Journal of the Royal Statistical Society
SERIES B Statistical Methodology

J. R. Statist. Soc. B (2013)
75, *Part* 1, *pp.* 103–122

Estimation of the mean of functional time series and a two-sample problem

Lajos Horváth,

University of Utah, Salt Lake City, USA

Piotr Kokoszka

Colorado State University, Fort Collins, USA

and Ron Reeder

University of Utah, Salt Lake City, USA

[Received January 2011. Final revision January 2012]

Summary. The paper is concerned with inference based on the mean function of a functional time series. We develop a normal approximation for the functional sample mean and then focus on the estimation of the asymptotic variance kernel. Using these results, we develop and asymptotically justify testing procedures for the equality of means in two functional samples exhibiting temporal dependence. Evaluated by means of a simulation study and application to a real data set, these two-sample procedures enjoy good size and power in finite samples.

Keywords: Eurodollar futures; Functional data analysis; Long-run variance; Time series; Two-sample problem

1. Introduction

Functional time series form a class of data structures which occurs in many applications, but several important aspects of estimation and testing for such data have not received as much attention as for functional data derived from randomized experiments. In the latter case, the curves can often be assumed to form a simple random sample; in particular, the functional observations are independent. This paper focuses on the methodology and theory for the estimation of the mean function of a functional time series, and on inference for the mean of two functional time series. Despite their central importance, these issues have not yet been studied. The contribution of this paper is thus twofold:

(a) we develop a methodology and an asymptotic theory for the estimation of the variance of the sample mean of temporally dependent curves under model-free assumptions;
(b) we propose procedures for testing equality of two mean functions in functional samples exhibiting temporal dependence.

A central issue in the analysis of functional time series is to take into account the temporal dependence of the observations. Bosq (2000) studied the theory of linear functional time series, focusing on the functional auto-regressive (FAR) model. For many functional time series it is,

Address for correspondence: Piotr Kokoszka, Department of Statistics, Colorado State University, Fort Collins, CO 80523-1877, USA.
E-mail: Piotr.Kokoszka@colostate.edu

© 2012 Royal Statistical Society 1369–7412/13/75103

however, not clear what specific model they follow, and for many statistical procedures it is not necessary to assume a specific model. In this paper, we assume that the functional time series is stationary, but we do not impose any specific model on it. We assume that the curves are dependent in a very broad sense, which is made precise in Section 1.1. The dependence condition that we use is satisfied by many examples of functional time series, including the linear and auto-regressive conditional heteroscedasticity type of processes; see Hörmann and Kokoszka (2010) and Aue *et al.* (2012).

A direct motivation for the research that is presented in this paper comes from a two-sample problem in which we wish to test whether the mean functions of two functional time series are equal. A specific problem, which is studied in greater detail in Section 4, is to test whether the mean curves of certain financial assets are equal over certain periods. This in turn allows us to conclude whether the expectations of future market conditions are the same or different at specific time periods. In general, if the same mean is assumed for the whole time series, whereas, in fact, it is different for disjoint segments, the inference or exploratory analysis that follows will be faulty, as all prediction and model fitting procedures for functional time series start with subtracting the sample mean, which is viewed as an estimate of the unique population mean function. The same holds true for independent curves; if two subsamples have different mean functions, subtracting the sample mean function based on the whole data set will lead to spurious results. Despite the importance of two-sample problems for functional data, they have received little attention. Recent papers of Horváth *et al.* (2009) and Panaretos *et al.* (2010) are the only contributions to a two-sample problem in a functional setting which develop inferential methodology. Horváth *et al.* (2009) compared linear operators in two functional regression models. Panaretos *et al.* (2010) focused on testing the equality of the covariance operators in two samples of independent and identically distributed (IID) Gaussian functional observations; our paper focuses on the means of dependent (and non-Gaussian) observations. We develop the methodology required, justify it by asymptotic arguments and describe its practical implementation.

Any inference involving mean functions requires estimates of the variability of the sample mean. In IID functional samples, the sample covariance operator is used, but for functional time series this problem is more complicated. For scalar and vector-valued time series, the variance of the sample mean is asymptotically approximated by the long-run variance. Convergence of various estimators of the long-run variance has been established under several types of assumption, including broad model specifications (e.g. linear processes), cumulant conditions and various mixing conditions. Hörmann and Kokoszka (2010) and Gabrys *et al.* (2010) advocated use of the notion of L^p m-approximability for functional time series, as this condition is intuitively appealing and is easy to verify for functional time series models. We therefore develop a general framework for the estimation of the long-run covariance kernel in this setting.

The long-run covariance kernel corresponds to the asymptotic variance in a normal approximation for the sample mean of a scalar time series, but no central limit theorem for a general functional time series has been established yet (results for linear processes were established in Bosq (2000)). We provide such a generally applicable result as well (theorem 1).

The remainder of the paper is organized as follows. We conclude the introduction by defining the notion of dependence for functional time series which we use throughout this paper. Then, in Section 2, we state the asymptotic results for the mean of a single functional time series, with proofs developed in Appendix A. Section 3 focuses on the problem of testing the equality of means of two functional samples which exhibit temporal dependence. In Section 4, we evaluate the finite sample performance of the procedures that are proposed in Section 3 by means of a simulation study and application to real data.

Throughout the paper, the integral sign without the limits of integration refers to the integration over the unit interval $[0, 1]$.

1.1. Approximable functional time series

We consider a stationary functional time series $\{\varepsilon_i(t), i \in \mathbb{Z}, t \in [0, 1]\}$, which we can view as an error sequence in a more complex functional model, e.g. a regression model, as in Gabrys *et al.* (2010). We assume that these errors are non-linear moving averages (Bernoulli shifts) $\varepsilon_i = f(\delta_i, \delta_{i-1}, \ldots)$, for some measurable function $f : S^\infty \to L^2$, and IID elements δ_i of a measurable space S. In all models that are used in practice $S = L^2$. To motivate the construction below, it is useful to write the ε_i as

$$\varepsilon_i = f(\delta_i, \ldots, \delta_{i-m+1}, \delta_{i-m}, \delta_{i-m-1} \ldots). \tag{1.1}$$

Under equation (1.1), the sequence $\{\varepsilon_i\}$ is stationary and ergodic. The function f must decay sufficiently fast to ensure that the sequence $\{\varepsilon_i\}$ is weakly dependent. The weak dependence condition is stated in terms of an approximation by m-dependent sequences, namely, we require that

$$\sum_{m \geqslant 1} \left(E\left[\int \{\varepsilon_i(t) - \varepsilon_{i,m}(t)\}^2 \, \mathrm{d}t \right] \right)^{1/2} < \infty, \tag{1.2}$$

where

$$\varepsilon_{i,m} = f(\delta_i, \ldots, \delta_{i-m+1}, \delta^{(m)}_{i,i-m}, \delta^{(m)}_{i,i-m-1}, \ldots), \tag{1.3}$$

with the sequences $\{\delta^{(m)}_{i,k}\}$ being independent copies of the sequence $\{\delta_i\}$. Note that the sum in condition (1.2) does not depend on i.

The idea behind the above construction is that the function f decays so fast that the effect of the innovations δ_i far back in the past becomes negligible; they can be replaced by different, fully independent innovations. If the ε_i follow a linear model $\varepsilon_i = \Sigma_{j \geqslant 0} c_j(\delta_{i-j})$, condition (1.2) intuitively means that the approximations by the finite moving averages $\varepsilon_{i,m} = \Sigma_{0 \leqslant j \leqslant m} c_j(\delta_{i-j})$ become increasingly precise. This means that the operators c_j must decay sufficiently fast in an appropriate operator norm. The conditions that are stated in this section are very similar to those used by Hörmann and Kokoszka (2010) and Aue *et al.* (2011), who also present examples of non-linear functional time series satisfying condition (1.2).

2. Normal approximation and long-run variance for functional time series

The purpose of this section is to develop the central limit theorem for the sample mean of a functional time series and an estimation technique for the long-run covariance kernel. Both are needed to develop procedures for comparing the mean functions and can be used in other contexts.

We assume that $\{\varepsilon_i\}$ is an L^2 m-approximable (and hence stationary) functional time series satisfying

$$E(\varepsilon_0) = 0, \tag{2.1}$$

in L^2, and

$$\int E\{\varepsilon_0^2(t)\} \, \mathrm{d}t < \infty. \tag{2.2}$$

Theorem 1. If conditions (1.1), (1.2), (2.1) and (2.2) hold, then

$$N^{-1/2}\sum_{i=1}^{N}\varepsilon_i \xrightarrow{\mathrm{d}} Z, \tag{2.3}$$

in L^2, where Z is a Gaussian process with

$$E\{Z(t)\}=0,$$
$$E\{Z(t)\,Z(s)\}=c(t,s);$$
$$c(t,s)=E\{\varepsilon_0(t)\,\varepsilon_0(s)\}+\sum_{i\geqslant 1}E\{\varepsilon_0(t)\,\varepsilon_i(s)\}+\sum_{i\geqslant 1}E\{\varepsilon_0(s)\,\varepsilon_i(t)\}. \tag{2.4}$$

The infinite sums in the definition of the kernel c converge in $L^2([0,1]\times[0,1])$, i.e. c is a square integrable function on the unit square.

Theorem 1 is proved in Appendix A.1.

