Statistica Sinica **29** (2019), 2055-2081
doi:https://doi.org/10.5705/ss.202017.0208

A SIMPLE METHOD TO CONSTRUCT CONFIDENCE BANDS IN FUNCTIONAL LINEAR REGRESSION

Masaaki Imaizumi and Kengo Kato

The Institute of Statistical Mathematics and Cornell University

Abstract: This study develops a simple method for constructing confidence bands centered at a principal component analysis (PCA)-based estimator of the slope function in a functional linear regression model with a scalar response variable and a functional predictor variable. A PCA-based estimator is a series estimator with estimated basis functions; thus, constructing these valid confidence bands is a nontrivial challenge. We propose a confidence band that covers most of the slope function with a prespecified probability (level), and prove its asymptotic validity under suitable regularity conditions. To the best of our knowledge, this is the first study to derive that derives confidence bands with theoretical justifications for the PCA-based estimator. We also propose a practical method for choosing the cutoff level used in the PCA-based estimation, and conduct numerical studies to verify the finite-sample performance of the proposed bands. Finally, we apply our methodology to spectrometric data, and discuss extensions of our methodology to cases where additional vector-valued regressors are present.

Key words and phrases: Confidence band, functional linear regression, functional principal component analysis.

1. Introduction

Data collected on dense grids can be viewed as realizations of a random function. These data are called *functional data*, and the statistical methodology used to analyze the data is called a *functional data analysis*. Such analyses are widely used in areas such as chemometrics, econometrics, and biomedical studies; see, for example, Ramsey and Silverman (2005); Ferraty and Vieu (2006); Hsing and Eubank (2015). One of the most basic models used in these analysis is the functional linear regression model, where researchers typically focus on estimations of, and inferences on the slope function. Such estimations are often based on functional principal component analysis (PCA) (cf. Cardot, Ferraty and Sarda (1999); Ramsey and Silverman (2005); Yao, Müller and Wang (2005a); Cai and Hall (2006); Hall and Horowitz (2007)).

Here, we develop a simple method for constructing confidence bands for the slope function in a functional linear regression model tha can be applied to a PCA-based estimator. Specifically, we work with the following setting. Let Y be a scalar response variable and let X be a predictor variable, which we assume to be an $L^2(I)$-valued random variable (random function) such that $\int_I \mathrm{E}\{X^2(t)\}dt < \infty$, where I is a compact interval. Consider a functional linear model with a scalar response variable

$$Y = a + \int_I b(t)[X(t) - \mathrm{E}\{X(t)\}]dt + \varepsilon,\ \mathrm{E}(\varepsilon) = 0,\ \mathrm{E}(\varepsilon^2) = \sigma^2 \in (0, \infty), \quad (1.1)$$

where a is an unknown constant ($a = \mathrm{E}(Y)$), $b \in L^2(I)$ is an unknown slope function, and X and ε are independent. The error variance σ^2 is also unknown. We are interested in constructing confidence bands for the slope function b centered at a PCA-based estimator. In spite of extensive studies on functional linear regression models, to the best of our knowledge, none provide a formal result on confidence bands for the slope function b that can be applied to a PCA-based estimator (see the literature review). The purpose of this paper is to fill this important void.

Quantifying the uncertainty of an estimator is a pivotal part or a statistical analysis. Confidence bands provide a simple-to-interpret graphical description of the accuracy of nonparametric estimators. Several techniques are available for constructing confidence bands for kernel estimations of density and regression functions (Smirnov (1950); Bickel and Rosenblatt (1973); Claeskens and Van Keilegom (2003); Chernozhukov, Chetverikov and Kato (2014a,b)), as well as series estimation using *nonstochastic* basis functions (Chernozhukov, Chetverikov and Kato (2014a); Belloni et al. (2015); Chen and Christensen (2015)). See also Wasserman (2006) and Giné and Nickl (2016) for more information on nonparametric inferences. However, PCA estimations of functional linear models use the eigenfunctions of the empirical covariance function. Because the latter function is stochastic, the eigenfunctions are stochastic as well. Thus, the randomness of these eigenfunctions must be considered, which presents a new and nontrivial challenge. Of course, in principle, it is possible to show that the effects of estimation errors in the empirical eigenfunctions is negligible and, thus, to apply existing tools (e.g., those developed by Belloni et al. (2015)) to construct confidence bands for the population eigenfunction. However, translating the required regularity conditions into primitive conditions is highly nontrivial because a functional PCA is essentially an L^2-theory, where the confidence bands require controlling the sup-norm error of the estimator. Furthermore, the required regularity conditions are technically involved. For instance, the regularity conditions given in Belloni et al. (2015, Thm. 5.6) for the confidence bands for series estimators (with known basis functions) involve L^∞-approximation properties and smoothness of the basis functions. Thus, translating these conditions on the basis functions to the covariance function to the eigenfunctions are used as the basis functions in nontrivial. In addition, in reality, we use the empirical eigenfunctions, which are stochastic. Hence, the bias is stochastic as well and the sup-norm control of the bias is nontrivial. Note that the eigenfunctions of the covariance function depend intrinsically on the distribution of X. Thus, placing restrictions on the eigenfunctions narrows the admissible class of distributions of X, which, in turn, restricts the applicability of the resulting method.

The aim of this study is to propose a simple method for constructing confidence bands centered at a PCA-based estimator that works under the regularity conditions standard in the literature on functional linear regressions. To this end, we slightly relax the coverage requirements of the confidence bands, as in Cai, Low and Ma (2014), and require that our confidence band cover the slope function b at "most" of the points $t \in I$ with a prespecified probability, say 90% or 95%. We then propose a confidence band centered at the PCA-based estimator and show that under suitable regularity conditions, the proposed confidence band satisfies this new requirement asymptotically. For the proposed confidence band to work in practice, the choice of the cutoff level is crucial. In theory, we should choose the cutoff level in such a way that it "undersmoothes" the PCA-based estimator. To this end, we propose choosing a level slightly larger than the optimal level that minimizes the estimate of the L^2-risk of the PCA-based estimator. The proposed confidence band, asymptotic validity of the band, and selection rule of the cutoff level are new. We investigate the finite-sample performance of the proposed confidence band using numerical simulations, showing that the proposed band and the selection rule for the cutoff level work well in practice. Finally, we apply our methodology to spectrometric data, and discuss extensions of our methodology to cases where additional vector-valued regressors are present (see Appendix A).

There are many studies on estimations and predictions in functional linear regression models; see Cardot, Ferraty and Sarda (1999, 2003), Yao, Müller and Wang (2005a), Cai and Hall (2006), Hall and Horowitz (2007), Li and Hsing (2007), Crambes, Kneip and Sarda (2009), James, Wang and Zhu (2009), Cardot and Johannes (2010), Yuan and Cai (2010), Meister (2011), Delaigle and Hall (2012), and Cai and Yuan (2012). Statistical inferences, such as hypothesis testing and the construction of (pointwise) confidence intervals, for functional linear models is studied in Müller and Stadmüller (2005), Cardot, Mas and Sarda (2007), González-Manteiga and Martínez-Calvo (2011), Hilgert, Mas and Verzelen (2013), Lei (2014), Shang and Cheng (2015), and Khademnoe and Hosseini-Nasab (2016). Except for Müller and Stadmüller (2005), these works do not address confidence bands for the slope function. Cardot, Mas and Sarda (2007), González-Manteiga and Martínez-Calvo (2011), and Khademnoe and Hosseini-Nasab (2016) examine confidence intervals for a scalar parameter $\int_I b(t)x(t)dt$ for a fixed $x \in L^2(I)$, and Hilgert, Mas and Verzelen (2013) and Lei (2014) examine testing the hypothesis that $b = 0$ against suitable alternatives. These topics are related to, but substantially different to that examined in our study. Shang and Cheng (2015) develop a number of important inference results for a generalized functional linear model, which includes our model (1.1) as a special case. In particular, they prove a pointwise asymptotic normality result for an estimator based on a reproducing kernel Hilbert space approach (see their Corollary 3.7), which leads to valid pointwise confidence intervals for the slope function. However, they do not consider confidence bands for the slope function, and work with a different estimator to our PCA-based estimator. Müller and Stadmüller (2005) is an important and pioneering work on confidence bands for the slope function in a generalized functional linear model. However, they work with nonstochastic basis functions and, strictly speaking, prove only that their band is a valid confidence band for the surrogate function, but not for the slope function itself. Hence, they do not formally show whether their band is valid when the estimated eigenfunctions are used. See Section 2.3 for a detailed comparison between our bands and the confidence band of Müller and Stadmüller (2005). Our numerical studies in Section 5 show that the confidence band of Müller and Stadmüller (2005), when applied to the PCA-based estimator, tends to have a coverage probability far below the nominal level. Babii (2016) studies a generic (but conservative) method for constructing honest confidence bands for ill-posed inverse problems, which include functional linear regressions as a special case. However, Babii (2016) focuses on the Tikhonov regularization estimation (and, thus, does not cover PCA-based estimations), and makes substantially different assumptions to ours (see his Assumption 5). Other works on confidence bands for functional data include Bunea, Ivanescu and Wegkamp (2011), Degras (2011), Cao, Yanga and Todemc (2012), Ma, Yang and Carroll (2012), Chang, Lin and Ogden (2017). However, these studies do not examine with their functional linear regression model (1.1), and the methodologies and techniques differ substantially from ours. For example, Chang, Lin and Ogden (2017) consider a functional regression model where the response variable is a function and the predictor variable is a vector, which is the opposite setting to that in our study.

The rest of the paper is organized as follows. In Section 2, we informally present our methodology for constructing confidence bands for b using a PCA-based estimator. In Section 3, we present theoretical guarantees of the proposed confidence band. In Section 4, we propose a practical method for choosing the cutoff level used in the PCA-based estimation. In Section 5, we present numerical results to verify the finite-sample performance of the proposed confidence band. Section 6 concludes the paper. The Appendix, available in the Supplementary Material, contains an extension of the methodology to cases with additional regressors, as well as additional numerical experiments and all proofs.

1.1. Notation

We use the following notation. For any measurable functions $f : I \to \mathbb{R}$ and $R : I^2 \to \mathbb{R}$, let $\|f\| = \left\{\int_I f^2(t)dt\right\}^{1/2}$ and $|||R||| = \left\{\iint_{I^2} R^2(s,t)dsdt\right\}^{1/2}$. 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.

