Electronic Journal of Statistics
Vol. 9 (2015) 1779–1798
ISSN: 1935-7524
DOI: 10.1214/15-EJS1052

Simultaneous inference of the mean of functional time series*

Ming Chen and Qiongxia Song†

Department of Mathematical Sciences, The University of Texas at Dallas
Richardson, TX, 75080, USA
e-mail: mxc104720@utdallas.edu; song@utdallas.edu

Abstract: For functional time series with physical dependence, we construct confidence bands for its mean function. The physical dependence is a general dependence framework, and it slightly relaxes the conditions of m-approximable dependence. We estimate functional time series mean functions via a spline smoothing technique. Confidence bands have been constructed based on a long-run variance and a strong approximation theorem, which is satisfied with mild regularity conditions. Simulation experiments provide strong evidence that corroborates the asymptotic theories. Additionally, an application to S&P500 index data demonstrates a non-constant volatility mean function at a certain significance level.

MSC 2010 subject classifications: Primary 62G08, 62G15.
Keywords and phrases: Confidence bands, functional time series, high-frequency data, long-run variance, nonparametric regression, spline.

Received April 2015.

Contents

1. Introduction

Functional data often arise from measurements obtained by separating an almost continuous time period into natural consecutive intervals, for example days. The functions thus obtained form a time series $\{Y_i, i \in Z\}$, where each Y_i is

*We want to thank the referees for their helpful comments.
†Corresponding author.

a curve $Y_i(t), t \in [a,b]$. Such data structures are referred as functional time series. Without loss of generality, we assume that the domain of the functions is $[0,1]$. A central issue in functional time series analysis is taking into account the temporal dependence of the observations, i.e. the dependence between curves of $\{Y_i(t), i \le m\}$ and $\{Y_j(t), j \ge m+l\}$ for any $l \in Z$. Further, we assume that $Y_i(\cdot)$, $i \in Z$, is strictly stationary (stationary in short); see [3]. If a functional time series is not stationary, it is assumed that it can be transformed into a stationary time series by a preprocessing procedure.

The functional data analysis focuses mostly on i.i.d. functional observations; see [18, 22, 23, 17, 21] and references therein. In spite of the methodological advancement with independent observations, the work on functional time series has been sparse and of a more theoretical nature; see, for example, [3]. Functional time series retains the merit of functional observations, while it is more flexible than purely i.i.d curves by allowing a dependence structure. Applications of functional time series analysis are as wide as in the i.i.d. case. For example, we consider S&P500 data. We interpret the intraday information as curves, and the dependence among curves is natural. Recently, more and more researchers work on this type of functional data. To the best of our knowledge, the available inference methods focus mostly on nonparametric estimations of mean and covariance functions of functional series; see Part IV of [10, 1, 13, 14] among the others. A simultaneous confidence band is lacking in the literature.

Two distinct types of functional data have been studied. [22, 23, 16] studied sparse longitudinal data. While [5, 19] and [10] concerned dense functional data. As in the i.i.d. case, functional time series has the same two scenarios, and we only deal with equally-spaced dense design in this paper.

Denote by Y_{ij} the jth observation of the random curve $\xi_i(\cdot)$ at the time point X_{ij}, $1 \le i \le n$, $1 \le j \le N_i$. For the equally-spaced dense design, $N_1 = N_2 = \cdots = N_n = N$, $X_{ij} = j/N$, $1 \le i \le n$, $1 \le j \le N$, with N going to infinity. For the ith subject, $i = 1, 2, \ldots, n$, its sample path $\{j/N, Y_{ij}\}$ is a noisy realization of the continuous time stochastic process $\xi_i(t)$ in the sense that

$$Y_{ij} = \xi_i(j/N) + \sigma(j/N)\epsilon_{ij}, \tag{1.1}$$

with i.i.d. errors ϵ_{ij} satisfying $E(\epsilon_{ij}) = 0, E(\epsilon_{ij}^2) = 1$, and $\{\xi_i(t), t \in [0,1]\}$, independent of errors, is L^2, i.e., $E\int_{[0,1]} \xi^2(t)dt < +\infty$.

The basic unit of information of functional data analysis is the entire observed curve, and an important theoretical and practical issue is the necessity of detecting the global trend of its mean function. Nonparametric simultaneous confidence bands are powerful tools of global inferences for functions, see [6, 9, 11, 4, 20, 24, 25] for the theory and applications. For i.i.d. functional data, [4] constructed simultaneous confidence bands for its mean function.

The paper seeks to construct simultaneous confidence bands of the mean function of dense functional time series data. Our goal is to develop a simple but flexible nonparametric method with a well-justified theory and a fast algorithm to implement in practice. This is done by approximating the nonparametric components via spline estimation. Our approach allows for formal derivation

of the asymptotic properties of the proposed estimators. We establish the $\sqrt{n}$-consistency of the proposed estimator for the mean function.

We organize our paper as follows. In Section 2 we describe the physical dependence and asymptotics under it. In Section 3 we state our main results of confidence bands. In Section 4 we provide the procedure to implement the construction of the proposed confidence band. In Section 5 we report findings of a simulation study. An empirical example in Section 6 illustrates how we use the confidence bands for statistical inferences. We end the paper with a concluding section. Technical proofs are in the Appendix.

2. Weak dependence

In this paper we use the following notation. Let F denote $L^2([0,1])$; the Hilbert space on a compact domain of square integrable functions with the norm $||x||_2^2 = \int_0^1 x^2(s)ds$ and the inner product $\langle x;y\rangle = \int_0^1 x(s)y(s)ds$ for $x;y \in F$. For a Lebesgue measurable function ϕ on $[0,1]$, denote $\|\phi\|_r = \{\int_0^1 |\phi(t)|^r\, dt\}^{1/r}$, $1 \le r < +\infty$ and $\|\phi\|_\infty = \sup_{t\in[0,1]} |\phi(t)|$. For any vector $\zeta = (\zeta_1, \ldots, \zeta_s) \in R^s$, denote the norm $\|\zeta\|_r = \left(|\zeta_1|^r + \cdots + |\zeta_s|^r\right)^{1/r}$, $1 \le r < +\infty$, $\|\zeta\|_\infty = \max\left(|\zeta_1|, \ldots, |\zeta_s|\right)$. $||\cdot||$ means $||\cdot||_2$ when r is not specified.

Now we focus on the temporal dependence of functional time series, and we extend the physical dependence for random variables ([15] and [2]) to a functional setup. Recall that, in our paper, the strictly stationary time series $\{Y_i\}_{i=1}^n$ is functional valued rather than real valued, i.e. each Y_i is a random curve.

In the following, we directly use the description of the physical dependence in [2] to a functional setup. Assume $EY_1 = 0$ with $E\,\|Y_1\|_2^p < \infty$, $2 < p \le 4$. For a stationary sequence allowing the representation

$$Y_i = g(\ldots, \epsilon_{i-1}, \epsilon_i, \epsilon_{i+1}, \ldots),$$

where $\epsilon_i, i \in Z$ are i.i.d random elements, and $g : F^Z \to F$ is a measurable function with $F = L^2([0,1])$. For $i \in Z$ define the shift process $\mathcal{F}_i = (\epsilon_{l+i}, l \in Z)$. The central element of $\mathcal{F}_i$ is ϵ_i, and $Y_i = g(\mathcal{F}_i)$. Let $\{\epsilon_i'; i \in Z\}$ be an i.i.d. copy of $\{\epsilon_i; i \in Z\}$, and for $i, j \in Z$, let $\mathcal{F}_{i,j}$ denote the process obtained from $\mathcal{F}_i$ by replacing the coordinate ϵ_j by ϵ_j'. Define $Y_{i,0} = g(\mathcal{F}_{i,\{0\}})$. Let the physical dependence measure $\theta_{i,p} = (E||Y_i - Y_{i,0}||^p)^{1/p}$. This quantity can be interpreted as the dependence of Y_i on ϵ_0 and $Y_{i,0}$ is a coupled version of Y_i with ϵ_0 in the latter replaced by ϵ_0'. Throughout the paper we assume

$$\Theta_p = \sum_{i=1}^{\infty} \theta_{i,p} < \infty. \tag{2.1}$$

We provide an example with the functional physical dependence.

Example 1. *Define the nonlinear autoregressive model by*

$$X_n = f(X_{n-1}) + \epsilon_n, n \in Z, \tag{2.2}$$

where $\|f(x) - f(y)\| \leq \rho||x - y||, 0 < \rho < 1$. *Special cases of (2.2) include the functional AR model. If* $E||\epsilon_0||_2 < \infty$, *then* X_n *can be represented as* $g(\ldots, \epsilon_{n-1}, \epsilon_n)$ *and satisfies* $\theta_{n,p} \leq c\rho^n$ *for some* $0 < \rho < 1$. *Therefore (2.1) is satisfied.*

In the literature in functional time series, m-approximable dependence has been extensively used, see [1] and [13] among the others. In their framework, weak dependence is quantified by a summability condition which intuitively states that the function g decays so fast that the impact of shocks far back in the past is so small that they can be replaced by their independent copies, with only a small change in the distribution of the process. To be specific, define

$$Y_{i,m} = g(\ldots, \epsilon_{i-m-1}, \epsilon_{i,i-m}^{(m)}, \epsilon_{i,i-m+1}^{(m)}, \ldots), \tag{2.3}$$

then the L^p-m-approximable dependence condition is $\sum_{m\geq 1} v_p\left(Y_i - Y_{i,m}\right) < \infty$, here $v_p\left(X\right) = \left(E\left\|X\right\|_2^p\right)^{1/p}$. The next lemma shows that L^p-m-approximable dependence is ensured by the physical dependence for stationary functional time series.

