The Annals of Statistics
2010, Vol. 38, No. 3, 1845–1884
DOI: 10.1214/09-AOS768

WEAKLY DEPENDENT FUNCTIONAL DATA

BY SIEGFRIED HÖRMANN AND PIOTR KOKOSZKA[1]

Université Libre de Bruxelles and Utah State University

Functional data often arise from measurements on fine time grids and are obtained by separating an almost continuous time record into natural consecutive intervals, for example, days. The functions thus obtained form a functional time series, and the central issue in the analysis of such data consists in taking into account the temporal dependence of these functional observations. Examples include daily curves of financial transaction data and daily patterns of geophysical and environmental data. For scalar and vector valued stochastic processes, a large number of dependence notions have been proposed, mostly involving mixing type distances between σ-algebras. In time series analysis, measures of dependence based on moments have proven most useful (autocovariances and cumulants). We introduce a moment-based notion of dependence for functional time series which involves m-dependence. We show that it is applicable to linear as well as nonlinear functional time series. Then we investigate the impact of dependence thus quantified on several important statistical procedures for functional data. We study the estimation of the functional principal components, the long-run covariance matrix, change point detection and the functional linear model. We explain when temporal dependence affects the results obtained for i.i.d. functional observations and when these results are robust to weak dependence.

1. Introduction. The assumption of independence is often too strong to be realistic in many applications, especially if data are collected sequentially over time. It is then natural to expect that the current observation depends to some degree on the previous observations. This remains true for functional data and has motivated the development of appropriate functional time series models. The most popular model is the autoregressive model of Bosq [14]. This model and its various extensions are particularly useful for prediction (see, e.g., Besse, Cardot and Stephenson [11] Damon and Guillas [23], Antoniadis and Sapatinas [4]). For many functional time series it is, however, not clear what specific model they follow, and for many statistical procedures it is not necessary to assume a specific model. In such cases, it is important to know what the effect of the dependence on a given procedure is. Is it robust to temporal dependence, or does this type of dependence introduce a serious bias? To answer questions of this type, it is essential to quantify

Received June 2009; revised November 2009.

[1]Supported in part by NSF Grants DMS-08-04165 and DMS-09-31948 at Utah State University.

AMS 2000 subject classifications. Primary 62M10; secondary 60G10, 62G05.

Key words and phrases. Asymptotics, change points, eigenfunctions, functional principal components, functional time series, long-run variance, weak dependence.

the notion of temporal dependence. For scalar and vector time series, this question has been approached from a number of angles, but, except for the linear model of Bosq [14], for functional time series data no general framework is available. Our goal in this paper is to propose such a framework, which applies to both linear and nonlinear dependence, develop the requisite theory and apply it to selected problems in the analysis of functional time series. Our examples are chosen to show that some statistical procedures for functional data are robust to temporal dependence as quantified in this paper, while other require modifications that take this dependence into account.

While we focus here on a general theoretical framework, this research has been motivated by our work with functional data arising in space physics and environmental science. For such data, especially for the space physics data, no validated time series models are currently available, so to justify any inference drawn from them, they must fit into a general, one might say, nonparametric, dependence scheme. An example of space physics data is shown in Figure 1. Temporal dependence from day to day can be discerned, but has not been modeled.

The paper is organized as follows. In Section 2 we introduce our dependence condition and illustrate it with several examples. In particular, we show that the linear functional processes fall into our framework, and present some nonlinear models that also do. It is now recognized that the *functional principal components* (FPCs) play a far greater role than their multivariate counterparts (Yao and

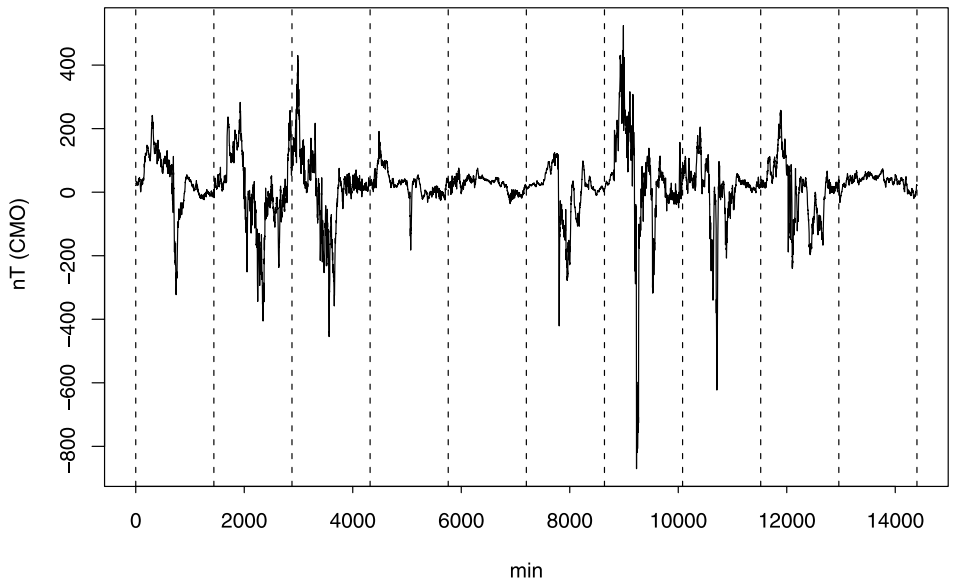

FIG. 1. *Ten consecutive functional observations of a component of the magnetic field recorded at College, Alaska. The vertical lines separate days. Long negative spikes lasting a few hours correspond to the aurora borealis.*

Lee [64], Hall and Hosseini-Nasab [33], Reiss and Ogden [51], Benko, Härdle and Kneip [6], Müller and Yao [45]). To develop theoretical justification for procedures involving the FPCs, it is necessary to use the convergence of the estimated FPCs to their population counterparts. Results of this type are available only for independent observations (Dauxois, Pousse and Romain [24], and linear processes, Bosq [14], Bosq and Blanke [15]). We show in Section 3 how the consistency of the estimators for the eigenvalues and eigenfunctions of the covariance operator extends to dependent functional data. Next, in Section 4, we turn to the estimation of an appropriately defined long-run variance matrix for functional data. For most time series procedures, the long-run variance plays a role analogous to the variance–covariance matrix for independent observations. Its estimation is therefore of fundamental importance, and has been a subject of research for many decades (Anderson [1], Andrews [3] and Hamilton [34] provide the background and numerous references). In Sections 5 and 6, we illustrate the application of the results of Sections 3 and 4 on two problems of recent interest: change point detection for functional data and the estimation of kernel in the functional linear model. We show that the detection procedure of Berkes et al. [7] must be modified if the data exhibit dependence, but the estimation procedure of Yao, Müller and Wang [65] is robust to mild dependence. Section 5 also contains a small simulation study and a data example. The proofs are collected in the Appendix.

2. Approximable functional time series. The notion of weak dependence has, over the past decades, been formalized in many ways. Perhaps the most popular are various mixing conditions (see Doukhan [25], Bradley [16]), but in recent years several other approaches have also been introduced (see Doukhan and Louhichi [26] and Wu [62], [63], among others). In time series analysis, moment based measures of dependence, most notably autocorrelations and cumulants, have gained a universal acceptance. The measure we consider below is a moment-type quantity, but it is also related to the mixing conditions as it considers σ-algebras m time units apart, with m tending to infinity.

A most direct relaxation of independence is the m-dependence. Suppose $\{X_n\}$ is a sequence of random elements taking values in a measurable space S. Denote by $\mathcal{F}_k^- = \sigma\{\ldots, X_{k-2}, X_{k-1}, X_k\}$ and $\mathcal{F}_k^+ = \sigma\{X_k, X_{k+1}, X_{k+2}, \ldots\}$ the σ-algebras generated by the observations up to time k and after time k, respectively. Then the sequence $\{X_n\}$ is said to be m-dependent if for any k, the σ-algebras $\mathcal{F}_k^-$ and $\mathcal{F}_{k+m}^+$ are independent.

Most time series models are not m-dependent. Rather, various measures of dependence decay sufficiently fast, as the distance m between the σ-algebras $\mathcal{F}_k^-$ and $\mathcal{F}_{k+m}^+$ increases. However, m-dependence can be used as a tool to study properties of many nonlinear sequences (see, e.g., Hörmann [35] and Berkes, Hörmann and Schauer [8] for recent applications). The general idea is to approximate $\{X_n, n \in \mathbb{Z}\}$ by m-dependent processes $\{X_n^{(m)}, n \in \mathbb{Z}\}$, $m \geq 1$. The goal is to establish that for every n the sequence $\{X_n^{(m)}, m \geq 1\}$ converges in some sense to X_n, if

we let $m \to \infty$. If the convergence is fast enough, then one can obtain the limiting behavior of the original process from corresponding results for m-dependent sequences. Definition 2.1 formalizes this idea and sets up the necessary framework for the construction of such m-dependent approximation sequences. The idea of approximating scalar sequences by m-dependent nonlinear moving averages appears already in Section 21 of Billingsley [12], and it was developed in several directions by Pötscher and Prucha [48].

In the sequel we let $H = L^2([0,1], \mathcal{B}_{[0,1]}, \lambda)$ be the Hilbert space of square integrable functions defined on $[0,1]$. For $f \in H$ we set $\|f\|^2 = \int_0^1 |f(t)|^2 dt$. All our random elements are assumed to be defined on some common probability space $(\Omega, \mathcal{A}, P)$. For $p \geq 1$ we denote by $L^p = L^p(\Omega, \mathcal{A}, P)$ the space of (classes of) real valued random variables such that $\|X\|_p = (E|X|^p)^{1/p} < \infty$. Further we let $L_H^p = L_H^p(\Omega, \mathcal{A}, P)$ be the space of H valued random variables X such that $\nu_p(X) = (E\|X\|^p)^{1/p} < \infty$.

DEFINITION 2.1. A sequence $\{X_n\} \in L_H^p$ is called *L^p–m-approximable* if each X_n admits the representation,

$$X_n = f(\varepsilon_n, \varepsilon_{n-1}, \ldots), \tag{2.1}$$

where the ε_i are i.i.d. elements taking values in a measurable space S, and f is a measurable function $f : S^\infty \to H$. Moreover we assume that if $\{\varepsilon_i'\}$ is an independent copy of $\{\varepsilon_i\}$ defined on the same probability space, then letting

$$X_n^{(m)} = f(\varepsilon_n, \varepsilon_{n-1}, \ldots, \varepsilon_{n-m+1}, \varepsilon_{n-m}', \varepsilon_{n-m-1}', \ldots), \tag{2.2}$$

we have

$$\sum_{m=1}^{\infty} \nu_p\big(X_m - X_m^{(m)}\big) < \infty. \tag{2.3}$$

For our applications, choosing $p = 4$ will be convenient, but any $p \geq 1$ can be used, depending on what is needed. (Our definition makes even sense if $p < 1$, but then ν_p is no longer a norm.) Definition 2.1 implies that $\{X_n\}$ is strictly stationary. It is clear from the representation of X_n and $X_n^{(m)}$ that $E\|X_m - X_m^{(m)}\|^p = E\|X_1 - X_1^{(m)}\|^p$, so that condition (2.3) could be formulated solely in terms of X_1 and the approximations $X_1^{(m)}$. Obviously the sequence $\{X_n^{(m)}, n \in \mathbb{Z}\}$ as defined in (2.2) is *not* m-dependent. To this end we need to define for each n an independent copy $\{\varepsilon_k^{(n)}\}$ of $\{\varepsilon_k\}$ (this can always be achieved by enlarging the probability space) which is then used instead of $\{\varepsilon_k'\}$ to construct $X_n^{(m)}$; that is, we set

$$X_n^{(m)} = f\big(\varepsilon_n, \varepsilon_{n-1}, \ldots, \varepsilon_{n-m+1}, \varepsilon_{n-m}^{(n)}, \varepsilon_{n-m-1}^{(n)}, \ldots\big). \tag{2.4}$$

We will call this method the *coupling construction*. Since this modification leaves condition (2.3) unchanged, we will assume from now on that the $X_n^{(m)}$ are defined

by (2.4). Then, for each $m \geq 1$, the sequences $\{X_n^{(m)}, n \in \mathbb{Z}\}$ are strictly stationary and m-dependent, and each $X_n^{(m)}$ is equal in distribution to X_n.

The coupling construction is only one of a variety of possible m-dependent approximations. In most applications, the measurable space S coincides with H, and the ε_n represent model errors. In this case, we can set

$$\tilde{X}_n^{(m)} = f(\varepsilon_n, \varepsilon_{n-1}, \ldots, \varepsilon_{n-m+1}, 0, 0, \ldots). \tag{2.5}$$

The sequence $\{\tilde{X}_n^{(m)}, n \in \mathbb{Z}\}$ is strictly stationary and m-dependent, but $X_n^{(m)}$ is no longer equal in distribution to X_n. This is not a big problem but requires additional lines in the proofs. For the *truncation construction* (2.5), condition (2.3) is replaced by

$$\sum_{m=1}^{\infty} \nu_p\big(X_m - \tilde{X}_m^{(m)}\big) < \infty. \tag{2.6}$$

Since $E\|\tilde{X}_m^{(m)} - X_m^{(m)}\|^p = E\|\tilde{X}_m^{(m)} - X_m\|^p$, (2.6) implies (2.3), but not vice versa. Thus the coupling construction allows to study a slightly broader class of time series.