2. Methodology

2.1. Functional PCA

We begin by reviewing an approach for estimating b based on a functional PCA. Let $K(s,t)$ denote the covariance function of X, namely, $K(s,t) = \mathrm{Cov}\{X(s), X(t)\}$ for $s, t \in I$. We assume that the integral operator from $L^2(I)$ onto itself with kernel K, namely, the covariance operator of X, is injective. The covariance operator is self-adjoint and positive-definite. The Hilbert–Schmidt theorem (see, e.g., Reed and Simon (1980, Thm. VI.16)) then ensures that K admits the spectral expansion

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

in $L^2(I^2)$, where $\kappa_1 \geq \kappa_2 \geq \cdots > 0$ is a nonincreasing sequence of eigenvalues tending to zero and $\{\phi_j\}_{j=1}^{\infty}$ is an orthonormal basis of $L^2(I)$ consisting of the eigenfunctions of the integral operator, that is,

$$\int_I K(s,t)\phi_j(t)dt = \kappa_j \phi_j(s),\ j = 1, 2, \ldots.$$

Because $\{\phi_j\}_{j=1}^\infty$ is an orthonormal basis of $L^2(I)$, we have the following expansions in $L^2(I)$: $b(t) = \sum_{j=1}^\infty b_j\phi_j(t)$ and $X(t) = \mathrm{E}\{X(t)\} + \sum_{j=1}^\infty \xi_j\phi_j(t)$, where b_j and ξ_j are defined by $b_j = \int_I b(t)\phi_j(t)dt$ and $\xi_j = \int_I [X(t) - \mathrm{E}\{X(t)\}]\phi_j(t)dt$, respectively. Then, we obtain the following alternative expression of the regression model (1.1): $Y = a + \sum_{j=1}^\infty b_j\xi_j + \varepsilon$. Now, observe that $\mathrm{E}(\xi_j) = 0$ for all $j = 1, 2, \ldots$ and

$$\mathrm{E}(\xi_j\xi_k) = \iint_{I^2} K(s,t)\phi_j(s)\phi_k(t)dsdt = \begin{cases} \kappa_j & \text{if } j = k, \\ 0 & \text{if } j \neq k, \end{cases}$$

which yields that $\mathrm{E}(\xi_j Y) = \kappa_j b_j$ for each $j = 1, 2, \ldots$, that is,

$$b_j = \frac{\mathrm{E}(\xi_j Y)}{\kappa_j}. \tag{2.1}$$

This characterization leads to a method for estimating b.

Let $(Y_1, X_1), \ldots, (Y_n, X_n)$ be independent copies of (Y, X). First, we estimate K using the empirical covariance function $\widehat{K}$, defined as $\widehat{K}(s,t) = n^{-1}\sum_{i=1}^n \{X_i(s) - \overline{X}(s)\}\{X_i(t) - \overline{X}(t)\}$ for $s, t \in I$, where $\overline{X} = n^{-1}\sum_{i=1}^n X_i$. Let $\widehat{K}(s,t) = \sum_{j=1}^\infty \widehat{\kappa}_j\widehat{\phi}_j(s)\widehat{\phi}_j(t)$ 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 $\{\widehat{\phi}_j\}_{j=1}^\infty$ is an orthonormal basis of $L^2(I)$ consisting eigenfunctions of the integral operator with kernel $\widehat{K}$, that is,

$$\int_I \widehat{K}(s,t)\widehat{\phi}_j(t)dt = \widehat{\kappa}_j\widehat{\phi}_j(s),\ j = 1, 2, \ldots.$$

The spectral expansion of $\widehat{K}$ is possible since the integral operator with kernel $\widehat{K}$ is of finite rank (at most $(n-1)$). Thus, in addition to an orthonormal system of $L^2(I)$ consisting of the eigenfunctions corresponding to the positive eigenvalues, we can add functions such that the augmented system of functions $\{\widehat{\phi}_j\}_{j=1}^\infty$ becomes an orthonormal basis of $L^2(I)$. Now, let

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

Using the characterization in (2.1), we estimate each b_j by $\widehat{b}_j = n^{-1}\sum_{i=1}^n \widehat{\xi}_{i,j}Y_i/\widehat{\kappa}_j$, and consider an estimator for b of the form

$$\widehat{b}(t) = \sum_{j=1}^{m_n} \widehat{b}_j\widehat{\phi}_j(t),$$

where m_n is the cutoff level such that $m_n \to \infty$ as $n \to \infty$. Hall and Horowitz (2007) study the properties of the PCA-based estimator $\widehat{b}$ in detail and provide conditions under which the estimator is rate-optimal.

2.2. Construction of confidence bands

For a given $\tau \in (0,1)$, a confidence band for b with level $1-\tau$ is a collection of random intervals $\mathcal{C} = \{\mathcal{C}(t) = [\ell(t), u(t)] : t \in I\}$ such that

$$\mathrm{P}\{b(t) \in [\ell(t), u(t)] \text{ for all } t \in I\} \geq 1 - \tau. \tag{2.2}$$

Here, we focus on confidence bands centered at the PCA-based estimator $\widehat{b}$, thereby quantifying the uncertainty of the PCA-based estimator $\widehat{b}$. However, as discussed in the Introduction, the requirement given in (2.2) is too stringent for our problem. Thus, and we consider a less demanding requirement. That is, instead of requiring (2.2), we construct a confidence band $\mathcal{C} = \{\mathcal{C}(t) = [\ell(t), u(t)] : t \in I\}$ such that for given $\tau_1, \tau_2 \in (0,1)$, with probability at least $1-\tau_1$, the proportion of the set of t at which b is not covered by $\mathcal{C}$ is at most τ_2; that is,

$$\mathrm{P}\left\{\lambda\left(\{t \in I : b(t) \notin [\ell(t), u(t)]\}\right) \leq \tau_2 \lambda(I)\right\} \geq 1 - \tau_1, \tag{2.3}$$

where λ denotes the Lebesgue measure. If the band $\mathcal{C}$ satisfies the new requirement (2.3), then it covers b over more than $100(1-\tau_2)\%$ of points in I with probability at least $1-\tau_1$. Thus, as long as τ_2 is close to 0, the band $\mathcal{C}$ covers b over "most" points in I, with probability at least $1-\tau_1$. Hence, the new requirement (2.3) is a reasonable relaxation of the former requirement (2.2).

A relaxed coverage requirement similar to (2.3) appears in Cai, Low and Ma (2014) for the purpose of constructing adaptive confidence bands in a nonparametric regression. We employ the relaxed coverage requirement (2.3) to resolve a different challenge, namely, that of constructing confidence bands for a series estimator with estimated basis functions.

In what follows, we informally present our methodology for constructing a confidence band for the PCA-based estimator $\widehat{b}$ that satisfies (2.3) asymptotically. Under some regularity conditions, we show that

$$\begin{aligned} &n\|\widehat{b} - b\|^2 \\ &= \sum_{j=1}^{m_n}\left(n^{-1/2}\sum_{i=1}^{n}\frac{\varepsilon_i\widehat{\xi}_{i,j}}{\widehat{\kappa}_j}\right)^2 + O_{\mathrm{P}}(m_n^{\alpha/2+1} + \sqrt{n}m_n^{-\beta+\alpha/2+1} + nm_n^{-2\beta+1}), \end{aligned} \tag{2.4}$$

where $\varepsilon_i = Y_i - a - \int_I b(t)[X_i(t) - \mathrm{E}\{X(t)\}]dt$ for $i = 1, \ldots, n$, where the last term on the right-hand side of (2.4) is (suitably) negligible relative to the first term (the parameters α and β are given in the next section). Observe that, by definition,

$$n^{-1}\sum_{i=1}^{n}\widehat{\xi}_{i,j}\widehat{\xi}_{i,k}=\iint_{I^2}\widehat{K}(s,t)\widehat{\phi}_j(s)\widehat{\phi}_k(t)dsdt=\begin{cases}\widehat{\kappa}_j & \text{if } j=k,\\ 0 & \text{if } j\neq k.\end{cases}$$

Hence, conditional on $X_1^n=\{X_1,\dots,X_n\}$,

$$\left(n^{-1/2}\sum_{i=1}^{n}\frac{\varepsilon_i\widehat{\xi}_{i,j}}{\widehat{\kappa}_j}\right)_{j=1}^{m_n} \tag{2.5}$$

is the sum of independent random vectors with mean zero, and the covariance matrix of the random vector (2.5) conditional on X_1^n is $\sigma^2\Lambda_n$, where $\Lambda_n=\operatorname{diag}(1/\widehat{\kappa}_1,\dots,1/\widehat{\kappa}_{m_n})$. We show that, under some regularity conditions, the distribution of the random vector (2.5) can be approximated by that of $N(0,\sigma^2\Lambda_n)$. Therefore the distribution of the first term on the right-hand side of (2.4) can be approximated by $\sigma^2\sum_{j=1}^{m_n}\eta_j/\widehat{\kappa}_j$, where $\eta_1,\dots,\eta_{m_n}$ are independent $\chi^2(1)$ random variables independent of X_1^n. Note that when ε is Gaussian, the random vector (2.5) has the same distribution as that of $N(0,\sigma^2\Lambda_n)$. Thus, for a given $\tau\in(0,1)$, let

$$\widehat{c}_n(1-\tau)=\text{conditional }(1-\tau)\text{-quantile of }\sqrt{\sum_{j=1}^{m_n}\frac{\eta_j}{\widehat{\kappa}_j}}\text{ given }X_1^n,$$

which can be computed using simulations, and consider an L^2-confidence ball for b of the form

$$\mathcal{B}_n(1-\tau)=\left\{b:\|\widehat{b}-b\|\leq\frac{\widehat{\sigma}\widehat{c}_n(1-\tau)}{\sqrt{n}}\right\}, \tag{2.6}$$

where $\widehat{\sigma}^2=n^{-1}\sum_{i=1}^{n}(Y_i-\overline{Y}-\sum_{j=1}^{m_n}\widehat{b}_j\widehat{\xi}_{i,j})^2$ with $\overline{Y}=n^{-1}\sum_{i=1}^{n}Y_i$, and $\widehat{\sigma}=\sqrt{\widehat{\sigma}^2}$. We show that, under some regularity conditions, this confidence ball contains the slope function b with probability $1-\tau+o(1)$ as $n\to\infty$. However, it is well known that an L^2-confidence ball is difficult to visualize/interpret. Thus, instead, we construct a confidence band for b by modifying the confidence ball, borrowing from Juditsky and Lambert-Lacroix (2003); see also Section 5.8 in Wasserman (2006). Specifically, we propose the following confidence band for b:

$$\widehat{\mathcal{C}}=\left\{\widehat{\mathcal{C}}(t)=\left[\widehat{b}(t)-\frac{\widehat{\sigma}\widehat{c}_n(1-\tau_1)}{\sqrt{n}}\sqrt{\frac{1}{\tau_2\lambda(I)}},\widehat{b}(t)+\frac{\widehat{\sigma}\widehat{c}_n(1-\tau_1)}{\sqrt{n}}\sqrt{\frac{1}{\tau_2\lambda(I)}}\ \right]:t\in I\right\}, \tag{2.7}$$

where τ_1 and τ_2 are constants such that $\tau_1,\tau_2\in(0,1)$.