The kernel c is defined analogously to the long-run variance of a scalar time series. It is directly related to the covariance operator of the sample mean defined by

$$\hat{C}_N(x)=N\,E\left(\left\langle\frac{1}{N}\sum_{i=1}^{N}\varepsilon_i,x\right\rangle\frac{1}{N}\sum_{j=1}^{N}\varepsilon_j\right)=\frac{1}{N}\sum_{i,j=1}^{N}E(\langle\varepsilon_i,x\rangle\varepsilon_j).$$

If the ε_i are independent, then $\hat{C}_N(x)=N^{-1}\Sigma_{i=1}^{N}E(\langle\varepsilon_i,x\rangle\varepsilon_i)$ becomes the usual sample (empirical) covariance operator, which plays a central role in many exploratory and inferential tools of functional data analysis of IID functional observations, mostly through the empirical functional principal components defined as its eigenfunctions. For any stationary functional time series $\{\varepsilon_i\}$,

$$\hat{C}_N(x)(t)=\int\left[\frac{1}{N}\sum_{i,j=1}^{N}E\{\varepsilon_i(s)\,\varepsilon_j(t)\}\right]x(s)\,\mathrm{d}s=\int c_N(t,s)\,x(s)\,\mathrm{d}s,$$

where

$$c_N(t,s)=\sum_{|k|<N}\left(1-\frac{|k|}{N}\right)E\{\varepsilon_0(s)\,\varepsilon_k(t)\}. \tag{2.5}$$

The summands in equation (2.5) converge to those in equation (2.4), but the estimation of the long-run covariance kernel c is far from trivial.

To enhance the applicability of our result, we state it for the case of a non-zero-mean function, which is estimated by the sample mean. We thus assume that

$$X_i(t)=\mu(t)+\varepsilon_i(t),\qquad 1\leqslant i\leqslant N, \tag{2.6}$$

with the series $\{\varepsilon_i\}$ satisfying the assumptions of theorem 1.

Let K be a kernel (weight) function defined on the line and satisfying the conditions

$$K(0)=1, \tag{2.7}$$

$$K \text{ is continuous}, \tag{2.8}$$

$$K \text{ is bounded}, \tag{2.9}$$

$$K(u)=0,\qquad \text{if } |u|>c, \text{ for some } c>0. \tag{2.10}$$

Condition (2.10) is assumed only to simplify the proofs; a sufficiently fast decay could be assumed instead.

Next we define the empirical (sample) correlation functions

$$\hat{\gamma}_i(t,s) = \frac{1}{N}\sum_{j=i+1}^{N}\{X_j(t)-\bar{X}_N(t)\}\{X_{j-i}(s)-\bar{X}_N(s)\}, \qquad 0 \leqslant i \leqslant N-1, \tag{2.11}$$

where $\bar{X}_N(t) = (1/N)\Sigma_{1\leqslant i\leqslant N} X_i(t)$. The estimator for c is given by

$$\hat{c}_N(t,s) = \hat{\gamma}_0(t,s) + \sum_{i=1}^{N-1} K\left(\frac{i}{h}\right)\{\hat{\gamma}_i(t,s)+\hat{\gamma}_i(s,t)\} \tag{2.12}$$

where $h = h(N)$ is the smoothing bandwidth satisfying

$$h(N)\to\infty, \qquad h(N)/N \to 0, \qquad \text{as } N\to\infty. \tag{2.13}$$

In addition to condition (1.2), we also assume that

$$\lim_{m\to\infty} m\left(E\left[\int\{\varepsilon_n(t)-\varepsilon_{n,m}(t)\}^2\,\mathrm{d}t\right]\right)^{1/2} = 0. \tag{2.14}$$

Condition (2.14) is very natural because if, in condition (1.2),

$$\left(E\left[\int\{\varepsilon_i(t)-\varepsilon_{i,m}(t)\}^2\,\mathrm{d}t\right]\right)^{1/2} = O(m^{-\alpha}),$$

then we need $\alpha > 1$, and for such α condition (2.14) holds. Condition (2.14) is satisfied by all the examples that were considered by Hörmann and Kokoszka (2010) and Aue *et al.* (2011).

Theorem 2. Suppose that the functional time series $\{X_i\}$ follows model (2.6). Under conditions (1.1), (1.2), (2.1), (2.2), (2.7)–(2.10), (2.13) and (2.14),

$$\int\int\{\hat{c}_N(t,s)-c(t,s)\}^2\,\mathrm{d}t\,\mathrm{d}s \xrightarrow{\mathrm{P}} 0, \tag{2.15}$$

with $c(t,s)$ defined by equation (2.4) and $\hat{c}_N(t,s)$ by equation (2.12).

Theorem 2 is proved in Appendix A.2. We now use the results of this section in the problem of testing the equality of means in two functional samples.

3. Testing the equality of mean functions

We consider two samples of curves, $X_1, X_2, \ldots, X_N$ and $X_1^*, X_2^*, \ldots X_M^*$, satisfying the location models

$$\begin{aligned} X_i(t) &= \mu(t)+\varepsilon_i(t), \\ X_j^*(t) &= \mu^*(t)+\varepsilon_j^*(t). \end{aligned} \tag{3.1}$$

The error functions ε_i are assumed to satisfy the conditions that were stated in Sections 1 and 2. The functions ε_j^* satisfy exactly the same conditions and their long-run covariance kernel $c^*(t,s)$ is defined analogously.

We assume that

$$\{\varepsilon_i, 1\leqslant i\leqslant N\} \quad \text{and} \quad \{\varepsilon_j^*, 1\leqslant j\leqslant M\} \quad \text{are independent.} \tag{3.2}$$

We are interested in testing

$$H_0\colon\ \mu = \mu^* \tag{3.3}$$

against the alternative

$$H_{\mathrm{A}}\colon\ \mu \neq \mu^*. \tag{3.4}$$

The equality in hypothesis (3.3) is in the space $L^2 = L^2([0,1])$, i.e. $\mu = \mu^*$ means that $\int\{\mu(t) - \mu^*(t)\}^2\,\mathrm{d}t = 0$, and the alternative means that $\int\{\mu(t) - \mu^*(t)\}^2\,\mathrm{d}t > 0$.

Since the statistical inference is about the mean functions of the observations, our procedures are based on the sample mean curves

$$\bar{X}_N(t) = \frac{1}{N}\sum_{1\leqslant i\leqslant N} X_i(t),$$

$$\bar{X}^*_M = \frac{1}{M}\sum_{1\leqslant i\leqslant M} X_j(t).$$

The sample means $\bar{X}_N$ and $\bar{X}^*_M$ are unbiased estimators of μ and μ^* respectively, so H_0 will be rejected if

$$U_{N,M} = \frac{NM}{N+M}\int\{\bar{X}_N(t) - \bar{X}^*_M(t)\}^2\,\mathrm{d}t$$

is large.

Before introducing the test procedures, we state two results which describe the asymptotic behaviour of the statistic $U_{N,M}$ under hypotheses H_0 and H_A. They motivate and explain the development that follows. The proofs of all results of this section follow from the theorems of Section 2. Details are presented in Reeder (2011).

Theorem 3. Suppose that hypothesis H_0, the assumptions of theorem 1 (and analogous assumptions for the ε^*_j) and assumption (3.2) hold. If

$$\frac{N}{N+M} \to \theta, \qquad \text{for some } 0\leqslant\theta\leqslant 1, \text{ as } \min(M,N)\to\infty, \tag{3.5}$$

then

$$U_{N,M} \xrightarrow{\mathrm{d}} \int \Gamma^2(t)\,\mathrm{d}t,$$

where $\{\Gamma(t), 0\leqslant t\leqslant 1\}$ is a mean 0 Gaussian process with covariances

$$E\{\Gamma(t)\Gamma(s)\} = d(t,s) := (1-\theta)\,c(t,s) + \theta\,c^*(t,s).$$

Theorem 4. If hypothesis H_A and the remaining assumptions of theorem 3 hold, then

$$\frac{N+M}{NM}U_{N,M} \xrightarrow{\mathrm{P}} \int\{\mu(t) - \mu^*(t)\}^2\,\mathrm{d}t.$$

In particular, if $0<\theta<1$, then $U_{N,M}\to^{\mathrm{P}}\infty$.

The kernel $d(t,s)$ in theorem 3 defines a covariance operator D. The eigenvalues of D are non-negative and are denoted by $\lambda_1 \geqslant \lambda_2 \geqslant \ldots$. By the Karhunen–Loève expansion, we have

$$\int \Gamma^2(t)\,\mathrm{d}t = \sum_{i=1}^{\infty}\lambda_i N_i^2, \tag{3.6}$$

where $\{N_i, 1\leqslant i<\infty\}$ are independent standard normal random variables.

Since the eigenvalues λ_i are unknown, the right-hand side of equation (3.6) cannot be used directly to simulate the distribution of $\int\Gamma^2(t)\,\mathrm{d}t$. We shall now explain how to estimate the λ_is.

Suppose that $\hat{D}_{N,M}$ is an L^2-consistent estimator of D, i.e.

$$\int\int\{\hat{d}_{N,M}(t,s) - d(t,s)\}^2\,\mathrm{d}t\,\mathrm{d}s \xrightarrow{\mathrm{P}} 0, \qquad \text{as } \min(M,N)\to\infty. \tag{3.7}$$

We discuss the construction of estimators $\hat{D}_{N,M}$ that satisfy condition (3.7) below. For the estimators that we propose, relationship (3.7) holds regardless of whether H_0 or H_A holds; nor do they depend on μ or μ^*.