Lemma 1. *For stationary functional time series,* L^p*-m-approximable dependence ensures the physical dependence. That is to say,*

$$\sum_{m\geq 1} v_p\left(Y_i - Y_{i,m}\right) \leq \infty \Rightarrow \sum_{i\geq 1} v_p\left(Y_i - Y_{i,\{0\}}\right) \leq \infty.$$

The proof is in the appendix. The opposite of the statement in lemma 1 is not true, and a simple counter example is constructed.

Example 2. *Consider that* $Y_i = \sum_{n=0}^{\infty} \frac{\epsilon_{i-n}}{(n+1)^{3/2}}$, *here* $\{\epsilon_i\}_i$ *are i.i.d. Gaussian processes. According to (2.3),* $Y_i - Y_{i,m} = \sum_{n=m}^{\infty} \frac{1}{(n+1)^{3/2}}\left(\epsilon_{i-n} - \epsilon'_{i-n}\right)$ *with* $\{\epsilon'_n; n \in Z\}$ *being an i.i.d. copy of* $\{\epsilon_n; n \in Z\}$. *Taking expectation on the inner product, we have*

$$\begin{aligned}
& \left\{E\left\langle Y_i - Y_{i,m}, Y_i - Y_{i,m}\right\rangle\right\}^{1/2} \\
= & \left\{\sum_{k=m}^{\infty}\sum_{n=m}^{\infty} \frac{1}{(n+1)^{3/2}(k+1)^{3/2}} E\int \left(\epsilon_{i-n} - \epsilon'_{i-n}\right)\left(\epsilon_{i-k} - \epsilon'_{i-k}\right) dt\right\}^{1/2} \\
= & C\sqrt{\left(\sum_{k=m}^{\infty} \frac{1}{(n+1)^3}\right)} \geq C\frac{1}{m},
\end{aligned}$$

for a positive constant C. *Noticing that* $\sum_{m=1}^{\infty}\left\{E\left\langle Y_i - Y_{im}, Y_i - Y_{im}\right\rangle\right\}^{1/2} \geq \sum_{m=1}^{\infty}\frac{1}{m} = \infty$, Y_i *is not of* L^2*-m-approximable dependence. However,* $Y_i - Y_{i,\{0\}} = \frac{1}{(i+1)^{3/2}}\left(\epsilon_0 - \epsilon'_0\right)$, *then* $\left\{E\left\langle Y_i - Y_{i,\{0\}}, Y_i - Y_{i,\{0\}}\right\rangle\right\}^{1/2} = \frac{C}{(i+1)^{3/2}}$, *which entails the physical dependence.*

3. Main results

For a standard process $\{\xi(t), t \in [0,1]\}$, one defines its mean function $m(t) = E\{\xi(t)\}$. A stationary functional time series $\{\xi_i(t), t \in [0,1]\}_{i=1}^n$ allows Karhunen-Loève L^2 representation

$$\xi_i(t) = m(t) + \sum_{k=1}^{\infty} \xi_{i,k}\phi_k(t), \tag{3.1}$$

where the random coefficients $\xi_{\cdot k}$ are of mean 0 and variance 1, and the functions $\phi_k = \sqrt{\lambda_k}\psi_k$, here sequences $\{\lambda_k\}_{k=1}^{\infty}$ and $\{\psi_k(t)\}_{k=1}^{\infty}$ are the eigenvalues and eigenfunctions of the covariance function $C(t,s) = cov\{\xi_1(t), \xi_1(s)\}$ respectively. Recalling (1.1), the data generating process is now written as

$$Y_{ij} = m(j/N) + \sum_{k=1}^{\infty} \xi_{i,k}\phi_k(j/N) + \sigma(j/N)\epsilon_{ij}.$$

Practically, we express a functional observation Y_{ij} as

$$Y_{ij} = m(j/N) + \sum_{k=1}^{\kappa} \xi_{i,k}\phi_k(j/N) + \sigma(j/N)\epsilon_{ij}, \tag{3.2}$$

where κ is assume to be a finite positive integer. The number κ of the basis functions impacts the performance of some procedures. For the data studied in this paper, we generally choose κ so that the plotted functional objects resemble original data with some smoothing that eliminates the most obvious noise. A practical selection of κ is given in Section 4.

3.1. Confidence bands

In this subsection, we present the confidence band result. In the first step, we estimate the mean function via spline smoothing. To describe the spline functions, we first introduce a sequence of equally spaced points $\{t_j\}_{j=1}^{N_p}$, called interior knots which divide the interval $[0,1]$ into (N_p+1) equal subintervals $I_j = [t_j, t_{j+1})$, $j = 0, \ldots, N_p - 1$, $I_{N_p} = [t_{N_p}, 1]$.

$$0 = t_0 < t_1 < \cdots < t_{N_p+1} = 1,$$

where $h_p = \frac{1}{N_p+1}$ is the distance between neighboring knots. Denote S_p the space of pth order spline space, i.e. S_p is spanned by B-spline basis $\{B_{j,p}\}$ of order p as described in [7].

Let the "infeasible estimator" of function m

$$\overline{m}(t) = \overline{\eta}(t) = n^{-1}\sum_{i=1}^{n} \eta_i(t)$$

with η_i defined in (3.1). The term "infeasible" refers to the fact that $\overline{m}(t)$ is computed from the unknown quantites $\eta_i(t)$, while $\overline{m}(t)$ would be the natural estimator of $m(t)$ if all the i.i.d. random curves $\eta_i(t)$, $1 \le i \le n$, were observed, a view taken in [10]. We propose to estimate the mean function $m(t)$ by

$$\hat{m}(t) = \underset{g(\cdot)\in S_p}{\operatorname{argmin}} \sum_{i=i}^{n}\sum_{j=1}^{N} \{Y_{ij} - g(j/N)\}^2, \tag{3.3}$$

in which S_p is a p-th degree spline space.

In i.i.d. functional samples, the sample covariance operator is used, but for functional time series this problem is more complicated with a dependence structure. For vector-valued time series, the variance of its sample mean is asymptotically approximated by the long-run variance $G(t,s)$, here

$$G(t,s) = E\eta_1(t)\eta_1(s) + \sum_{l=2}^{\infty} E\eta_1(t)\eta_l(s) + \sum_{l=2}^{\infty} E\eta_1(s)\eta_l(t). \tag{3.4}$$

It is proven in [14] that the series in (3.4) is convergent in L^2 under some weak dependence assumptions.

Denote by $\zeta(t)$ a standardized Gaussian process such that $E\zeta(t)\equiv 0$, $E\zeta^2(t)\equiv 1$ with the correlation function

$$E\zeta(t)\zeta(s) = G(t,s)\{G(t,t)G(s,s)\}^{-1/2}, t,s \in [0,1].$$

We denote by $Q_{1-\alpha}$ the $100(1-\alpha)^{th}$ percentile of the absolute maximum distribution of $\zeta(t)$, i.e., $P[\sup_{t\in[0,1]}|\zeta(t)| \le Q_{1-\alpha}] = 1-\alpha, \forall\alpha \in (0,1)$. Denote by $z_{1-\alpha/2}$ the $100(1-\alpha/2)^{th}$ percentile of a standard normal distribution.

For a continuous function ϕ on $[0,1]$, denote the modulus of continuity as $\omega(\phi,\delta) = \max_{t,s\in[0,1],|t-s|\le\delta}|\phi(t)-\phi(s)|$. For any $\beta \in (0,1]$, we denote by $C^{0,\beta}[0,1]$ the space of order β Hölder continuous function on $[0,1]$, i.e.,

$$C^{0,\beta}[0,1] = \left\{\phi : \|\phi\|_{0,\beta} = \sup_{t\neq s,t,s\in[0,1]} \frac{|\phi(t)-\phi(s)|}{|t-s|^{\beta}} < +\infty\right\},$$

in which $\|\phi\|_{0,\beta}$ is called the $C^{0,\beta}$-norm of ϕ. Clearly, $C^{0,\beta}[0,1] \subset C[0,1]$ and if $\phi \in C^{0,\beta}[0,1]$, then $\omega(\phi,\delta) \le \|\phi\|_{0,\beta}\delta^{\beta}$.