It follows from an argument similar to that in Wasserman (2006, p.95) that, with probability at least $1-\tau_1+o(1)$, the proportion of the set of t at which b is not covered by $\widehat{\mathcal{C}}$ is at most τ_2, that is,

$$\mathrm{P}\left\{\lambda\left(\left\{t\in I: b(t)\notin\widehat{\mathcal{C}}(t)\right\}\right)\le\tau_2\lambda(I)\right\}\ge 1-\tau_1+o(1), \tag{2.8}$$

so that the proposed confidence band (2.7) satisfies requirement (2.3) asymptotically. In fact, let U be a uniform random variable on I independent of the data, and let P_U denote the probability with respect to U only. Then ,

$$\lambda\left(\left\{t\in I: b(t)\notin\widehat{\mathcal{C}}(t)\right\}\right)=\lambda(I)\mathrm{P}_U\left\{\sqrt{n\tau_2\lambda(I)}|\widehat{b}(U)-b(U)|>\widehat{\sigma}\widehat{c}_n(1-\tau_1)\right\},$$

and Markov's inequality yields that the right-hand side is bounded by $n\tau_2\lambda(I)\|\widehat{b}-b\|^2/\{\widehat{\sigma}^2\widehat{c}_n^2(1-\tau_1)\}$. Therefore,

$$\begin{aligned}&\mathrm{P}\left\{\lambda\left(\{t\in I: b(t)\notin\widehat{\mathcal{C}}(t)\}\right)\le\tau_2\lambda(I)\right\}\\ &\ge\mathrm{P}\left\{n\|\widehat{b}-b\|^2\le\widehat{\sigma}^2\widehat{c}_n^2(1-\tau_1)\right\}=1-\tau_1+o(1),\end{aligned}$$

which yields the desired result.

The values of τ_1 and τ_2 are chosen by users, where $1-\tau_1$ is the nominal level. Thus, a popular choice of τ_1 is 0.1 or 0.05. The value of τ_2 is the proportion of the set of points not covered by the confidence band. In practice, we should choose τ_2 to be small (but not too small since this would make the width of the band too large). In the numerical studies in Section 5, we take $\tau_2=0.1$. In theory, it is relatively straightforward to see that we may choose τ_2 in such a way that $\tau_2\downarrow 0$, so that the proportion of the excluded domain is asymptotically vanishing. See also Remark 5.

To compute the quantile $\widehat{c}_n(1-\tau_1)$, we propose using simulations. An alternative way to approximate the quantile $\widehat{c}_n(1-\tau_1)$ is to apply the central limit theorem to $\sum_{j=1}^{m_n}\eta_j/\widehat{\kappa}_j$. In fact, under some regularity conditions, it holds that $1/(\sqrt{2\sum_{k=1}^{m_n}\widehat{\kappa}_k^{-2}})\sum_{j=1}^{m_n}\widehat{\kappa}_j^{-1}(\eta_j-1)\xrightarrow{d}N(0,1)$. Therefore, $\widehat{c}_n^2(1-\tau_1)$ can be approximated as $\sum_{j=1}^{n}\widehat{\kappa}_j^{-1}+\Phi^{-1}(1-\tau_1)\sqrt{2\sum_{k=1}^{m_n}\widehat{\kappa}_k^{-2}}$, where Φ is the distribution function of the standard normal distribution. However, in applications, m_n is often small compared with n, and the above normal approximation can be imprecise. Therefore, we recommend simulating the quantile $\widehat{c}_n(1-\tau_1)$ directly instead of relying on the central limit theorem.

Note that our confidence band (2.7) is, in general, conservative; namely, $\liminf_{n\to\infty}\mathrm{P}\{\lambda(\{t\in I: b(t)\notin\widehat{\mathcal{C}}(t)\})\le\tau_2\lambda(I)\}$ is, in general, larger than $1-\tau_1$, which is clear from the discussion above. However, the numerical studies in Section 5 suggest that the width of our band, with the cutoff level chosen by the rule suggested in Section 4, is reasonably narrow in practice.

The proposed confidence bands allow a small portion of the domain to be excluded from the confidence bands. Despite the proposed confidence bands not covering all points in the domain with a given level, they are able to capture the global shape of the slope function, which helps practitioners to make inferences on the slope function. Furthermore, partly because of the conservative nature of our bands, in our numerical studies, we find that our bands tend to have reasonably good uniform coverage probabilities. Hence, we believe that the proposed methodology adds a valuable option for inferences on functional linear regressions.

Remark 1 (Inference on sub-regions). Note that our confidence band over the full domain can be applied to construct an estimate of a sub-domain on which the slope function b satisfies a certain property, for instance, a subset of I on which $\nu_\ell \leq b(t) \leq \nu_u$, for some thresholds $\nu_\ell < \nu_u$. More formally, let $I_{b,\nu}$ with $\nu = (\nu_\ell, \nu_u)$ be a subset of I defined by $I_{b,\nu} := \{t \in I : \nu_\ell \leq b(t) \leq \nu_u\}$. Based on our confidence band $\widehat{\mathcal{C}} = \{\mathcal{C}(t) = [\widehat{\ell}(t), \widehat{u}(t)] : t \in I\}$, we can construct an estimate for the set $I_{b,\nu}$ by $\widehat{I}_{b,\nu} = \left\{t \in I : \nu_\ell \leq \widehat{u}(t), \widehat{\ell}(t) \leq \nu_u\right\}$. From equation (2.8), we have

$$\lambda(I_{b,\nu} \cap \widehat{I}_{b,\nu}) \geq \lambda(I_{b,\nu}) - \lambda(\{t \in I : b(t) \notin \widehat{\mathcal{C}}(t)\}) \geq \lambda(I_{b,\nu}) - \tau_2 \lambda(I),$$

with probability at least $1 - \tau_1 + o(1)$. Therefore, we have

$$\mathrm{P}\left\{\frac{\lambda\left(I_{b,\nu} \cap \widehat{I}_{b,\nu}\right)}{\lambda(I_{b,\nu})} \geq 1 - \tau_2 \frac{\lambda(I)}{\lambda(I_{b,\nu})}\right\} \geq 1 - \tau_1 + o(1).$$

As $\tau_1 \downarrow 0$ and $\tau_2 \downarrow 0$ as $n \to \infty$, we have $(\lambda(I_{b,\nu} \cap \widehat{I}_{b,\nu}))/(\lambda(I_{b,\nu})) \xrightarrow{P} 1$, which shows a version of consistency of the estimate $\widehat{I}_{b,\nu}$.

Next, our confidence band is a global band, and hence, is not suitable for constructing "local" confidence bands, that is, confidence bands on a chosen subset of I, say a neighborhood of an interior point $t_0 \in I$. However, we believe that our methodology is a useful addition to functional data analyses. To construct "local" confidence bands, one approach would be to take a convolution of $\widehat{b}$ with a kernel. That is, let $L : \mathbb{R} \to \mathbb{R}$ be a sufficiently regular kernel function that integrates to one and is supported in $[-1, 1]$, and let $h > 0$ be a sufficiently small bandwidth such that $[t_0 \pm h] \subset I$. In addition, let $\widehat{b}_h(t_0) = \langle L_h(\cdot - t_0), \widehat{b}\rangle$, and $b_h(t_0) = \langle L_h(\cdot - t_0), b\rangle$ where $L_h(\cdot) = h^{-1}L(\cdot/h)$. Then, under (possibly restrictive) regularity conditions, it is expected that $a_n(\widehat{b}_h(t_0) - b_h(t_0))$ converges in distribution to a normal distribution for a suitable norming constant a_n; see Cardot, Mas and Sarda (2007). This leads to a method for constructing a confidence interval for $b_h(t_0)$. However, in theory, we need to take $h = h_n \to 0$ sufficiently fast to ensure that the bias $b_h(t_0) - b(t_0)$ is negligible, which requires an additional smoothness assumption on b. Because this approach is tangential to the methodology developed here, an analysis of such "local" confidence bands is left for future research.

Remark 2 (Equivariance of the band). Note that our confidence band (2.7) is equivariant under location-scale changes to the index t. Suppose that $I = [\underline{c}, \overline{c}]$, and consider a change of variable $t = \underline{c} + u(\overline{c} - \underline{c})$ for $u \in [0,1]$. Let $X_i^\dagger(u) = X_i(\underline{c} + u(\overline{c} - \underline{c}))$ and $b^\dagger(u) = (\overline{c} - \underline{c})b(\underline{c} + u(\overline{c} - \underline{c}))$ for $u \in [0,1]$, and observe that $\int_I b(t)X_i(t)dt = \int_0^1 b^\dagger(u)X_i^\dagger(u)du$. Furthermore, let $\widehat{\kappa}_j^\dagger = \widehat{\kappa}_j/(\overline{c} - \underline{c})$ and $\widehat{\phi}_j^\dagger(u) = \sqrt{\overline{c} - \underline{c}}\widehat{\phi}_j(\underline{c} + u(\overline{c} - \underline{c}))$ for $u \in [0,1]$. Then, $\{(\widehat{\kappa}_j^\dagger, \widehat{\phi}_j^\dagger)\}_{j=1}^\infty$ are eigenvalue/eigenfunction pairs for the empirical covariance function $\widehat{K}^\dagger$ of $\{X_i^\dagger\}_{i=1}^n$, that is, $\int_0^1 \widehat{K}^\dagger(v,u)\widehat{\phi}_j^\dagger(u)du = \widehat{\kappa}_j^\dagger\widehat{\phi}_j^\dagger(v)$. It is not difficult to see that the PCA-based estimator with cutoff level m_n for $b^\dagger$ based on the data $\{(Y_i, X_i^\dagger)\}_{i=1}^n$ is $\widehat{b}^\dagger(u) = (\overline{c} - \underline{c})\widehat{b}(\underline{c} + u(\overline{c} - \underline{c}))$ for $u \in [0,1]$. Next, the conditional $(1-\tau_1)$-quantile of $\sqrt{\sum_{j=1}^{m_n} \eta_j/\widehat{\kappa}_j^\dagger} = \sqrt{\overline{c} - \underline{c}}\sqrt{\sum_{j=1}^{m_n} \eta_j/\widehat{\kappa}_j}$, denoted by $\widehat{c}_n^\dagger(1-\tau_1)$, is identical to $\sqrt{\overline{c} - \underline{c}}\widehat{c}_n(1-\tau_1)$. Thus, our confidence band applied to the data $\{(Y_i, X_i^\dagger)\}_{i=1}^n$ is

$$\begin{aligned}\widehat{\mathcal{C}}^\dagger(u) &= \left[\widehat{b}^\dagger(u) \pm \frac{\widehat{\sigma}\widehat{c}_n^\dagger(1-\tau_1)}{\sqrt{n}}\sqrt{\frac{1}{\tau_2}}\ \right] \\ &= (\overline{c} - \underline{c})\left[\widehat{b}(\underline{c} + u(\overline{c} - \underline{c})) \pm \frac{\widehat{\sigma}\widehat{c}_n(1-\tau_1)}{\sqrt{n}}\sqrt{\frac{1}{\tau_2(\overline{c} - \underline{c})}}\ \right],\end{aligned}$$

for $u \in [0,1]$. Therefore, we conclude that $b^\dagger(u) \in \widehat{\mathcal{C}}^\dagger(u) \Leftrightarrow b(\underline{c} + u(\overline{c} - \underline{c})) \in \widehat{\mathcal{C}}(\underline{c} + u(\overline{c} - \underline{c}))$ for $u \in [0,1]$, and thus $\lambda(\{u \in [0,1] : b^\dagger(u) \notin \widehat{\mathcal{C}}^\dagger(u)\} = \lambda(\{t \in I : b(t) \notin \widehat{\mathcal{C}}(t)\})/(\overline{c} - \underline{c})$.