Let

$$\hat{\lambda}_1 = \hat{\lambda}_1(N, M) \geqslant \hat{\lambda}_2 = \hat{\lambda}_2(N, M) \geqslant \dots$$

denote the eigenvalues of $\hat{D}_{N,M}$, i.e.

$$\int \hat{d}_{N,M}(t, s)\,\hat{\varphi}_i(s)\,\mathrm{d}s = \hat{\lambda}_i \hat{\varphi}_i(t), \tag{3.8}$$

where the $\hat{\varphi}_i(t) = \hat{\varphi}_i(t; N, M)$ are the corresponding eigenfunctions satisfying $\int \hat{\varphi}_i^2(t)\,\mathrm{d}t = 1$. Choosing p so large that $\Sigma_{i=1}^{p}\hat{\lambda}_i$ is a large percentage of $\Sigma_{i=1}^{N+M}\hat{\lambda}_i$, in light of equation (3.6), we can approximate the distribution of $\int \Gamma^2(t)\,\mathrm{d}t$ by that of $\Sigma_{i=1}^{p}\hat{\lambda}_i N_i^2$.

The statistical inference is based on the difference $\bar{X}_N - \bar{X}_M^*$. Observe that

$$\frac{MN}{M+N} E[\{\bar{X}_N(t) - \bar{X}_M^*(t)\}\{\bar{X}_N(s) - \bar{X}_M^*(s)\}] \to d(t, s), \qquad \text{as } \min(M, N) \to \infty,$$

i.e. d is the asymptotic covariance kernel of the difference $\bar{X}_N - \bar{X}_M^*$. We therefore use projections onto the eigenfunctions $\varphi_1, \varphi_2, \dots, \varphi_p$ that are associated with the p largest eigenvalues of D.

The choice of a projection space might be different if we knew the direction of the alternative. For example, if we knew that $\mu - \mu^* = g$, we would use just one projection in the direction of g. In the absence of such information, we project on $\varphi_1, \varphi_2, \dots, \varphi_p$. As we discuss in what follows, in certain situations the test may have little power.

Without any loss of generality, we assume that the $\varphi_1, \varphi_2, \dots, \varphi_p$ form an orthonormal system (the φ_i are orthogonal under condition (3.13), so only a normalization to unit norm is required). We define the projections

$$a_i = \langle \bar{X}_N - \bar{X}_M^*, \varphi_i \rangle, \qquad 1 \leqslant i \leqslant p, \tag{3.9}$$

and the vectors

$$\mathbf{a} = (a_1, a_2, \dots, a_p)^{\mathrm{T}}.$$

We show in the proof of theorem 5 that

$$\left(\frac{MN}{M+N}\right)^{1/2} \mathbf{a} \xrightarrow{\mathrm{d}} \mathbf{N}_p(\mathbf{0}, \mathbf{Q}), \tag{3.10}$$

where $N_p(\mathbf{0}, \mathbf{Q})$ stands for the p-variate normal random vector with mean $\mathbf{0}$ and covariance matrix $\mathbf{Q} = \mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$. Since the operator D is unknown, we cannot compute the φ_i. However, any estimator for D satisfying condition (3.7) can be used to find estimates for the φ_i. Let $\hat{\varphi}_i$ be the empirical eigenfunctions that are defined by expression (3.8), and set

$$\hat{a}_i = \langle \bar{X}_N - \bar{X}_M^*, \hat{\varphi}_i \rangle, \qquad 1 \leqslant i \leqslant p.$$

The limit relation (3.10) suggests the following statistics:

$$U_{N,M}^{(1)} = \frac{MN}{M+N} \sum_{i=1}^{p} \hat{a}_i{}^2 \tag{3.11}$$

and

$$U_{N,M}^{(2)} = \frac{MN}{M+N} \sum_{i=1}^{p} \frac{\hat{a}_i{}^2}{\hat{\lambda}_i}. \tag{3.12}$$

The following theorem establishes the limits of $U^{(1)}_{N,M}$ and $U^{(2)}_{N,M}$ under hypothesis H_0.

Theorem 5. Suppose that hypothesis H_0, the remaining assumptions of theorem 3, condition (3.7) and

$$\lambda_1 > \lambda_2 > \ldots > \lambda_p > \lambda_{p+1} \tag{3.13}$$

hold. Then

$$U^{(1)}_{N,M} \xrightarrow{\mathrm{d}} \sum_{i=1}^{p} \lambda_i N_i^2, \tag{3.14}$$

where $N_1, N_2, \ldots, N_p$ are independent standard normal random variables, and

$$U^{(2)}_{N,M} \xrightarrow{\mathrm{d}} \chi^2(p), \tag{3.15}$$

where $\chi^2(p)$ stands for a χ^2-distributed random variable with p degrees of freedom.

We note that $U^{(1)}_{N,M}$ is essentially the first p terms in the Karhunen–Loève expansion of the integral in the definition of $U_{N,M}$. Thus, the limit in condition (3.14) is exactly the random variable that we used to approximate the distribution of $U_{N,M}$. The limit in condition (3.15) is distribution free.

The consistency of the tests based on $U^{(1)}_{N,M}$ and $U^{(2)}_{N,M}$ follows from the following result.

Theorem 6. Suppose that hypothesis H_A, the remaining assumptions of theorem 3 and conditions (3.7) and (3.13) hold. Then

$$\frac{N+M}{NM} U^{(1)}_{N,M} \xrightarrow{\mathrm{P}} \sum_{i=1}^{p} \langle \mu - \mu^*, \varphi_i \rangle^2$$

and

$$\frac{N+M}{NM} U^{(2)}_{N,M} \xrightarrow{\mathrm{P}} \sum_{i=1}^{p} \frac{\langle \mu - \mu^*, \varphi_i \rangle^2}{\lambda_i}.$$

In particular, if $0 < \theta < 1$ in condition (3.5), then $U^{(1)}_{N,M} \to^{\mathrm{P}} \infty$ and $U^{(2)}_{N,M} \to^{\mathrm{P}} \infty$, provided that $\langle \mu - \mu^*, \varphi_i \rangle \neq 0$ for at least one $1 \leqslant i \leqslant p$.

We see that the condition for the consistency is that $\mu - \mu^*$ is not orthogonal to the linear subspace of L^2 spanned by the eigenfunctions $\varphi_i, 1 \leqslant i \leqslant p$.

The implementation of the tests based on theorems 5 and 6 depends on the existence of an estimator of the kernel $d(t,s)$ which satisfies condition (3.7). The remainder of this section is dedicated to this issue.

The estimation of D is very simple if the ε_i are IID, and the ε^*_j are IID. In this case, setting

$$\hat{\theta} = \frac{N}{N+M},$$

we can use

$$\hat{d}_{N,M}(t,s) = (1-\hat{\theta})\, \hat{c}_N(t,s) + \hat{\theta}\, \hat{c}^*_M(t,s), \tag{3.16}$$

where

$$\hat{c}_N(t,s) = \frac{1}{N} \sum_{i=1}^{N} \{X_i(t) - \bar{X}_N(t)\}\{X_i(s) - \bar{X}_N(s)\},$$

$$\hat{c}^*_M(t,s) = \frac{1}{M} \sum_{j=1}^{M} \{X^*_j(t) - \bar{X}^*_M(t)\}\{X^*_j(s) - \bar{X}^*_N(s)\}.$$

By condition (2.2), we can use the weak law of large numbers in a Hilbert space to establish condition (3.7). The estimation of D is much more difficult if only condition (1.2) is assumed, and its asymptotic justification relies on theorem 2. Recall the definition of the estimator $\hat{c}_N(t,s)$ that is given in equation (2.12), and define the estimator $\hat{c}^*_M(t,s)$ fully analogously. Our estimator for $d(t,s)$ is then equation (3.16) with $\hat{c}_N(t,s)$ and $\hat{c}^*_M(t,s)$ so defined. The following result then follows directly from theorem 2.

Theorem 7. Suppose that the functional time series $\{X_i\}$ satisfies the assumptions of theorem 2, and the series $\{X^*_j\}$ satisfies the same assumptions stated in terms of the ε^*_j. If condition (3.2) holds, then condition (3.7) holds.

We emphasize that under the conditions of theorem 7 relation (3.7) holds under both H_0 and H_A.

The $\hat{a}_i$ and the $\hat{\lambda}_i$ require the computation of the eigenfunctions and the eigenvalues of the operator $\hat{D}$. Except in the case of independent observations in each of the two samples, these quantities cannot be computed by using existing software because $\hat{D}$ is not an empirical covariance operator of a functional IID sample. We have developed an algorithmic procedure to compute $\hat{a}_i$ and $\hat{\lambda}_i$ which for brevity we do not present here; it is described in Reeder (2011).

To complete the description of the test procedures, we must specify how the value of p in the definitions of $U^{(1)}_{N,M}$ and $U^{(2)}_{N,M}$ is selected. This issue has been extensively studied in one-sample problems, and several approaches have been put forward, including cross-validation and penalty criteria; see for example Müller and Stadtmüller (2005) and Yao *et al.* (2005). In our experience (for smooth densely recorded curves), the simple cumulative variance method that was advocated by Ramsay and Silverman (2005) has been satisfactory. We therefore recommend to use p such that the first p empirical functional principal components in each sample explain about 85% of the variance. Scree plots of eigenvalues also offer valuable guidance, and it is useful to look at the P-values for a range of ps. We emphasize, however, that the issue of selecting the optimal p in two-sample problems is important and deserves a more careful study, which for brevity is not pursued in this paper. In particular, extensions of cross-validation and penalty approaches to two-sample problems have not been investigated. An approach that we have found useful is to apply a statistical procedure for functional data by using several values of p; typically not more than five will be reasonable. It often happens that the conclusions of a test are the same, no matter what value of p is taken. This approach was applied, for example, in Gromenko *et al.* (2012). It can, however, happen (see for example the data example in Kokoszka *et al.* (2008)) that the conclusions depend on the number of the functional principal components that are used to perform the test. A safe solution, which is recommended in a science problem, is to say that the results of the test are not conclusive, but such problems do emphasize the need for more sophisticated approaches to selecting p.

