}\right| \lesssim \left(\|\widehat{K}-K\|\sum_{p:p\neq j}\frac{p^{-\gamma}}{|\kappa_j-\kappa_p|} + j^{-\gamma}\|\widehat{\phi}_j - \phi_j\|\right) \times \left(\|\widehat{K}-K\|\sum_{q:q\neq k}\frac{q^{-\beta}}{|\kappa_k-\kappa_q|} + k^{-\beta}\|\widehat{\phi}_k-\phi_k\|\right)$$
$$\lesssim \{\|\widehat{K}-K\|(1+j^{\alpha-\gamma+1}\ln j) + j^{-\gamma}\|\widehat{\phi}_j-\phi_j\|\} \times \{\|\widehat{K}-K\|(1+k^{\alpha-\beta+1}\ln k) + k^{-\beta}\|\widehat{\phi}_k-\phi_k\|\}.$$

Therefore, we have

$$\sum_{j=1}^{m_{n,1}}\sum_{k=1}^{m_{n,2}}\left(\sum_{p,q} b_{p,q}\widehat{w}_{j,p}\widehat{w}_{k,q}\right)^2 = O_{\mathrm{P}}\left[n^{-2}\{m_{n,1} + m_{n,1}^{2\alpha-2\gamma+3}(\ln m_{n,1})^2\}\{m_{n,2} + m_{n,2}^{2\alpha-2\beta+3}(\ln m_{n,2})^2\}\right].$$

Since $\beta > \alpha/2+1$ and $\gamma > \alpha/2+1$, the last expression is $o_{\mathrm{P}}(n^{-2}m_{n,1}^{\alpha+1}m_{n,2}^{\alpha+1}) = o_{\mathrm{P}}(n^{-1})$. So we conclude that

$$\sum_{j=1}^{m_{n,1}}\sum_{k=1}^{m_{n,2}}(\check{b}_{j,k}^{*} - b_{j,k})^2 = O_{\mathrm{P}}(n^{-1}).$$

Next, since conditionally on $X_1^n = \{X_1, \ldots, X_n\}$, $\widehat{\varepsilon}_{1,j}, \ldots, \widehat{\varepsilon}_{n,j}$ are independent with mean zero, we have

$$\mathrm{E}\left\{\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{\varepsilon}_{i,j}\widehat{\xi}_{i,k}\right)^2 \,\middle|\, X_1^n\right\} = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{E}(\widehat{\varepsilon}_{i,j}^2 \mid X_1^n)\widehat{\xi}_{i,k}^2.$$

Further, since by Bessel's inequality,

$$\sum_{j=1}^{m_{n,1}}\mathrm{E}(\widehat{\varepsilon}_{i,j}^2 \mid X_1^n) \leq \mathrm{E}\left(\sum_{j=1}^{m_{n,1}}\widehat{\varepsilon}_{i,j}^2 \mid X_1^n\right) \leq \mathrm{E}(\|\mathcal{E}_i\|^2 \mid X_1^n) = \mathrm{E}(\|\mathcal{E}_i\|^2 \mid X_i) \leq C_1,$$

we have, using the fact that $\max_{1\leq k\leq m_{n,2}}|\kappa_k/\widehat{\kappa}_k - 1| = o_{\mathrm{P}}(1)$,

$$\mathrm{E}\left\{\sum_{j=1}^{m_{n,1}}\sum_{k=1}^{m_{n,2}}\frac{1}{\widehat{\kappa}_k^2}\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{\varepsilon}_{i,j}\widehat{\xi}_{i,k}\right)^2 \,\middle|\, X_1^n\right\} \lesssim n^{-1}\sum_{k=1}^{m_{n,2}}\widehat{\kappa}_k^{-1} \leq n^{-1}\{1+o_{\mathrm{P}}(1)\}\sum_{k=1}^{m_{n,2}}\kappa_k^{-1} = O_{\mathrm{P}}(n^{-1}m_{n,2}^{\alpha+1}).$$

This yields

$$\sum_{j=1}^{m_{n,1}}\sum_{k=1}^{m_{n,2}}\widehat{\kappa}_k^{-2}\left(n^{-1}\sum_{i=1}^{n}\widehat{\varepsilon}_{i,j}\widehat{\xi}_{i,k}\right)^2 = O_{\mathrm{P}}(n^{-1}m_{n,2}^{\alpha+1}).$$

Summarizing, we conclude that $\sum_{j=1}^{m_{n,1}}\sum_{k=1}^{m_{n,2}}(\widetilde{b}_{j,k} - b_{j,k})^2 = O_{\mathrm{P}}(n^{-1}m_{n,2}^{\alpha+1})$.