We impose the following technical assumptions:

(A1) *The mean function* $m \in C^{(p)}[0,1]$, class of functions with pth continuous derivatives.
(A2) *The standard deviation function* $\sigma(t) \in C^{0,\mu}[0,1]$ for $\mu \in [0,1]$.
(A3) *There exist* $C_C, C_G > 0$ *such that* $C(t,t) \ge C_C$ and $G(t,t) \ge C_G$, $t \in [0,1]$, *for* $1 \le k \le \kappa$, $\phi_k(t) \in C^{0,\mu}[0,1]$, *for* $\mu \in [0,1]$.
(A4) $N = O(n^{\theta})$ for some $\theta > 1/2p$. The number of interior knots satisfies $NN_p^{-1} \to \infty$, $N_p^{-p}n^{1/2} \to 0$ and $N^{-1/2}N_p^{1/2}\log n \to 0$ as $n \to \infty$.
(A5) *There exists* $\nu > 4+2\theta$ *such that* $E|\epsilon_{ij}|^{\nu} < \infty$, *for* $1 \le i < \infty$, $1 \le j < \infty$.

Assumptions (A1) and (A2) are typical for spline smoothing ([20]). Assumption (A3) ensures that the principle components have collectively bounded smoothness. Assumption (A4) concerns the number of observations for each subject and the number of knots of B-splines, which are needed to ensure the asymptotic result. Assumption (A5) provides the Gaussian approximation of estimation error process. Besides the above assumptions, we also assume regularity conditions for the physical dependence measure about functional time series Y, which will be stated in the Appendix.

We first present the asymptotic property of $\overline{m}(t)$, then we show that $\hat{m}(t)$ has the same asymptotic property as $\overline{m}(t)$.

Theorem 1. *Under Assumptions (A1)–(A5), for any* $\alpha \in (0,1)$*, as* $n \to \infty$*, the "infeasible estimator"* $\overline{m}(t)$ *converges at the* $\sqrt{n}$ *rate with*

$$P\left\{\sup_{t\in[0,1]} n^{1/2}\left|\overline{m}(t)-m(t)\right| G\left(t,t\right)^{-1/2} \leq Q_{1-\alpha}\right\} \to 1-\alpha,$$

and for any $t \in [0,1]$,

$$P\left\{n^{1/2}\left|\overline{m}(t)-m(t)\right| G\left(t,t\right)^{-1/2} \leq z_{1-\alpha/2}\right\} \to 1-\alpha.$$

The spline estimator $\hat{m}$ *is asymptotically equivalent to* $\overline{m}$*, that is*

$$\sup_{t\in[0,1]} n^{1/2}\left|\overline{m}(t)-\hat{m}(t)\right| = o_p\left(1\right).$$

The explicit expression of the confidence band for $m(t)$ is presented in the following corollary, which is a direct result of the theorem above.

Corollary 1. *Under Assumptions (A1)–(A5), for any* $\alpha \in (0,1)$*, as* $n \to \infty$*, an asymptotic* $100\left(1-\alpha\right)\%$ *exact confidence band for* $m(t)$ *is*

$$\hat{m}(t) \pm G\left(t,t\right)^{1/2} Q_{1-\alpha} n^{-1/2}, t \in [0,1],$$

while an asymptotic $100\left(1-\alpha\right)\%$ *pointwise confidence interval for* $m(t)$ *is*

$$\hat{m}(t) \pm G\left(t,t\right)^{1/2} z_{1-\alpha/2} n^{-1/2}.$$

4. Implementation

In this section, we show in detail the kernel bandwidth selection and how we choose the number of interior knots when we apply a spline smoothing technique. Given any data set $(j/n, Y_{ij})_{j=1,i=1}^{N,n}$ from model (3.2), we obtain the spline estimator $\hat{m}(t)$ through (3.3). According to Assumption (A4), the number of interior knots for estimating $m(t)$ is taken to be $N_p = c[n^{1/2p}\log(n)]$, where $[a]$ denotes the integer part of a. To select the number of knots N_p, we use BIC method as follows,

$$\mathrm{BIC}\left(\mathbf{N_p}\right) = \log\left(\frac{1}{n_{\mathrm{T}}}\sum_{i=1}^{n}\sum_{j=1}^{N}\widehat{r}_{ij}^{2}\right) + \frac{\log\left(n_{\mathrm{T}}\right)}{n_{\mathrm{T}}}(N_p+p),$$

here $\widehat{r}_{ij}$s are residuals and n_T is the total number of observations. We choose knots with the smallest BIC within the range of $[.5n^{1/2p}\log(n), 5n^{1/2p}\log(n)]$.

When constructing the confidence bands, one needs to evaluate the function $\sigma_n^2\left(t\right)$ by estimating the unknown functions $f\left(t\right)$ and $\sigma_Y^2\left(t\right)$ and then plugging in these estimators. The same approach from [20] will be taken for these estimators. Also, one needs to estimate the unknown functions $C\left(t,s\right)$, $G\left(t,s\right)$ and the quantile $Q_{1-\alpha}$.

For two curves with lag deviation i, the empirical covariance function is defined as

$$\hat{\gamma}_i(t,s) = \frac{1}{n}\sum_{j=i+1}^{n}\left(Y_j(t)-\bar{Y}_n(t)\right)\left(Y_{j-i}(s)-\bar{Y}_n(s)\right).$$

where $\bar{Y}_n(t) = 1/n \sum_{1 \le i \le n} Y_i(t)$. Then, the estimation of G is given by

$$\hat{G}_n(t,s) = \hat{\gamma}_0(t,s) + \sum_{i=1}^{n-1} K(i/h)\{\hat{\gamma}_i(t,s) + \hat{\gamma}_i(s,t)\}, \tag{4.1}$$

where $K(\cdot)$ is a kernel function and $h = h(n)$ is a smoothing bandwidth satisfying $h(n) \to \infty$ and $h(n)/n \to 0$. Under m-approximable dependence assumption, [14] showed that $\hat{G}$ is a consistent estimator of the long-run covariance function G in the sense of $\int\int\{\hat{G}_n(t,s) - G(t,s)\}^2 dtds \to 0$ in probability as n goes to infinity. Similarly, this consistency holds under the physical dependence.

We use the flat top kernel as suggested by [14].

$$K(t) = \begin{cases} 1 & 0 \le |t| < 0.1, \\ 1.1 - |t| & 0.1 \le |t| < 1.1, \\ 0 & 1.1 \le |t|. \end{cases}$$

Equation (4.1) provides a pilot estimator of the covariance function $G(\cdot,\cdot)$. The smoothed estimator of covariance function $G(t,s)$ is through

$$\hat{G}_{sm}(t,s) = \underset{g \in S_p \otimes S_p}{\operatorname{argmin}} \sum\nolimits_{j,j'=1}^{N_i} \{G_{\cdot jj'} - g(j/N, j'/N)\}^2, \tag{4.2}$$

where $G_{\cdot jj'} = \hat{G}(j/N, j'/N)$, $1 \le j, j' \le N_i$, and $S_p \otimes S_p$ is a tensor product spline space. We then estimate $C(t,s)$. A pilot estimator is by the sample covariance

$$\hat{C}(t,s) = \frac{1}{n}\sum_{i=1}^{n}(Y_i(t) - \bar{Y}_n(t))(Y_i(s) - \bar{Y}_n(s)).$$

Then we obtain $\hat{C}_{sm}(t,s)$ through bivariate spline smoothing similarly as in (4.2). Eigenfunction decomposition of $\hat{C}_{sm}(t,s)$ gives eigenvalues $\{\hat{\lambda}_1 \ge \hat{\lambda}_2 \ge \hat{\lambda}_3 \ge \ldots\}$ as well as eigenfunctions $\{\hat{\phi}_k\}$ correspondingly. Choose the number κ of eigenfunctions by using an empirical criterion, i.e. $\kappa = \operatorname{argmin}_{1 \le l \le T}\{\sum_{k=1}^{l} \hat{\lambda}_k / \sum_{k=1}^{T} \hat{\lambda}_k > .99\}$, where $\{\hat{\lambda}_k\}$ are the first T estimated positive eigenvalues. Finally, we simulate $\hat{\varsigma}_b(t) = \hat{G}_{sm}(t,t)^{-1/2} \sum_{k=1}^{\kappa} Z_{k,b}\hat{\phi}_k(t)$, where $Z_{k,b}$ are i.i.d. standard normal variables with $1 \le k \le \kappa$ and $1 \le b \le B$ with $B = 1000$ a preset large number. We take the maximum absolute value of each copy of $\hat{\varsigma}_b$ and estimate $Q_{1-\alpha}$ by the empirical quantile $\hat{Q}_{1-\alpha}$.