Remark 3 (Additional smoothing steps). To simplify the theoretical analysis, we assume that entire trajectories of X_i are observed without measurement errors. In applications, functional predictor variables are often discrete and contain measurement errors. In such cases, a standard approach is to first estimate X_i using a smoothing technique (Yao, Müller and Wang (2005b); Hall, Müller and Wang (2006)). Our methodology includes over to cases where such additional smoothing steps are involved, since the proofs of Theorems 1 and 2 rely essentially on the fact that 1) the distribution of the vector $(n^{-1/2}\sum_{i=1}^n \varepsilon_i\widehat{\xi}_{i,j}/\widehat{\kappa}_j)_{j=1}^{m_n}$

conditional on X_1^n is $N(0,\sigma^2\Lambda_n)$ or can be approximated by $N(0,\sigma^2\Lambda_n)$, and 2) the empirical eigenfunctions $\widehat{\phi}_j$ are sufficiently accurate estimates of the population eigenfunctions of ϕ_j, without relying on how we estimate the eigenfunctions ϕ_j.

Remark 4 (Extension to a multidimensional domain). The extension to a compact multi-dimensional domain (i.e., to the case where I is a compact subset of $\mathbb{R}^d$ with $d \geq 2$) is possible without making any significant changes. The spectral expansion of the covariance function holds for the multi-dimensional case and, hence, it is not difficult to see that the other analysis carries over to the multi-dimensional case as well. However, since it is more difficult to visualize confidence bands in the multi-dimensional case, we focus on one-dimensional intervals here.

2.3. Comparison with the confidence band of Müller and Stadmüller (2005)

Müller and Stadmüller (2005) provided a pioneering work on confidence bands for the slope function in a generalized functional linear model. In the context of our model (1.1), their proposal reads as follows. Suppose for the sake of simplicity that $\mathrm{E}(Y) = 0$ and $\mathrm{E}\{X(t)\} = 0$ for all $t \in I$. For a given, nonstochastic orthonormal basis $\{\rho_j\}_{j=1}^\infty$ of $L^2(I)$, expand X_i and b as $X_i = \sum_j \zeta_{i,j}\rho_j$ and $b = \sum_j \theta_j\rho_j$, respectively, with $\zeta_{i,j} = \int_I X_i(t)\rho_j(t)dt$ and $\theta_j = \int_I b(t)\rho_j(t)dt$. Now, observe that $Y_i = \sum_j \zeta_{i,j}\theta_j + \varepsilon_i$ and obtain an estimator $\widehat{\theta} = (\widehat{\theta}_1,\ldots,\widehat{\theta}_{m_n})^T$ of $\theta = (\theta_1,\ldots,\theta_{m_n})^T$ by regressing Y_i on $(\zeta_{i,1},\ldots,\zeta_{i,m_n})^T$, where $m_n \to \infty$ as $n \to \infty$. Müller and Stadmüller (2005) show that, under some regularity conditions, $(n(\widehat{\theta}-\theta)^T(\Gamma/\sigma^2)(\widehat{\theta}-\theta) - m_n)/\sqrt{2m_n} \xrightarrow{d} N(0,1)$, where $\Gamma = \{\mathrm{E}(\zeta_{1,j}\zeta_{1,k})\}_{1\leq j,k\leq m_n}$. Based on this result, they propose the following confidence band: denote by $(e_1,\lambda_1),\ldots,(e_{m_n},\lambda_{m_n})$ the eigenvectors/eigenvalues of the matrix Γ, and consider

$$\widetilde{b}(t) \pm \sigma\sqrt{\frac{\widetilde{c}_n(1-\tau)}{n}\sum_{j=1}^{m_n}\frac{\omega_j^2(t)}{\lambda_j}},\ t \in I, \tag{2.9}$$

where $\omega_j(t) = \sum_{k=1}^{m_n}\rho_k(t)e_{j,k}$ with $e_j = (e_{j,1},\ldots,e_{j,m_n})^T$, and $\widetilde{c}_n(1-\tau) = m_n + \sqrt{2m_n}\Phi^{-1}(1-\tau)$. To compare our band (2.7) with (2.9), assume that the covariance function K is known for (2.9) and use the eigenfunctions $\{\phi_j\}_{j=1}^\infty$ for $\{\rho_j\}_{j=1}^\infty$. In that case, the band (2.9) is of the form

$$\widetilde{b}(t) \pm \sigma \sqrt{\frac{\widetilde{c}_n(1-\tau)}{n} \sum_{j=1}^{m_n} \frac{\phi_j^2(t)}{\kappa_j}},\ t \in I. \tag{2.10}$$

In their theoretical analysis, Müller and Stadmüller (2005) work with nonstochastic basis functions and, strictly speaking, prove only that the band (2.9) is a valid confidence band for the surrogate function $\sum_{j=1}^{m_n} \theta_j \rho_j$ (i.e., they prove that the band (2.9) contains $\sum_{j=1}^{m_n} \theta_j \rho_j(t)$ for all $t \in I$ with probability at least $1 - \tau + o(1)$), but not for the slope function b itself. Hence, they do not formally show whether the band (2.10) is valid for b when the estimated eigenfunctions $\{\widehat{\phi}_j\}_{j=1}^{\infty}$ are used. It is possible to show that, under suitable regularity conditions, the band (2.10), with (κ_j, ϕ_j) replaced by $(\widehat{\kappa}_j, \widehat{\phi}_j)$, contains the (random) surrogate function $\sum_{j=1}^{m_n} \breve{b}_j \widehat{\phi}_j$ with probability at least $1 - \tau + o(1)$, where $\breve{b}_j = \int_I b(t)\widehat{\phi}_j(t)dt$. However, to show that the band is valid for b (i.e., to show that the band contains $b(t)$ for all $t \in I$ with probability at least $1-\tau+o(1)$), we would have to show that the supremum bias $\sup_{t\in I} |b(t) - \sum_{j=1}^{m_n} \breve{b}_j \widehat{\phi}_j(t)|$ (which is random) is negligible relative to the infimum width of the band, which is nontrivial.

3. Theoretical Guarantees

In this section, we investigate the validity of our confidence band. We separately consider the cases where the error distribution is Gaussian and non-Gaussian.

3.1. Case with Gaussian errors

We first consider the case where the error distribution is Gaussian. In this case, we make the following assumptions.

Assumption 1. *There exist constants $\alpha > 1, \beta > \alpha/2 + 3/2$, and $C_1 > 1$ such that (i) $\mathrm{E}(\|X\|^2) < \infty$ and $\mathrm{E}(\xi_j^4) \le C_1\kappa_j^2$ for all $j = 1, 2, \ldots$; (ii) $\kappa_j \le C_1 j^{-\alpha}$ and $\kappa_j - \kappa_{j+1} \ge C_1^{-1} j^{-\alpha-1}$ for all $j = 1, 2, \ldots$; (iii) $|b_j| \le C_1 j^{-\beta}$ for all $j = 1, 2, \ldots$; and (iv) $m_n^{2\alpha+2}/n \to 0$ and $m_n^{\alpha+2\beta-1}/n \to \infty$.*

Conditions (i)–(iii) are adapted from Hall and Horowitz (2007) and are (more or less) standard in theoretical analyses of PCA-based estimators (Cai and Hall (2006); Meister (2011); Lei (2014); Kong et al. (2016)). The estimation of the slope function b is an ill-posed inverse problem (as discussed in Hall and Horowitz (2007)), and the value of α that appears in Condition (ii) measures the "ill-posedness" of the estimation problem (the larger α is, the more difficult the estimation of b is). The second part of Condition (ii), which ensures the sufficient estimation accuracy of the empirical eigenfunctions, also implies that $\kappa_j \geq j^{-\alpha}/(C_1\alpha)$ for all $j = 1, 2, \ldots$ Condition (iii) is concerned with the smoothness of b. The requirement that $m_n^{2\alpha+2}/n \to 0$ is a technical condition used to control for estimation errors of the empirical eigenfunctions. The last condition, $m_n^{\alpha+2\beta-1}/n \to 0$, can be interpreted as an "undersmoothing" condition. From Hall and Horowitz (2007), the optimal rate of m_n for an estimation is $m_n \sim n^{1/(\alpha+2\beta)}$. However, the last condition requires that m_n has to be of a larger order than the optimal value in order for the bias to be negligible relative to the "variance" term. This undersmoothing condition is commonly used to construct confidence bands. See Section 5.7 in Wasserman (2006) for a related discussion. We discuss a practical choice of the cutoff level in the next section. Note that to ensure that Condition (iv) is nonvoid, we need that $\beta > \alpha/2 + 3/2$.

Theorem 1. *For given $\tau_1, \tau_2 \in (0, 1)$, consider the confidence band $\widehat{\mathcal{C}}$ defined in* (2.7)*. Let $\varepsilon \sim N(0, \sigma^2)$. Then, under Assumption* 1*, the result* (2.8) *holds as $n \to \infty$. Furthermore, the width of the band $\widehat{\mathcal{C}}$ is $O_{\mathrm{P}}(\sqrt{m_n^{\alpha+1}/n})$.*

The proof of Theorem 1 consists of approximating the distribution of $n\|\widehat{b} - b\|^2$ by that of $\sigma^2 \sum_{j=1}^{m_n} \eta_j/\widehat{\kappa}_j$, where $\eta_1, \ldots, \eta_{m_n}$ are independent $\chi^2(1)$ random variables independent of X_1^n. However, since the approximating distribution also depends on n (and random), the proof of the theorem is nontrivial. To formally show that the error of the stochastic approximation in (2.4) is negligible for the distributional approximation, we rely on concentration and anti-concentration inequalities for a weighted sum of independent $\chi^2(1)$ random variables; see Lemma 1 in the Supplementary Material.

Remark 5. An inspection of the proof shows that the result (2.8) holds even if we choose $\tau_2 \downarrow 0$ (the rate at which $\tau_2 \downarrow 0$ as $n \to \infty$ can be arbitrary). The width of the band is then $O_{\mathrm{P}}\{\sqrt{m_n^{\alpha+1}/(n\tau_2)}\}$.