Recall that $\widetilde{b} = \sum_{j=1}^{m_{n,1}}\sum_{k=1}^{m_{n,2}}\widetilde{b}_{j,k}(\widehat{\phi}_j \otimes \widehat{\phi}_k)$, and observe that

$$\widetilde{b} - b = \sum_{j=1}^{m_{n,1}}\sum_{k=1}^{m_{n,2}}(\widetilde{b}_{j,k} - b_{j,k})(\widehat{\phi}_j \otimes \widehat{\phi}_k) + \sum_{j=1}^{m_{n,1}}\sum_{k=1}^{m_{n,2}} b_{j,k}(\widehat{\phi}_j \otimes \widehat{\phi}_k - \phi_j \otimes \phi_k) + B_n',$$

where $B_n' = b - \sum_{j=1}^{m_{n,1}}\sum_{k=1}^{m_{n,2}} b_{j,k}(\phi_j \otimes \phi_k)$. So

$$\|\widetilde{b} - b\|^2 \lesssim \sum_{j=1}^{m_{n,1}}\sum_{k=1}^{m_{n,2}}(\widetilde{b}_{j,k} - b_{j,k})^2 + \iint\left\{\sum_{j=1}^{m_{n,1}}\sum_{k=1}^{m_{n,2}} b_{j,k}(\widehat{\phi}_j \otimes \widehat{\phi}_k - \phi_j \otimes \phi_k)\right\}^2$$
$$+ \left(\sum_{j>m_{n,1}}\sum_{k=1}^{m_{n,2}} + \sum_{j=1}^{m_{n,1}}\sum_{k>m_{n,2}} + \sum_{j>m_{n,1}}\sum_{k>m_{n,2}}\right) b_{j,k}^2$$
$$= O_{\mathrm{P}}\left(n^{-1}m_{n,2}^{\alpha+1} + m_{n,1}^{-2\gamma+1} + m_{n,2}^{-2\beta+1}\right) + \iint\left\{\sum_{j=1}^{m_{n,1}}\sum_{k=1}^{m_{n,2}} b_{j,k}(\widehat{\phi}_j \otimes \widehat{\phi}_k - \phi_j \otimes \phi_k)\right\}^2.$$

Using the decomposition

$$\widehat{\phi}_j \otimes \widehat{\phi}_k - \phi_j \otimes \phi_k = (\widehat{\phi}_j - \phi_j) \otimes \phi_k + \phi_j \otimes (\widehat{\phi}_k - \phi_k) + (\widehat{\phi}_j - \phi_j) \otimes (\widehat{\phi}_k - \phi_k),$$

we have

$$\sum_{j=1}^{m_{n,1}}\sum_{k=1}^{m_{n,2}} b_{j,k}(\widehat{\phi}_j\otimes\widehat{\phi}_k-\phi_j\otimes\phi_k)=\sum_{j=1}^{m_{n,1}}(\widehat{\phi}_j-\phi_j)\otimes\left(\sum_{k=1}^{m_{n,2}} b_{j,k}\phi_k\right)$$
$$+\sum_{k=1}^{m_{n,2}}\left(\sum_{j=1}^{m_{n,1}} b_{j,k}\phi_j\right)\otimes(\widehat{\phi}_k-\phi_k)+\sum_{j=1}^{m_{n,1}}\sum_{k=1}^{m_{n,2}} b_{j,k}\{(\widehat{\phi}_j-\phi_j)\otimes(\widehat{\phi}_k-\phi_k)\}.$$

Observe that

$$\iint\left\{\sum_{j=1}^{m_{n,1}}(\widehat{\phi}_j-\phi_j)\otimes\left(\sum_{k=1}^{m_{n,2}} b_{j,k}\phi_k\right)\right\}^2\le m_{n,1}\sum_{j=1}^{m_{n,1}}\|\widehat{\phi}_j-\phi_j\|^2\sum_{k=1}^{m_{n,2}} b_{j,k}^2\lesssim m_{n,1}\sum_{j=1}^{m_{n,1}} j^{-2\gamma}\|\widehat{\phi}_j-\phi_j\|^2=O_{\mathrm{P}}(n^{-1}m_{n,1}).$$