4. Simulation study and data examples

We begin by presenting the results of a simulation study that is intended to evaluate the empirical size and power of the testing procedures that were introduced in Section 3. We then illustrate their properties on a data example.

4.1. Simulation study

In this section, we compare the performance of the tests based on statistics $U^{(1)}_{N,M}$ and $U^{(2)}_{N,M}$ by using simulated Gaussian functional data. We consider all combinations of sample sizes $N, M = 50, 100, 200$, and each pair of data-generated processes was replicated 3000 times. To

investigate the empirical size, without loss of generality, we set $\mu(t) = \mu^*(t) = 0$. Under the alternative, we set $\mu(t) = 0$ and $\mu^*(t) = at(1-t)$. The power is then a function of the parameter a. We considered two settings for the errors:

(a) both the $\varepsilon_i(t)$ and the $\varepsilon_j^*(t)$ are IID Brownian bridges;
(b) both the $\varepsilon_i(t)$ and the $\varepsilon_j^*(t)$ are FAR(1) processes with the kernel

$$\psi(t,s) = \frac{\exp\{-(t^2+s^2)/2\}}{4\int \exp(-x^2)\,\mathrm{d}x},$$

i.e. the error terms $\varepsilon_i(t)$ follow the model

$$\varepsilon_i(t) = \int \psi(t,s)\,\varepsilon_{i-1}(s)\,\mathrm{d}s + B_i(t),$$

where $B_i(t)$ are IID Brownian bridges.

We calculated the test statistics $U_{N,M}^{(1)}$ and $U_{N,M}^{(2)}$ as explained in Section 3. These statistics depend on the choice of the weight functions K and K^*, and the bandwidth functions h and h^*. Much attention has been devoted over several decades to the optimal selection of these functions for scalar and vector-valued time series, and we cannot address this issue within the limits of this paper. We follow the recommendation of Politis and Romano (1996) and use, for both samples, the flat top kernel

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

with $h = N^{1/3}$ and $h^* = M^{1/3}$. We emphasize that this full estimation procedure was used for all data-generating processes, including those with independent errors.

The results of the simulation study can be summarized as follows. The empirical size is slightly larger than nominal and is larger in the case of FAR(1) errors. When a increases to 0.2 or larger, the empirical power of the test is smaller in the case of FAR(1) errors. Thus increasing the dependence in the error terms increases the size and decreases the power of the test. In both cases the tests have a slightly larger-than-nominal size and very good power. These observations are illustrated in Tables 1 and 2. On the basis of the whole simulation study, we can conclude that the performance of both tests is better if the sample sizes N and M are about equal. For example, for $N = M = 100$, the empirical sizes are closer to the nominal sizes than in the case $N = 100$ and $M = 200$ that is shown in Tables 1 and 2. The power is very high even for small sample sizes. This is illustrated in Fig. 1 which shows the samples with $N = M = 50$ and with slightly different means ($a = 0.8$). Visual inspection does not readily lead to the conclusion that the samples in Figs 1(a) and 1(b) have different means, yet our tests can detect it with a very high probability. Both Table 1 and Table 2 use $p = 3$, which explains about 85% of the variance. The rejection rates are very similar for $p = 4$ and $p = 5$. Neither of the two test statistics clearly dominates the other for the simulated Gaussian data, but a difference in behaviour can be seen when the tests are applied to real data sets, to which we now turn.

4.2. Eurodollar futures contracts

Our example uses financial data that were kindly made available by Vladislav Kargin. These data are used as an example of modelling with the functional AR(1) process in Kargin and Onatski (2008). The curves, one curve per day, are constructed from the prices of eurodollar futures contracts with decreasing expiration dates. The seller of a eurodollar futures contract

Table 1. Power of the test using $U^{(1)}_{100,200}$ and $U^{(2)}_{100,200}$ with Brownian bridge errors

a	*Powers (%) for the following values of* α*:*					
	$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.10$	
	$U^{(1)}_{100,200}$	$U^{(2)}_{100,200}$	$U^{(1)}_{100,200}$	$U^{(2)}_{100,200}$	$U^{(1)}_{100,200}$	$U^{(2)}_{100,200}$
0.0	1.5	1.5	6.3	6.2	11.4	11.6
0.1	2.5	3	7.4	8.0	13.2	13.6
0.2	6.0	4.4	16.7	13.0	24.8	20.2
0.3	14.2	9.2	30.4	23.2	41.3	33.0
0.4	26.4	17.0	48.5	36.1	60.8	48.0
0.5	44.0	31.4	64.7	53.5	75.2	64.3
0.6	59.4	45.9	80.3	68.0	87.9	78.4
0.7	78.0	64.7	91.8	82.4	96.0	89.0
0.8	88.0	78.0	95.9	90.9	98.0	94.6
0.9	94.2	88.1	98.7	95.8	99.4	97.9
1.0	98.0	94.8	99.5	98.5	99.9	99.3
1.1	99.6	98.4	100.0	99.8	100.0	100.0
1.2	99.9	99.4	100.0	99.9	100.0	100.0
1.3	100.0	99.8	100.0	100.0	100.0	100.0

Table 2. Power of the test using $U^{(1)}_{100,200}$ and $U^{(2)}_{100,200}$ with FAR(1) errors

a	*Powers (%) for the following values of* α*:*					
	$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.10$	
	$U^{(1)}_{100,200}$	$U^{(2)}_{100,200}$	$U^{(1)}_{100,200}$	$U^{(2)}_{100,200}$	$U^{(1)}_{100,200}$	$U^{(2)}_{100,200}$
0.0	1.8	1.9	6.6	7.2	12.2	13.5
0.1	2.4	2.2	7.9	7.7	13.5	14.5
0.2	5.1	3.3	13.6	11.6	21.6	18.7
0.3	9.8	6.3	23.6	17.6	34.5	26.8
0.4	19.4	12.3	35.9	26.5	46.7	36.3
0.5	26.8	19.5	47.9	38.6	60.4	49.7
0.6	42.1	29.6	62.2	51.8	73.1	62.5
0.7	56.4	42.8	75.4	63.8	83.2	74.0
0.8	68.6	53.8	85.7	74.6	91.5	83.1
0.9	80.8	67.6	92.7	85.9	96.4	91.9
1.0	87.4	78.7	95.9	90.8	98.1	94.5
1.1	93.7	86.8	97.9	95.8	99.1	97.6
1.2	97.6	93.7	99.5	98.1	99.8	99.2
1.3	98.5	96.4	99.7	98.9	99.9	99.6

takes on an obligation to deliver a deposit of 1 million US dollars to a bank account outside the USA i months from the current date. The price that the buyer is willing to pay for this contract depends on the prevailing interest rate. These contracts are traded on the Chicago Mercantile Exchange and provide a way to lock in an interest rate. Eurodollar futures are a liquid asset and are responsive to the Federal Reserve policy, inflation and economic indicators.

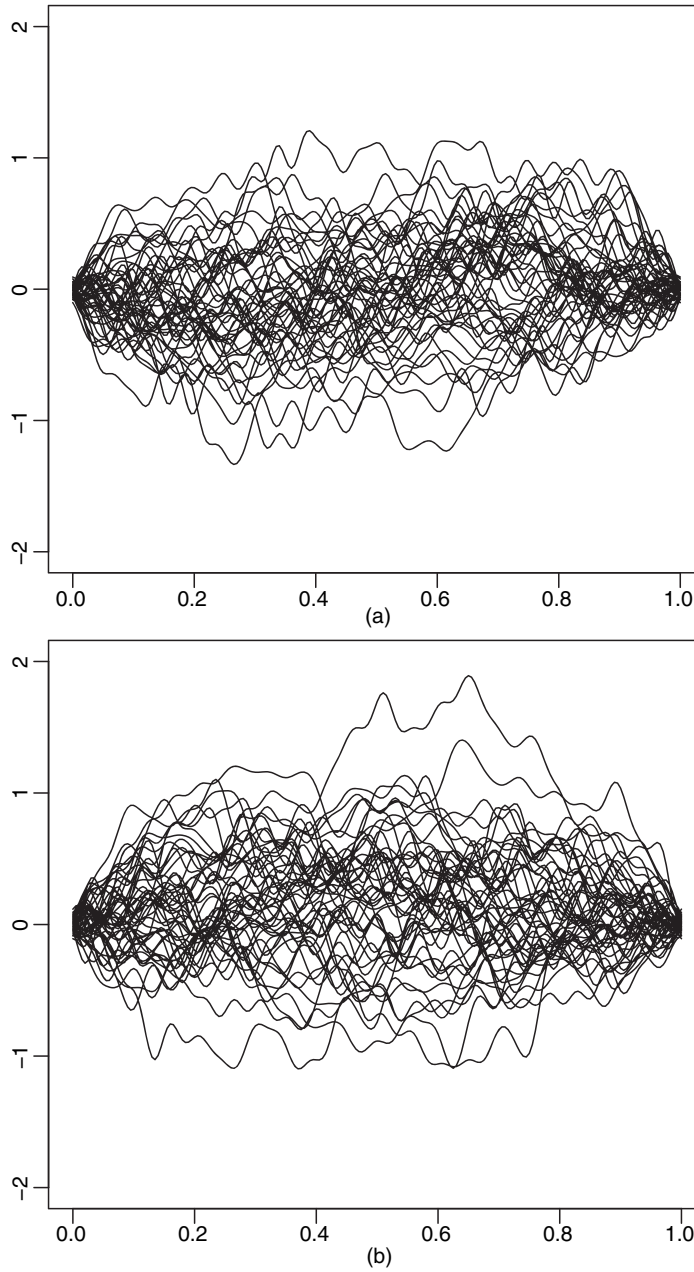

Fig. 1. (a) 50 trajectories of the Brownian bridge and (b) 50 independent trajectories of the Brownian bridge plus $\mu^*(t) = 0.8t(1-t)$: the tests can detect the different means with probability close to 90%