An important question that needs to be addressed at this point is how our notion of weak dependence compares to other existing ones. The closest relative of L^p–m-approximability is the notion of L^p-approximability studied by Pötscher and Prucha [48] for scalar and vector-valued processes. Since our definition applies with an obvious modification to sequences with values in any normed vector spaces H (especially $\mathbb{R}$ or $\mathbb{R}^n$), it can been seen as a generalization of L^p-approximability. There are, however, important differences. By definition, L^p-approximability only allows for approximations that are, like the truncation construction, measurable with respect to a finite selection of basis vectors, $\varepsilon_n, \ldots, \varepsilon_{n-m}$, whereas the coupling construction does not impose this condition. On the other hand, L^p-approximability is not based on independence of the innovation process. Instead independence is relaxed to certain mixing conditions. Clearly, m-dependence implies the CLT, and so our L^p–m-approximability implies central limit theorems for practically all important time series models. As we have shown in previous papers [5, 8, 35, 36], a scalar version of this notion has much more potential than solely giving central limit theorems.

The concept of weak dependence introduced in Doukhan and Louhichi [26] is defined for scalar variables in a very general framework and has been successfully used to prove (empirical) FCLTs. Like our approach, it does not require smoothness conditions. Its extensions to problems of functional data analysis have not been studied yet.

Another approach to weak dependence is a martingale approximation, as developed in Gordin [31] and Philipp and Stout [47]. In the context of sequences $\{X_k\}$ of the form (2.1), particularly complete results have been proved by Wu [62,

63]. Again, L^p–m-approximability cannot be directly compared to approximating martingale conditions; the latter hold for a very large class of processes, but, unlike L^p–m-approximability, they apply only in the context of partial sums.

The classical approach to weak dependence, developed in the seminal papers of Rosenblatt [54] and Ibragimov [37], uses the strong mixing property and its variants like β, ϕ, ρ and ψ mixing. The general idea is to measure the maximal dependence between two events lying in the "past" $\mathcal{F}_k^-$ and in the "future" $\mathcal{F}_{k+m}^+$, respectively. The fading memory is described by this maximal dependence decaying to zero for m growing to ∞. For example, the α-mixing coefficient is given by

$$\alpha_m = \sup\{|P(A \cap B) - P(A)P(B)| A \in \mathcal{F}_k^-, B \in \mathcal{F}_{k+m}^+, k \in \mathbb{Z}\}.$$

A sequence is called α-mixing (strong mixing) if $\alpha_m \to 0$ for $m \to \infty$.

This method yields very sharp results (for a complete account of the classical theory (see Bradley [16]), but verifying mixing conditions of the above type is not easy, whereas the verification of L^p–m-approximability is almost immediate as our examples below show. This is because the L^p–m-approximability condition uses directly the model specification $X_n = f(\varepsilon_n, \varepsilon_{n-1}, \ldots)$. Another problem is that even when mixing applies (e.g., for Markov processes), it typically requires strong smoothness conditions. For example, for the AR(1) process

$$Y_k = \tfrac{1}{2} Y_{k-1} + \varepsilon_k$$

with Bernoulli innovations, strong mixing fails to hold (cf. Andrews [2]). Since c-mixing, where c is either of ψ, ϕ, β or ρ, implies α-mixing, $\{Y_k\}$ above satisfies none of these mixing conditions, whereas Example 2.1 shows that the AR(1) process is L^p–m-approximable without requiring any smoothness properties for the innovations process. Consequently our condition does not imply strong mixing. On the other hand, L^p–m-approximability is restricted to a more limited class of processes, namely processes allowing the representation $X_n = f(\varepsilon_n, \varepsilon_{n-1}, \ldots)$. We emphasize, however, that all time series models used in practice (scalar, vector or functional) have this representation (cf. [49, 59, 60]), as an immediate consequence of their "forward" dynamics, for example, their definitions by a stochastic recurrence equations. See the papers of Rosenblatt [55–57] for sufficient criteria.

We conclude that *L^p–m-approximability is not directly comparable with classical mixing coefficients.*

The following lemma shows how L^p–m-approximability is unaffected by linear transformations, whereas independence assumptions are needed for product type operations.

LEMMA 2.1. *Let $\{X_n\}$ and $\{Y_n\}$ be two L^p–m-approximability sequences in L_H^p. Define*:

- $Z_n^{(1)} = A(X_n)$, *where* $A \in \mathcal{L}$;

- $Z_n^{(2)} = X_n + Y_n$;
- $Z_n^{(3)} = X_n \circ Y_n$ (*point-wise multiplication*);
- $Z_n^{(4)} = \langle X_n, Y_n \rangle$;
- $Z_n^{(5)} = X_n \otimes Y_n$.

Then $\{Z_n^{(1)}\}$ *and* $\{Z_n^{(2)}\}$ *are* L^p*–m-approximable sequences in* L_H^p. *If* X_n *and* Y_n *are independent, then* $\{Z_n^{(4)}\}$ *and* $\{Z_n^{(5)}\}$ *are* L^p*–m-approximable sequences in the respective spaces. If* $E \sup_{t\in[0,1]} |X_n(t)|^p + E \sup_{t\in[0,1]} |Y_n(t)|^p < \infty$, *then* $\{Z_n^{(3)}\}$ *is* L^p*–m-approximable in* L_H^p.

PROOF. The first two relations are immediate. We exemplify the rest of the simple proofs for $Z_n = Z_n^{(5)}$. For this we set $Z_m^{(m)} = X_m^{(m)} \otimes Y_m^{(m)}$ and note that Z_m and $Z_m^{(m)}$ are (random) kernel operators, and thus Hilbert–Schmidt operators. Since

$$\begin{aligned}\|Z_m - Z_m^{(m)}\|_{\mathcal{L}} &\le \|Z_m - Z_m^{(m)}\|_S \\ &\le \left(\int\int \left(X_m(t)Y_m(s) - X_m^{(m)}(t)Y_m^{(m)}(s)\right)^2 dt\, ds\right)^{1/2} \\ &\le \sqrt{2}\big(\|X_m\|\|Y_m - Y_m^{(m)}\| + \|Y_m^{(m)}\|\|X_m - X_m^{(m)}\|\big),\end{aligned}$$

the proof follows from the independence of X_n and Y_n. □

The proof shows that our assumption can be modified and independence is not required. However, if X, Y are not independent, then $E|XY| \neq E|X|E|Y|$. We have then to use the Cauchy–Schwarz inequality and obviously need $2p$ moments.

We want to point out that only a straightforward modification is necessary in order to generalize the theory of this paper to noncausal processes $X_n = f(\ldots, \varepsilon_{n+1}, \varepsilon_n, \varepsilon_{n-1}, \ldots)$. Our framework can be also extended to nonstationary sequences, for example, those of the form (2.1) where $\{\varepsilon_k\}$ is a sequence of independent, but not necessarily identically distributed, or random variables where

$$X_n = f_n(\varepsilon_n, \varepsilon_{n-1}, \ldots).$$

The m-dependent coupled process can be defined in the exact same way as in the stationary case

$$X_n^{(m)} = f_n\big(\varepsilon_n, \varepsilon_{n-1}, \ldots, \varepsilon_{n-m+1}, \varepsilon_{n-m}^{(n)}, \varepsilon_{n-m-1}^{(n)}, \ldots\big).$$

A generalization of our method to nonstationarity would be useful, especially when the goal is to develop methodology for locally stationary data. Such work is, however, beyond the intended scope of this paper.

We now illustrate the applicability of Definition 2.1 with several examples. Let $\mathcal{L} = \mathcal{L}(H, H)$ be the set of bounded linear operators from H to H. For $A \in \mathcal{L}$

we define the operator norm $\|A\|_{\mathcal{L}} = \sup_{\|x\|\le 1} \|Ax\|$. If the operator is Hilbert–Schmidt, then we denote with $\|A\|_{\mathcal{S}}$ its Hilbert–Schmidt norm. Recall that for any Hilbert–Schmidt operator $A \in \mathcal{L}$, $\|A\|_{\mathcal{L}} \le \|A\|_{\mathcal{S}}$.

EXAMPLE 2.1 (Functional autoregressive process). Suppose $\Psi \in \mathcal{L}$ satisfies $\|\Psi\|_{\mathcal{L}} < 1$. Let $\varepsilon_n \in L_H^2$ be i.i.d. with mean zero. Then there is a unique stationary sequence of random elements $X_n \in L_H^2$ such that

$$X_n(t) = \Psi(X_{n-1})(t) + \varepsilon_n(t). \tag{2.7}$$

For details see Chapter 3 of Bosq [14]. The AR(1) sequence (2.7) admits the expansion $X_n = \sum_{j=0}^{\infty} \Psi^j(\varepsilon_{n-j})$ where Ψ^j is the jth iterate of the operator Ψ. We thus set $X_n^{(m)} = \sum_{j=0}^{m-1} \Psi^j(\varepsilon_{n-j}) + \sum_{j=m}^{\infty} \Psi^j(\varepsilon_{n-j}^{(n)})$. It is easy to verify that for every A in $\mathcal{L}$, $\nu_p(A(Y)) \le \|A\|_{\mathcal{L}} \nu_p(Y)$. Since $X_m - X_m^{(m)} = \sum_{j=m}^{\infty} (\Psi^j(\varepsilon_{m-j}) - \Psi^j(\varepsilon_{m-j}^{(m)}))$, it follows that $\nu_p(X_m - X_m^{(m)}) \le 2\sum_{j=m}^{\infty} \|\Psi\|_{\mathcal{L}}^j \nu_p(\varepsilon_0) = O(1) \times \nu_p(\varepsilon_0)\|\Psi\|_{\mathcal{L}}^m$. By assumption $\nu_2(\varepsilon_0) < \infty$ and therefore $\sum_{m=1}^{\infty} \nu_2(X_m - X_m^{(m)}) < \infty$, so condition (2.6) holds with $p \ge 2$, as long as $\nu_p(\varepsilon_0) < \infty$.

The argument in the above example shows that a sufficient condition to obtain L^p–m-approximability is

$$\begin{aligned} &\|f(a_m, \ldots, a_1, x_0, x_{-1}, \ldots) - f(a_m, \ldots, a_1, y_0, y_{-1}, \ldots)\| \\ &\quad \le c_m \|f(x_0, x_{-1}, \ldots) - f(y_0, y_{-1}, \ldots)\|, \end{aligned}$$

where $\sum_{m\ge 1} c_m < \infty$. This holds for a functional AR(1) process and offers an attractive sufficient and distribution-free condition for L^p–m-approximability. The interesting question, whether one can impose some other, more general conditions on the function f that would imply L^p–m-approximability remains open. For example, the simple criterion above does not apply to general linear processes. We recall that a sequence $\{X_n\}$ is said to be a *linear process in H* if $X_n = \sum_{j=0}^{\infty} \Psi_j(\varepsilon_{n-j})$ where the errors $\varepsilon_n \in L_H^2$ are i.i.d. and zero mean, and each Ψ_j is a bounded operator. If $\sum_{j=1}^{\infty} \|\Psi_j\|_{\mathcal{L}}^2 < \infty$, then the series defining X_n converges a.s. and in L_H^2 (see Section 7.1 of Bosq [14]).

A direct verification, following the lines of Example 2.1, yields sufficient conditions for a general linear process to be L^p–m-approximable.

PROPOSITION 2.1. *Suppose $\{X_n\} \in L_H^2$ is a linear process whose errors satisfy $\nu_p(\varepsilon_0) < \infty$, $p \ge 2$. The operator coefficients satisfy*

$$\sum_{m=1}^{\infty} \sum_{j=m}^{\infty} \|\Psi_j\| < \infty. \tag{2.8}$$

Then $\{X_n\}$ is L^p–m-approximable.

We note that condition (2.8) is comparable to the usual assumptions made in the scalar case. For a scalar linear process the weakest possible condition for weak dependence is

$$\sum_{j=0}^{\infty} |\psi_j| < \infty. \tag{2.9}$$

If it is violated, the resulting time series are referred to as strongly dependent, long memory, long-range dependent or persistent. Recall that (2.9) merely ensures the existence of fundamental population objects like an absolutely summable autocovariance sequence or a bounded spectral density. It is, however, too weak to establish any statistical results. For example, for the asymptotic normality of the sample autocorrelations we need $\sum j\psi_j^2 < \infty$, for the convergence of the periodogram ordinates $\sum \sqrt{j}|\psi_j| < \infty$. Many authors assume $\sum j|\psi_j| < \infty$ to be able to use all these basic results. The condition $\sum j|\psi_j| < \infty$ is equivalent to (2.8).

We next give a simple example of a nonlinear L^p–m-approximable sequence. It is based on the model used by Maslova et al. [44] to simulate the so-called solar quiet (Sq) variation in magnetometer records (see also Maslova et al. [43]). In that model, $X_n(t) = U_n(S(t) + Z_n(t))$ represents the part of the magnetometer record on day n which reflects the magnetic field generated by ionospheric winds of charged particles driven by solar heating. These winds flow in two elliptic cells, one on each day-side of the equator. Their position changes from day to day, causing a different appearance of the curves, $X_n(t)$, with changes in the amplitude being most pronounced. To simulate this behavior, $S(t)$ is introduced as the typical pattern for a specific magnetic observatory, $Z_n(t)$, as the change in shape on day n and the scalar random variable U_n as the amplitude on day n. With this motivation, we formulate the following example.