Likewise, we have

$$\iint\left\{\sum_{k=1}^{m_{n,2}}\left(\sum_{j=1}^{m_{n,1}} b_{j,k}\phi_j\right)\otimes(\widehat{\phi}_k-\phi_k)\right\}^2=O_{\mathrm{P}}(n^{-1}m_{n,2}).$$

Finally, we have

$$\iint\left[\sum_{j=1}^{m_{n,1}}\sum_{k=1}^{m_{n,2}} b_{j,k}\{(\widehat{\phi}_j-\phi_j)\otimes(\widehat{\phi}_k-\phi_k)\}\right]\le m_{n,1}m_{n,2}\sum_{j=1}^{m_{n,1}}\sum_{k=1}^{m_{n,2}} b_{j,k}^2\|\widehat{\phi}_j-\phi_j\|^2\|\widehat{\phi}_k-\phi_k\|^2=O_{\mathrm{P}}(n^{-2}m_{n,1}m_{n,2}).$$

Therefore, we conclude that

$$\|\widetilde{b}-b\|^2=O_{\mathrm{P}}\{n^{-1}(m_{n,1}+m_{n,2}^{\alpha+1})+m_{n,1}^{-2\gamma+1}+m_{n,2}^{-2\beta+1}\}.$$

The second assertion follows directly from the first assertion. This completes the proof. □

A.2. Proof of Theorem 3

The proof is inspired by that of (3.6) in [19]; the current proof relies on Assouad's lemma [36, Lemma 2.12] and Theorem 2.12(iv) in [36]. To apply those results, we must construct a sequence of conditional distributions of Y given X, to which end we employ the theory of Gaussian measures on Banach spaces; see, e.g., [35], Chapter VIII. The proof proceeds as follows. We first construct a conditional density of Y given X with respect to a suitable dominating measure. Next, we construct a family of joint distributions of (Y,X) indexed by 0–1 sequences $\theta=(\theta_1,\ldots,\theta_{\nu_n})\in\{0,1\}^{\nu_n}$ with $\nu_n=[n^{1/(\alpha+2\beta)}]$, and reduce the problem of estimating b under the $L^2(I^2)$ norm into that of estimating θ under the Hamming loss. The final step is to apply Assouad's lemma [36, Lemma 2.12] and Theorem 2.12(iv) in [36] to derive minimax lower bounds. We divide the rest of the proof into two steps.

Step 1 (Preliminary). We first construct a conditional density of Y given X with respect to a suitable dominating measure. Recall that $\lambda_1\ge\lambda_2\ge\cdots>0$ are eigenvalues of the covariance function R of the error term $\mathcal{E}$. For any $b\in L^2(I^2)$ and $x\in L^2(I)$, let $P_{b,x}$ denote the distribution of $\int_I b(\cdot,t)x(t)dt+\mathcal{E}(\cdot)$, and let P_0 denote the distribution of $\mathcal{E}$. Those distributions are defined on the Borel σ-field of $L^2(I)$. Associated to $\mathcal{E}$, the Cameron–Martin space – or the reproducing kernel Hilbert space (RKHS) associated with R – is given by

$$H=\left\{h=\sum_j h_j\phi_j:\sum_j\frac{h_j^2}{\lambda_j}<\infty\right\}$$

equipped with the inner product

$$\langle h,g\rangle_H=\sum_j\frac{h_jg_j}{\lambda_j},\quad h=\sum_j h_j\phi_j,\quad g=\sum_j g_j\phi_j\in H.$$

Let $b=\sum_{j,k}b_{j,k}\phi_j\otimes\phi_k$ and $x=\sum_k x_k\phi_k$; then $P_{b,x}$ is absolutely continuous with respect to P_0 if and only if

$$\left\|\int_I b(\cdot,t)x(t)dt\right\|_H^2=\sum_j\frac{1}{\lambda_j}\left(\sum_k b_{j,k}x_k\right)^2<\infty,$$

and its Radon–Nikodym derivative is given by the Cameron–Martin formula

$$p_{b,x}(y)=\frac{dP_{b,x}}{dP_0}(y)=\exp\left\{-\sum_j\frac{(\sum_k b_{j,k}x_k)^2}{2\lambda_j}+\sum_j\frac{y_j\sum_k b_{j,k}x_k}{\lambda_j}\right\},$$

where $y = \sum_j y_j \phi_j$. See Theorem 8.2.9 in [35], or Section 2.6 in [18] on RKHS. Denote by Q the distribution of X; then the joint distribution of (X, Y) is given by $p_{b,x}(y)dP_0(y)dQ(x)$.