The data that we study consist of 114 points per day; point i corresponds to the price of a contract with closing date i months from the current date. We consider four samples, each consisting of 100 days of these data:

(a) *sample 1*, curves from September 7th, 1999, to January 27th, 2000;
(b) *sample 2*, curves from January 24th, 1997, to June 17th, 1997;

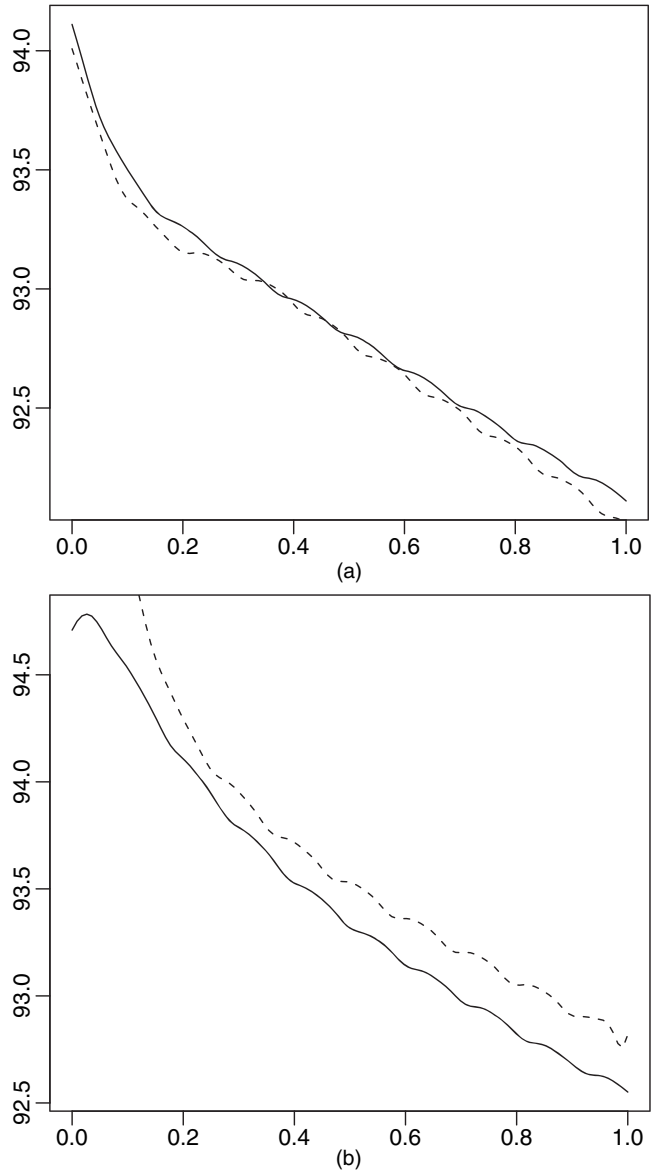

Fig. 2. Sample means of the eurodollar curves: (a) sample 1 (———) and sample 2 (- - - -); (b) sample 3 (———) and sample 4 (- - - -)

(c) *sample 3*, curves from December 4th, 1995, to April 24th, 1996;
(d) *sample 4*, curves from March 6th, 2001, to July 26th, 2001.

Fig. 2 shows the sample mean functions for the four samples. If a significance test does not reject hypothesis H_0, we can conclude that the expectations of the future evolution of interest rates are the same for the two periods over which the samples were taken. A rejection means that these expectations are significantly different. As the analysis below reveals, we can conclude that expectations of future interest rates were different in spring 1996 from those in summer 2001.

Table 3. *P*-values of the statistics applied to the eurodollar data for samples 1–4

Statistic	*P-values (%) for the following values of p:*				
	1	*2*	*3*	*4*	*5*
Samples 1 and 2					
$U^{(1)}$	38.49	39.17	37.12	37.15	35.50
$U^{(2)}$	38.49	68.52	0.00	0.00	0.00
Samples 3 and 4					
$U^{(1)}$	0.81	0.23	0.10	0.01	0.07
$U^{(2)}$	0.81	0.00	0.00	0.00	0.00

Table 3 shows the P-values as a function of p when the test is applied to samples 1 and 2, and also when it is applied to samples 3 and 4. In both sample 1 and sample 2, $p=1$ explains more than 94% of the variance; in both sample 3 and sample 4, $p=1$ explains more than 84% of the variance. Thus, following the recommendation of Section 3, we use the P-values that are obtained with $p=1$. They lead to the acceptance of the null hypothesis of the equality of mean functions for periods corresponding to samples 1 and 2, and to its rejection for periods corresponding to samples 3 and 4 (note that the 0.81 in the lower panel of Table 3 is 0.81%). These conclusions agree with a visual evaluation of the sample mean functions in Fig. 2. They also confirm the observation that was made in Section 4.1 that the tests have very good power, as the curves in Fig. 2(b) are not far apart. Both graphs in Fig. 2(b) give us an idea of what kind of differences in the sample mean functions are statistically significant, and which are not.

4.3. Conclusions

The simulations and the data examples of this section and of Reeder (2011) show that the tests that we propose enjoy good finite sample properties. Tests of this type allow us to quantify statistical significance of conjectures made on the basis of exploratory analysis.

In many procedures of functional data analysis, both exploratory and inferential, the issue of choosing an optimal dimension reduction parameter, like the p in our setting, is delicate. Therefore, procedures that are less sensitive to such a choice are preferable. From this angle, the Monte Carlo test based $U^{(1)}$ is preferable. The test based on $U^{(2)}$ is, however, easier to apply and in our examples and simulations leads to the same conclusions if p is chosen according to the cumulative variance rule. Data examples also show that the optimal value of p is typically a small single-digit number, 1 or 2 in our examples. Therefore, developing asymptotics as $p \to \infty$ is not necessary and may, in fact, be misleading because for larger values of p the tests may yield counterintuitive results.

Acknowledgements

The research was partially supported by National Science Foundation grants DMS-0905400 and DMS-0931948. Constructive comments of the referees and the Associate Editor helped us to improve the substance and the presentation of the results.

Appendix A: Proofs of the results of Section 2

A.1. Proof of theorem 1

The proof of theorem 1 is done in two steps. First we approximate $N^{-1/2}\sum_{i=1}^{N}\varepsilon_i(t)$ with m-dependent sequences. In the second step, the infinite dimensional m-dependent variables are replaced with finite dimensional m-dependent sequences. This reduces the proof to proving the result for the finite dimensional m-dependent sequences.

As the first step, we show that

$$\limsup_{m\to\infty}\limsup_{N\to\infty}E\left(\int\left[N^{-1/2}\sum_{i=1}^{N}\{\varepsilon_i(t)-\varepsilon_{i,m}(t)\}\right]^2\right)\mathrm{d}t=0, \tag{A.1}$$

where the variables $\varepsilon_{i,m}$ are defined in equation (1.3). By stationarity,

$$\begin{aligned}E\Big[\sum_{1\leqslant i\leqslant N}\{\varepsilon_i(t)-\varepsilon_{i,m}(t)\}\Big]^2&=\sum_{1\leqslant i\leqslant N}\sum_{1\leqslant j\leqslant N}E[\{\varepsilon_i(t)-\varepsilon_{i,m}(t)\}\{\varepsilon_j(t)-\varepsilon_{j,m}(t)\}]\\&=N\,E[\{\varepsilon_0(t)-\varepsilon_{0,m}(t)\}^2]+2\sum_{1\leqslant i<j\leqslant N}E[\{\varepsilon_i(t)-\varepsilon_{i,m}(t)\}\{\varepsilon_j(t)-\varepsilon_{j,m}(t)\}].\end{aligned}$$

In the proof, we shall repeatedly use independence relations which follow from representations (1.1) and (1.3). First observe that, if $j>i$, then $(\varepsilon_i,\varepsilon_{i,m})$ is independent of $\varepsilon_{j,j-i}$ because

$$\varepsilon_{j,j-i}=f(\delta_j,\ldots,\delta_{i+1},\delta_{j,i}^{(j-i)},\delta_{j,i-1}^{(j-i)},\ldots).$$

Consequently, $E[\{\varepsilon_i(t)-\varepsilon_{i,m}(t)\}\varepsilon_{j,j-i}(t)]=0$, and so

$$\sum_{1\leqslant i<j\leqslant N}E[\{\varepsilon_i(t)-\varepsilon_{i,m}(t)\}\,\varepsilon_j(t)]=\sum_{1\leqslant i<j\leqslant N}E[\{\varepsilon_i(t)-\varepsilon_{i,m}(t)\}\{\varepsilon_j(t)-\varepsilon_{j,j-i}(t)\}].$$