EXAMPLE 2.2 (Product model). Suppose $\{Y_n\} \in L_H^p$ and $\{U_n\} \in L^p$ are both L^p–m-approximable sequences, independent of each other. The respective representations are $Y_n = g(\eta_1, \eta_2, \ldots)$ and $U_n = h(\gamma_1, \gamma_2, \ldots)$. Each of these sequences could be a linear sequence satisfying the assumptions of Proposition 2.1, but they need not be. The sequence $X_n(t) = U_n Y_n(t)$ is then a nonlinear L^p–m-approximable sequence with the underlying i.i.d. variables $\varepsilon_n = (\eta_n, \gamma_n)$. This follows by after a slight modification from Lemma 2.1.

Example 2.2 illustrates the principle that in order for products of L^p–m-approximable sequences to be L^p–m-approximable, independence must be assumed. It does not have to be assumed as directly as in Example 2.2; the important point being that appropriately-defined functional Volterra expansions should not contain diagonal terms so that moments do not pile up. Such expansions exist (see, e.g., Giraitis, Kokoszka and Leipus [28], for all nonlinear scalar processes used to model financial data). The model $X_n(t) = Y_n(t)U_n$ is similar to the popular

scalar stochastic volatility model $r_n = v_n \varepsilon_n$ used to model returns r_n on a speculative asset. The dependent sequence $\{v_n\}$ models volatility, and the i.i.d. errors ε_n, independent of the v_n, generate unpredictability in returns.

Our next examples focus on functional extensions of popular nonlinear models, namely the bilinear model of [32] and the celebrated ARCH model of Engle [27]. Both models will be treated in more detail in forthcoming papers. Proofs of Propositions 2.2 and 2.3 are available upon request.

EXAMPLE 2.3 (Functional bilinear process). Let (ε_n) be an H-valued i.i.d. sequence and let $\psi \in H \otimes H$ and $\phi \in H \otimes H \otimes H$. Then the process defined as the recurrence equation,

$$X_{n+1}(t) = \int \psi(t,s) X_n(s)\, ds + \iint \phi(t,s,u) X_n(s) \varepsilon_n(u)\, ds\, du + \varepsilon_{n+1}(t),$$

is called *functional bilinear process*.

A neater notation can be achieved by defining $\psi : H \to H$, the kernel operator with the kernel function $\phi(t,s)$, and $\phi_n : H \to H$, the random kernel operator with kernel

$$\phi_n(t,s) = \int \phi(t,s,u) \varepsilon_n(u)\, du.$$

In this notation, we have

$$X_{n+1} = (\psi + \phi_n)(X_n) + \varepsilon_{n+1} \tag{2.10}$$

with the usual convention that $(A+B)(x) = A(x) + B(x)$ for operators A, B. The product of two operators $AB(x)$ is interpreted as successive application $A(B(x))$.

A formal solution to (2.10) is

$$X_{n+1} = \sum_{k=0}^{\infty} \prod_{j=0}^{k-1} (\psi + \phi_{n-j})(\varepsilon_{n+1-k}) \tag{2.11}$$

and the approximating sequence is defined by

$$\tilde{X}_n^{(m)} = \sum_{k=0}^{m} \prod_{j=0}^{k-1} (\psi + \phi_{n-j})(\varepsilon_{n+1-k}).$$

The following proposition establishes sufficient conditions for the L^p–m-approximability.

PROPOSITION 2.2. *Let $\{X_n\}$ be the functional bilinear process defined in (2.10). If $E \log \|\varepsilon_0\| < \infty$ and $E \log \|\psi + \phi_0\| < 0$, then a unique strictly stationary solution for this equation exists. The solution has (L^2-)representation (2.11). If $v_p((\psi + \phi_0)(\varepsilon_0)) < \infty$ and $E\|\psi + \phi_0\|_S^p < 1$, the process is L^p–m-approximable.*

EXAMPLE 2.4 (Functional ARCH). Let $\delta \in H$ be a positive function and let $\{\varepsilon_k\}$ an i.i.d. sequence in L_H^4. Further, let $\beta(s,t)$ be a nonnegative kernel function in $L^2([0,1]^2, \mathcal{B}^2_{[0,1]}, \lambda^2)$. Then we call the process

$$y_k(t) = \varepsilon_k(t)\sigma_k(t), \qquad t \in [0,1], \tag{2.12}$$

where

$$\sigma_k^2(t) = \delta(t) + \int_0^1 \beta(t,s) y_{k-1}^2(s)\, ds, \tag{2.13}$$

the *functional* ARCH(1) *process*.

Proposition 2.3 establishes conditions for the existence of a strictly stationary solution to (2.12) and (2.13) and its L^p–m-approximability.

PROPOSITION 2.3. *Define* $K(\varepsilon_1^2) = (\iint \beta^2(t,s)\varepsilon_1^4(s)\, ds\, dt)^{1/2}$. *If there is some* $p > 0$ *such that* $E\{K(\varepsilon_1^2)\}^p < 1$ *then* (2.12) *and* (2.13) *have a unique strictly stationary and causal solution and the sequence* $\{y_k\}$ *is* L^p–m*-approximable.*

3. Convergence of eigenvalues and eigenfunctions. Denote by $C = E[\langle X, \cdot\rangle X]$ the covariance operator of some $X \in L_H^2$. The eigenvalues and eigenfunctions of C are a fundamental ingredient for principal component analysis which is a key technique in functional data analysis. In practice, C and its eigenvalues/eigenfunctions are unknown and must be estimated. The purpose of this section is to prove consistency of the corresponding estimates for L^4-m-approximable sequences. The results derived below will be applied in the following sections. We start with some preliminary results.

Consider two compact operators $C, K \in \mathcal{L}$ with singular value decompositions

$$C(x) = \sum_{j=1}^{\infty} \lambda_j \langle x, v_j\rangle f_j, \qquad K(x) = \sum_{j=1}^{\infty} \gamma_j \langle x, u_j\rangle g_j. \tag{3.1}$$

The following lemma is proven in Section VI.1 of (see Gohberg, Golberg and Kaashoek [30], Corollary 1.6, page 99).

LEMMA 3.1. *Suppose* $C, K \in \mathcal{L}$ *are two compact operators with singular value decompositions* (3.1). *Then, for each* $j \geq 1$, $|\gamma_j - \lambda_j| \leq \|K - C\|_{\mathcal{L}}$.

We now tighten the conditions on the operator C by assuming that it is Hilbert–Schmidt, symmetric and positive definite. These conditions imply that $f_j = v_j$ in (3.1), $C(v_j) = \lambda_j v_j$ and $\sum_j \lambda_j^2 < \infty$. Consequently λ_j are eigenvalues of C and v_j the corresponding eigenfunctions. We also define

$$v_j' = \hat{c}_j v_j, \qquad \hat{c}_j = \operatorname{sign}(\langle u_j, v_j\rangle).$$

Using Lemma 3.1, the next lemma can be established by following the lines of the proof of Lemma 4.3 of Bosq [14].

LEMMA 3.2. *Suppose $C, K \in \mathcal{L}$ are two compact operators with singular value decompositions (3.1). If C is Hilbert–Schmidt, symmetric and positive definite, and its eigenvalues satisfy*

$$\lambda_1 > \lambda_2 > \cdots > \lambda_d > \lambda_{d+1}, \tag{3.2}$$

then

$$\|u_j - v_j'\| \le \frac{2\sqrt{2}}{\alpha_j}\|K - C\|_{\mathcal{L}}, \qquad 1 \le j \le d,$$

where $\alpha_1 = \lambda_1 - \lambda_2$ and $\alpha_j = \min(\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1}), 2 \le j \le d$.

Let $\{X_n\} \in L_H^2$ be a stationary sequence with covariance operator C. In principle we could now develop a general theory for H valued sequences, where H is an arbitrary separable Hilbert space. In practice, however, the case $H = L^2([0,1], \mathcal{B}_{[0,1]}, \lambda)$ is most important. In order to be able to fully use the structure of H and and not to deal with technical assumptions, we need the two basic regularity conditions below, which will be assumed throughout the paper without further notice.

ASSUMPTION 3.1. (i) Each X_n is measurable $(\mathcal{B}_{[0,1]} \times \mathcal{A})/\mathcal{B}_{\mathbb{R}}$.
(ii) $\sup_{t\in[0,1]} E|X(t)|^2 < \infty$.

Assumption 3.1(i) is necessary in order that the sample paths of X_n are measurable. Together with (ii) it also implies that C is an integral operator with kernel $c(t,s) = \mathrm{Cov}(X_1(t), X_1(s))$ whose estimator is

$$\hat{c}(t,s) = N^{-1}\sum_{n=1}^{N}\big(X_n(t) - \bar{X}_N(t)\big)\big(X_n(s) - \bar{X}_N(s)\big). \tag{3.3}$$

Then natural estimators of the eigenvalues λ_j and eigenfunctions v_j of C are the eigenvalues $\hat{\lambda}_j$ and eigenfunctions $\hat{v}_j$ of $\hat{C}$, the operator with the kernel (3.3). By Lemmas 3.1 and 3.2 we can bound the estimation errors for eigenvalues and eigenfunctions by $\|C - \hat{C}\|_{\mathcal{S}}^2$. Mas and Mennetau [42] show that transferring asymptotic results from the operators to the eigenelements holds quite generally, including a.s. convergence, weak convergence or large deviation principles. This motivates the next result.

THEOREM 3.1. *Suppose $\{X_n\} \in L_H^4$ is an L^4–m-approximable sequence with covariance operator C. Then there is some constant $U_X < \infty$, which does not depend on N, such that*

$$NE\|\hat{C} - C\|_{\mathcal{S}}^2 \le U_X. \tag{3.4}$$

If the X_n have zero mean, then we can choose

$$U_X = \nu_4^4(X) + 4\sqrt{2}\nu_4^3(X)\sum_{r=1}^{\infty}\nu_4(X_r - X_r^{(r)}). \tag{3.5}$$

The proof of Theorem 3.1 is given in Section A.1. Let us note that by Lemma 3.1 and Theorem 3.1,

$$NE[|\lambda_j - \hat{\lambda}_j|^2] \le NE\|\hat{C} - C\|_{\mathcal{L}}^2 \le NE\|\hat{C} - C\|_{\mathcal{S}}^2 \le U_X.$$

Assuming (3.2), by Lemma 3.2 and Theorem 3.1, $[\hat{c}_j = \operatorname{sign}(\langle \hat{v}_j, v_j\rangle)]$,

$$NE[\|\hat{c}_j\hat{v}_j - v_j\|^2] \le \left(\frac{2\sqrt{2}}{\alpha_j}\right)^2 NE\|\hat{C} - C\|_{\mathcal{L}}^2 \le \frac{8}{\alpha_j^2}NE\|\hat{C} - C\|_{\mathcal{S}}^2 \le \frac{8U_X}{\alpha_j^2}$$

with the α_j defined in Lemma 3.2.

These inequalities establish the following result.

THEOREM 3.2. *Suppose $\{X_n\} \in L_H^4$ is an L^4–m-approximable sequence and assumption (3.2) holds. Then, for $1 \le j \le d$,*

$$\limsup_{N\to\infty} NE[|\lambda_j - \hat{\lambda}_j|^2] < \infty, \qquad \limsup_{N\to\infty} NE[\|\hat{c}_j\hat{v}_j - v_j\|^2] < \infty. \tag{3.6}$$

Relations (3.6) have become a fundamental tool for establishing asymptotic properties of procedures for functional simple random samples which are based on the functional principal components. Theorem 3.2 shows that in many cases one can expect that these properties will remain the same under weak dependence; an important example is discussed in Section 6. The empirical covariance kernel (3.3) is, however, clearly designed for simple random samples, and may not be optimal for representing dependent data in the most "useful" way. The term "useful" depends on the application. Kargin and Onatski [38] show that a basis different than the eigenfunctions v_k is optimal for prediction with a functional AR(1) model. An interesting open problem is how to construct a basis optimal in some general sense for dependent data. In Section 4 we focus on a related, but different, problem of constructing a matrix which "soaks up" the dependence in a manner that allows the extension of many multivariate time series procedures to a functional setting. The construction of this matrix involves *arbitrary* basis vectors v_k estimated by $\hat{v}_k$ in such a way that (3.6) holds.

4. Estimation of the long-run variance. The main results of this section are Corollary 4.1 and Proposition 4.1 which state that the long-run variance matrix obtained by projecting the data on the functional principal components can be consistently estimated. The concept of the long-run variance, while fundamental in time series analysis, has not been studied for functional data, and not even for scalar

approximable sequences. It is therefore necessary to start with some preliminaries which lead to our main results and illustrate the role of the L^p–m-approximability.

Let $\{X_n\}$ be a scalar (weakly) stationary sequence. Its long-run variance is defined as $\sigma^2 = \sum_{j\in\mathbb{Z}} \gamma_j$, where $\gamma_j = \mathrm{Cov}(X_0, X_j)$, provided this series is absolutely convergent. Our first lemma shows that this is the case for L^2–m-approximable sequences.