5. Simulation

We carry out a set of simulation studies to illustrate the finite sample behavior of the proposed confidence bands obtained in Section 3. We consider all combinations of subject sizes $n = 200, 500, 1000, 2000$ and within subject observation sizes $N = 50, 100, 200$; each pair of data-generated processes was replicated 1000 times. We generate data from an autoregressive (FAR(1)) process

$$y_i - m = \beta(y_{i-1} - m) + \sigma\epsilon_i,$$

TABLE 1
Coverage frequency for $m = \sin(x)$ *under Setting A from* 1000 *replications*

		σ=1			σ=0.5		
n	$1-\alpha$	$N=50$	$N=100$	$N=200$	$N=50$	$N=100$	$N=200$
200	0.90	0.859	0.839	0.841	0.823	0.809	0.841
	0.95	0.910	0.912	0.898	0.894	0.886	0.905
	0.99	0.969	0.972	0.960	0.962	0.963	0.968
500	0.90	0.867	0.872	0.869	0.873	0.873	0.872
	0.95	0.924	0.925	0.921	0.923	0.924	0.928
	0.99	0.985	0.973	0.985	0.977	0.981	0.982
1000	0.90	0.886	0.878	0.894	0.866	0.871	0.882
	0.95	0.938	0.926	0.946	0.926	0.921	0.934
	0.99	0.983	0.977	0.987	0.983	0.972	0.985

TABLE 2
Coverage frequency for $m = \sin(x)$ *under Setting B from* 1000 *replications*

		σ=1			σ=0.5		
n	$1-\alpha$	$N=50$	$N=100$	$N=200$	$N=50$	$N=100$	$N=200$
200	0.90	0.851	0.855	0.867	0.858	0.828	0.867
	0.95	0.898	0.907	0.924	0.912	0.879	0.910
	0.99	0.963	0.970	0.979	0.974	0.958	0.970
500	0.90	0.882	0.873	0.880	0.876	0.870	0.843
	0.95	0.934	0.931	0.933	0.926	0.922	0.904
	0.99	0.980	0.977	0.983	0.983	0.970	0.978
1000	0.90	0.881	0.882	0.881	0.876	0.869	0.885
	0.95	0.929	0.940	0.937	0.928	0.932	0.942
	0.99	0.978	0.988	0.985	0.988	0.986	0.989

here $\beta(\cdot)$ is a (bounded) kernel operator defined by $\beta(x)(t) = \int \beta(t,s)x(s)ds$. We consider the following two kernel settings for the dependence structure

$$\text{Setting A}: \quad \beta(t,s) = 16s(1-s)t(1-t),$$
$$\text{Setting B}: \quad \beta(t,s) = \frac{\exp\left\{-(t^2+s^2)/2\right\}}{4\int \exp\left(-x^2\right)dx}.$$

For the mean function $m(\cdot)$, we used a $C^{(\infty)}[0,1]$ function $m(t) = sin(2\pi t)$, for $0 \le t \le 1$. Suggested by a referee, we also considered a $C^{(3)}[0,1]$ mean function

$$m(t) = \begin{cases} -8(t-0.5)^4, & 0 \le t \le 0.5 \\ 8(t-0.5)^4 & 0.5 \le t \le 1. \end{cases} \tag{5.1}$$

We used $\epsilon_i(t) = B_i(t) + N_i\sqrt{1-t(1-t)}$, where $B_i(t)'s$ are i.i.d. standard Brownian bridges and $N_i's$ are i.i.d. standard Normals. Note that this gives $E(\epsilon_i^2(t)) = 1$ for all $0 \le t \le 1$, which is assumed by our estimation procedure. Different noise levels $\sigma = 0.5, 1$ were used to interpret the result. The first 200 trajectories were discarded to ensure stationarity.

We used cubic spline to estimate the mean functions in our simulation studies. For the long-run covariance functions estimation, we used the flat top kernel as described in section 4 with bandwidth $h = [n^{1/3}]$. This estimation was used for all data-generating processes. Tables 1, 2, 4 and 5 show the coverage frequencies from 1000 replications for the nominal confidence levels $1-\alpha = 0.90, 0.95$ and

TABLE 3
Maximum and averaged widths of confidence bands for $m = \sin(x)$ (Setting A)

			$\sigma = 1$			$\sigma = 0.5$		
n	$1-\alpha$		$N = 50$	$N = 100$	$N = 200$	$N = 50$	$N = 100$	$N = 200$
200	0.90	max	0.689	0.908	0.691	0.520	0.535	0.535
		ave	0.548	0.556	0.560	0.406	0.418	0.419
	0.95	max	0.784	0.792	0.794	0.593	0.610	0.609
		ave	0.631	0.639	0.634	0.463	0.476	0.476
500	0.90	max	0.450	0.454	0.455	0.339	0.343	0.351
		ave	0.362	0.367	0.368	0.265	0.270	0.275
	0.95	max	0.518	0.522	0.525	0.386	0.390	0.399
		ave	0.417	0.422	0.424	0.302	0.307	0.313
1000	0.90	max	0.321	0.326	0.328	0.245	0.249	0.250
		ave	0.259	0.264	0.265	0.192	0.196	0.197
	0.95	max	0.370	0.376	0.377	0.279	0.284	0.285
		ave	0.298	0.304	0.305	0.218	0.223	0.224

TABLE 4
Coverage frequency for $m(\cdot)$ in (5.1) under Setting A from 1000 replications

		σ=1			σ=0.5		
n	$1-\alpha$	$N = 50$	$N = 100$	$N = 200$	$N = 50$	$N = 100$	$N = 200$
200	0.90	0.908	0.892	0.909	0.884	0.908	0.875
	0.95	0.945	0.930	0.939	0.918	0.946	0.927
	0.99	0.984	0.973	0.978	0.963	0.975	0.969
500	0.90	0.935	0.921	0.932	0.885	0.903	0.873
	0.95	0.963	0.957	0.959	0.936	0.935	0.920
	0.99	0.989	0.989	0.990	0.975	0.975	0.975
1000	0.90	0.924	0.929	0.902	0.903	0.928	0.901
	0.95	0.951	0.962	0.947	0.954	0.964	0.943
	0.99	0.990	0.988	0.982	0.986	0.986	0.990

TABLE 5
Coverage frequency for $m(\cdot)$ in (5.1) under Setting B from 1000 replications

		σ=1			σ=0.5		
n	$1-\alpha$	$N = 50$	$N = 100$	$N = 200$	$N = 50$	$N = 100$	$N = 200$
200	0.90	0.907	0.914	0.899	0.918	0.918	0.908
	0.95	0.943	0.948	0.936	0.918	0.938	0.946
	0.99	0.976	0.986	0.983	0.980	0.986	0.980
500	0.90	0.928	0.918	0.942	0.897	0.907	0.889
	0.95	0.964	0.956	0.972	0.954	0.951	0.929
	0.99	0.989	0.989	0.989	0.989	0.981	0.970
1000	0.90	0.942	0.921	0.917	0.893	0.910	0.902
	0.95	0.965	0.958	0.951	0.944	0.959	0.953
	0.99	0.989	0.989	0.988	0.982	0.986	0.991

0.99 of two different mean functions under settings A and B respectively. The empirical frequency is measured as that the true curve $m(t)$, $0 \le t \le 1$, is covered by the spline confidence band at $N+1$ grid points $\{0/N, 1/N, \ldots, N/N\}$. Tables 3 and 6 provide both the maximum and averaged widths of the confidence bands over the grid points under different combinations.

The results of the simulation studies can be summarized as follows. The empirical coverage rates go to the nominal ones as the sample size increases in all cases. When the sample size is as large as $n = 1000$, the coverage percentages of the confidence bands are very close to the nominal confidence levels $0.90, 0.95$

TABLE 6
Maximum and averaged widths of confidence bands for $m(\cdot)$ in (5.1) (Setting A)

			$\sigma = 1$			$\sigma = 0.5$		
n	$1-\alpha$		$N = 50$	$N = 100$	$N = 200$	$N = 50$	$N = 100$	$N = 200$
200	0.90	max	0.692	0.694	0.701	0.349	0.351	0.346
		ave	0.546	0.552	0.558	0.276	0.278	0.275
	0.95	max	0.776	0.780	0.786	0.390	0.394	0.388
		ave	0.612	0.620	0.626	0.309	0.313	0.309
500	0.90	max	0.456	0.452	0.455	0.226	0.227	0.228
		ave	0.362	0.361	0.364	0.179	0.181	0.183
	0.95	max	0.512	0.508	0.511	0.253	0.254	0.256
		ave	0.406	0.405	0.409	0.201	0.202	0.205
1000	0.90	max	0.325	0.325	0.326	0.163	0.163	0.164
		ave	0.259	0.259	0.261	0.129	0.130	0.131
	0.95	max	0.365	0.364	0.366	0.182	0.183	0.183
		ave	0.291	0.291	0.293	0.145	0.146	0.147