Using the Cauchy–Schwarz inequality and equation (1.2), we conclude that

$$\begin{aligned}&\Big|\int\sum_{1\leqslant i<j\leqslant N}E[\{\varepsilon_i(t)-\varepsilon_{i,m}(t)\}\{\varepsilon_j(t)-\varepsilon_{j,j-i}(t)\}]\,\mathrm{d}t\Big|\\&\qquad\leqslant\sum_{1\leqslant i<j\leqslant N}\int E[\{\varepsilon_i(t)-\varepsilon_{i,m}(t)\}^2]^{1/2}\,E[\{\varepsilon_j(t)-\varepsilon_{j,j-i}(t)\}^2]^{1/2}\,\mathrm{d}t\\&\qquad\leqslant\sum_{1\leqslant i<j\leqslant N}\left[\int E[\{\varepsilon_i(t)-\varepsilon_{i,m}(t)\}^2]\mathrm{d}t\right]^{1/2}\left(\int E[\{\varepsilon_j(t)-\varepsilon_{j,j-i}(t)\}^2]\mathrm{d}t\right)^{1/2}\\&\qquad=\sum_{1\leqslant i<j\leqslant N}\left(\int E[\{\varepsilon_0(t)-\varepsilon_{0,m}(t)\}^2]\mathrm{d}t\right)^{1/2}\left(\int E[\{\varepsilon_0(t)-\varepsilon_{0,j-i}(t)\}^2]\mathrm{d}t\right)^{1/2}\\&\qquad\leqslant N\left(\int E[\{\varepsilon_{0,m}(t)-\varepsilon_0(t)\}^2]\mathrm{d}t\right)^{1/2}\sum_{k\geqslant 1}\left[\int\{\varepsilon_0(t)-\varepsilon_{0,k}(t)\}^2\mathrm{d}t\right]^{1/2}.\end{aligned}$$

Hence

$$\limsup_{m\to\infty}\limsup_{N\to\infty}\frac{1}{N}\Big|\int\sum_{1\leqslant i<j\leqslant N}E[\{\varepsilon_{i,m}(t)-\varepsilon_i(t)\}\,\varepsilon_j(t)]\,\mathrm{d}t\Big|=0.$$

Similar arguments give

$$\limsup_{m\to\infty}\limsup_{N\to\infty}\frac{1}{N}\Big|\int\sum_{1\leqslant i<j\leqslant N}E[\{\varepsilon_{i,m}(t)-\varepsilon_i(t)\}\,\varepsilon_{j,m}(t)]\,\mathrm{d}t\Big|=0,$$

completing the verification of expression (A.1).

The next step is to show that $N^{-1/2}\Sigma_{1\leqslant i\leqslant N}\varepsilon_{i,m}$ converges to a Gaussian process Z_m with covariances defined analogously to expression (2.4). Recall that, for every integer $m\geqslant 1$, $\{\varepsilon_{i,m}\}$ is an m-dependent sequence of functions.

Let $K>1$ be an integer and $\psi_i=\psi_{i,m}$ be an orthonormal basis determined by the eigenfunctions of

$$c_m(t,s) = E\{\varepsilon_{0,m}(t)\,\varepsilon_{0,m}(s)\} + \sum_{i=1}^{m} E\{\varepsilon_{0,m}(t)\,\varepsilon_{i,m}(s)\} + \sum_{i=1}^{m} E\{\varepsilon_{0,m}(s)\,\varepsilon_{i,m}(t)\}.$$

The corresponding eigenvalues are denoted by $\nu_i = \nu_{i,m}$. Then, by the Karhunen–Loève expansion, we have

$$\varepsilon_{i,m}(t) = \sum_{l \geqslant 1} \langle \varepsilon_i, \psi_l \rangle\, \psi_l(t).$$

Next we define

$$\varepsilon_{i,m}^{(K)}(t) = \sum_{1 \leqslant l \leqslant K} \langle \varepsilon_i, \psi_l \rangle \psi_l(t).$$

By the triangle inequality we have that

$$E\left(\int \Big[\sum_{1 \leqslant i \leqslant N} \{\varepsilon_{i,m}(t) - \varepsilon_{i,m}^{(K)}(t)\} \Big]^2 \mathrm{d}t \right)^{1/2} \leqslant E\left(\int \Big[\sum_{i \in V(0)} \{\varepsilon_{i,m}(t) - \varepsilon_{i,m}^{(K)}(t)\} \Big]^2 \mathrm{d}t \right)^{1/2}$$
$$+ \ldots + E\left(\int \Big[\sum_{i \in V(m-1)} \{\varepsilon_{i,m}(t) - \varepsilon_{i,m}^{(K)}(t)\} \Big]^2 \mathrm{d}t \right)^{1/2},$$

where $V(k) = \{i : 1 \leqslant i \leqslant N, i = k (\mathrm{mod}\ m)\}$, $0 \leqslant k \leqslant m-1$. Owing to the m dependence of the sequence $\{\varepsilon_{i,m}\}$, $\Sigma_{i \in V(k)} \{\varepsilon_{i,m}(t) - \varepsilon_{i,m}^{(K)}(t)\}$ is a sum of IID random variables, and thus we obtain that

$$E\left(\int \Big[\sum_{i \in V(m-1)} \{\varepsilon_{i,m}(t) - \varepsilon_{i,m}^{(K)}(t)\} \Big]^2 \mathrm{d}t \right) \leqslant N \sum_{l \geqslant K} E(\langle \varepsilon_{0,m}, \psi_l \rangle^2).$$

Utilizing

$$\lim_{K \to \infty} \sum_{l \geqslant K} E(\langle \varepsilon_{0,m}, \psi_l \rangle^2) = 0$$

we conclude that for any $x > 0$

$$\limsup_{K \to \infty} \limsup_{N \to \infty} P\left(\int \Big[\frac{1}{N^{1/2}} \sum_{1 \leqslant i \leqslant N} \{\varepsilon_{i,m}(t) - \varepsilon_{i,m}^{(K)}(t)\} \Big]^2 \mathrm{d}t > x \right) = 0.$$

The sum of the $\varepsilon_{i,m}^{(K)}$s can be written as

$$\frac{1}{N^{1/2}} \sum_{1 \leqslant i \leqslant N} \varepsilon_{i,m}^{(K)}(t) = \sum_{1 \leqslant l \leqslant K} \psi_l(t) \frac{1}{N^{1/2}} \sum_{1 \leqslant i \leqslant N} \langle \varepsilon_{i,m}, \psi_l \rangle.$$

Next, we use the central limit theorem for stationary m-dependent sequences of random vectors (see Lehmann (1999) and the Cramér-Wold theorems in DasGupta (2008), pages 9 and 120) and obtain that

$$\left(\frac{1}{N^{1/2}} \sum_{1 \leqslant i \leqslant N} \langle \varepsilon_{i,m}, \psi_l \rangle, 1 \leqslant l \leqslant K \right)^{\mathrm{T}} \xrightarrow{\mathrm{d}} N_K(\mathbf{0}, \mathbf{\Delta}_K),$$

where $N_K(\mathbf{0}, \mathbf{\Delta}_K)$ is a K-dimensional normal random variable with zero mean and covariance matrix $\mathbf{\Delta}_K = \mathrm{diag}(\nu_1, \ldots, \nu_K)$. Thus we proved that for all $K > 1$

$$N^{-1/2} \sum_{1 \leqslant i \leqslant N} \varepsilon_{i,m}^{(K)}(t) \xrightarrow{\mathrm{d}} \sum_{1 \leqslant l \leqslant K} \nu_l^{1/2} N_l\, \psi_l(t)$$

in L^2, where $N_i, i \geqslant 1$, are independent standard normal random variables.

It is easy to see that

$$\int \left\{ \sum_{K < l < \infty} \nu_l^{1/2} N_l\, \psi_l(t) \right\}^2 \mathrm{d}t = \sum_{K < l < \infty} \nu_l N_l^2 \xrightarrow{\mathrm{P}} 0,$$

as $K \to \infty$. Hence, for every m, $N^{-1/2} \Sigma_{1 \leqslant i \leqslant N} \varepsilon_{i,m}$ converges to Γ_m in distribution in L^2, where Γ_m is a mean 0 Gaussian process with $\mathrm{cov}\{\Gamma_m(t), \Gamma_m(s)\} = c_m(t,s)$.

Since $\iint \{c_m(t,s) - c(t,s)\}^2 \mathrm{d}t\, \mathrm{d}s \to 0$, as $m \to \infty$, we obtain that Γ_m converges in distribution to Γ, where Γ is a mean 0 Gaussian process with $\mathrm{cov}\{\Gamma(t), \Gamma(s)\} = c(t,s)$, as required.

A.2. Proof of theorem 2

First we reduce expression (2.15) to equation (A.2). Then, we reduce equation (A.2) to equation (A.6) below. Since

$$\hat{\gamma}_0(t,s)=\frac{1}{N}\sum_{i=1}^{N}\{X_i(t)-\mu(t)\}\{X_i(s)-\mu(s)\}-\{\bar{X}_N(t)-\mu(t)\}\{\bar{X}_N(s)-\mu(s)\},$$

we obtain that

$$\begin{aligned}\int\int[\hat{\gamma}_0(t,s)-E\{\varepsilon_0(t)\,\varepsilon_0(s)\}]^2\,\mathrm{d}t\,\mathrm{d}s &\leqslant 4\int\int\left[\frac{1}{N}\sum_{i=1}^{N}\{X_i(t)-\mu(t)\}\{X_i(s)-\mu(s)\}-E\{\varepsilon_0(t)\,\varepsilon_0(s)\}\right]^2\mathrm{d}t\,\mathrm{d}s\\ &\quad+4\left[\int\{\bar{X}_N(t)-\mu(t)\}^2\,\mathrm{d}t\right]^2\\ &=o_P(1),\end{aligned}$$

using the ergodic theorem for random variables in a Hilbert space.