LEMMA 4.1. *Suppose $\{X_n\}$ is a scalar L^2–m-approximable sequence. Then its autocovariance function $\gamma_j = \mathrm{Cov}(X_0, X_j)$ is absolutely summable, that is, $\sum_{j=-\infty}^{\infty} |\gamma_j| < \infty$.*

PROOF. Observe that for $j > 0$,

$$\mathrm{Cov}(X_0, X_j) = \mathrm{Cov}\big(X_0, X_j - X_j^{(j)}\big) + \mathrm{Cov}\big(X_0, X_j^{(j)}\big).$$

Since

$$X_0 = f(\varepsilon_0, \varepsilon_{-1}, \ldots), \qquad X_j^{(j)} = f^{(j)}\big(\varepsilon_j, \varepsilon_{j-1}, \ldots, \varepsilon_1, \varepsilon_0^{(j)}, \varepsilon_{-1}^{(j)}, \ldots\big),$$

the random variables X_0 and $X_j^{(j)}$ are independent, so $\mathrm{Cov}(X_0, X_j^{(j)}) = 0$, and

$$|\gamma_j| \le [EX_0^2]^{1/2}\big[E\big(X_j - X_j^{(j)}\big)^2\big]^{1/2}. \qquad \square$$

The summability of the autocovariances is the fundamental property of weak dependence because then $N\,\mathrm{Var}[\bar{X}_N] \to \sum_{j=-\infty}^{\infty} \gamma_j$; that is, the variance of the sample mean converges to zero at the rate N^{-1}, the same as for i.i.d. observations. A popular approach to the estimation of the long-run variance is to use the kernel estimator

$$\hat{\sigma}^2 = \sum_{|j|\le q} \omega_q(j)\hat{\gamma}_j, \qquad \hat{\gamma}_j = \frac{1}{N}\sum_{i=1}^{N-|j|} (X_i - \bar{X}_N)\big(X_{i+|j|} - \bar{X}_N\big).$$

Various weights $\omega_q(j)$ have been proposed and their optimality properties studied (see Anderson [1] and Andrews [3], among others). In theoretical work, it is typically assumed that the bandwith q is a deterministic function of the sample size such that $q = q(N) \to \infty$ and $q = o(N^r)$, for some $0 < r \le 1$. We will use the following assumption:

ASSUMPTION 4.1. The bandwidth $q = q(N)$ satisfies $q \to \infty, q^2/N \to 0$, and the weights satisfy $\omega_q(j) = \omega_q(-j)$ and

$$|\omega_q(j)| \le b \tag{4.1}$$

and, for every fixed j,

$$\omega_q(j) \to 1. \tag{4.2}$$

All kernels used in practice have symmetric weights and satisfy conditions (4.1) and (4.2).

The absolute summability of the autocovariances is not enough to establish the consistency of the kernel estimator $\hat{\sigma}^2$. Traditionally, summability of the cumulants has been assumed to control the fourth order structure of the data. Denoting $\mu = EX_0$, the fourth order cumulant of a stationary sequence is defined by

$$\kappa(h,r,s) = \operatorname{Cov}\big((X_0-\mu)(X_h-\mu),(X_r-\mu)(X_s-\mu)\big) - \gamma_r\gamma_{h-s} - \gamma_s\gamma_{h-r}.$$

The ususal sufficient condition for the consistency of $\hat{\sigma}$ is

$$\sum_{h=-\infty}^{\infty}\sum_{r=-\infty}^{\infty}\sum_{s=-\infty}^{\infty}|\kappa(h,r,s)| < \infty. \tag{4.3}$$

Recently, Giraitis et al. [29] showed that condition (4.3) can be replaced by a weaker condition,

$$\sup_h \sum_{r=-\infty}^{\infty}\sum_{s=-\infty}^{\infty}|\kappa(h,r,s)| < \infty. \tag{4.4}$$

A technical condition we need is

$$N^{-1}\sum_{k,l=0}^{q(N)}\sum_{r=1}^{N-1}\big|\operatorname{Cov}\big(X_0(X_k - X_k^{(k)}), X_r^{(r)}X_{r+\ell}^{(r+\ell)}\big)\big| \to 0. \tag{4.5}$$

By analogy to condition (4.4), it can be replaced by a much stronger, but a more transparent condition,

$$\sup_{k,l\geq 0}\sum_{r=1}^{\infty}\big|\operatorname{Cov}\big(X_0(X_k - X_k^{(k)}), X_r^{(r)}X_{r+\ell}^{(r+\ell)}\big)\big| < \infty. \tag{4.6}$$

To explain the intuition behind conditions (4.5) and (4.6), consider the linear process $X_k = \sum_{j=0}^{\infty} c_j X_{k-j}$. For $k \geq 0$,

$$X_k - X_k^{(k)} = \sum_{j=k}^{\infty} c_j\varepsilon_{k-j} - \sum_{j=k}^{\infty} c_j\varepsilon_{k-j}^{(k)}.$$

Thus $X_0(X_k - X_k^{(k)})$ depends on

$$\varepsilon_0, \varepsilon_{-1}, \varepsilon_{-2}, \ldots \quad \text{and} \quad \varepsilon_0^{(k)}, \varepsilon_{-1}^{(k)}, \varepsilon_{-2}^{(k)}, \ldots \tag{4.7}$$

and $X_r^{(r)}X_{r+\ell}^{(r+\ell)}$ depends on

$$\varepsilon_{r+\ell}, \ldots, \varepsilon_1, \varepsilon_0^{(r)}\varepsilon_{-1}^{(r)}, \varepsilon_{-2}^{(r)}, \ldots \quad \text{and} \quad \varepsilon_0^{(r+\ell)}\varepsilon_{-1}^{(r+\ell)}, \varepsilon_{-2}^{(r+\ell)}, \ldots.$$

Consequently, the covariances in (4.6) vanish except when $r = k$ or $r + \ell = k$, so condition (4.6) always holds for linear processes.

For general nonlinear sequences, the difference

$$X_k - X_k^{(k)} = f(\varepsilon_k, \ldots, \varepsilon_1, \varepsilon_0, \varepsilon_{-1}, \ldots) - f\big(\varepsilon_k, \ldots, \varepsilon_1, \varepsilon_0^{(k)}, \varepsilon_{-1}^{(k)}, \ldots\big),$$

cannot be expressed only in terms of the errors (4.7), but the errors $\varepsilon_k, \ldots, \varepsilon_1$ should approximately cancel, so that the difference $X_k - X_k^{(k)}$ is small and very weakly correlated with $X_r^{(r)} X_{r+\ell}^{(r+\ell)}$.

With this background, we now formulate the following result.

THEOREM 4.1. *Suppose $\{X_n\} \in L^4$ is an L^4–m-approximable and assume condition (4.5) holds. If Assumption 4.1 holds, then $\hat{\sigma}^2 \xrightarrow{P} \sum_{j=-\infty}^{\infty} \gamma_j$.*

Theorem 4.1 is proven in Section A.1. The general plan of the proof is the same as that of the proof of Theorem 3.1 of Giraitis et al. [29], but the verification of the crucial relation (A.5) uses a new approach based on L^4–m-approximability. The arguments preceding (A.5) show that replacing $\bar{X}_N$ by $\mu = EX_0$ does not change the limit. We note that the condition $q^2/N \to 0$ we assume is stronger than the condition $q/N \to 0$ assumed by Giraitis et al. [29]. This difference is of little practical consequence, as the optimal bandwidths for the kernels used in practice are typically of the order $O(N^{1/5})$. Finally, we notice that by further strengthening conditions on the behavior of the bandwidth function $q = q(N)$, the convergence in probability in Theorem 4.1 could be replaced by the almost sure convergence, but we do not pursue this research here. The corresponding result under condition (4.4) was established by Berkes et al. [9]; it is also stated without proof as part of Theorem A.1 of Berkes et al. [10].

We now turn to the vector case in which the data are of the form

$$\mathbf{X}_n = [X_{1n}, X_{2n}, \ldots, X_{dn}]^T, \qquad n = 1, 2, \ldots, N.$$

Just as in the scalar case, the estimation of the mean by the sample mean does not affect the limit of the kernel long-run variance estimators, so *we assume that* $EX_{in} = 0$ and define the autocovariances as

$$\gamma_r(i, j) = E[X_{i0} X_{jr}], \qquad 1 \le i, j \le d.$$

If $r \ge 0$, $\gamma_r(i, j)$ is estimated by $N^{-1} \sum_{n=1}^{N-r} X_{in} X_{j,n+r}$, but if $r < 0$, it is estimated by $N^{-1} \sum_{n=1}^{N-|r|} X_{i,n+|r|} X_{j,n}$. We therefore define the autocovariance matrices

$$\hat{\mathbf{\Gamma}}_r = \begin{cases} N^{-1} \displaystyle\sum_{n=1}^{N-r} \mathbf{X}_n \mathbf{X}_{n+r}^T, & \text{if } r \ge 0, \\ N^{-1} \displaystyle\sum_{n=1}^{N-|r|} \mathbf{X}_{n+|r|} \mathbf{X}_n^T, & \text{if } r < 0. \end{cases}$$

The variance $\operatorname{Var}[N^{-1}\bar{\mathbf{X}}_n]$ has (i, j)-entry

$$N^{-2}\sum_{m,n=1}^{N} E[X_{im}X_{jn}] = N^{-1}\sum_{|r|<N}\left(1-\frac{|r|}{N}\right)\gamma_r(i,j),$$

so the long-run variance is

$$\boldsymbol{\Sigma} = \sum_{r=-\infty}^{\infty} \boldsymbol{\Gamma}_r, \qquad \boldsymbol{\Gamma}_r := [\gamma_r(i,j), 1 \le i, j \le d],$$

and its kernel estimator is

$$\hat{\boldsymbol{\Sigma}} = \sum_{|r|\le q} \omega_q(r)\hat{\boldsymbol{\Gamma}}_r. \tag{4.8}$$

The consistency of $\hat{\boldsymbol{\Sigma}}$ can be established by following the lines of the proof of Theorem 4.1 for every fixed entry of the matrix $\hat{\boldsymbol{\Sigma}}$. Conditition (4.5) must be replaced by

$$N^{-1}\sum_{k,l=0}^{q(N)}\sum_{r=1}^{N-1}\max_{1\le i,j\le d}\left|\operatorname{Cov}\left(X_{i0}\left(X_{jk}-X_{jk}^{(k)}\right), X_{ir}^{(r)}X_{j,r+\ell}^{(r+\ell)}\right)\right| \to 0. \tag{4.9}$$

Condition (4.9) is analogous to cumulant conditions for vector processes which require summability of fourth order cross-cumulants of all scalar components (see, e.g., Andrews [3], Assumption A, page 823).

For ease of reference we state these results as a theorem.

THEOREM 4.2. (a) *If $\{\mathbf{X}_n\} \in L^2_{\mathbb{R}^d}$ is an L^2–m-approximable sequence, then the series $\sum_{r=-\infty}^{\infty}\boldsymbol{\Gamma}_r$ converges absolutely.* (b) *Suppose $\{\mathbf{X}_n\} \in L^4_{\mathbb{R}^d}$ an L^4–m-approximable sequence such that condition* (4.9) *holds. If Assumption* 4.1 *holds, then $\hat{\boldsymbol{\Sigma}} \xrightarrow{P} \boldsymbol{\Sigma}$.*

We are now able to turn to functional data. Suppose $\{X_n\} \in L^2_H$ is a zero mean sequence, and $v_1, v_2, \ldots, v_d$ is any set of orthonormal functions in H. Define $X_{in} = \int X_n(t)v_i(t)\,dt$, $\mathbf{X}_n = [X_{1n}, X_{2n}, \ldots, X_{dn}]^T$ and $\Gamma_r = \operatorname{Cov}(\mathbf{X}_0, \mathbf{X}_r)$. A direct verification shows that if $\{X_n\}$ is L^p–m-approximable, then so is the vector sequence $\{\mathbf{X}_n\}$. We thus obtain the following corollary.

COROLLARY 4.1. (a) *If $\{X_n\} \in L^2_H$ is an L^2–m-approximable sequence, then the series $\sum_{r=-\infty}^{\infty}\boldsymbol{\Gamma}_r$ converges absolutely.* (b) *If, in addition, $\{X_n\}$ is L^4–m-approximable and Assumption* 4.1 *and condition* (4.9) *hold, then $\hat{\boldsymbol{\Sigma}} \xrightarrow{P} \boldsymbol{\Sigma}$.*

In Corollary 4.1, the functions $v_1, v_2, \ldots, v_d$ form an arbitrary orthonormal deterministic basis. In many applications, a random basis consisting of the estimated

principal components $\hat{v}_1, \hat{v}_2, \ldots, \hat{v}_d$ is used. The scores with respect to this basis are defined by

$$\hat{\eta}_{\ell i} = \int (X_i(t) - \bar{X}_N(t))\hat{v}_\ell(t)\,dt, \qquad 1 \le \ell \le d.$$

To use the results established so far, it is convenient to decompose the stationary sequence $\{X_n\}$ into its mean and a zero mean process; that is, we set $X_n(t) = \mu(t) + Y_n(t)$, where $EY_n(t) = 0$. We introduce the unobservable quantities

$$\beta_{\ell n} = \int Y_n(t)v_\ell(t)\,dt, \qquad \hat{\beta}_{\ell n} = \int Y_n(t)\hat{v}_\ell(t)\,dt, \qquad 1 \le \ell \le d. \tag{4.10}$$

We then have the following proposition which will be useful in most statistical procedures for functional time series. An application to change point detection is developed in Section 5.