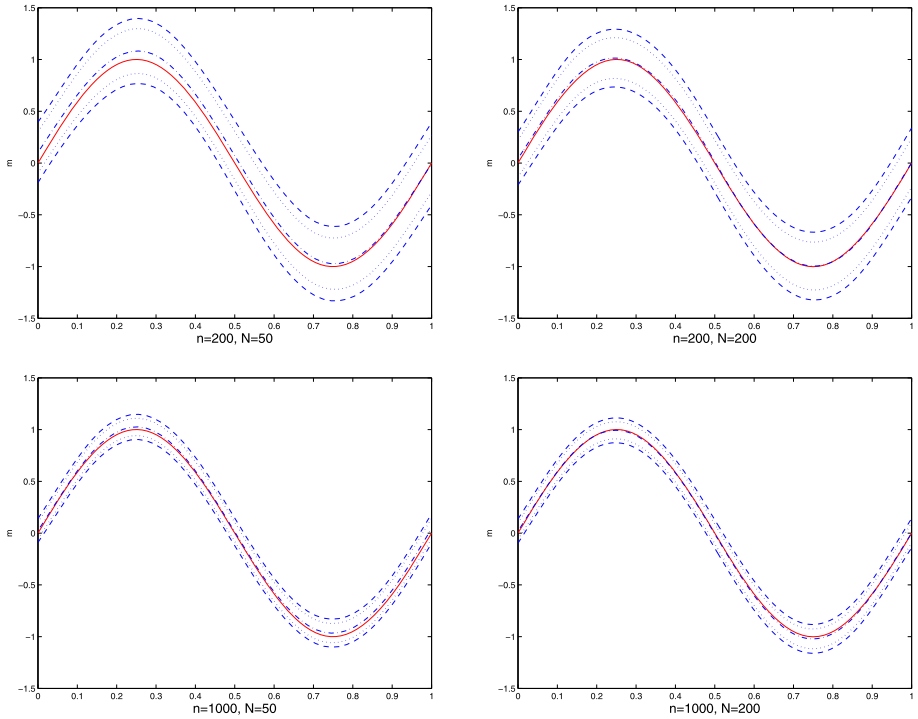

FIG 1. *Plots of* 90% *(dotted lines) and* 99% *(dashed lines) confidence bands as well as their estimators (dash-dotted lines) for combinations of* $n = 200, 1000$ *and* $N = 50, 200$ *for* $m(x) = \sin(x)$ *(Setting A,* $\sigma = 1$*). True function* $m(x)$ *is plotted in solid line in each plot.*

and 0.99. The coverage frequencies under different noise levels ($\sigma = 0.5$ and 1) are comparable for both settings. However, for higher noise levels, the confidence bands are wider. One can also see that both the maximum and averaged widths decrease as the sample size increases. Visualizations of such confidence bands are shown in Figures 1 and 2 for combinations of different subject sizes

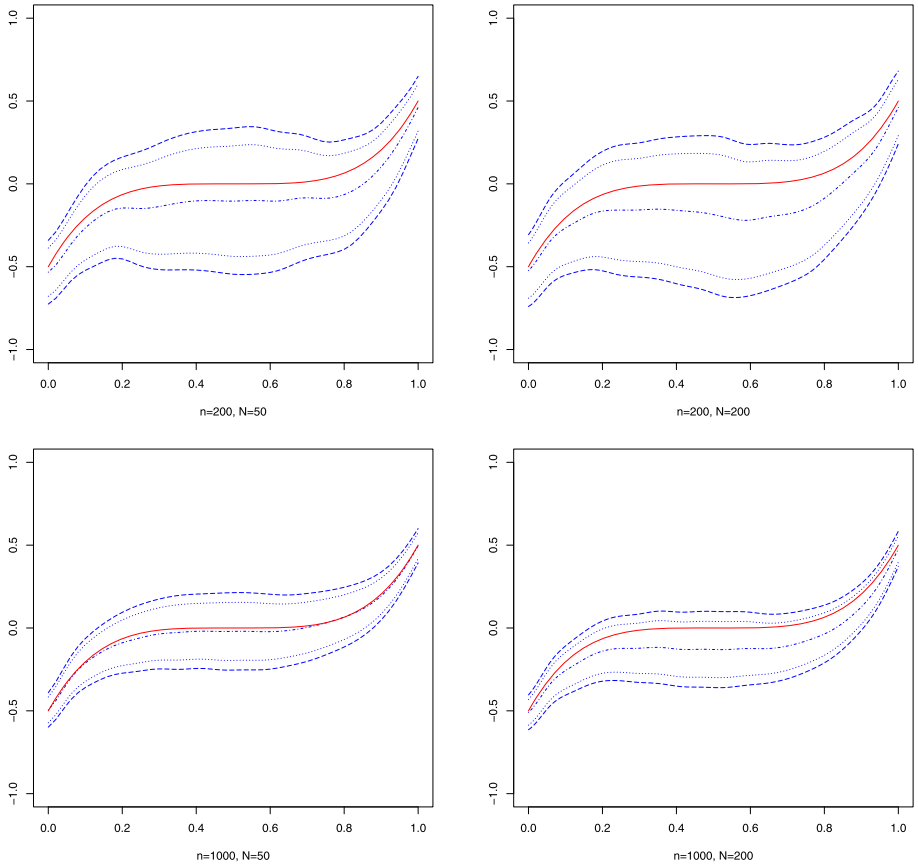

FIG 2. *Plots of* 90% *(dotted lines) and* 99% *(dashed lines) confidence bands as well as their estimators (dash-dotted lines) for combinations of* $n = 200, 1000$ *and* $N = 50, 200$ *for* $m(x)$ *in (5.1) (Setting A,* $\sigma = 1$*). True function* $m(x)$ *is plotted in solid line in each plot.*

$n = 200, 1000$ and observation sizes $N = 50, 200$ for both mean functions under kernel setting A with $\sigma = 1$.

6. Application

In this section, we apply the confidence band to S&P500 intraday returns data. Let $I_k(t)$ denote S&P500 index on the day k at the time t. Then $y_k(t)$ can be viewed as the log-return of the stock, $y_k(t) = \log I_k(t) - \log I_k(t-h)$, during period h, where h is typically 1 or 5 minutes. We used $h = 5$ for 5-minute returns, which was recorded throughout each trading day. The volatility of the stock is then represented by $\sigma_k^2(t) = Var(y_k(t)|\mathcal{F}_{k-1})$ with $\mathcal{F}_{k-1}$ being the information up to day $k-1$. The left panel in Figure 3 shows S&P500 indices for 10 consecutive days. It is natural to choose one trading day as the underlying time

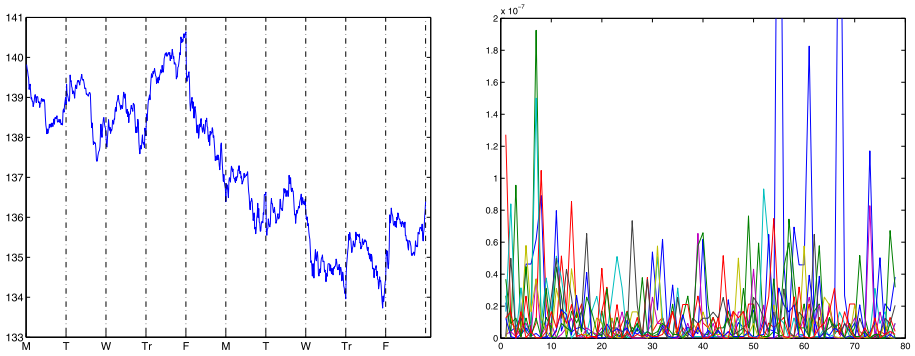

FIG 3. *Left: S&P500 index for 10 consecutive days. Right: functional trajectories of volatility for 10 days.*

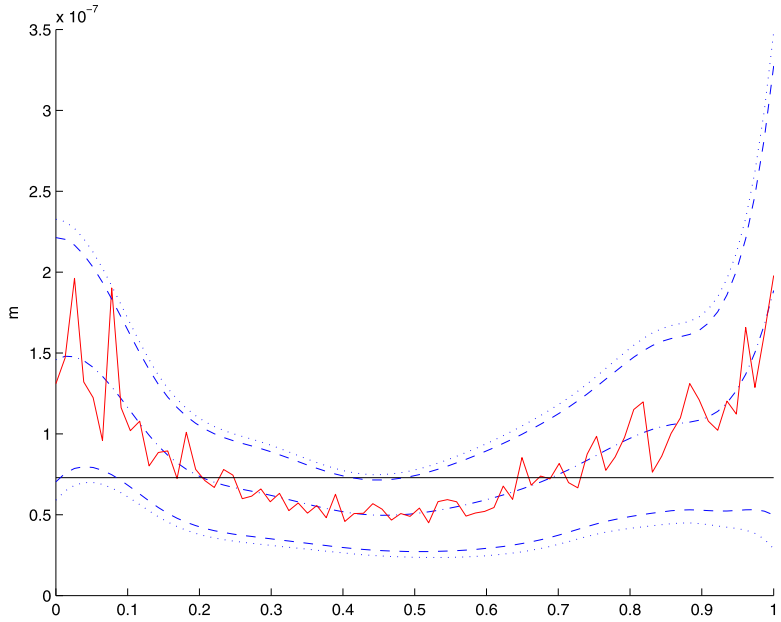

FIG 4. *Plot of daily volatility mean (solid line) and its spline regression (dash-dotted line), 90% confidence band (upper and lower dashed lines), 95% confidence band (upper and lower dotted lines) and a constant line (solid line).*

interval. We show, in the right panel of Figure 3, the corresponding trajectories of the 10 daily volatility curves. Every day, the volatility curve exhibits a certain pattern, typically with some upward or downward momentum. To quantify such behavior, we use the confidence band of functional time series. Our data record is from January 2006 to December 2011, containing 1511 trading days. For each day, there are 79 observations from 9:30 am to 4 pm.