Next observe that

$$\hat{\gamma}_i(t,s)=\frac{1}{N}\sum_{j=i+1}^{N}\varepsilon_j(t)\,\varepsilon_{j-i}(s)+\frac{N-i}{N}\bar{\varepsilon}_N(t)\,\bar{\varepsilon}_N(s)-\left\{\frac{1}{N}\sum_{j=i+1}^{N}\varepsilon_j(t)\right\}\bar{\varepsilon}_N(s)-\bar{\varepsilon}_N(t)\left\{\frac{1}{N}\sum_{j=i+1}^{N}\varepsilon_{j-i}(s)\right\}.$$

Therefore, setting

$$\bar{\gamma}_i(t,s)=\frac{1}{N}\sum_{j=i+1}^{N}\varepsilon_j(t)\,\varepsilon_{j-i}(s),$$

we obtain

$$\begin{aligned}\sum_{i=1}^{N-1}K\left(\frac{i}{h}\right)\hat{\gamma}_i(t,s)&=\sum_{i=1}^{N-1}K\left(\frac{i}{h}\right)\bar{\gamma}_i(t,s)-\sum_{i=1}^{N-1}K\left(\frac{i}{h}\right)\left\{\frac{1}{N}\sum_{j=i+1}^{N}\varepsilon_j(t)\right\}\bar{\varepsilon}_N(s)\\ &\quad-\sum_{i=1}^{N-1}K\left(\frac{i}{h}\right)\left\{\frac{1}{N}\sum_{j=i+1}^{N}\varepsilon_{j-i}(s)\right\}\bar{\varepsilon}_N(t)+\sum_{i=1}^{N-1}K\left(\frac{i}{h}\right)\frac{N-i}{N}\,\bar{\varepsilon}_N(t)\,\bar{\varepsilon}_N(s).\end{aligned}$$

By stationarity we conclude that, for any $1\leqslant i\leqslant N$,

$$\begin{aligned}E\left[\int\left\{\frac{1}{N}\sum_{j=i+1}^{N}\varepsilon_j(t)\right\}^2\mathrm{d}t\right]&=\frac{1}{N^2}\sum_{j=i+1}^{N}\int E\{\varepsilon_0^2(t)\}\,\mathrm{d}t+\frac{2}{N^2}\sum_{j=i+1}^{N}(N-j)\int E\{\varepsilon_0(t)\,\varepsilon_j(t)\}\,\mathrm{d}t\\ &\leqslant\frac{1}{N}\int E\{\varepsilon_0^2(t)\}\,\mathrm{d}t+\frac{4}{N}\sum_{j=1}^{\infty}\left|\int E\{\varepsilon_0(t)\,\varepsilon_j(t)\}\,\mathrm{d}t\right|.\end{aligned}$$

Since ε_0 and $\varepsilon_{j,j}$ are independent, we obtain by equation (1.2)

$$\begin{aligned}\sum_{j=1}^{\infty}\left|\int E\{\varepsilon_0(t)\,\varepsilon_j(t)\}\,\mathrm{d}t\right|&=\sum_{j=1}^{\infty}\left|\int E[\varepsilon_0(t)\{\varepsilon_j(t)-\varepsilon_{j,j}(t)\}]\mathrm{d}t\right|\\ &\leqslant\sum_{j=1}^{\infty}E\left(\left\{\int\varepsilon_0^2(t)\,\mathrm{d}t\right\}^{1/2}\left[\int\{\varepsilon_j(t)-\varepsilon_{j,j}(t)\}^2\,\mathrm{d}t\right]^{1/2}\right)\\ &\leqslant E\left\{\int\varepsilon_0^2(t)\,\mathrm{d}t\right\}^{1/2}\sum_{j=1}^{\infty}\left(\int E[\{\varepsilon_j(t)-\varepsilon_{j,j}(t)\}^2]\mathrm{d}t\right)^{1/2}<\infty.\end{aligned}$$

Thus, we have

$$\max_{1\leqslant i\leqslant N}\left(E\left[\int\left\{\frac{1}{N}\sum_{j=i+1}^{N}\varepsilon_j(t)\right\}^2\mathrm{d}t\right]\right)=O(1).$$

Consequently, using the triangle inequality,

$$
\begin{aligned}
E\Bigg(\int\int\Bigg[\sum_{i=1}^{N-1}K\Big(\frac{i}{h}\Big)\Big\{\frac{1}{N}\sum_{j=i+1}^{N}\varepsilon_j(t)\Big\}\bar{\varepsilon}_N(s)\Bigg]^2\,\mathrm{d}t\,\mathrm{d}s\Bigg)^{1/2}&\\
\leqslant\sum_{i=1}^{N-1}\Big|K\Big(\frac{i}{h}\Big)\Big|\,E\Bigg(\Big[\int\Big\{\frac{1}{N}\sum_{j=i+1}^{N}\varepsilon_j(t)\Big\}^2\mathrm{d}t\Big]^{1/2}\Big\{\int\bar{\varepsilon}_N^2(s)\,\mathrm{d}s\Big\}^{1/2}\Bigg)&\\
\leqslant\sum_{i=1}^{N-1}\Big|K\Big(\frac{i}{h}\Big)\Big|\,E\Big[\int\Big\{\frac{1}{N}\sum_{j=i+1}^{N}\varepsilon_j(t)\Big\}^2\Big]^{1/2}\Big\{\int\bar{\varepsilon}_N^2(s)\,\mathrm{d}s\Big\}^{1/2}&\\
=\frac{h}{N}O(1)=o(1),&
\end{aligned}
$$

on account of expression (2.13).

Hence, to establish expression (2.15), it is enough to prove that

$$
\int\int\Big\{\sum_{i=1}^{N-1}K\Big(\frac{i}{h}\Big)\,\tilde{\gamma}_i(t,s)-c_1(t,s)\Big\}^2\,\mathrm{d}t\,\mathrm{d}s=o_P(1),\tag{A.2}
$$

where

$$
c_1(t,s)=\sum_{i\geqslant 1}E\{\varepsilon_0(s)\,\varepsilon_i(t)\}.
$$

Let $\{\varepsilon_{n,m},\ -\infty<n<\infty\}$ be the random variables that are defined in equation (1.3), where m is a fixed number. Let

$$
\tilde{\gamma}_{i,m}(t,s)=\frac{1}{N}\sum_{j=i+1}^{N}\varepsilon_{j,m}(t)\,\varepsilon_{j-i,m}(s).
$$

We show that, for every $m\geqslant 1$,

$$
\int\int\Big\{\sum_{i=1}^{N-1}K\Big(\frac{i}{h}\Big)\,\tilde{\gamma}_{i,m}(t,s)-c_1^{(m)}(t,s)\Big\}^2\,\mathrm{d}t\,\mathrm{d}s=o_P(1),\tag{A.3}
$$

where

$$
c_1^{(m)}(t,s)=\sum_{i=1}^{\infty}E\{\varepsilon_{1,m}(s)\,\varepsilon_{i+1,m}(t)\}.
$$

We also note that expressions (1.3) and (1.2) imply that

$$
\lim_{m\to\infty}\int\int\{c_1^{(m)}(t,s)-c_1(t,s)\}^2\,\mathrm{d}t\,\mathrm{d}s=0.\tag{A.4}
$$

Since $\{\varepsilon_{n,m},\ -\infty<n<\infty\}$ is an m-dependent sequence,

$$
c_1^{(m)}(t,s)=\sum_{i=1}^{m}E\{\varepsilon_{1,m}(s)\,\varepsilon_{i+1,m}(t)\}.
$$

Using equations (2.7), (2.8) and (2.13), we obtain

$$
\max_{1\leqslant i\leqslant m}\Big|K\Big(\frac{i}{h}\Big)-1\Big|\to 0,\qquad\qquad\text{as } N\to\infty.
$$

By the ergodic theorem,

$$
\int\int[\tilde{\gamma}_{i,m}(t,s)-E\{\varepsilon_{1,m}(s)\,\varepsilon_{i+1,m}(t)\}]^2\,\mathrm{d}t\,\mathrm{d}s=o_P(1),
$$

for any fixed i. Hence result (A.3) is proved, once we have shown that

$$\int\int\left\{\sum_{i=m+1}^{N-1}K\left(\frac{i}{h}\right)\tilde{\gamma}_{i,m}(t,s)\right\}^2\mathrm{d}t\,\mathrm{d}s=o_P(1). \tag{A.5}$$

It is easy to see that

$$E\left[\int\int\left\{\sum_{i=m+1}^{N-1}K\left(\frac{i}{h}\right)\tilde{\gamma}_{i,m}(t,s)\right\}^2\mathrm{d}t\,\mathrm{d}s\right]$$
$$=\int\int\left\{\frac{1}{N^2}\sum_{i=m+1}^{h}\sum_{l=m+1}^{h}\sum_{k=i+1}^{N-1}\sum_{n=l+1}^{N-1}K\left(\frac{i}{h}\right)K\left(\frac{l}{h}\right)E(\varepsilon_{k,m}\varepsilon_{k-i,m}\varepsilon_{n,m}\varepsilon_{n-l,m})\right\},$$

provided that $h\leqslant N-1$. The sequence $\{\varepsilon_{n,m},\ -\infty<n<\infty\}$ is an m-dependent sequence, and therefore $\varepsilon_{k,m}$ and $\varepsilon_{k-i,m}$ are independent, since $i\geqslant m+1$. Similarly, $\varepsilon_{n,m}$ and $\varepsilon_{n-l,m}$ are independent. Hence the number of terms when $E(\varepsilon_{k,m}\varepsilon_{k-i,m}\varepsilon_{n,m}\varepsilon_{n-l,m})$ is different from 0 is $O(Nh)$. Consequently,

$$E\left[\int\int\left\{\sum_{i=m+1}^{N-1}K\left(\frac{i}{h}\right)\tilde{\gamma}_{i,m}(t,s)\right\}^2\mathrm{d}t\,\mathrm{d}s\right]=O\left(\frac{h}{N}\right)=o(1).$$

This completes the verification of result (A.5).