PROPOSITION 4.1. *Let* $\hat{\mathbf{C}} = \operatorname{diag}(\hat{c}_1, \ldots, \hat{c}_d)$, *with* $\hat{c}_i = \operatorname{sign}(\langle v_i, \hat{v}_i\rangle)$. *Suppose* $\{X_n\} \in L_H^4$ *is* L^4*–m-approximable and that* (3.2) *holds. Assume further that Assumption* 4.1 *holds with a stronger condition* $q^4/N \to 0$. *Then*

$$|\hat{\boldsymbol{\Sigma}}(\boldsymbol{\beta}) - \hat{\boldsymbol{\Sigma}}(\hat{\mathbf{C}}\hat{\boldsymbol{\beta}})| = o_P(1) \quad \textit{and} \quad |\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\eta}}) - \hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\beta}})| = o_P(1).$$

The proof of Proposition 4.1 is delicate and is presented in Section A.1. We note that condition (4.9) does not appear in the statement of Proposition 4.1. Its point is that if $\hat{\boldsymbol{\Sigma}}(\boldsymbol{\beta})$ is consistent under some conditions, then so is $\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\eta}})$.

5. Change point detection. Functional time series are obtained from data collected sequentially over time, and it is natural to expect that conditions under which observations are made may change. If this is the case, procedures developed for stationary series will produce spurious results. In this section, we develop a procedure for the detection of a change in the mean function of a functional time series, the most important possible change. In addition to its practical relevance, the requisite theory illustrates the application of the results developed in Sections 3 and 4. The main results of this Section, Theorems 5.1 and 5.2, are proven in Section A.2.

We thus consider testing the null hypothesis,

$$H_0 : EX_1(t) = EX_2(t) = \cdots = EX_N(t), \qquad t \in [0, 1].$$

Note that under H_0, we do not specify the value of the common mean.

Under the alternative, H_0 does not hold. The test we construct has a particularly good power against the alternative in which the data can be divided into several consecutive segments, and the mean is constant within each segment but changes from segment to segment. The simplest case of only two segments (one change

point) is specified in Assumption 5.2. First we note that under the null hypothesis, we can represent each functional observation as

$$X_i(t) = \mu(t) + Y_i(t), \qquad EY_i(t) = 0. \tag{5.1}$$

The following assumption specifies conditions on $\mu(\cdot)$ and the errors $Y_i(\cdot)$ needed to establish the convergence of the test statistic under H_0.

ASSUMPTION 5.1. The mean μ in (5.1) is in H. The error functions $Y_i \in L^4_H$ are L^4–m-approximable mean zero random elements such that the eigenvalues of their covariance operator satisfy (3.2).

Recall that the L^4–m-approximability implies that the Y_i are identically distributed with $\nu_4(Y_i) < \infty$. In particular, their covariance function,

$$c(t,s) = E[Y_i(t)Y_i(s)], \qquad 0 \le t, s \le 1,$$

is square integrable, that is, is in $L^2([0,1] \times [0,1])$.

We develop the theory under the alternative of exactly one change point, but the procedure is applicable to multiple change points by using a segmentation algorithm described in Berkes et al. [7] and dating back at least to Vostrikova [61].

ASSUMPTION 5.2. The observations follow the model

$$X_i(t) = \begin{cases} \mu_1(t) + Y_i(t), & 1 \le i \le k^*, \\ \mu_2(t) + Y_i(t), & k^* < i \le N, \end{cases}$$

in which the Y_i satisfy Assumption 5.1, the mean functions μ_1 and μ_2 are in L^2 and

$$k^* = [n\theta] \qquad \text{for some } 0 < \theta < 1.$$

The general idea of testing is similar to that developed in Berkes et al. [7] for independent observations, the central difficulty is in accommodating the dependence. To define the test statistic, recall that bold symbols denote d-dimensional vectors, for example, $\hat{\boldsymbol{\eta}}_i = [\hat\eta_{1i}, \hat\eta_{2i}, \ldots, \hat\eta_{di}]^T$. To lighten the notation, define the partial sums process, $\mathbf{S}_N(x, \boldsymbol{\xi}) = \sum_{n=1}^{\lfloor Nx \rfloor} \boldsymbol{\xi}_n$, $x \in [0,1]$, and the process, $\mathbf{L}_N(x, \boldsymbol{\xi}) = \mathbf{S}_N(x, \boldsymbol{\xi}) - x\mathbf{S}_N(1, \boldsymbol{\xi})$, where $\{\boldsymbol{\xi}_n\}$ is a generic R^d-valued sequence. Denote by $\boldsymbol{\Sigma}(\boldsymbol{\xi})$ the long-run variance of the sequence $\{\boldsymbol{\xi}_n\}$, and by $\hat{\boldsymbol{\Sigma}}(\boldsymbol{\xi})$ its kernel estimator (see Section 4). The proposed test statistic is then

$$T_N(d) = \frac{1}{N} \int_0^1 \mathbf{L}_N(x, \hat{\boldsymbol{\eta}})^T \hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\eta}})^{-1} \mathbf{L}_N(x, \hat{\boldsymbol{\eta}})\, dx. \tag{5.2}$$

Our first theorem establishes its asymptotic null distribution.

THEOREM 5.1. *Suppose H_0 and Assumption 5.1 hold. If the estimator $\hat{\mathbf{\Sigma}}(\hat{\boldsymbol{\eta}})$ is consistent, then*

$$T_N(d) \xrightarrow{d} T(d) := \sum_{\ell=1}^{d} \int_0^1 B_\ell^2(x)\,dx, \tag{5.3}$$

where $\{B_\ell(x), x \in [0,1]\}$, $1 \le \ell \le d$ are independent Brownian bridges.

The distribution of the random variable $T(d)$ was derived by Kiefer [39]. The limit distribution is the same as in the case of independent observations; this is possible because the long-run variance estimator $\hat{\mathbf{\Sigma}}(\hat{\boldsymbol{\eta}})$ soaks up the dependence. Sufficient conditions for its consistency are stated in Section 4, and, in addition to the assumptions of Theorem 5.1, they are: Assumption 4.1 with $q^4/N \to 0$, and condition (4.9).

The next result shows that our test has asymptotic power 1. Our proof requires the following condition:

$$\hat{\mathbf{\Sigma}}(\hat{\boldsymbol{\eta}}) \xrightarrow{\text{a.s.}} \mathbf{\Omega} \qquad \text{where } \mathbf{\Omega} \text{ is some positive definite matrix.} \tag{5.4}$$

Condition (5.4) could be replaced by weaker technical conditions, but we prefer it, as it leads to a transparent, short proof. Essentially, it states that the matrix $\hat{\mathbf{\Sigma}}(\hat{\boldsymbol{\eta}})$ does not become degenerate in the limit, and the matrix $\mathbf{\Omega}$ has only positive eigenvalues. A condition like (5.4) is not needed for independent Y_i because that case does not require normalization with the long-run covariance matrix. To formulate our result, introduce vectors $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \in \mathbb{R}^d$ with coordinates

$$\int \mu_1(t) v_\ell(t)\,dt \quad \text{and} \quad \int \mu_2(t) v_\ell(t)\,dt, \qquad 1 \le \ell \le d.$$

THEOREM 5.2. *Suppose Assumption 5.2 and condition (5.4) hold. If the vectors $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are not equal, then $T_N(d) \xrightarrow{P} \infty$.*

We conclude this section with two numerical examples which illustrate the effect of dependence on our change point detection procedure. Example 5.1 uses synthetic data while Example 5.2 focuses on particulate pollution data. Both show that using statistic (5.2) with $\hat{\mathbf{\Sigma}}(\hat{\boldsymbol{\eta}})$ being the estimate for just the covariance, not the long-run covariance matrix, leads to spurious rejections of H_0, a nonexistent change point can be detected with a large probability.

EXAMPLE 5.1. We simulate 200 observations of the functional AR(1) process of Example 2.1, when Ψ has the parabolic integral kernel $\psi(t,s) = \gamma \cdot (2 - (2x - 1)^2 - (2y - 1)^2)$. We chose the constant γ such that $\|\Psi\|_{\mathcal{S}} = 0.6$ (the Hilbert–Schmidt norm). The innovations $\{\varepsilon_n\}$ are standard Brownian bridges. The first 3 principal components explain approximately 85% of the total variance, so we

compute the test statistic $T_{200}(3)$ given in (5.2). For the estimation of the long-run covariance matrix $\mathbf{\Sigma}$ we use the Bartlett kernel

$$\omega_q^{(1)}(j) = \begin{cases} 1 - |j|/(1+q), & \text{if } |j| \le q; \\ 0, & \text{otherwise.} \end{cases}$$

We first let $q = 0$ which corresponds to using just the sample covariance of $\{\hat{\boldsymbol{\eta}}_n\}$ in the normalization for the test statistic (5.2) (dependence is ignored). We use 1000 replications and the 5% confidence level. The rejection rate is 23.9%, much higher than the nominal level of 5%. In contrast, using an appropriate estimate for the long-run variance, the reliability of the test improves dramatically. Choosing an optimal bandwidth q is a separate problem which we do not pursue here. Here we adapt the formula $q \approx 1.1447(aN)^{1/3}$, $a = \frac{4\psi^2}{(1+\psi)^4}$ valid for a a scalar AR(1) process with the autoregressive coefficient ψ (Andrews [3]). Using this formula with $\psi = \|\Psi\|_{\mathcal{S}} = 0.6$ results in $q = 4$. This choice gives the empirical rejection rate of 3.7%, much closer to the nominal rate of 5%.

EXAMPLE 5.2. This example, which uses `pm10` (particulate matter with diameter < 10 μm, measured in μg/m^3) data, illustrates a similar phenomenon as Example 5.1. For the analysis we use `pm10` concentration data measured in the Austrian city of Graz during the winter of 2008/2009 (N=151). The data are given in 30 minutes resolution, yielding an intraday frequency of 48 observations. As in Stadtlober, Hörmann and Pfeiler [58] we use a square root transformation to reduce heavy tails. Next we remove possible weekly periodicity by subtracting the corresponding mean vectors obtained from the different weekdays. A time series plot of this new sequence is given in Figure 2. The data look relatively stable, although a shift appears to be possible in the center of the time series. It should be

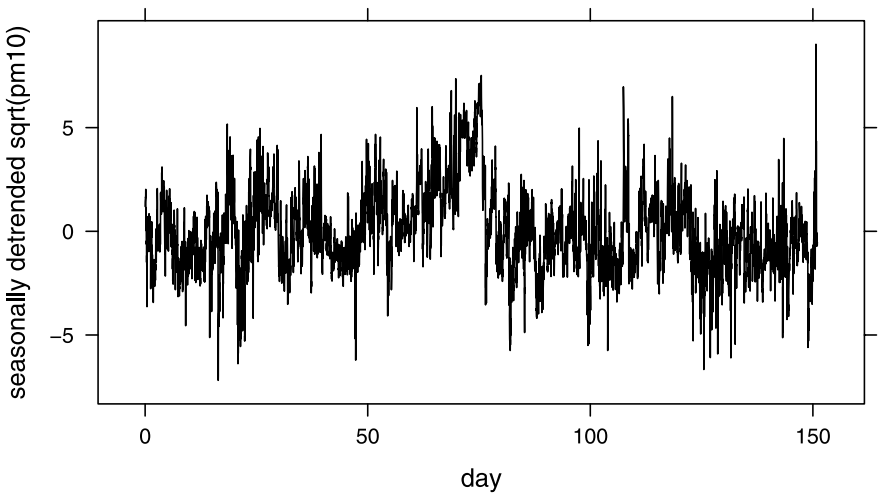

FIG. 2. *Seasonally detrended* $\sqrt{\texttt{pm10}}$, *Nov* 1, 2008–*Mar* 31, 2009.

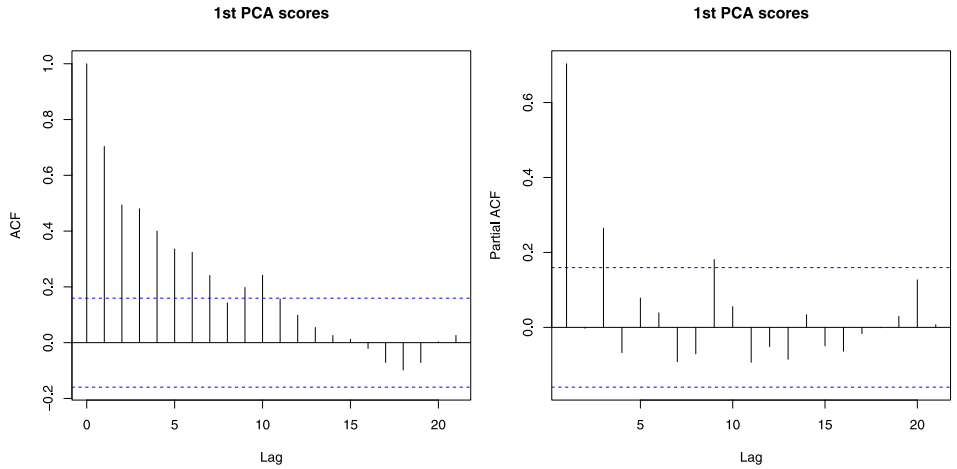

FIG. 3. *Left panel*: *sample autocorrelation function of the first empirical PC scores*. *Right panel*: *sample partial autocorrelation function of the first empirical PC scores*.

emphasized, however, that `pm10` data, like many geophysical time series, exhibit a strong, persistent, positive autocorrelation structure. These series are stationary over long periods of time with an appearance of local trends or shifts at various time scales (random self-similar or fractal structure).