We estimated the mean function of daily volatility curves using cubic spline, shown as dashed-dotted line in the middle of Figure 4. Notice from there that

the volatility, $\sigma_k^2(t)$, tends to be larger at the beginning and the end of a trading day. We, therefore, constructed 90% and 95% confidence bands (upper and lower dashed lines and dotted ones in Figure 4) for the mean daily volatility function. To test the hypothesis that the mean daily curve is constant, one would like to see if a constant line can be fitted in the confidence bands. We failed to do it under 90% confidence level since the maximum value of the lower bound curve is larger than the minimum value of the upper bound curve as shown in Figure 4. Therefore, we conclude non-constant mean daily volatility with .10 significance level. However, we cannot reject constant daily volatility using .05 significance level as also shown in Figure 4 (The constant horizontal line with value 7.3×10^{-8} goes through the 95% band completely). Higher volatility at the beginning and the end of a trading day has been observed by several authors (e.g. [12] and [8]). We confirm this phenomenon with statistical inference results.

7. Concluding remarks

In this article, we have constructed simultaneous confidence bands of mean of functional time series. It is not a simple extension of confidence bands for functional data analysis with independence assumption. We have rigorously established the consistency and asymptotic normality of the proposed estimators. The assumptions of asymptotics are based on a physical measure dependence structure. A widely used m-approximable dependence can be ensured in our dependence structure.

The current article has focused on fixed and dense design, and it can be used in financial data. For instance, the intraday return data are mostly obtained and recorded at fixed time points with high frequency. Scientific experiments may also be designed and recorded this way. However, in some applications, practitioners may look for random and/or sparsely designed functional data, and this is a future topic for our research.

As one referee pointed out, our results in Theorem 1 and Corollary 1 could be improved with the estimated $\hat{G}(t,t)$ instead of the theoretical $G(t,t)$. The distribution theory of mean function $\mu(t)$ with $\hat{G}(t,t)$ is much more challenging than the case with $G(t,t)$. At this stage, we don't even have a nice expression of $\hat{G}(t,t)$. However, we believe this would be an interesting future research topic.

Appendix

In this section, we first provide the proof of Lemma 1. Then we include in the subsections A.1 and A.2 the preliminaries and proof of Theorem 1 respectively. Throughout this section, C means a positive nonzero constant in a generic sense.

Proof of Lemma 1. Because of stationarity, $Y_i - Y_{i,m}$ and $Y_m - Y_{m,m}$ are of the same distribution. Notice that $Y_m - Y_{m,m}$ is identical to $Y_{m,\{0\}} - Y_{m,m}$. Thus,

$$\sum_{m\geq 1} v_p\left(Y_i - Y_{i,m}\right) = \sum_{m\geq 1} v_p\left(Y_{m,\{0\}} - Y_{m,m}\right).$$

Combining the triangle inequality and Minkowski inequality, we have

$$\begin{aligned}
& \textstyle\sum_{i\geq 1}\left(E\left\|Y_i - Y_{i,\{0\}}\right\|^p\right)^{1/p} \\
= \; & \textstyle\sum_{i\geq 1}\left(E\left\|Y_i - Y_{i,m} + Y_{i,m} - Y_{i,\{0\}}\right\|^p\right)^{1/p} \\
\leq \; & \textstyle\sum_{i\geq 1}\left\{E\left(\|Y_i - Y_{i,m}\| + \left\|Y_{i,m} - Y_{i,\{0\}}\right\|\right)^p\right\}^{1/p} \\
\leq \; & C\left(\textstyle\sum_{i\geq 1}\left(E\left\|Y_i - Y_{i,m}\right\|^p\right)^{1/p} + \sum_{i\geq 1}\left(E\left\|Y_{i,m} - Y_{i,\{0\}}\right\|^p\right)^{1/p}\right) \\
= \; & C\left(\textstyle\sum_{m\geq 1} v_p\left(Y_m - Y_{m,m}\right) + \sum_{m\geq 1} v_p\left(Y_{m,m} - Y_{m,\{0\}}\right)\right) \\
= \; & C\textstyle\sum_{m\geq 1} v_p\left(Y_i - Y_{i,m}\right) < \infty.
\end{aligned}$$

We get the desired result immediately. □

A.1. Preliminaries

First, we state a strong approximation lemma for a real valued stationary process $\{X_i\}_{i=1}^n$ with physical dependence, which is used in the proofs of Lemma A.3 and Theorem 1.

(C1) Let $2 < p \leq 4$, suppose that there exists a constant C satisfying $0 < C \leq 1/p$ such that for every $0 < \delta < C$ and every $\epsilon > 0$

$$\sum_{j=0}^{\infty} 2^{j(1-\delta)} P\Big(\sum_{i=1}^{2^{\delta j}} |X_i| \geq \epsilon 2^{j/p}\Big) < \infty.$$

(C2) Define $\theta_{i,p} = (E|X_i - X_{i,0}|^p)^{1/p}$ and $\Theta_{n,p} = \sum_{i=1}^n \theta_{i,p}$, then $\Theta_{n,p} = O(n^{-(p-2)/(2(4-p))-\tau})$ for some $\tau > 0$. Here $X_{i,0}$ is a coupled version of X_i.

Lemma A.1 (Theorem 2.1 of [15])**.** *Let $2 < p < 4$. Suppose that $\{X_i\}_{i=1}^n$ is stationary with $E(X_1) = 0, E(X_1^p) < \infty$, under conditions (C1) and (C2), then on a richer probability space, setting $S_n = \sum_{i=1}^n X_i$,*

$$|S_n - \sigma B(n)| = o_{a.s.}(n^{1/p}),$$

here $B(n)$ is a standard Brownian motion and $\sigma^2 = \sum_{i\in Z} E(X_0 X_i)$.

Based on Karhunen-Loève decomposition in (3.2), for $1 \leq k \leq \kappa$, consider the coefficient sequence $\{\xi_{i,k}\}_{i=1}^n$. Let the physical dependence measure $\theta_{\xi_k,i,p} = (E|\xi_{i,k} - \xi_{i,k}^*|^p)^{1/p}$, $1 \leq i \leq n$, where $\xi_{i,k}^*$ is the corresponding Karhunen-Loève coefficient for $\{Y_{i,0}\}_{i=1}^n$. Correspondingly, $\Theta_{\xi_k,n,p} = \sum_{i=1}^n \theta_{\xi_k,i,p}$. We show the conditions for Lemma A.1 are satisfied for $\{\xi_{i,k}\}_{i=1}^n$ if the conditions hold for $\{Y_i\}_{i=1}^n$.

Lemma A.2. *Conditions (C1) and (C2) hold for time series $\{\xi_{i,k}\}_{i=1}^n$ if they hold for functional series $\{Y_i\}_{i=1}^n$ in functional settings correspondingly.*

Proof. For $1 \le k \le \kappa$, $\|\xi_{i,k} - \xi^*_{i,k}\|_p = \|\langle Y_i, \phi_k\rangle - \langle Y_{i,0}, \phi_k\rangle\|_p$. Using Cauchy-Swartz inequality, we have

$$\begin{aligned} & \xi_{i,k} - \xi^*_{i,k} = \langle Y_i - Y_{i,0}, \phi_k\rangle \\ \le \;& \langle Y_i - Y_{i,0}, Y_i - Y_{i,0}\rangle^{1/2} \langle \phi_k, \phi_k\rangle^{1/2} = \lambda_k^{1/2} \|Y_i - Y_{i,0}\|_2 . \end{aligned}$$

Thus $\Theta_{\xi_k,n,p} = O(n^{-(p-2)/(2(4-p))-\tau})$ follows immediately under the assumption that $\Theta_{n,p} = O(n^{-(p-2)/(2(4-p))-\tau})$. Similarly, condition (C1) holds for $\{\xi_{i,k}\}_{i=1}^n$ if it holds for $\{Y_i\}_{i=1}^n$. □

In the following, we assume that the conditions in Lemma A.1 hold for $\{Y_i\}_{i=1}^n$.

Lemma A.3. *Under Assumptions (A2)–(A5), there exist i.i.d.* $N(0,1)$ *random variables* $Z_{ik,\xi}, Z_{ij,\epsilon}, 1 \le i \le n, 1 \le j \le N, 1 \le k \le \kappa$ *and* $2 < p < 4$, $\beta \in (0, 1/2)$ *such that as* $n \to \infty$

$$\max_{1\le k\le\kappa} \left|\overline{\xi}_{.,k} - \overline{Z}_{.,k,\xi}\right| = o_{a.s.}\left(n^{1/p-1}\right) \tag{A.1}$$

$$\max_{1\le j\le N} \left|\overline{\epsilon}_{.,j} - \overline{Z}_{.,j,\epsilon}\right| = O_{a.s.}\left(n^{\beta-1}\right) \tag{A.2}$$

in which $\overline{Z}_{.,k,\xi} = n^{-1}\sum_{i=1}^n Z_{ik,\xi}, \overline{Z}_{.,j,\epsilon} = n^{-1}\sum_{i=1}^n Z_{ij,\epsilon}, 1 \le j \le N, 1 \le k \le \kappa$.