Step 2 (Derivation of minimax lower bounds). Now, let $\nu_n = [n^{1/(\alpha+2\beta)}]$, and

$$b^\theta = \sum_{k=\nu_n+1}^{2\nu_n} k^{-\beta}\theta_{k-\nu_n}(\phi_1 \otimes \phi_k),$$

where $\theta = (\theta_1, \ldots, \theta_{\nu_n}) \in \{0, 1\}^{\nu_n}$. Then $b^\theta \in \mathcal{B}(\beta, \gamma, C_1)$ and $b^\theta_{j,k} = 0$ for every integer $j \geq 2$, so that

$$p_{b^\theta,x}(y) = \exp\left\{ -\frac{(\sum_{k=\nu_n+1}^{2\nu_n} k^{-\beta}\theta_{k-\nu_n}x_k)^2}{2\lambda_1} + \frac{y_1 \sum_{k=\nu_n+1}^{2\nu_n} k^{-\beta}\theta_{k-\nu_n}x_k}{\lambda_1} \right\}.$$

Define $\widetilde{p}_{\theta,x}(y) = p_{b^\theta,x}(y)$ and $d\widetilde{P}_\theta(x, y) = \widetilde{p}_{\theta,x}(y)dP_0(y)dQ(x)$ for each $\theta = (\theta_1, \ldots, \theta_{\nu_n}) \in \{0, 1\}^{\nu_n}$. Next, let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be i.i.d. from $\widetilde{P}_\theta$.

For any estimator $\overline{b}^n = \sum_{j,k}\overline{b}^n_{j,k}(\phi_j \otimes \phi_k)$ of b^θ, we have by Bessel's inequality,

$$\|\overline{b}^n - b^\theta\|^2 \geq \sum_{k=\nu_n+1}^{2\nu_n} (\overline{b}^n_{1,k} - k^{-\beta}\theta_{k-\nu_n})^2 \geq \frac{1}{4}\sum_{k=\nu_n+1}^{2\nu_n} k^{-2\beta}(\overline{\theta}^n_{k-\nu_n} - \theta_{k-\nu_n})^2 \geq \frac{(2\nu_n)^{-2\beta}}{4}\sum_{k=1}^{\nu_n}|\overline{\theta}^n_k - \theta_k|,$$

where

$$\overline{\theta}^n_{k-\nu_n} = \arg\min_{\vartheta\in\{0,1\}} (k^\beta \overline{b}^n_{1,k} - \vartheta)^2 = \begin{cases} 0 & \text{if } k^\beta\overline{b}^n_{1,k} \leq 1/2, \\ 1 & \text{if } k^\beta\overline{b}^n_{1,k} > 1/2. \end{cases}$$

Indeed, since $(k^\beta\overline{b}^n_{1,k} - \overline{\theta}^n_{k-\nu_n})^2 \leq (k^\beta\overline{b}^n_{1,k} - \theta_{k-\nu_n})^2$ by the definition of $\overline{\theta}^n_{k-\nu_n}$,

$$(\overline{\theta}^n_{k-\nu_n} - \theta_{k-\nu_n})^2 \leq 2(k^\beta\overline{b}^n_{1,k} - \overline{\theta}^n_{k-\nu_n})^2 + 2(k^\beta\overline{b}^n_{1,k} - \theta_{k-\nu_n})^2 \leq 4(k^\beta\overline{b}^n_{1,k} - \theta_{k-\nu_n})^2.$$

For any $\theta, \theta' \in \{0, 1\}^{\nu_n}$, let $\rho(\theta, \theta') = \sum_{k=1}^{\nu_n}|\theta_k - \theta'_k|$ denote the Hamming distance. Then we have

$$\mathrm{P}_\theta\left\{\|\overline{b}^n - b^\theta\|^2 \geq \frac{(2\nu_n)^{-2\beta}}{4}c\right\} \geq \mathrm{P}_\theta\{\rho(\overline{\theta}^n, \theta) \geq c\}$$

for any $\theta \in \{0, 1\}^{\nu_n}$ and any constant $c > 0$, where P_θ denotes the probability under θ. To bound the right-hand side from below, we compute the Kullback–Leibler divergence