The daily measurement vectors are transformed into smooth functional data using 15 B-splines functions of order 4. The functional principal component analysis yields that the first three principal components explain $\approx 84\%$ of the total variability, so we use statistic (5.2) with $d = 3$. A look at the `acf` and `pacf` of the first empirical PC scores (Figure 3) suggests an AR(1), maybe AR(3) behavior. The second and third empirical PC scores show no significant autocorrelation structure. We use the formula given in Example 5.1 with $\psi = 0.70$ (`acf` at lag 1) and $N = 151$ and obtain $q \approx 4$. This gives $T_{151}(3) = 0.94$ which is close to the critical value 1.00 when testing at a 95% confidence level but does not support rejection of the no-change hypothesis. In contrast, using only the sample covariance matrix in (5.3) gives $T_{151}(3) = 1.89$ and thus a clear and possibly wrongful rejection of the null hypothesis.

6. Functional linear model with dependent regressors. The functional linear model is one of the most widely used tools of FDA. Its various forms are introduced in Chapters 12–17 of Ramsay and Silverman [50]. To name a few recent references we mention Cuevas, Febrero and Fraiman [22], Malfait and Ramsay [41], Cardot et al. [18], Cardot, Ferraty and Sarda [19], Chiou, Müller and Wang [21], Müller and Stadtmüller [46], Yao, Müller and Wang [65], Cai and Hall [17], Chiou and Müller [20], Li and Hsing [40], Reiss and Ogden [51], Reiss and Ogden [52, 53].

We focus on the fully functional model of the form

$$Y_n(t) = \int \psi(t,s)X_n(s) + \varepsilon_n(t), \qquad n = 1, 2 \ldots, N, \tag{6.1}$$

in which both the regressors and the responses are functions. The results of this section can be easily specialized to the case of scalar responses.

In (6.1), the regressors are random functions, assumed to be independent and identically distributed. As explained in Section 1, for functional time series the assumption of the independence of the X_n is often questionable, so it is important to investigate if procedures developed and theoretically justified for independent regressors can still be used if the regressors are dependent.

We focus here on the estimation of the kernel $\psi(t,s)$. Our result is motivated by the work of Yao, Müller and Wang [65] who considered functional regressors and responses obtained from sparce *independent* data measured with error. The data that motivates our work are measurements of physical quantities obtained with negligible errors or financial transaction data obtained without error. In both cases the data are available at fine time grids, and the main concern is the presence of temporal dependence between the curves X_n. We therefore merely assume that the sequence $\{X_n\} \in L^4_H$ is L^4–m-approximable, which, as can be easily seen, implies the L^4–m-approximability of $\{Y_n\}$. To formulate additional technical assumptions, we need to introduce some notation.

We assume that the errors ε_n are i.i.d. and independent of the X_n, and denote by X and Y random functions with the same distribution as X_n and Y_n, respectively. We work with their expansions

$$X(s) = \sum_{i=1}^{\infty} \xi_i v_i(s), \qquad Y(t) = \sum_{j=1}^{\infty} \zeta_j u_j(t),$$

where the v_j are the FPCs of X and the u_j the FPCs of Y, and $\xi_i = \langle X, v_i\rangle$, $\zeta_j = \langle Y, u_j\rangle$. Indicating with the "hat" the corresponding empirical quantities, an estimator of $\psi(t,s)$ proposed by Yao, Müller and Wang [65] is

$$\hat{\psi}_{KL}(t,s) = \sum_{k=1}^{K}\sum_{\ell=1}^{L} \hat{\lambda}_\ell^{-1}\hat{\sigma}_{\ell k}\hat{u}_k(t)\hat{v}_\ell(s),$$

where $\hat{\sigma}_{\ell k}$ is an estimator of $E[\xi_\ell \zeta_k]$. We will work with the simplest estimator,

$$\hat{\sigma}_{\ell k} = \frac{1}{N}\sum_{i=1}^{N} \langle X_i, \hat{v}_\ell\rangle\langle Y_i, \hat{u}_k\rangle, \tag{6.2}$$

but any estimator for which Lemma A.1 holds can be used without affecting the rates.

Let λ_j and γ_j be the eigenvalues corresponding to v_j and u_j. Define α_j as in Lemma 3.2, and define α'_j accordingly with γ_j instead of λ_j. Set

$$h_L = \min\{\alpha_j, 1 \le j \le L\}, \qquad h'_L = \min\{\alpha'_j, 1 \le j \le L\}.$$

To establish the consistency of the estimator $\hat{\psi}_{KL}(t,s)$ we assume that

$$\Psi := \sum_{k=1}^{\infty}\sum_{\ell=1}^{\infty}\frac{(E[\xi_\ell \zeta_k])^2}{\lambda_\ell^2} < \infty \tag{6.3}$$

and that the following assumption holds:

ASSUMPTION 6.1. (i) We have $\lambda_1 > \lambda_2 > \cdots$ and $\gamma_1 > \gamma_2 > \cdots$.
(ii) We have $K = K(N)$, $L = L(N) \to \infty$ and $\frac{KL}{\lambda_L \min\{h_K, h'_L\}} = o(N^{1/2})$.

For model (6.1), condition (6.3) is equivalent to the assumption that $\psi(t,s)$ is a Hilbert–Schmidt kernel, that is, $\int\int \psi^2(t,s)\,dt\,ds < \infty$. It is formulated in the same way as in Yao, Müller and Wang [65] because this form is convenient in the theoretical arguments. Assumption 6.1 is much shorter than the corresponding assumptions of Yao, Müller and Wang [65] which take up over two pages. This is because we do not deal with smoothing and so can isolate the impact of the magnitude of the eigenvalues on the bandwidths K and L.

THEOREM 6.1. *Suppose $\{X_n\} \in L_H^4$ is a zero mean L^4–m-approximable sequence independent of the sequence of i.i.d. errors $\{\varepsilon_n\}$. If (6.3) and Assumption 6.1 hold, then*

$$\int\int [\hat{\psi}_{KL}(t,s) - \psi(t,s)]^2\,dt\,ds \xrightarrow{P} 0, \qquad (N \to \infty). \tag{6.4}$$

The proposition of Theorem 6.1 is comparable to the first part of Theorem 1 in Yao, Müller and Wang [65]. Both theorems are established under (6.3) and finite fourth moment conditions. Otherwise the settings are quite different. Yao, Müller and Wang [65] work under the assumption that the subject (Y_i, X_i), $i = 1, 2, \ldots$ are independent and sparsely observed whereas the crucial point of our approach is that we allow dependence. Thus Theorems 1 and 2 in the related paper Yao, Müller and Wang [66], which serve as the basic ingredients for their results, cannot be used here and have to be replaced directly with the theory developed in Section 3 of this paper. Furthermore, our proof goes without complicated assumptions on the resolvents of the covariance operator, in particular without the very technical assumptions (B.5) of Yao, Müller and Wang [65]. In this sense, our short alternative proof might be of value even in the case of independent observations.

APPENDIX

We present the proofs of results stated in Sections 3–6. Throughout we will agree on the following conventions. All $X_n \in L_H^2$ satisfy Assumption 3.1. A generic X, which is assumed to be equal in distribution to X_1, will be used at some places. Any constants occurring will be denoted by $\kappa_1, \kappa_2, \ldots$. The κ_i may change their values from proof to proof.

A.1. Proofs of the results of Sections 3 and 4.

PROOF OF THEOREM 3.1. We assume for simplicity that $EX = 0$ and set

$$\hat{c}(t,s) = N^{-1}\sum_{n=1}^{N} X_n(t)X_n(s), \qquad c(t,s) = E[X(t)X(s)].$$

The proof with a general mean function $\mu(t)$ requires some additional but similar arguments. The Cauchy–Schwarz inequality shows that $\hat{c}(\cdot,\cdot)$ and $c(\cdot,\cdot)$ are Hilbert–Schmidt kernels, so $\hat{C} - C$ is a Hilbert–Schmidt operator with the kernel $\hat{c}(t,s) - c(t,s)$. Consequently,

$$NE\|\hat{C} - C\|_{\mathcal{S}}^2 = N\iint \operatorname{Var}\left[N^{-1}\sum_{n=1}^{N}\bigl(X_n(t)X_n(s) - E[X_n(t)X_n(s)]\bigr)\right]dt\,ds.$$

For fixed s and t, set $Y_n = X_n(t)X_n(s) - E[X_n(t)X_n(s)]$. Due the stationarity of the sequence $\{Y_n\}$ we have

$$\operatorname{Var}\left(N^{-1}\sum_{n=1}^{N} Y_n\right) = N^{-1}\sum_{|r|<N}\left(1 - \frac{|r|}{N}\right)\operatorname{Cov}(Y_1, Y_{1+r})$$

and so

$$N\operatorname{Var}\left(N^{-1}\sum_{n=1}^{N} Y_n\right) \le \operatorname{Var}(Y_1) + 2\sum_{r=1}^{\infty}|\operatorname{Cov}(Y_1, Y_{1+r})|.$$

Setting $Y_n^{(m)} = X_n^{(m)}(t)X_n^{(m)}(s) - E[X_n(t)X_n(s)]$, we obtain

$$|\operatorname{Cov}(Y_1, Y_{1+r})| = \bigl|\operatorname{Cov}\bigl(Y_1, Y_{1+r} - Y_{1+r}^{(r)}\bigr)\bigr| \le [\operatorname{Var}(Y_1)]^{1/2}\bigl[\operatorname{Var}\bigl(Y_{1+r} - Y_{1+r}^{(r)}\bigr)\bigr]^{1/2}.$$

Consequently, $NE\|\hat{C} - C\|_{\mathcal{S}}^2$ is bounded from above by

$$\begin{aligned}&\iint \operatorname{Var}[X(t)X(s)]\,dt\,ds\\ &\quad + 2\sum_{r=1}^{\infty}\iint [\operatorname{Var}(X(t)X(s))]^{1/2}\\ &\qquad\qquad \times \bigl[\operatorname{Var}\bigl(X_{1+r}(t)X_{1+r}(s) - X_{1+r}^{(r)}(t)X_{1+r}^{(r)}(s)\bigr)\bigr]^{1/2}dt\,ds.\end{aligned}$$

For the first summand we have the upper bound $\nu_4^4(X)$ because

$$\iint E[X^2(t)X^2(s)]\,dt\,ds = E\int X^2(t)\,dt\int X^2(s)\,ds = \nu_4^4(X). \tag{A.1}$$

To find upper bounds for the summands in the infinite sum, we use the inequality

$$|ab - cd|^2 \le 2a^2(b-d)^2 + 2d^2(a-c)^2, \tag{A.2}$$

which yields

$$\begin{aligned}
&\iint [\operatorname{Var}(X(t)X(s))]^{1/2}\big[\operatorname{Var}\big(X_{1+r}(t)X_{1+r}(s) - X_{1+r}^{(r)}(t)X_{1+r}^{(r)}(s)\big)\big]^{1/2}\,dt\,ds \\
&\quad\le \iint [E(X^2(t)X^2(s))]^{1/2}\big[E\big(X_{1+r}(t)X_{1+r}(s) \\
&\qquad\qquad - X_{1+r}^{(r)}(t)X_{1+r}^{(r)}(s)\big)^2\big]^{1/2}\,dt\,ds \\
&\quad\le \sqrt{2}\iint [E(X^2(t)X^2(s))]^{1/2}\big[EX_{1+r}^2(t)\big(X_{1+r}(s) - X_{1+r}^{(r)}(s)\big)^2\big]^{1/2}\,dt\,ds \\
&\qquad + \sqrt{2}\iint [E(X^2(t)X^2(s))]^{1/2} \\
&\qquad\qquad \times \big[EX_{1+r}^{(r)2}(s)\big(X_{1+r}(t) - X_{1+r}^{(r)}(t)\big)^2\big]^{1/2}\,dt\,ds.
\end{aligned}$$

For the first term, using the Cauchy–Schwarz inequality and (A.1), we obtain

$$\begin{aligned}
&\iint [E(X^2(t)X^2(s))]^{1/2}\big[EX_{1+r}^2(t)\big(X_{1+r}(s) - X_{1+r}^{(r)}(s)\big)^2\big]^{1/2}\,dt\,ds \\
&\quad\le \nu_4^2(X)\left\{E\left[\int X_{1+r}^2(t)\,dt \int \big(X_{1+r}(s) - X_{1+r}^{(r)}(s)\big)^2\,ds\right]\right\}^{1/2} \\
&\quad\le \nu_4^2(X)\{E\|X_{1+r}\|^4\}^{1/4}\big\{E\big\|X_{1+r}(s) - X_{1+r}^{(r)}(s)\big\|\big\}^{1/4} \\
&\quad= \nu_4^3(X)\nu_4\big(X_1 - X_1^{(r)}\big).
\end{aligned}$$

The exact same argument applies for the second term. The above bounds imply (3.4). □