Proof. Applying Lemmas A.1 and A.2, there exist $\overline{Z}_{.,k,\xi}$ to ensure $|\overline{\xi}_{.,k} - \overline{Z}_{.,k,\xi}| = o_{\text{a.s.}}(n^{1/p-1})$ for $1 \le k \le \kappa$. For independent ϵ_{ik}, as in [4], we have (A.2) under Assumption (A5). □

A.2. Proof of Theorem 1

With spline basis defined in section 3, denote

$$\mathbf{B} = \begin{pmatrix} B_{1-p}(1/N) & \cdots & B_{N_p}(1/N) \\ \vdots & \ddots & \vdots \\ B_{1-p}(N/N) & \cdots & B_{N_p}(N/N) \end{pmatrix}.$$

Then the solution of $\hat{m}(t)$ in (3.3) can be expressed as

$$\hat{m}(t) = \left\{B_{1-p}(t), \ldots, B_{N_p}(t)\right\} \left(\mathbf{B}^{\mathrm{T}}\mathbf{B}\right)^{-1} \mathbf{B}^{\mathrm{T}}\mathbf{Y},$$

in which $\mathbf{Y} = \left(\bar{Y}_{.,1}, \ldots, \bar{Y}_{.,N}\right)^T, \bar{Y}_{.,j} = n^{-1}\sum_{i=1}^n Y_{ij}, 1 \le j \le N$. To prove the theorem, we break the estimation error $\hat{m}(t) - m(t)$ by the representation of Y_{ij}, and we obtain the following crucial decomposition

$$\hat{m}(t) = \widetilde{m}(t) + \widetilde{e}(t) + \widetilde{\xi}(t), \tag{A.3}$$

with

$$\begin{aligned}
\widetilde{m}(t) &= \left\{B_{1-p,p}(t),\ldots,B_{N_p}(t)\right\}\left(\mathbf{B}^{\mathrm{T}}\mathbf{B}\right)^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{m} \\
\widetilde{e}(t) &= \left\{B_{1-p,p}(t),\ldots,B_{N_p}(t)\right\}\left(\mathbf{B}^{\mathrm{T}}\mathbf{B}\right)^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{e} \\
\widetilde{\xi}(t) &= \left\{B_{1-p,p}(t),\ldots,B_{N_p}(t)\right\}\left(\mathbf{B}^{\mathrm{T}}\mathbf{B}\right)^{-1}\mathbf{B}^{\mathrm{T}}\xi, \qquad \text{(A.4)}
\end{aligned}$$

in which $\mathbf{m}=(m(1/N),\ldots m(N/N))^{T}$ is the signal vector, $\mathbf{e}=(\sigma(1/N)\bar{\epsilon}_{\cdot,1},\ldots,\sigma(N/N)\bar{\epsilon}_{\cdot,N})^{T}$ is the noise vector with $\overline{\epsilon}_{\cdot,j}=n^{-1}\sum_{i=1}^{n}\epsilon_{ij}, 1\le j\le N$, and $\xi=\left(\sum_{k=1}^{\kappa}\bar{\xi}_{,k}\phi_k(1/N),\ldots,\sum_{k=1}^{\kappa}\bar{\xi}_{,k}\phi_k(N/N)\right)^{T}$ is the eigenfunction vector with $\bar{\xi}_{,k}=n^{-1}\sum_{i=1}^{n}\xi_{ik}, 1\le k\le\kappa$.

Now, $\hat{m}(t)-m(t)=\widetilde{m}(t)-m(t)+\widetilde{\xi}(t)+\widetilde{e}(t)$ with the bias term $\widetilde{m}(t)-m(t)+\widetilde{\xi}(t)$ and the noise term $\widetilde{e}(t)$. The next three lemmas concern $\widetilde{m}(t)$, $\widetilde{e}(t)$ and $\widetilde{\xi}(t)$ given in (A.4), and Theorem 1 follows immediately.

Lemma A.4. *Under Assumptions (A1) and (A4), as* $n\to\infty$,

$$\sup_{t\in[0,1]} n^{1/2}\left|\widetilde{m}(t)-m(t)\right| G(t,t)^{-1/2}=o_p(1).$$

Proof. By theorems of spline approximations in de Boor and Assumption (A1), $\left\|\widetilde{m}(t)-m(t)\right\|_{\infty}=O\left(h_p^{p}\right)$, thus as $n\to\infty$, $\sup_{t\in[0,1]} n^{1/2}|\widetilde{m}(t)-m(t)|G(t,t)^{-1/2}=O\left(h_p^{p}n^{1/2}\right)=o(1)$, here the last equation is satisfied with Assumption (A4). □

Denote next $\widetilde{\phi}_k=\left\{B_{1-p}(t),\ldots,B_{N_p}(t)\right\}\left(\mathbf{B}^{T}\mathbf{B}\right)^{-1}\mathbf{B}^{T}\phi_k$ and $\widetilde{Z}_k(t)=\overline{Z}_{\cdot,k,\xi}\widetilde{\phi}_k$, $k=1,\ldots,\kappa$, similar to the definitions of $\widetilde{m}(t)$ and $\widetilde{\xi}_k(t)$ in (A.4). Also denote $\widetilde{\zeta}_k(t)=\overline{Z}_{\cdot,k,\xi}\phi_k$, $k=1,\ldots,\kappa$ and define

$$\widetilde{\zeta}(t)=n^{1/2}G(t,t)^{-1/2}\sum_{k=1}^{\kappa}\widetilde{\zeta}_k(t) \qquad \text{(A.5)}$$

here $G(t,t)$ is the long-run variance.

Lemma A.5. *Under Assumptions (A2)–(A5), as* $n\to\infty$, *there is an approximate version* $\widetilde{\zeta}(t)$ *of* $\zeta(t)$ *such that*

$$\sup_{t\in[0,1]}\left|\widetilde{\zeta}(t)-n^{1/2}G(t,t)^{-1/2}\widetilde{\xi}(t)\right|=o_p(1)$$

hence for any $\alpha\in(0,1)$

$$\begin{aligned}
P\left\{\sup_{t\in[0,1]} n^{1/2}\left|\widetilde{\xi}(t)\right| G(t,t)^{-1/2}\le Q_{1-\alpha}\right\} &\to 1-\alpha, \\
P\left\{\sup_{t\in[0,1]} n^{1/2}\left|\overline{m}(t)-m(t)\right| G(t,t)^{-1/2}\le Q_{1-\alpha}\right\} &\to 1-\alpha.
\end{aligned}$$

Proof. First note the fact that $\overline{Z}_{\cdot,k,\xi}$ are independent $N\left(0,n^{-1}\right)$ variables implies that $\max_{1\le k\le\kappa}\left|\overline{Z}_{\cdot,k,\xi}\right|=O_p\left(n^{-1/2}\right)$. Similarly as in [4], we have, under Assumption (A3),

$$\max_{1\le k\le\kappa}\left\|\phi_k-\widetilde{\phi}_k\right\|_{\infty}\le C_g\left(\max_{1\le k\le\kappa}\|\phi_k\|_{0,\mu}\right)h_p^{\mu}.$$

The definition of $\widetilde{\zeta}(t)$ in (A.5), together with the definition of $\overline{m}(t)$ in (3.1), the strong approximation in (A.1) and the above bound on $\max_{1\leq k\leq\kappa}\left|\overline{Z}_{.,k,\xi}\right|$ entail that

$$\begin{aligned}
&\sup_{t\in[0,1]}\left|\overline{m}(t)-m(t)-\widetilde{\xi}(t)\right| \\
&\leq \kappa\max_{1\leq k\leq\kappa}\left|\overline{\xi}_{,k}\right|\sup_{t\in[0,1]}\left\|\phi_k(t)-\widetilde{\phi}_k(t)\right\|_\infty \leq C\kappa\max_{1\leq k\leq\kappa}\left(\left|\overline{Z}_{.,k,\xi}\right|+\left|\overline{\xi}_{,k}-\overline{Z}_{.,k,\xi}\right|\right)h_p^\mu \\
&= O_p\left(n^{-1/2}h_p^\mu+n^{1/p-1}h_p^\mu\right)=o_p\left(n^{-1/2}\right) \\
&\sup_{t\in[0,1]}\left|\overline{m}(t)-m(t)-n^{-1/2}G(t,t)^{1/2}\widetilde{\zeta}(t)\right| \\
&\leq \max_{1\leq k\leq\kappa}\left|\overline{\xi}_{,k}-\overline{Z}_{.,k,\xi}\right|\sup_{t\in[0,1]}\left\|\phi_k(t)\right\|_\infty = O_p\left(n^{1/p-1}\right)=o_p\left(n^{-1/2}\right).
\end{aligned}$$