$$K(\widetilde{P}_\theta, \widetilde{P}_{\theta'}) = \int \ln\frac{d\widetilde{P}_\theta}{d\widetilde{P}_{\theta'}}d\widetilde{P}_\theta$$

for any $\theta, \theta' \in \{0, 1\}^{\nu_n}$ with $\rho(\theta, \theta') = 1$. Suppose that $\theta_k \neq \theta'_k$ for some $1 \leq k \leq \nu_n$ and $\theta_\ell = \theta'_\ell$ for all $\ell \neq k$. Then a straightforward calculation shows that

$$K(\widetilde{P}_\theta, \widetilde{P}_{\theta'}) = \mathrm{E}_\theta\left\{\ln\frac{\widetilde{p}_{\theta,X}(Y)}{\widetilde{p}_{\theta',X}(Y)}\right\} = \frac{(\nu_n + k)^{-\alpha-2\beta}}{2\lambda_1} \leq \frac{(\nu_n+1)^{-\alpha-2\beta}}{2\lambda_1} \leq \frac{1}{2\lambda_1 n},$$

which yields that

$$K(\widetilde{P}^{\otimes n}_\theta, \widetilde{P}^{\otimes n}_{\theta'}) = nK(\widetilde{P}_\theta, \widetilde{P}_{\theta'}) \leq \frac{1}{2\lambda_1}.$$

Now, applying Assouad's lemma and Theorem 2.12(iv) in [36], we have

$$\max_{\theta\in\{0,1\}^{\nu_n}} \mathrm{E}_\theta\{\rho(\overline{\theta}^n, \theta)\} \geq \frac{\nu_n}{4}e^{-1/(2\lambda_1)},$$

where E_θ denotes the expectation under θ. Choose $\theta \in \{0, 1\}^{\nu_n}$ at which the maximum on the left-hand side is attained, and observe that $\rho(\overline{\theta}^n, \theta) \leq \nu_n$. The Paley–Zygmund inequality then yields

$$\mathrm{P}_\theta\left\{\rho(\overline{\theta}^n, \theta) \geq \frac{\nu_n}{8}e^{-1/(2\lambda_1)}\right\} \geq \mathrm{P}_\theta\left[\rho(\overline{\theta}^n, \theta) \geq \frac{1}{2}\mathrm{E}_\theta\{\rho(\overline{\theta}^n, \theta)\}\right] \geq \frac{1}{4}\frac{[\mathrm{E}_\theta\{\rho(\overline{\theta}^n, \theta)\}]^2}{\mathrm{E}\{\rho(\overline{\theta}^n, \theta)^2\}} \geq \frac{1}{16}e^{-1/(2\lambda_1)}.$$

Therefore,

$$\max_{\theta\in\{0,1\}^{\nu_n}} \mathrm{P}_\theta\left\{\|\overline{b}^n - b^\theta\|^2 \geq \frac{\nu_n^{-2\beta+1}}{2^{2\beta+5}}e^{-1/(2\lambda_1)}\right\} \geq \frac{1}{16}e^{-1/(2\lambda_1)}.$$

Since $\nu_n^{-2\beta+1} \sim n^{-(2\beta-1)/(\alpha+2\beta)}$, the proof is complete. □