Next we show that, for all $\varepsilon>0$,

$$\lim_{m\to\infty}\limsup_{N\to\infty}P\left(\int\int\left[\sum_{i=1}^{N-1}K\left(\frac{i}{h}\right)\{\bar{\gamma}_i(t,s)-\tilde{\gamma}_{i,m}(t,s)\}\right]^2\mathrm{d}t\,\mathrm{d}s>\varepsilon\right)=0. \tag{A.6}$$

Using the definitions of the covariances $\bar{\gamma}_i(t,s)$ and $\tilde{\gamma}_{i,m}(t,s)$, we consider the decompositions

$$\frac{1}{N}\sum_{i=1}^{N-1}K\left(\frac{i}{h}\right)\sum_{j=i+1}^{N}\{\varepsilon_j(t)\,\varepsilon_{j-i}(s)-\varepsilon_{j,m}(t)\,\varepsilon_{j-i,m}(s)\}$$
$$=\frac{1}{N}\left(\sum_{i=1}^{m}+\sum_{i=m+1}^{h}\right)K\left(\frac{i}{h}\right)\sum_{j=i+1}^{N}\{\varepsilon_j(t)\varepsilon_{j-i}(s)-\varepsilon_{j,m}(t)\,\varepsilon_{j-i,m}(s)\}$$

and

$$\varepsilon_j(t)\,\varepsilon_{j-i}(s)-\varepsilon_{j,m}(t)\,\varepsilon_{j-i,m}(s)=\{\varepsilon_j(t)-\varepsilon_{j,m}(t)\}\,\varepsilon_{j-i}(s)+\{\varepsilon_{j-i}(s)-\varepsilon_{j-i,m}(s)\}\,\varepsilon_{j,m}(t).$$

Clearly,

$$\left(\int\int\left[\frac{1}{N}\sum_{i=1}^{m}K\left(\frac{i}{h}\right)\sum_{j=i+1}^{N}\{\varepsilon_j(t)-\varepsilon_{j,m}(t)\}\,\varepsilon_{j-i}(s)\right]^2\mathrm{d}t\,\mathrm{d}s\right)^{1/2}$$
$$\leqslant\frac{1}{N}\sum_{i=1}^{m}\left|K\left(\frac{i}{h}\right)\right|\left[\int\{\varepsilon_j(t)-\varepsilon_{j,m}(t)\}^2\mathrm{d}t\right]^{1/2}\left\{\int\varepsilon_{j-i}^2(s)\,\mathrm{d}s\right\}^{1/2},$$

so, by equation (2.14),

$$E\left(\int\int\left[\frac{1}{N}\sum_{i=1}^{m}K\left(\frac{i}{h}\right)\sum_{j=i+1}^{N}\{\varepsilon_j(t)-\varepsilon_{j,m}(t)\}\,\varepsilon_{j-i}(s)\right]^2\mathrm{d}t\,\mathrm{d}s\right)^{1/2}$$
$$\leqslant m\left(E\left[\int\{\varepsilon_0(t)-\varepsilon_{0,m}(t)\}^2\mathrm{d}t\right]E\left\{\int\varepsilon_0^2(s)\,\mathrm{d}s\right\}\right)^{1/2}$$
$$\leqslant Am\left(E\left[\int\{\varepsilon_0(t)-\varepsilon_{0,m}(t)\}^2\mathrm{d}t\right]\right)^{1/2}\to 0,\qquad\text{as } m\to\infty,$$

according to equation (2.14), where A is a constant.

Next we use the decomposition

$$\varepsilon_j(t)\,\varepsilon_{j-i}(s)=\varepsilon_{j,i}(t)\,\varepsilon_{j-i}(s)+\{\varepsilon_j(t)-\varepsilon_{j,i}(t)\}\,\varepsilon_{j-i}(s)$$

to obtain

$$\left(\int\int\left[\frac{1}{N}\sum_{i=m+1}^{h}K\left(\frac{i}{h}\right)\sum_{j=i+1}^{N}\{\varepsilon_j(t)-\varepsilon_{j,i}(t)\}\,\varepsilon_{j-i}(s)\right]^2\mathrm{d}t\,\mathrm{d}s\right)^{1/2}$$
$$\leqslant\frac{1}{N}\sum_{i=m+1}^{\infty}\sum_{j=i+1}^{N}\left[\int\{\varepsilon_j(t)-\varepsilon_{j,i}(t)\}^2\,\mathrm{d}t\right]^{1/2}\left\{\int\varepsilon_{j-i}^2(s)\,\mathrm{d}s\right\}^{1/2}.$$

Therefore, by expressions (2.2) and (1.2), we have

$$E\left(\int\int\left[\frac{1}{N}\sum_{i=m+1}^{h}K\left(\frac{i}{h}\right)\sum_{j=i+1}^{N}\{\varepsilon_j(t)-\varepsilon_{j,i}(t)\}\,\varepsilon_{j-i}(s)\right]^2\mathrm{d}t\,\mathrm{d}s\right)^{1/2}$$
$$\leqslant\frac{1}{N}\sum_{i=m+1}^{\infty}\sum_{j=i+1}^{N}E\left(\left[\int\{\varepsilon_j(t)-\varepsilon_{j,i}(t)\}^2\,\mathrm{d}t\right]^{1/2}\left\{\int\varepsilon_{j-i}^2(s)\,\mathrm{d}s\right\}^{1/2}\right)$$
$$\leqslant\frac{1}{N}\sum_{i=m+1}^{\infty}\sum_{j=i+1}^{N}\left[\int\{\varepsilon_j(t)-\varepsilon_{j,i}(t)\}^2\,\mathrm{d}t\right]^{1/2}\left\{\int\varepsilon_{j-i}^2(s)\,\mathrm{d}s\right\}^{1/2}$$
$$\leqslant A\sum_{i=m+1}^{\infty}\left[\int\{\varepsilon_0(t)-\varepsilon_{0,i}(t)\}^2\,\mathrm{d}t\right]^{1/2}\to 0,\qquad\text{as } m\to\infty.$$

We have shown so far that for, any $\varepsilon>0$,

$$\lim_{m\to\infty}\limsup_{N\to\infty}P\left[\int\int\left\{\frac{1}{N}\sum_{i=m+1}^{h}K\left(\frac{i}{h}\right)\sum_{j=i+1}^{N}\varepsilon_j(t)\varepsilon_{j-i}(s)\right\}^2\mathrm{d}t\,\mathrm{d}s>\varepsilon\right]=0.$$

Similar arguments give

$$\lim_{m\to\infty}\limsup_{N\to\infty}P\left[\int\int\left\{\frac{1}{N}\sum_{i=m+1}^{h}K\left(\frac{i}{h}\right)\sum_{j=i+1}^{N}\varepsilon_{j,m}(t)\,\varepsilon_{j-i,m}(s)\right\}^2\mathrm{d}t\,\mathrm{d}s>\varepsilon\right]=0.$$

This completes the verification of result (A.6), so result (A.2) is proved.

References

Aue, A., Hörmann, S., Horváth, L., Hušková, M. and Steinebach, J. (2012) Sequential testing for the stability of high-frequency portfolio betas. *Econmetr. Theor.*, **28**, 1–34.

Bosq, D. (2000) *Linear Processes in Function Spaces*. New York: Springer.

DasGupta, A. (2008) *Asymptotic Theory of Statistics and Probability*. New York: Springer.

Gabrys, R., Horváth, L. and Kokoszka, P. (2010) Tests for error correlation in the functional linear model. *J. Am. Statist. Ass.*, **105**, 1113–1125.

Gromenko, O., Kokoszka, P., Zhu, L. and Sojka, J. (2012) Estimation and testing for spatially indexed curves with application to ionospheric and magnetic field trends. *Ann. Appl. Statist.*, to be published.

Hörmann, S. and Kokoszka, P. (2010) Weakly dependent functional data. *Ann. Statist.*, **38**, 1845–1884.

Horváth, L., Kokoszka, P. and Reimherr, M. (2009) Two sample inference in functional linear models. *Can. J. Statist.*, **37**, 571–591.

Kargin, V. and Onatski, A. (2008) Curve forecasting by functional autoregression. *J. Multiv. Anal.*, **99**, 2508–2526.

Kokoszka, P., Maslova, I., Sojka, J. and Zhu, L. (2008) Testing for lack of dependence in the functional linear model. *Can. J. Statist.*, **36**, 207–222.

Lehmann, E. L. (1999) *Elements of Large Sample Theory*. New York: Springer.

Müller, H.-G. and Stadtmüller, U. (2005) Generalized functional linear models. *Ann. Statist.*, **33**, 774–805.

Panaretos, V. M., Kraus, D. and Maddocks, J. H. (2010) Second-order comparison of Gaussian random functions and the geometry of DNA minicircles. *J. Am. Statist. Ass.*, **105**, 670–682.

Politis, D. N. and Romano, J. P. (1996) On flat-top spectral density estimators for homogeneous random fields. *J. Statist. Planng Inf.*, **51**, 41–53.

Ramsay, J. O. and Silverman, B. W. (2005) *Functional Data Analysis*. New York: Springer.

Reeder, R. (2011) Limit theorems in functional data analysis with applications. *PhD Thesis*. University of Utah, Salt Lake City.

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