Now for any $t\in[0,1]$, $\widetilde{\zeta}(t)$ is Gaussian with $E\widetilde{\zeta}(t)\equiv 0$, and the covariance $E\widetilde{\zeta}(t)\widetilde{\zeta}(s)$ equal to

$$\begin{aligned}
&n^{1/2}G(t,t)^{-1/2}n^{1/2}G(s,s)^{-1/2}\,cov\left[\left\{\sum\nolimits_{k=1}^{\kappa}\overline{Z}_{.,k,\xi}\phi_k(t)\right\},\left\{\sum\nolimits_{k=1}^{\kappa}\overline{Z}_{.,k,\xi}\phi_k(s)\right\}\right] \\
&\approx n^{1/2}G(t,t)^{-1/2}n^{1/2}G(s,s)^{-1/2}\,cov\left[\left\{\sum\nolimits_{k=1}^{\kappa}\overline{\xi}_{.,k,\xi}\phi_k(t)\right\},\left\{\sum\nolimits_{k=1}^{\kappa}\overline{\xi}_{.,k,\xi}\phi_k(s)\right\}\right] \\
&= G(t,t)^{-1/2}n^{1/2}G(s,s)^{-1/2}G(t,s),\forall t,s\in[0,1],
\end{aligned}$$

so $\mathcal{L}\{\widetilde{\zeta}(t),t\in[0,1]\}=\mathcal{L}\{\zeta(t),t\in[0,1]\}$. Lemma A.5 is proved. □

Lemma A.6. *Under Assumptions (A2)–(A5), as* $n\to\infty$,

$$\sup_{t\in[0,1]}n^{1/2}\left|\widetilde{e}(t)\right|G(t,t)^{-1/2}=o_p(1).$$

Proof. Denote that

$$\widetilde{\mathbf{Z}}_\epsilon(t)=\left\{B_{1-p}(t),\ldots,B_{N_p}(t)\right\}\left(\mathbf{B}^{\mathrm{T}}\mathbf{B}\right)^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{Z},$$

where $\mathbf{Z}=\left(\sigma(1/N)\overline{Z}_{.,1,\epsilon},\ldots,\sigma(N/N)\overline{Z}_{.,N,\epsilon}\right)^{\mathrm{T}}$. According to Lemma A.3 and Assumption (A5) , $\|\mathbf{Z}-\mathbf{e}\|_\infty=O_{\text{a.s.}}\left(n^{\beta-1}\right)$, while

$$\begin{aligned}
&\left\|N^{-1}\mathbf{B}^{\mathrm{T}}(\mathbf{Z}-\mathbf{e})\right\|_\infty \leq \|\mathbf{Z}-\mathbf{e}\|_\infty\max_{1-p\leq J\leq N_p}\langle B_J,1\rangle_{2,N} \\
\leq\ & \|\mathbf{Z}-\mathbf{e}\|_\infty\max_{1-p\leq J\leq N_p}\#\{j:B_J(j/N)>0\}N^{-1}\leq C\|\mathbf{Z}-\mathbf{e}\|_\infty h_p.
\end{aligned}$$

Since for any fixed $t\in[0,1]$, all values $B_{1-p}(t),\ldots,B_{N_p}(t)\in[0,1]$ and at most p of the which are positive

$$\begin{aligned}
\left\|\widetilde{\mathbf{Z}}_\epsilon(t)-\widetilde{e}(t)\right\|_\infty &= \left\|\left\{B_{1-p}(t),\ldots,B_{N_p}(t)\right\}\left(\mathbf{B}^{\mathrm{T}}\mathbf{B}\right)^{-1}\mathbf{B}^{\mathrm{T}}(\mathbf{Z}-\mathbf{e})\right\|_\infty \\
&\leq C\|\mathbf{Z}-\mathbf{e}\|_\infty=O_{\text{a.s.}}\left(n^{\beta-1}\right).
\end{aligned}$$

Note next that the random vector $(\mathbf{B}^{\mathrm{T}}\mathbf{B})^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{Z}$ is (N_p+p)-dimensional normal with the covariance matrix bounded above by

$$\max_{t\in[0,1]}\sigma^2(t)\,n^{-1}(\mathbf{B}^{\mathrm{T}}\mathbf{B})^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{B}\left(\mathbf{B}^{\mathrm{T}}\mathbf{B}\right)^{-1}\leq Cn^{-1}(\mathbf{B}^{\mathrm{T}}\mathbf{B})^{-1}\leq Cn^{-1},$$

bounding the tail probabilities of entries of $(\mathbf{B}^{\mathrm{T}}\mathbf{B})^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{Z}$ and applying Borel-Cantelli Lemma leads to

$$\begin{aligned}\left\|(\mathbf{B}^{\mathrm{T}}\mathbf{B})^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{Z}\right\|_{\infty} &= O_{\text{a.s.}}\left(N^{-1/2}n^{-1/2}h_p^{-1/2}\log^{1/2}(N_p+p)\right)\\ &= O_{\text{a.s.}}\left(N^{-1/2}n^{-1/2}h_p^{-1/2}\log n\right).\end{aligned}$$

Thus, $\sup_{t\in[0,1]}|n^{1/2}\widetilde{\mathbf{Z}}_{\epsilon}(t)\,G(t,t)^{-1/2}| = O_{\text{a.s.}}(N^{-1/2}h_p^{-1/2}\log n)$. Under Assumption (A4), $\sup_{t\in[0,1]}|n^{1/2}\widetilde{\mathbf{e}}(t)\,G(t,t)^{-1/2}| = O_{\text{a.s.}}(n^{\beta-1/2}+\frac{\log n}{(Nh_p)^{1/2}}) = o_{\text{a.s.}}(1)$. Lemma A.6 is thus proved. □

References

[1] Aue, A., Norinho, D. D., and Hörmann, S. (2012). On the prediction of functional time series. arXiv:1208.2892.

[2] Berkes, I., Liu, W., and Wu, W. B. (2014). Komlós-Major-Tusnády approximation under dependence. *The Annals of Probability*, 42(2):794–817. MR3178474

[3] Bosq, D. (2000). *Linear Processes in Function Spaces: Theory and Applications*, volume 149. Springer. MR1783138

[4] Cao, G., Yang, L., and Todem, D. (2012). Simultaneous inference for the mean function based on dense functional data. *Journal of Nonparametric Statistics*, 24(2):359–377. MR2921141

[5] Castro, P. E., Lawton, W. H., and Sylvestre, E. A. (1986). Principal modes of variation for processes with continuous sample curves. *Technometrics*, 28(4):329–337.

[6] Claeskens, G. and Keilegom, I. v. (2003). Bootstrap confidence bands for regression curves and their derivatives. *The Annals of Statistics*, 31(6):1852–1884. MR2036392

[7] de Boor, C. (2001). *A Practical Guide to Splines*. Springer-Verlag, Berlin. MR1900298

[8] Evans, K. P. and Speight, A. E. (2010). Intraday periodicity, calendar and announcement effects in euro exchange rate volatility. *Research in International Business and Finance*, 24(1):82–101.

[9] Fan, J. and Zhang, W. (2000). Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scandinavian Journal of Statistics*, 27(4):715–731. MR1804172

[10] Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. Springer. MR2229687

[11] Hall, P. and Titterington, D. (1988). On confidence bands in nonparametric density estimation and regression. *Journal of Multivariate Analysis*, 27(1):228–254. MR0971184

[12] Hörmann, S., Horváth, L., and Reeder, R. (2013). A functional version of the arch model. *Econometric Theory*, 29:267–288. MR3042756

[13] Hörmann, S. and Kokoszka, P. (2010). Weakly dependent functional data. *The Annals of Statistics*, 38(3):1845–1884. MR2662361

[14] Horváth, L., Kokoszka, P., and Reeder, R. (2013). Estimation of the mean of functional time series and a two-sample problem. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(1):103–122. MR3008273

[15] Liu, W. and Lin, Z. (2009). Strong approximation for a class of stationary processes. *Stochastic Processes and Their Applications*, 119(1):249–280. MR2485027

[16] Ma, S., Yang, L., and Carroll, R. J. (2012). A simultaneous confidence band for sparse longitudinal regression. *Statistica Sinica*, 22:95–122. MR2933169

[17] Müller, H.-G., Stadtmüller, U., and Yao, F. (2006). Functional variance processes. *Journal of the American Statistical Association*, 101(475):1007–1018. MR2324140

[18] Ramsay, J. O. and Silverman, B. (2005). *Functional Data Analysis*. Springer, 2nd edition. MR2168993

[19] Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53(1):233–243. MR1094283

[20] Wang, J. and Yang, L. (2009). Polynomial spline confidence bands for regression curves. *Statistica Sinica*, 19(1):325–342. MR2487893

[21] Yao, F. (2007). Functional principal component analysis for longitudinal and survival data. *Statistica Sinica*, 17(3):965–983. MR2408647

[22] Yao, F., Müller, H.-G., and Wang, J.-L. (2005a). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470):577–590. MR2160561

[23] Yao, F., Müller, H.-G., and Wang, J.-L. (2005b). Functional linear regression analysis for longitudinal data. *The Annals of Statistics*, 33(6):2873–2903. MR2253106

[24] Zhao, Z. and Wu, W. B. (2008). Confidence bands in nonparametric time series regression. *The Annals of Statistics*, 36(4):1854–1878. MR2435458

[25] Zhou, S., Shen, X., and Wolfe, D. (1998). Local asymptotics for regression splines and confidence regions. *The Annals of Statistics*, 26(5):1760–1782. MR1673277