References

[1] D. Benatia, M. Carrasco, J.-P. Florens, Functional linear regression with functional response, J. Econometrics (2017) (in press).
[2] D. Bosq, Linear Processes in Function Spaces: Theory and Applications, Springer, New York, 2000.
[3] Bureau of Labor Statistics, US Department of Labor, National Longitudinal Survey of Youth 1979 Cohort, 1979–2012 (rounds 1–25), produced and distributed by the Center for Human Resource Research, Ohio State University, Columbus, OH, 2012.
[4] T. Byczkowski, Gaussian measures on L^p spaces $0 \leq p < \infty$, Studia Math. 59 (1977) 249–261.
[5] T.T. Cai, P. Hall, Prediction in functional linear regression, Ann. Statist. 34 (2006) 2159–2179.
[6] T.T. Cai, M. Yuan, Minimax and adaptive prediction for functional linear regression, J. Amer. Statist. Assoc. 107 (2012) 1201–1216.
[7] D. Card, Estimating the return to schooling: Progress on some persistent econometric problems, Econometrica 69 (2001) 1127–1160.
[8] H. Cardot, F. Ferraty, P. Sarda, Functional linear model, Statist. Probab. Lett. 45 (1999) 11–22.
[9] H. Cardot, F. Ferraty, P. Sarda, Spline estimators for the functional linear models, Statist. Sinica 13 (2003) 571–591.
[10] H. Cardot, J. Johannes, Thresholding projection estimators in functional linear models, J. Multivariate Anal. 101 (2010) 395–408.
[11] J.M. Chiou, H.G. Müller, J.L. Wang, Functional response models, Statist. Sinica 14 (2004) 675–693.
[12] F. Comte, J. Johannes, Adaptive functional linear regression, Ann. Statist. 40 (2012) 2765–2797.
[13] C. Crambes, A. Kneip, P. Sarda, Smoothing splines estimators for functional linear regression, Ann. Statist. 37 (2009) 35–72.
[14] C. Crambes, A. Mas, Asymptotics of prediction in functional linear regression with functional outputs, Bernoulli 19 (2013) 2627–2651.
[15] A. Cuevas, M. Febrero, R. Fraiman, Linear functional regression: The case of fixed design and functional response, Canad. J. Statist. 30 (2002) 285–300.
[16] A. Delaigle, P. Hall, Methodology and theory for partial least squares applied to functional data, Ann. Statist. 40 (2012) 322–352.
[17] J.A. Doornik, Object-Oriented Matrix Programming using Ox, third ed., Timberlake Consultants Press, 2002.
[18] E. Giné, R. Nickl, Mathematical Foundations of Infinite-Dimensional Statistical Models, Cambridge University Press, 2016.
[19] P. Hall, J.L. Horowitz, Methodology and convergence rates for functional linear regression, Ann. Statist. 35 (2007) 70–91.
[20] G. He, H.-G. Müller, J.-L. Wang, W. Yang, Functional linear regression via canonical analysis, Bernoulli 16 (2010) 705–729.
[21] S. Hörmann, Ł. Kidzínski, A note on estimation in Hilbertian linear models, Scand. J. Statist. 42 (2015) 43–62.
[22] T. Hsing, R. Eubank, Theoretical Foundations of Functional Data Analysis with an Introduction To Linear Operators, Wiley, Chichester, 2015.
[23] G.M. James, J. Wang, J. Zhu, Functional linear regression that's interpretable, Ann. Statist. 37 (2009) 2083–2108.
[24] R. Kress, Linear Integral Equations, second ed., Springer, New York, 1999.
[25] Y. Li, T. Hsing, On rates of convergence in functional linear regression, J. Multivariate Anal. 98 (2007) 1782–1804.
[26] H. Lian, Minimax prediction for functional linear regression with functional responses in reproducing kernel Hilbert spaces, J. Multivariate Anal. 140 (2015) 395–402.
[27] D. Liebl, Modeling and forecasting electricity spot prices: A functional data perspective, Ann. Appl. Stat. 7 (2013) 1562–1592.
[28] A. Meister, Asymptotic equivalence of functional linear regression with a white noise inverse problem, Ann. Statist. 39 (2011) 1471–1495.
[29] J.-Y. Park, J. Qian, Functional regression of continuous state distributions, J. Econometrics 167 (2012) 397–412.
[30] T. Pham, V. Panaretos, Methodology and convergence rates for functional time series regression, 2016. arXiv:1612.07197.
[31] B.S. Rajput, Gaussian measures on L^p spaces, $1 \leq p < \infty$, J. Multivariate Anal. 2 (1972) 382–403.
[32] J.O. Ramsay, C.J. Dalzell, Some tools for functional data analysis, J. R. Stat. Soc. Ser. B Stat. Methodol. 53 (1991) 539–572.
[33] J.O. Ramsay, B.W. Silverman, Functional Data Analysis, second ed., Springer, New York, 2005.
[34] M. Reed, B. Simon, Methods of Modern Mathematical Physics I: Functional Analysis, Academic Press, New York, 1980 (Revised and Enlarged ed.).
[35] D.W. Stroock, Probability Theory: An Analytic View, second ed., Cambridge University Press, 2011.
[36] A.B. Tsybakov, Introduction To Nonparametric Estimation, Springer, New York, 2003.
[37] F. Yao, H.-G. Müller, J.-L. Wang, Functional linear regression analysis for longitudinal data, Ann. Statist. 33 (2005) 2873–2903.
[38] M. Yuan, T.T. Cai, A reproducing kernel Hilbert space approach to functional linear regression, Ann. Statist. 38 (2010) 3412–3444.