Remark 6 (Uniformity in distribution). The coverage result (2.8) holds uniformly over a certain class of distributions of (Y, X). For given $\alpha > 1, \beta > \alpha/2 + 3/2$, and $C_1 > 1$, let $\mathcal{F}_{\mathrm{Normal}}(\alpha, \beta, C_1)$ be the class of distributions of (Y, X) that verify (1.1) and Conditions (i)–(iii) in Assumption 1; furthermore $\varepsilon \sim N(0, \sigma^2)$ is independent of X with $C_1^{-1} \leq \sigma^2 \leq C_1$. Then, provided that $m_n^{2\alpha+2}/n \to 0$ and $m_n^{\alpha+2\beta-1}/n \to \infty$, we have

$$\liminf_{n\to\infty} \inf_{F \in \mathcal{F}_{\mathrm{Normal}}(\alpha,\beta,C_1)} \mathrm{P}_F \left\{ \lambda\left(\left\{ t \in I : b(t) \notin \widehat{\mathcal{C}}(t) \right\} \right) \leq \tau_2 \lambda(I) \right\} \geq 1 - \tau_1, \quad (3.1)$$

where P_F denotes the probability under F. In fact, to show (3.1), it is enough to verify that for any sequence $F_n \in \mathcal{F}_{\mathrm{Normal}}(\alpha, \beta, C_1)$, the result in (2.8) holds for $(Y_1, X_1), \ldots, (Y_n, X_n) \sim F_n$ i.i.d. for $n \geq 1$, which is not difficult to verify in view of the proof of Theorem 1. Furthermore, the result in (3.1) holds even if $\tau_2 \downarrow 0$. A similar comment applies to Theorem 2.

3.2. Case with non-Gaussian errors

Next, we consider the case where the error distribution is possibly non-Gaussian. Instead of Assumption 1, we make the following assumptions. For $q > 1$ and $\alpha > 0$, let $c(q, \alpha) = \max\{2\alpha + 2, 7/(2 - 2/q)\}$.

Assumption 2. *There exist an integer $q \geq 2$ and constants $\alpha > 1, \beta > \{c(q,\alpha) - \alpha + 1\}/2$, and $C_1 > 0$ such that*

$$\mathrm{E}(\xi_j^{2q}) \leq C_1 \kappa_j^q \text{ for all } j = 1, 2, \ldots, \tag{3.2}$$

and Conditions (ii) and (iii) in Assumption 1 *are satisfied. Furthermore, assume that*

$$\frac{m_n^{c(q,\alpha)}}{n} \to 0 \quad \text{and} \quad \frac{m_n^{\alpha+2\beta-1}}{n} \to \infty. \tag{3.3}$$

These conditions guarantee that all of the conclusions of Theorem 1 remain valid, even when the error is non-Gaussian.

Theorem 2. *Suppose that ε has mean zero and variance $\sigma^2 > 0$, and that $\mathrm{E}[\varepsilon^4] < \infty$. Then, under Assumption* 2*, all of the conclusions of Theorem* 1 *remain true.*

In comparison with the Gaussian error case, we require conditions that rae more restrictive (note that if $\mathrm{E}(\xi_j^{2q}) \leq C_1 \kappa_j^q$ for some $q \geq 2$, then $\mathrm{E}(\xi_j^4) \leq \{\mathrm{E}(\xi_j^{2q})\}^{2/q} \leq C_1^{2/q} \kappa_j^2$). These additional conditions are used to apply the high-dimensional central limit theorem of Bentkus (2005). Condition (3.2) is satisfied for all $q \geq 2$ if X is Gaussian. Conditions similar to (3.2) are also employed, for example, Cai and Hall (2006) and Hilgert, Mas and Verzelen (2013). If we choose q to be sufficiently large, that is, $q > (4\alpha + 4)/(4\alpha - 3)$, then the conditions on m_n reduce to those in the Gaussian error case.

4. Choice of Cutoff Level

For the proposed confidence band to work in practice, the choice of the cutoff level m_n is crucial. In theory, we should choose m_n so that it is of larger order than the optimal rate $n^{1/(\alpha+2\beta)}$ for an estimation under the L^2-risk. The idea is to construct an estimate of the L^2-risk of $\widehat{b}$ with the given cutoff level m, and to choose a cutoff level slightly larger than the optimal cutoff level that minimizes the estimate of the L^2-risk. The construction of the estimate of the L^2-risk of $\widehat{b}$ is inspired by Cavalier et al. (2002). Recall that b is expressed as $b(t) = \sum_{j=1}^{\infty} b_j\phi_j(t) = \sum_{j=1}^{\infty}(c_j/\kappa_j)\phi_j(t)$, where $c_j = \mathrm{E}(\xi_j Y)$. Suppose first that the covariance function K is known, and consider, for a given cutoff level m, the estimator

$$\widehat{b}^*(t;m) = \sum_{j=1}^{m} \widehat{b}_j^*\phi_j(t) = \sum_{j=1}^{m} \frac{\widehat{c}_j^*}{\kappa_j}\phi_j(t),$$

where $\widehat{c}_j^* = n^{-1}\sum_{i=1}^{n} \xi_{i,j}Y_i$ and $\widehat{b}_j^* = \widehat{c}_j^*/\kappa_j$. Let $R^*(m)$ denote the L^2-risk of the estimator $\widehat{b}^*(\cdot;m)$, that is,

$$\begin{aligned} R^*(m) &= \mathrm{E}[\|\widehat{b}^*(\cdot;m) - b\|^2] \\ &= \sum_{j>m} b_j^2 + \sum_{j=1}^{m} \frac{\mathrm{Var}(\widehat{c}_j^*)}{\kappa_j^2} = \|b\|^2 - \sum_{j=1}^{m} b_j^2 + \frac{1}{n}\sum_{j=1}^{m} \frac{\mathrm{Var}(\xi_j Y)}{\kappa_j^2}. \end{aligned}$$

Minimizing $R^*(m)$ is equivalent to minimizing

$$\check{R}^*(m) = -\sum_{j=1}^{m} b_j^2 + \frac{1}{n}\sum_{j=1}^{m} \frac{\mathrm{Var}(\xi_j Y)}{\kappa_j^2}.$$

Recall that $\check{R}^*(m)$ is still unknown. However, we may estimate $\check{R}^*(m)$ using

$$\widehat{R}^*(m) = -\sum_{j=1}^{m} (\widehat{b}_j^*)^2 + \frac{2}{n(n-1)}\sum_{j=1}^{m} \frac{\sum_{i=1}^{n}(\xi_{i,j}Y_i - \widehat{c}_j^*)^2}{\kappa_j^2}.$$

In fact, since $\mathrm{E}[(\widehat{b}_j^*)^2] = b_j^2 + \mathrm{Var}(\widehat{c}_j^*)/\kappa_j^2$, $\widehat{R}^*(m)$ is an unbiased estimator of $\check{R}^*(m)$.

In practice, K is unknown, and so we replace K with $\widehat{K}$, and for our estimator $\widehat{b}$, we use

$$\widehat{R}(m) = -\sum_{j=1}^{m} \widehat{b}_j^2 + \frac{2}{n(n-1)}\sum_{j=1}^{m} \frac{\sum_{i=1}^{n}(\widehat{\xi}_{i,j}Y_i - \widehat{c}_j)^2}{\widehat{\kappa}_j^2}$$

as an estimate of the L^2-risk of $\widehat{b}$ with cutoff level m, where $\widehat{c}_j = n^{-1}\sum_{i=1}^{n} \widehat{\xi}_{i,j}Y_i$. Now, let $\widehat{m}_n$ be a minimizer of $\widehat{R}(m)$ over a candidate set chosen by users; our recommendation is to choose either $\max\{\widehat{m}_n, 2\}$ or $\widehat{m}_n + 1$ for the construction of the proposed confidence band.

5. Numerical Results

5.1. Simulations

We consider the following data-generating process. Let $I = [0,1], \phi_1 \equiv 1$, and $\phi_{j+1}(t) = 2^{1/2}\cos(j\pi t)$, for $j = 1, 2, \ldots$, and generate (X, Y) as follows:

$$Y = \int_I b(t)X(t)dt + \varepsilon,\ X = \sum_{j=1}^{50} j^{-\alpha/2} U_j \phi_j,$$

$$b = \sum_{j=1}^{50} b_j \phi_j,\ b_1 = 1, b_j = 4(-1)^j j^{-\beta} \text{ for } j \neq 1,$$

where $U_j \sim \text{Unif.}[-3^{1/2}, 3^{1/2}]$ are independent. The distribution of the error term ε is either $N(0,1)$ or normalized $\chi^2(5)$. We consider the following configurations for (α, β): $\alpha \in \{1.1, 2\}$ and $\beta \in \{2.6, 3.2\}$. We construct confidence bands of the form given in (2.7), with $\tau_1 = \tau_2 = 0.1$, and examine the following sample sizes: $n \in \{100, 200, \ldots, 1{,}000\}$. We evaluate the confidence bands using

$$\text{UCP} = \text{P}\{b(t) \in \widehat{\mathcal{C}}(t) \text{ for all } t \in I\},$$

and

$$\text{MCP} = \text{P}\left\{\lambda\left(\left\{t \in I : b(t) \notin \widehat{\mathcal{C}}(t)\right\}\right) \leq \tau_2\right\},$$

where UCP signifies the "uniform coverage probability" and MCP signifies the "modified coverage probability." We compare the performance of our confidence band (2.7) with that of the Müller–Stadmüller (MS) band (2.10), where we replace (κ_j, ϕ_j) and σ by $(\widehat{\kappa}_j, \widehat{\phi}_j)$ and $\widehat{\sigma}$, respectively. Note that we have also examined a version of the MS band where we replaced $\widetilde{c}_n(1-\tau_1)$ with the $(1-\tau_1)$-quantile of the $\chi^2(m_n)$-distribution, trying to improve upon the performance of the MS band; however, the results where similar to those presented here. Recall that our band aims to control the MCP at level $1-\tau_1$, where the MS band aims to control the UCP at level $1-\tau_1$. The number of Monte Carlo repetitions in each of the following experiments is 2,000. The computations of the integrals and evaluations of the MCPs and UCPs are carried out by discretizing the unit interval [0, 1] into 50 equally spaced grids. For the computation of $\widehat{m}_n$, discussed in the previous section, we have to choose a set of candidate cutoff levels. In this simulation study, we select $\{1, \ldots, 10\}$ as the set of candidate cutoff levels.