PROOF OF THEOREM 4.1. As in Giraitis et al. [29], set $\mu = EX_0$ and

$$\tilde{\gamma}_j = \frac{1}{N}\sum_{i=1}^{N-|j|}(X_i - \mu)(X_{i+|j|} - \mu),$$

$$S_{k,\ell} = \sum_{i=k}^{\ell}(X_i - \mu).$$

Observe that

$$\hat{\gamma}_j - \tilde{\gamma}_j = \left(1 - \frac{|j|}{N}\right)(\bar{X}_N - \mu)^2 + \frac{1}{N}(\bar{X}_N - \mu)\big(S_{1,N-|j|} + S_{|j|+1,N}\big) =: \delta_j.$$

We therefore have the decomposition

$$\hat{\sigma}^2 = \sum_{|j|\le q}\omega_q(j)\tilde{\gamma}_j + \sum_{|j|\le q}\omega_q(j)\delta_j =: \hat{\sigma}_1^2 + \hat{\sigma}_2^2.$$

The proof will be complete once we have shown that

$$\hat{\sigma}_1^2 \xrightarrow{P} \sum_{j=-\infty}^{\infty} \gamma_j \tag{A.3}$$

and

$$\hat{\sigma}_2^2 \xrightarrow{P} 0. \tag{A.4}$$

We begin with the verification of the easier relation (A.4). By (4.1),

$$\begin{aligned} E|\hat{\sigma}_2^2| &\le b \sum_{|j|\le q} E|\delta_j| \\ &\le b \sum_{|j|\le q} E(\bar{X}_N - \mu)^2 \\ &\quad + \frac{b}{N}[E(\bar{X}_N - \mu)^2]^{1/2} \sum_{|j|\le q} \left[E\left(S_{1,N-|j|} + S_{|j|+1,N}\right)^2\right]^{1/2}. \end{aligned}$$

By Lemma 4.1,

$$E(\bar{X}_N - \mu)^2 = \frac{1}{N} \sum_{|j|\le N} \left(1 - \frac{|j|}{N}\right)\gamma_j = O(N^{-1}).$$

Similarly $E(S_{1,N-|j|} + S_{|j|+1,N})^2 = O(N)$. Therefore,

$$E|\hat{\sigma}_2^2| = O(qN^{-1} + N^{-1}N^{-1/2}qN^{1/2}) = O(q/N).$$

We now turn to the verification of (A.3). We will show that $E\hat{\sigma}_1^2 \to \sum_j \gamma_j$ and $\mathrm{Var}[\hat{\sigma}_1^2] \to 0$.

By (4.2),

$$E\hat{\sigma}_1^2 = \sum_{|j|\le q} \omega_q(j) \frac{N - |j|}{N} \gamma_j \to \sum_{j=-\infty}^{\infty} \gamma_j.$$

By (4.1), it remains to show that

$$\sum_{|k|,|\ell|\le q} |\mathrm{Cov}(\tilde{\gamma}_k, \tilde{\gamma}_\ell)| \to 0. \tag{A.5}$$

To lighten the notation, without any loss of generality, *we assume from now on that* $\mu = 0$, so that

$$\mathrm{Cov}(\tilde{\gamma}_k, \tilde{\gamma}_\ell) = \frac{1}{N^2} \mathrm{Cov}\left(\sum_{i=1}^{N-|k|} X_i X_{i+|k|}, \sum_{j=1}^{N-|\ell|} X_j X_{j+|\ell|} \right).$$

Therefore, by stationarity,

$$|\mathrm{Cov}(\tilde{\gamma}_k, \tilde{\gamma}_\ell)| \le \frac{1}{N^2} \sum_{i,j=1}^{N} |\mathrm{Cov}(X_i X_{i+|k|}, X_j X_{j+|\ell|})| = \frac{1}{N} \sum_{|r|<N} \left(1 - \frac{|r|}{N}\right) |\mathrm{Cov}(X_0 X_{|k|}, X_r X_{r+|\ell|})|.$$

The last sum can be split into three terms corresponding to $r = 0$, $r < 0$ and $r > 0$.

The contribution to the left-hand side of (A.5) of the term corresponding to $r = 0$ is

$$N^{-1} \sum_{|k|,|\ell|\le q} |\mathrm{Cov}(X_0 X_{|k|}, X_0 X_{|\ell|})| = O(q^2/N).$$

The terms corresponding to $r < 0$ and $r > 0$ are handled in the same way, so we focus on the contribution of the summands with $r > 0$ which is

$$N^{-1} \sum_{|k|,|\ell|\le q} \sum_{r=1}^{N-1} \left(1 - \frac{r}{N}\right) |\mathrm{Cov}(X_0 X_{|k|}, X_r X_{r+|\ell|})|.$$

We now use the decompositions

$$\mathrm{Cov}(X_0 X_{|k|}, X_r X_{r+|\ell|}) = \mathrm{Cov}(X_0 X_{|k|}, X_r^{(r)} X_{r+|\ell|}^{(r+|\ell|)}) + \mathrm{Cov}(X_0 X_{|k|}, X_r X_{r+|\ell|} - X_r^{(r)} X_{r+|\ell|}^{(r+|\ell|)})$$

and

$$\mathrm{Cov}(X_0 X_{|k|}, X_r^{(r)} X_{r+|\ell|}^{(r+|\ell|)}) = \mathrm{Cov}(X_0 X_{|k|}^{(|k|)}, X_r^{(r)} X_{r+|\ell|}^{(r+|\ell|)}) + \mathrm{Cov}(X_0 (X_{|k|} - X_{|k|}^{(|k|)}), X_r^{(r)} X_{r+|\ell|}^{(r+|\ell|)}).$$

By Definition 2.1, X_0 depends on $\varepsilon_0, \varepsilon_{-1}, \ldots$ while the random variables $X_{|k|}^{(k)}, X_r^{(r)}$ and $X_{r+|\ell|}^{(r+|\ell|)}$ depend on $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{k\vee(r+|\ell|)}$ and errors independent of the ε_i. Therefore $\mathrm{Cov}(X_0 X_{|k|}^{(|k|)}, X_r^{(r)} X_{r+|\ell|}^{(r+|\ell|)})$ is equal to

$$E[X_0 X_{|k|}^{(|k|)} X_r^{(r)} X_{r+|\ell|}^{(r+|\ell|)}] - E[X_0 X_{|k|}] E[X_r^{(r)} X_{r+|\ell|}^{(r+|\ell|)}] = E[X_0] E[X_{|k|}^{(|k|)} X_r^{(r)} X_{r+|\ell|}^{(r+|\ell|)}] - E[X_0] E[X_{|k|}^{(|k|)}][X_r^{(r)} X_{r+|\ell|}^{(r+|\ell|)}] = 0.$$

We thus obtain

$$\mathrm{Cov}(X_0 X_{|k|}, X_r X_{r+|\ell|}) = \mathrm{Cov}(X_0 (X_{|k|} - X_{|k|}^{(|k|)}), X_r^{(r)} X_{r+|\ell|}^{(r+|\ell|)}) + \mathrm{Cov}(X_0 X_{|k|}, X_r X_{r+|\ell|} - X_r^{(r)} X_{r+|\ell|}^{(r+|\ell|)}).$$

By Assumption (4.5), it remains to verify that

$$N^{-1}\sum_{|k|,|\ell|\le q}\sum_{r=1}^{N-1}\bigl|\operatorname{Cov}\bigl(X_0X_{|k|},\,X_rX_{r+|\ell|}-X_r^{(r)}X_{r+|\ell|}^{(r+|\ell|)}\bigr)\bigr|\to 0.$$

This is done using the technique introduced in the proof of Theorem 3.1. By the Cauchy–Schwarz inequality, the problem reduces to showing that

$$N^{-1}\sum_{|k|,|\ell|\le q}\sum_{r=1}^{N-1}\bigl\{E\bigl[X_0^2X_{|k|}^2\bigr]\bigr\}^{1/2}\bigl\{E\bigl[\bigl(X_rX_{r+|\ell|}-X_r^{(r)}X_{r+|\ell|}^{(r+|\ell|)}\bigr)^2\bigr]\bigr\}^{1/2}\to 0.$$

Using (A.2), this in turn is bounded by constant times

$$N^{-1}\sum_{|k|,|\ell|\le q}\sum_{r=1}^{\infty}\bigl\{E\bigl[X_r-X_r^{(r)}\bigr]^4\bigr\}^{1/4},$$

which tends to zero by L^4–m-approximability and the condition $q^2/N\to 0$. □

PROOF OF PROPOSITION 4.1. We only show the first part, the second is similar. Let $\omega_q(h)$ be the Bartlett estimates satisfying Assumption 4.1. Without loss of generality we will assume below that the constant b in (4.1) is 1. Then the element in the kth row and ℓth column of $\hat{\mathbf{\Sigma}}(\boldsymbol{\beta})-\hat{\mathbf{\Sigma}}(\hat{\mathbf{C}}\hat{\boldsymbol{\beta}})$ is

$$\begin{aligned}
&\sum_{|h|\le q}\frac{\omega_q(h)}{N}\sum_{1\le n\le N-|h|}\bigl(\beta_{kn}\beta_{\ell,n+|h|}-\hat c_k\hat\beta_{kn}\hat c_\ell\hat\beta_{\ell,n+|h|}\bigr)\\
&\quad=\sum_{|h|\le q}\frac{\omega_q(h)}{N}\sum_{1\le n\le N-|h|}\beta_{kn}\bigl(\beta_{\ell,n+|h|}-\hat c_\ell\hat\beta_{\ell,n+|h|}\bigr)\\
&\qquad+\sum_{|h|\le q}\frac{\omega_q(h)}{N}\sum_{1\le n\le N-|h|}\hat c_\ell\hat\beta_{\ell,n+|h|}(\beta_{kn}-\hat c_k\hat\beta_{kn})\\
&\quad=F_1(N,k,\ell)+F_2(N,k,\ell).
\end{aligned}$$

For reasons of symmetry it is enough to estimate $F_1(N,k,\ell)$. We have for any $t_N>0$

$$\begin{aligned}
&P\bigl(|F_1(N,k,\ell)|>\varepsilon\bigr)\\
&\quad\le\sum_{|h|\le q}P\biggl(\frac{\omega_q(h)}{N}\sum_{1\le n\le N-|h|}\beta_{kn}\bigl(\beta_{\ell,n+|h|}-\hat c_\ell\hat\beta_{\ell,n+|h|}\bigr)>\frac{\varepsilon}{2q+1}\biggr)\\
&\quad\le\sum_{|h|\le q}P\biggl(\sum_{1\le n\le N-|h|}\beta_{kn}^2\sum_{1\le n\le N-|h|}\bigl(\beta_{\ell,n+|h|}-\hat c_\ell\hat\beta_{\ell,n+|h|}\bigr)^2>\frac{\varepsilon^2N^2}{(2q+1)^2}\biggr)\\
&\quad\le(2q+1)P\biggl(\sum_{1\le n\le N}\beta_{kn}^2>N(2q+1)t_N\biggr)
\end{aligned}$$

$$+ (2q+1)P\Big(\sum_{1\le n\le N}(\beta_{\ell n}-\hat{c}_\ell\hat{\beta}_{\ell n})^2 > \frac{\varepsilon^2 N}{t_N(2q+1)^3}\Big)$$
$$= (2q+1)\big(P_1(k,N)+P_2(\ell,N)\big).$$

By the Markov inequality and the fact that the β_{kn}, $1\le n\le N$, are identically distributed, we get for all $k\in\{1,\dots,d\}$

$$(2q+1)P_1(k,N) \le \frac{E(\beta_{k1}^2)}{t_N} \le \frac{E\|Y_1\|^2}{t_N},$$

which tends to zero as long as $t_N\to\infty$.

The estimation of $P_2(\ell,N)$ requires a little bit more effort. We notice first that

$$\limsup_{N\to\infty}\frac{1}{N}\operatorname{Var}\Big(\sum_{1\le n\le N}\|Y_n\|^2\Big) \le \sum_{h\in\mathbb{Z}}|\operatorname{Cov}(\|Y_1\|^2,\|Y_h\|^2)| < \infty. \tag{A.6}$$

The summability of the latter series follows by now routine estimates from (2.3). For any $x,y>0$ we have

$$\begin{aligned}
&P\Big(\sum_{1\le n\le N}(\beta_{\ell n}-\hat{c}_\ell\hat{\beta}_{\ell n})^2 > x\Big)\\
&\quad= P\Big(\sum_{1\le n\le N}\Big(\int Y_n(t)\big(v_\ell(t)-\hat{c}_\ell\hat{v}_\ell(t)\big)\,dt\Big)^2 > x\Big)\\
&\quad\le P\Big(\sum_{1\le n\le N}\|Y_n\|^2\|v_\ell(t)-\hat{c}_\ell\hat{v}_\ell(t)\|^2 > x\Big)\\
&\quad\le P\Big(\sum_{1\le n\le N}\|Y_n\|^2 > xy\Big) + P\big(\|v_\ell(t)-\hat{c}_\ell\hat{v}_\ell(t)\|^2 > x/y\big)\\
&\quad= P_{21}(N)+P_{22}(\ell,N).
\end{aligned}$$

If we require that $y > NE\|Y_1\|^2/x$, then by the Markov inequality and (A.6) we have

$$P_{21}(N) \le \kappa_1\Big(\frac{xy}{\sqrt{N}}-\sqrt{N}E\|Y_1\|^2\Big)^{-2}$$

for some constant κ_1 which does not depend on N. By Theorem 3.2 and again the Markov inequality there exists a constant κ_2 such that for all $\ell\in\{1,\dots,d\}$

$$P_{22}(\ell,N) \le \kappa_2\frac{y}{xN}.$$

The x in the term $P_2(\ell,N)$ is given by

$$x = \frac{\varepsilon^2 N}{t_N(2q+1)^3}.$$

Set $y = 2NE\|Y_1\|^2/x$. Then for all $\ell \in \{1, \dots, d\}$

$$P_{21}(N) \le \kappa_1 \frac{1}{(E\|Y_1\|^2)^2 N} \quad \text{and} \quad P_{22}(\ell, N) \le \kappa_2 \frac{2E\|Y_1\|^2}{\varepsilon^4 N^2} t_N^2 (2q+1)^6. \tag{A.7}$$

Letting $t_N = (2q+1)^{1/2}$ shows that under $q^4/N \to 0$ the term $(2q+1)P_2(\ell, N) \to 0$. This finishes the proof of Proposition 4.1. □

A.2. Proofs of Theorems 5.1 and 5.2. The proof of Theorem 5.1 relies on Theorem A.1 of Aue et al. [5], which we state here for ease of reference.