Before investigating at the performance of the confidence bands, we examine how our selection rules for the cutoff level work in practice. For comparison, we also report the oracle cutoff level m_n^* that minimizes the L^2-risk of the PCA-

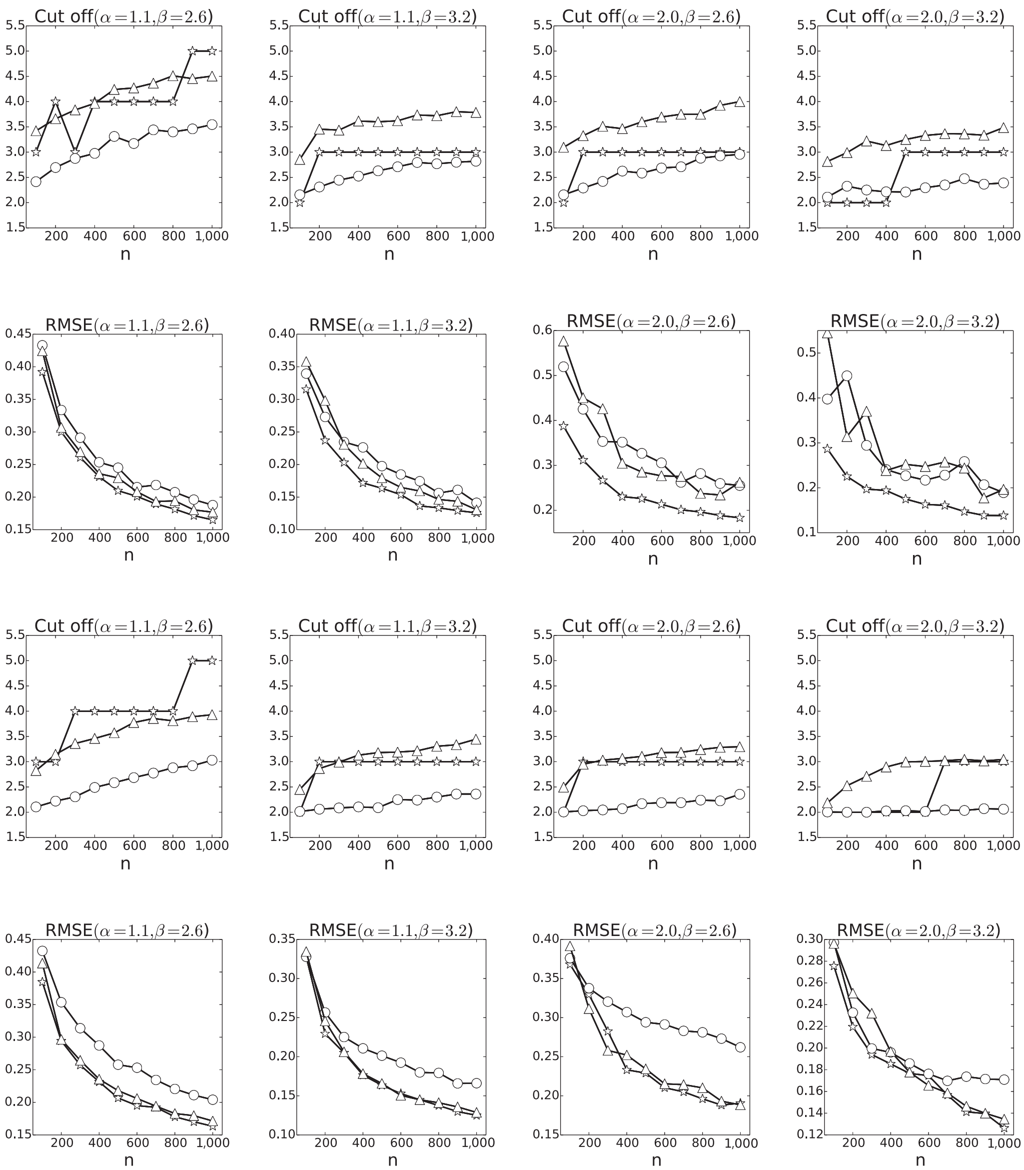

Figure 1. Values of the cutoff levels and of the RMSE for each level with Gaussian noise (upper 2 rows) and χ^2 noise (lower two rows). Stars correspond to cases with the oracle cutoff level m_n^*, and triangles and circles correspond to those with $\widehat{m}_n + 1$ and $\max\{\widehat{m}_n, 2\}$, respectively. THe Monte Carlo averages of cutoff levels $\widehat{m}_n + 1$ and $\max\{\widehat{m}_n, 2\}$ are also reported.

based estimator. That is, denoting as $\widehat{b}(\cdot; m)$ the PCA-based estimator with the given cutoff level, m_n^* is defined as

$$m_n^* = \arg\min\{\mathrm{E}[\|\widehat{b}(\cdot; m) - b\|^2] : m = 1, 2, \ldots, 10\}.$$

Figure 1 presents the values of the cutoff levels and the RMSE for each level. The RMSE is the square root of the L^2-risk, and the Monte Carlo averages of the cutoff levels $\widehat{m}_n + 1$ and $\max\{\widehat{m}_n, 2\}$ are also reported. For each of the parameter configurations, as expected, all of $\widehat{m}_n + 1, \max\{\widehat{m}_n, 2\}$, and m_n^* increase with n (with the exception for m_n^* when $(n, \alpha, \beta) = (300, 1.1, 2.6)$). Furthermore, $\widehat{m}_n + 1$ tends to be larger (on average) than the oracle level m_n^*, but $\max\{\widehat{m}_n, 2\}$ tends to be smaller (on average) than m_n^*. However, their differences are not significant. Figure 1 also shows that the optimal RMSE decreases monotonically in n, which is consistent with our intuitive expectation. However, the RMSE for the PCA-based estimator with a cutoff level of $\max\{\widehat{m}_n, 2\}$ or $\widehat{m}_n + 1$ need not decrease monotonically. We interpret that this is because $\max\{\widehat{m}_n, 2\}$ and $\widehat{m}_n + 1$ need not select the optimal cutoff level. Furthermore, because of the discreteness of the cutoff levels, the RMSE curve tends to be wiggly compared with, for example, the kernel estimation; that is, the bandwidth used in the kernel estimation can take a continuum of values while the cutoff level used in series estimation can take only positive integers. Therefore, the RMSE of the series estimation tends to be variable for different cutoff levels, resulting in the non-monotonicity of the RMSE with $\max\{\widehat{m}_n, 2\}$ or $\widehat{m}_n + 1$. However, overall, the RMSE tends to be smaller as n increases. In terms of the RMSE, both $\widehat{m}_n + 1$ and $\max\{\widehat{m}_n, 2\}$ work reasonably well compared with the optimal RMSE (i.e., the RMSE with m_n^*). Interestingly, $\widehat{m}_n + 1$ tends to yield better RMSEs than $\max\{\widehat{m}_n, 2\}$ does.

Next, we examine the performance of the confidence bands. The simulated coverage probabilities are plotted in Figure 2. The following observations can be drawn from the figure. First, the coverage probabilities of the MS band, either in the UCP or in the MCP, tend to be far below the nominal coverage probability of 90%. Secont, the MCPs of our band with the cutoff level $\widehat{m}_n + 1$ satisfy the nominal level in all cases, and those with the cutoff level $\max\{\widehat{m}_n, 2\}$ are reasonably close to the nominal level, except for a few cases. Note that our band with the cutoff level $\widehat{m}_n + 1$ appears to be conservative, but this is not inconsistent with the theory. Third, although our band is not designed to control the UCP, our band with the cutoff level $\widehat{m}_n + 1$ has reasonably good UCPs.

The simulation results for the expected maximum width and expected mean width of our confidence band (2.7) and the MS band (2.10) are plotted in Figures

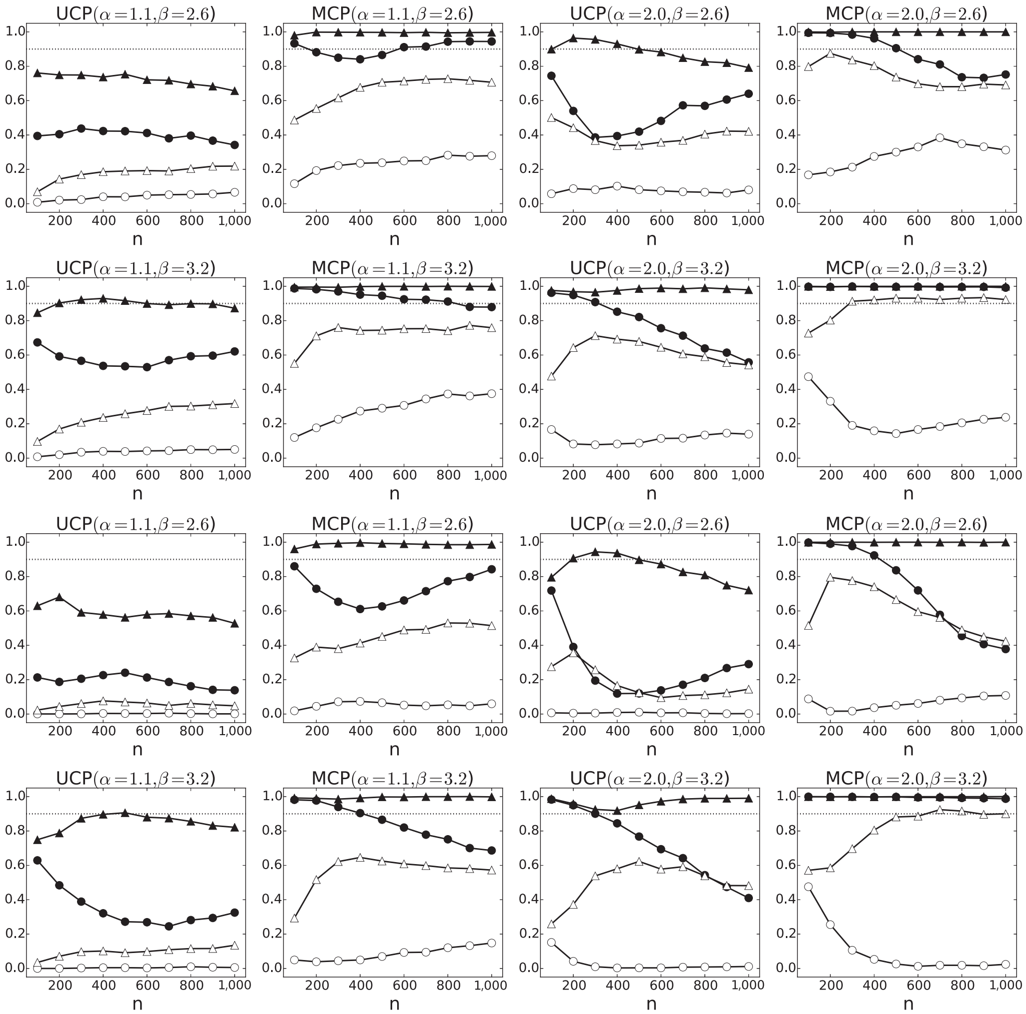

Figure 2. Coverage probabilities with Gaussian noise (upper two rows) and χ^2 noise (lower two rows). The black markers show the coverage probabilities of our band (2.7), and the white markers show those of the MS band (2.10). Circles correspond to cases with the cutoff level $\max\{\widehat{m}_n, 2\}$, and triangles to those with $\widehat{m}_n + 1$. The dashed line shows the value $1 - \tau_1 = 0.90$.

3 and 4, respectively. For a confidence band $\mathcal{C} = \{[\ell(t), u(t)] : t \in I\}$, the expected maximum width and expected mean width are defined by

$$\mathrm{E}\left[\max_{t \in I}\{u(t) - \ell(t)\}\right] \quad \text{and} \quad \mathrm{E}\left[\frac{1}{\lambda(I)}\int_I \{u(t) - \ell(t)\}dt\right],$$

respectively. Note that our confidence band has constant width; thus,the maximum and mean widths are identical for our band. The figures show that our

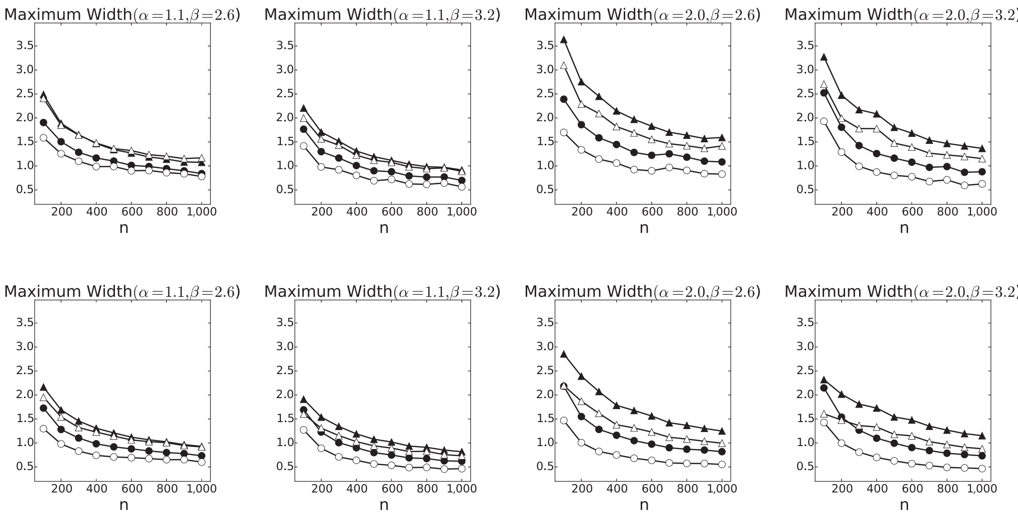

Figure 3. Expected maximum width of confidence bands with Gaussian noise (upper row) and χ^2 noise (lower row). The black markers correspond to our band (2.7), while and the markers to the MS band (2.10). Circles correspond to cases with the cutoff level $\max\{\widehat{m}_n, 2\}$, and triangles to those with $\widehat{m}_n + 1$.

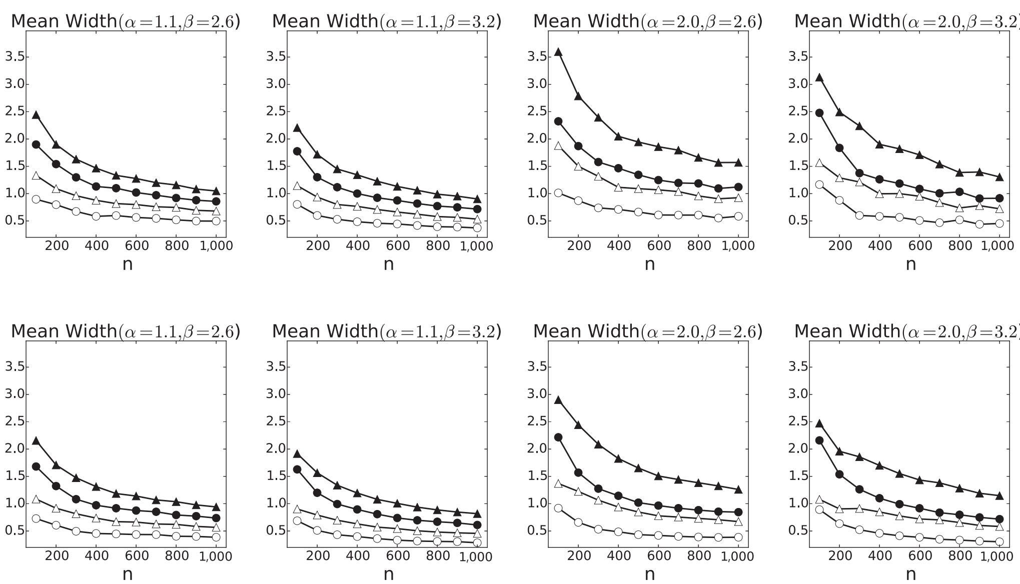

Figure 4. Expected mean width of confidence bands with Gaussian noise (upper row) and χ^2 noise (lower row). The black markers correspond to our band (2.7), while and the white markers to the MS band (2.10). Circles correspond to cases with the cutoff level $\max\{\widehat{m}_n, 2\}$, and triangles to those with $\widehat{m}_n + 1$.

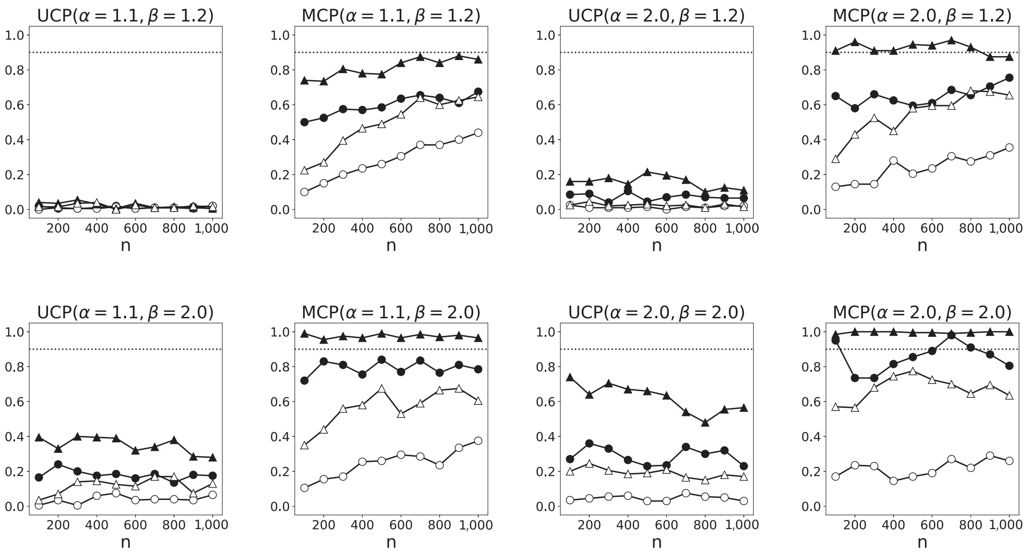

Figure 5. Coverage probabilities with Gaussian noise. The black markers show the coverage probabilities of our band, and the white markers show those of the MS band. Circles correspond to cases with the cutoff level $\max\{\widehat{m}_n, 2\}$, and triangles to those with $\widehat{m}_n + 1$. The dashed line shows the value $1 - \tau_1 = 0.90$.

confidence band (2.7) tends to have a larger width than that of the MS band (2.10), which is not surprising in view of the comparison of the coverage probabilities of the bands. In other words, the MS band has narrower widths, but at the cost of (severe) under-coverage. However, the width of our band is not excessively large compared with that of the MS band.

Note that in some cases, the UCPs and MCPs of our band with the cutoff level $\max\{\widehat{m}_n, 2\}$ decrease as n increases. This is partly because the bias has non-negligible effects relative to the width of the band, since the choice of $\max\{\widehat{m}_n, 2\}$ is not actually undersmoothing the function estimate.

In conclusion, the simulation results suggest that, in terms of the coverage probability, our confidence band with the cutoff level $\widehat{m}_n + 1$ is recommended, but that using the cutoff level $\max\{\widehat{m}_n, 2\}$ is a viable alternative if narrower confidence bands are preferred.

We also present additional experiments with parameter configurations of $\alpha \in \{1.1, 2\}$ and $\beta \in \{1.2, 2\}$, which violate the condition $\beta > \alpha/2 + 3/2$ in Assumption 1. The simulated coverage probabilities are plotted in Figure 5. Although the performance of our confidence band is worse than that shown in Figure 2, our confidence band has MCPs close to the nominal level, except for the case $(\alpha, \beta) = (1.1, 1.2)$.

5.2. Spectrometric data for predicting fat content

To demonstrate how our methodology works using real data, we apply our confidence band to a regression of fat content in meat on spectra of the light absorption of these substances. In chemometrics, we often observes a spectrum of light absorption for a substance, measured at different wavelengths. Such spectral curves can be regarded as functional data; see Borggaard and Thodberg (1992) and Chapter 5 of Ferraty and Vieu (2006). The analysis using spectrometric data is quick and non-destructive and thus, is often used to investigate the properties, of example, a food sample.

We use the spectrometric data from `http://lib.stat.cmu.edu/datasets/tecator`, which we use to predict the fat content in pure meat. In the dataset, we observe spectral curves from 215 pieces of finely chopped meat (recorded on a Tecator Infratec Food and Feed Analyzer) measured from wavelengths of 850 nm to 1,050 nm. Let $\{X_i(t) : t \in [850, 1{,}050]\}_{i=1}^{215}$ denote these spectral curves, the graphs of which are plotted in Figure 6. The dataset also contains the fat content Y_i in each piece of meat, measured by an analytical chemical processing.

The estimates $\widehat{b}$ and confidence bands $\widehat{\mathcal{C}}$ with cutoff levels 5 and 6 are plotted in the upper, right two panels in Figure 6, where we set $\tau_1 = \tau_2 = 0.1$ (note that we work with the original index set [850, 1,050]; if we normalize the index set to [0, 1], as in Remark 2, then the vertical axes in the right two panels in Figure 6 would have to be multiplied by $1{,}050 - 850 = 200$). Note that the value of $\widehat{m}_n$ is 5 for this dataset. The variance estimates are $\widehat{\sigma}^2 = 11.14$ for $m_n = 5$ and $\widehat{\sigma}^2 = 8.59$ for $m_n = 6$. For comparison, we also plot the 90% MS bands with cutoff levels of 5 and 6. The figure shows that both of our bands are reasonably narrow.

Confidence bands are useful to identify ranges of wavelengths that play a minor (or major) role in predicting the fat content. Figure 6 leads to the following two observations. First, there are some peaks in the estimates (negative at around 900 nm and 950 nm, and positive at around 930 nm) and our confidence bands at those peaks do not contain aero. Thus, the spectra at around those wavelengths certainly contribute to the fat content prediction. Second, for higher wavelengths (i.e., wavelengths higher than 970 nm), our confidence band with a cutoff level of 6 almost always contains zero, and our confidence band with a cutoff level of 5 also contains zero except at around 1,050 nm. This suggests that the spectra at higher wavelengths do not contribute much to the fat content prediction.

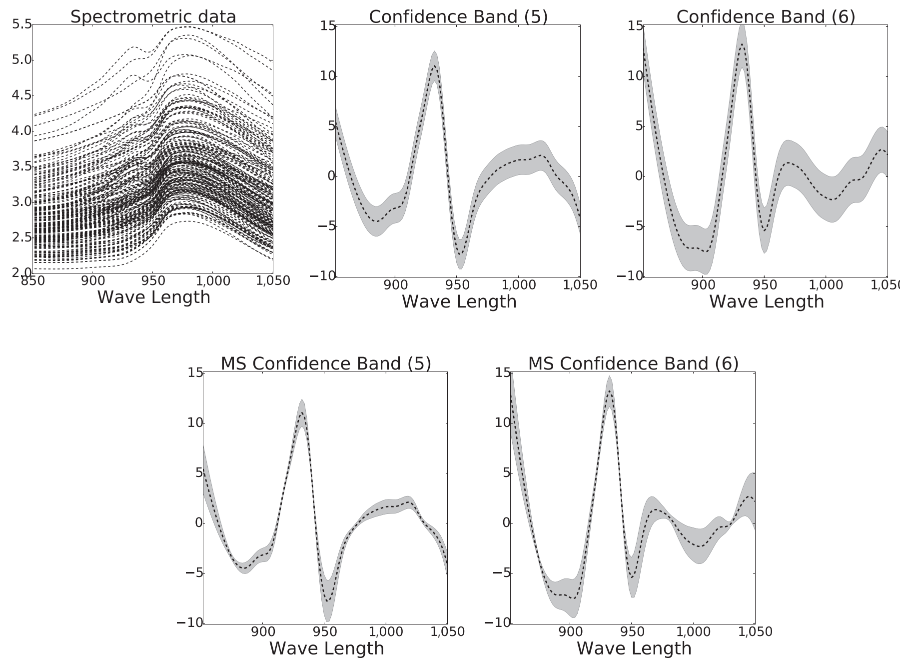

Figure 6. Spectrometric data (upper, far-left panel) and the estimates $\widehat{b}$ (dashed lines) and confidence bands (gray areas). The upper, right two panels depict our confidence bands with cutoff levels of 5 (left) and 6 (right), and lower two panels depict the MS bands with cutoff levels of 5 (left) and 6 (right).

6. Conclusion

We have proposed a simple method for constructing confidence bands, centered at a PCA-based estimator, for the slope function in a functional linear regression model. The proposed confidence band aims to cover the slope function at "most" of the points with a prespecified probability. Furthermore, we have proved its asymptotic validity under suitable regularity conditions. Importantly, to the best of our knowledge, this is the first study to derive confidence bands with theoretical justifications for the PCA-based estimator. We have also proposed a practical method for choosing the cutoff level. Our numeral simulations show that the proposed confidence band and selection rule for the cutoff level work well in practice.

Acknowledgements

M. Imaizumi is supported by Grant-in-Aid for JSPS Fellows (17J03208). K.

Kato is supported by Grant-in-Aid for Scientific Research (C) (15K03392) from the JSPS.

References

Babii, A. (2016). Honest confidence sets in nonparametric IV regression and other ill-posed models. arXiv:1611.03015.

Bickel, P. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1** 1071-1095. Correction (1975) **3**, 1370.

Belloni, A., Chernozhukov, V., Chetverikov, D. and Kato, K. (2015). Some new asymptotic theory for least squares series: pointwise and uniform results. *J. Econometrics* **187**, 345-366.

Bentkus, V. (2005). A Lyapunov-type bound in $\mathbb{R}^d$. *Theory Probab. Appl.* **49**, 311-323.

Borggaard, C. and Thodberg, H. H. (1992). Optimal minimal neural interpretation of spectra. *Anal. Chem.* **64**, 545-551.

Bosq, D. (2000). *Linear Processes in Function Spaces: Theory and Applications.* Springer.

Boucheron, S., Lugosi, G. and Massart, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence.* Oxford University Press.

Bunea, F., Ivanescu, A. E. and Wegkamp, M. H. (2011). Adaptive inference for the mean of gaussian process in functional data. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **73**, 531-558.

Cai, T. T. and Hall, P. (2006). Prediction in functional linear regression. *Ann. Statist.* **34**, 2159-2179.

Cai, T. T., Low, M. and Ma, Z. (2014). Adaptive confidence bands for nonparametric regression functions. *J. Amer. Statist. Assoc.* **109**, 1054-1070.

Cai, T. T. and Yuan, M. (2012). Minimax and adaptive prediction for functional linear regression. *J. Amer. Statist. Assoc.* **107**, 1201-1216.

Cao, G., Yanga, L. and Todemc, D. (2012). Simultaneous inference for the mean function based on dense functional data. *J. Nonparametr. Stat.* **24**, 359-377.

Cardot, H. and Johannes, J. (2010). Thresholding projection estimators in functional linear models. *J. Multivariate Anal.* **101**, 395-408.

Cardot, H., Ferraty, F. and Sarda, P. (1999). Functional linear model. *Statist. Probab. Lett.* **45**, 11-22.

Cardot, H., Ferraty, F. and Sarda, P. (2003). Spline estimators for the functional linear models. *Statist. Sinica* **13**, 571-591.

Cardot, H., Mas, A. and Sarda, P. (2007). CLT in functional linear regression models. *Probab. Theory Related Fields* **138**, 325-361.

Cavalier, L., Golubev, Y., Picard, D. and Tsybakov, A. B. (2002). Oracle inequalities for inverse problems. *Ann. Statist.* **30**, 843-874.

Chang, C., Lin, X. and Ogden, T. (2017). Simultaneous confidence bands for functional regression models. *J. Statist. Plann. Inference* **188**, 67-81.

Chen, X. and Christensen, T. (2015). Optimal sup-norm rates, adaptivity and inference in nonparametric instrumental variables estimation. arXiv:1508.03365.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2014a). Gaussian approximation of suprema of empirical processes. *Ann. Statist.* **42**, 1564-1597.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2014b). Anti-concentration and honest, adaptive confidence bands. *Ann. Statist.* **42**, 1787-1818.

Claeskens, G. and Van Keilegom, I. (2003). Bootstrap confidence bands for regression curves and their derivatives. *Ann. Statist.* **31**, 1852-1884.

Crambes, C., Kneip, A. and Sarda, P. (2009). Smoothing splines estimators for functional linear regression. *Ann. Statist.* **37**, 35-72.

Degras, D. A. (2011). Simultaneous confidence bands for nonparametric regression with functional data. *Statist. Sinica* **21**, 1735-1765.

Delaigle, A. and Hall, P. (2012). Methodology and theory for partial least squares applied to functional data. *Ann. Statist.* **40**, 322-352.

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice.* Springer.

Giné, E. and Nickl, R. (2016). *Mathematical Foundations of Infinite-Dimensional Statistical Models.* Cambridge University Press.

González-Manteiga, W. and Martínez-Calvo, A. (2011). Bootstrap in functional linear regression. *J. Statist. Plann. Inference* **141**, 453-461.

Hall, P. and Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *Ann. Statist.* **35**, 70-91.

Hall, P., Müller, H. and Wang, J. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist.* **34**, 1493-1517.

Hilgert, N., Mas, A. and Verzelen, N. (2013). Minimax adaptive tests for the functional linear model. *Ann. Statist.* **41**, 838-869.

Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis With An Introduction to Linear Operators.* Wiley.

Imaizumi, M. and Kato, K. (2016). PCA-based estimation for functional linear regression with functional responses. *Journal of Multivariate Analysis* **163**, 15-36.

James, G. M., Wang, J. and Zhu, J. (2009). Functional linear regression that's interpretable. *Ann. Statist.* **37**, 2083-2108.

Juditsky, A. and Lambert-Lacroix, S. (2003). On nonparametric confidence set estimation. *Math. Methods Statist.* **19**, 410-428.

Khademnoe, O. and Hosseini-Nasab, S. M. E. (2016). On properties of percentile bootstrap confidence intervals for prediction in functional linear regression. *J. Statist. Plann. Inference* **170**, 129-143.

Kong, D., Xue, K., Yao, F. and Zhang, H. H. (2016) Partially functional linear regression in high dimensions. *Biometrika* **103**, 147-159.

Lei, J. (2014). Adaptive global testing for functional linear models. *J. Amer. Statist. Assoc.* **109**, 624-634.

Li, Y. and Hsing, T. (2007). On rates of convergence in functional linear regression. *J. Multivariate Anal.* **98**, 1782-1804.

Ma, S., Yang, L. and Carroll, R.J. (2012). A simultaneous confidence band for sparse longitudinal regression. *Statist. Sinica* **22**, 95-122.

Meister, A. (2011). Asymptotic equivalence of functional linear regression with a white noise inverse problem. *Ann. Statist.* **39**, 1471-1495.

Müller, H.-G. and Stadtmüller, U. (2005). Generalized functional linear models. *Ann. Statist.*

33, 774-805.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis.* 2nd Edition. Springer.

Reed, M. and Simon, B. (1980). *Methods of Modern Mathematical Physics I: Functional Analysis (Revised and Enlarged Edition)*. Academic Press.

Shang, Z. and Cheng, G. (2015). Nonparametric inference in generalized functional linear models. *Ann. Statist.* **43**, 1742-1773.

Shin, H. (2009). Partial functional linear regression. *J. Statist. Plann. Inference* **139**, 3405-3418.

Smirnov, N. V. (1950). On the construction of confidence regions for the density of distribution of random variables. *Doklady Akad. Nauk SSSR* **74**, 189-191 (Russian).

Wasserman, L. (2006). *All of Nonparametric Statistics.* Springer.

Xu, M., Zhang, D. and Wu, W.B. (2014). L^2 asymptotics for high-dimensional data. arXiv:1405.7244.

Yao, F., Müller, H.-G. and Wang, J.-L. (2005a). Functional linear regression analysis for longitudinal data. *Ann. Statist.* **33**, 2873-2903.

Yao, F., Müller, H.-G. and Wang, J.-L. (2005b). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100**, 577-590.

Yuan, M. and Cai, T. (2010). A reproducing kernel Hilbert space approach to functional linear regression. *Ann. Statist.* **38**, 3412-3444.

The Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan.

E-mail: imaizumi@ism.ac.jp

1194 Comstock Hall, Ithaca, NY 14853, USA.

E-mail: kk976@cornell.edu

(Received May 2017; accepted February 2018)