THEOREM A.2. *Suppose $\{\boldsymbol{\xi}_n\}$ is a d-dimensional L^2–m-approximable mean zero sequence. Then*

$$N^{-1/2}\mathbf{S}_N(\cdot, \boldsymbol{\xi}) \xrightarrow{d} \mathbf{W}(\boldsymbol{\xi})(\cdot), \tag{A.8}$$

where $\{\mathbf{W}(\boldsymbol{\xi})(x), x \in [0,1]\}$ is a mean zero Gaussian process with covariances,

$$\operatorname{Cov}(\mathbf{W}(\boldsymbol{\xi})(x), \mathbf{W}(\boldsymbol{\xi})(y)) = \min(x, y)\boldsymbol{\Sigma}(\boldsymbol{\xi}).$$

The convergence in (A.8) is in the d-dimensional Skorokhod space $D_d([0,1])$.

PROOF OF THEOREM 5.1. Let

$$G_N(x, \boldsymbol{\xi}) = \frac{1}{N}\mathbf{L}_n(x, \boldsymbol{\xi})^T \hat{\boldsymbol{\Sigma}}(\boldsymbol{\xi})^{-1}\mathbf{L}_n(x, \boldsymbol{\xi})^T.$$

We notice that replacing the $\mathbf{L}_N(x, \hat{\boldsymbol{\eta}})$ with $\mathbf{L}_N(x, \hat{\boldsymbol{\beta}})$ does not change the test statistic in (5.2). Furthermore, since by the second part of Proposition 4.1 $|\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\eta}}) - \hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\beta}})| = o_P(1)$, it is enough to study the limiting behavior of the sequence $G_N(x, \hat{\boldsymbol{\beta}})$. This is done by first deriving the asymptotics of $G_N(x, \boldsymbol{\beta})$ and then analyzing the effect of replacing $\boldsymbol{\beta}$ with $\hat{\boldsymbol{\beta}}$.

Let $\boldsymbol{\beta}_i^{(m)}$ be the m-dependent approximations for $\boldsymbol{\beta}_i$ which are obtained by replacing $Y_i(t)$ in (4.10) by $Y_i^{(m)}(t)$. For a vector $\mathbf{v}$ in R^d we let $|\mathbf{v}|$ be its Euclidian norm. Then

$$\begin{aligned}
E|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_1^{(m)}|^2 &= E\sum_{\ell=1}^{d}(\beta_{\ell 1} - \beta_{\ell 1}^{(m)})^2 \\
&= \sum_{\ell=1}^{d} E\left(\int (Y_1(t) - Y_1^{(m)}(t))v_\ell(t)\,dt\right)^2 \\
&\le \sum_{\ell=1}^{d} E\int (Y_1(t) - Y_1^{(m)}(t))^2\,dt \int v_\ell^2(t)\,dt \\
&= d\nu_2^2(Y_1 - Y_1^{(m)}).
\end{aligned}$$

Since by Lyapunov's inequality we have $\nu_2(Y_1 - Y_1^{(m)}) \le \nu_4(Y_1 - Y_1^{(m)})$, (2.3) yields that $\sum_{m\ge1}(E|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_1^{(m)}|^2)^{1/2} < \infty$. Thus Theorem A.2 implies that

$$\frac{1}{\sqrt{N}}\mathbf{S}_N(x,\boldsymbol{\beta}) \overset{D^d[0,1]}{\longrightarrow} \mathbf{W}(\boldsymbol{\beta})(x).$$

The coordinatewise absolute convergence of the series $\boldsymbol{\Sigma}(\boldsymbol{\beta})$ follows from part (a) of Theorem 4.2. By assumption the estimator $\hat{\boldsymbol{\Sigma}}(\boldsymbol{\beta})$ is consistent, and consequently

$$\int G_N(x,\boldsymbol{\beta})\,dx \overset{D[0,1]}{\longrightarrow} \sum_{\ell=1}^{d}\int B_\ell^2(x)\,dx$$

follows from the continuous mapping theorem.

We turn now to the effect of changing $G_N(x,\boldsymbol{\beta})$ to $G_N(x,\hat{\boldsymbol{\beta}})$. Due to the quadratic structure of $G_N(x,\boldsymbol{\xi})$, we have $G_N(x,\hat{\boldsymbol{\beta}}) = G_N(x,\hat{\mathbf{C}}\hat{\boldsymbol{\beta}})$ when $\hat{\mathbf{C}} = \operatorname{diag}(\hat{c}_1,\hat{c}_2,\dots,\hat{c}_d)$. To finish the proof it is thus sufficient to show that

$$\sup_{x\in[0,1]}\frac{1}{\sqrt{N}}|\mathbf{S}_N(x,\boldsymbol{\beta}) - \mathbf{S}_N(x,\hat{\mathbf{C}}\hat{\boldsymbol{\beta}})| = o_P(1) \tag{A.9}$$

and

$$|\hat{\boldsymbol{\Sigma}}(\boldsymbol{\beta}) - \hat{\boldsymbol{\Sigma}}(\hat{\mathbf{C}}\hat{\boldsymbol{\beta}})| = o_P(1). \tag{A.10}$$

Relation (A.10) follows from Proposition 4.1. To show (A.9) we observe that by the Cauchy–Schwarz inequality and Theorem 3.2

$$\begin{aligned}
&\sup_{x\in[0,1]}\frac{1}{N}|\mathbf{S}_N(x,\boldsymbol{\beta}) - \mathbf{S}_N(x,\hat{\mathbf{C}}\hat{\boldsymbol{\beta}})|^2\\
&\quad= \sup_{x\in[0,1]}\frac{1}{N}\sum_{\ell=1}^{d}\left|\int\sum_{n=1}^{\lfloor Nx\rfloor} Y_n(t)\big(v_\ell(t) - \hat{c}_\ell\hat{v}_\ell(t)\big)\,dt\right|^2\\
&\quad\le \frac{1}{N}\sup_{x\in[0,1]}\int\left(\sum_{n=1}^{\lfloor Nx\rfloor} Y_n(t)\right)^2 dt \times \sum_{\ell=1}^{d}\int\big(v_\ell(t) - \hat{c}_\ell\hat{v}_\ell(t)\big)^2\,dt\\
&\quad\le \frac{1}{N}\int\max_{1\le k\le N}\left(\sum_{n=1}^{k} Y_n(t)\right)^2 dt \times O_P(N^{-1}).
\end{aligned}$$

Define

$$g(t) = E|Y_1(t)|^2 + 2(E|Y_1(t)|^2)^{1/2}\sum_{r\ge1}\big(E|Y_{1+r}(t) - Y_{1+r}^{(r)}(t)|^2\big)^{1/2}.$$

Then by similar arguments as in Section A.1 we have

$$E\left(\sum_{n=1}^{N} Y_n(t)\right)^2 \le Ng(t).$$

Hence by Menshov's inequality (see, e.g., Billingsley [13], Section 10) we infer that

$$E \max_{1\le k\le N}\left(\sum_{n=1}^{k} Y_n(t)\right)^2 \le (\log\log 4N)^2 N g(t).$$

Notice that (2.3) implies $\int g(t)\,dt < \infty$. In turn we obtain that

$$\frac{1}{N}\int \max_{1\le k\le N}\left(\sum_{n=1}^{k} Y_n(t)\right)^2 dt = O_P((\log\log N)^2),$$

which proves (A.9). □

PROOF OF THEOREM 5.2. Notice that if the mean function changes from $\mu_1(t)$ to $\mu_2(t)$ at time $k^* = \lfloor N\theta \rfloor$, then $\mathbf{L}_N(x, \hat{\boldsymbol{\eta}})$ can be written as

$$\mathbf{L}_N(x, \hat{\boldsymbol{\beta}}) + N \begin{cases} x(1-\theta)[\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2], & \text{if } x \le \theta; \\ \theta(1-x)[\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2], & \text{if } x > \theta, \end{cases} \tag{A.11}$$

where

$$\hat{\boldsymbol{\mu}}_1 = \left[\int \mu_1(t)\hat{v}_1(t)\,dt, \int \mu_1(t)\hat{v}_2(t)\,dt, \ldots, \int \mu_1(t)\hat{v}_d(t)\,dt\right]^T$$

and $\hat{\boldsymbol{\mu}}_2$ is defined analogously.

It follows from (A.11) that $T_N(d)$ can be expressed as the sum of three terms:

$$T_N(d) = T_{1,N}(d) + T_{2,N}(d) + T_{3,N}(d),$$

where

$$T_{1,N}(d) = \frac{1}{N}\int_0^1 \mathbf{L}_N(x, \hat{\boldsymbol{\beta}})^T \hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\eta}})^{-1} \mathbf{L}_N(x, \hat{\boldsymbol{\beta}})\,dx;$$

$$T_{2,N}(d) = \frac{N}{2}\theta(1-\theta)[\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2]^T \hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\eta}})^{-1}[\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2];$$

$$T_{3,N}(d) = \int_0^1 g(x,\theta)\mathbf{L}_N(x, \hat{\boldsymbol{\beta}})^T \hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\eta}})^{-1}[\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2]\,dx,$$

with $g(x,\theta) = 2\{x(1-\theta)I_{\{x\le\theta\}} + \theta(1-x)I_{\{x>\theta\}}\}$.

Since $\boldsymbol{\Omega}$ in (5.4) is positive definite (p.d.), $\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\eta}})$ is almost surely p.d. for large enough N (N is random). Hence for large enough N the term $T_{1,N}(d)$ is nonnegative. We will show that $N^{-1}T_{2,N}(d) \ge \kappa_1 + o_P(1)$, for a positive constant κ_1, and $N^{-1}T_{3,N}(d) = o_P(1)$. To this end we notice the following. Ultimately all eigenvalues of $\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\eta}})$ are positive. Let $\lambda^*(N)$ and $\lambda_*(N)$ denote the largest, respectively, the smallest eigenvalue. By Lemma 3.1, $\lambda^*(N) \to \lambda^*$ a.s. and $\lambda_*(N) \to \lambda_*$ a.s., where λ^* and λ_* are the largest and smallest eigenvalue of $\boldsymbol{\Omega}$. Next we claim that

$$|\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2| = |\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2| + o_P(1).$$

To obtain this, we use the relation $\|\hat{v}_i - \hat{c}_j v_j\| = o_P(1)$ which can be proven similarly as Lemma A.1 of Berkes et al. [7], but the law of large numbers in a Hilbert space must be replaced by the ergodic theorem. The ergodicity of $\{Y_n\}$ follows from the representation $Y_n = f(\varepsilon_n, \varepsilon_{n-1}, \ldots)$. Notice that because of the presence of a change point it cannot be claimed that $\|\hat{v}_i - \hat{c}_j v_j\| = O_P(N^{-1/2})$.

It follows that if N is large enough, then

$$[\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2]^T \hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\eta}})^{-1}[\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2] > \frac{1}{2\lambda^*}|\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2|^2 = \frac{1}{2\lambda^*}|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2|^2 + o_P(1).$$

To verify $N^{-1}T_{3,N}(d) = o_P(1)$, observe that

$$\begin{aligned}
&\sup_{x\in[0,1]} |\mathbf{L}_N(x, \hat{\boldsymbol{\beta}})^T \hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\eta}})^{-1}[\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2]| \\
&\quad \le \sup_{x\in[0,1]} |\mathbf{L}_N(x, \hat{\boldsymbol{\beta}})| \times |\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\eta}})^{-1}| \times |\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2| \\
&\quad = o_P(N)|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2|.
\end{aligned}$$

We used the matrix norm $|A| = \sup_{|x|\le 1}|Ax|$ and $|\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\eta}})^{-1}| \xrightarrow{\text{a.s}} |\boldsymbol{\Omega}^{-1}| < \infty$. □

A.3. Proof of Theorem 6.1. We first establish a technical bound which implies the consistency of the estimator $\hat{\sigma}_{\ell k}$ given in (6.2). Let $\hat{c}_\ell = \operatorname{sign}(\langle v_\ell, \hat{v}_\ell\rangle)$ and $\hat{d}_k = \operatorname{sign}(\langle u_k, \hat{u}_k\rangle)$.

LEMMA A.1. *Under the assumptions of Theorem 6.1 we have*

$$\limsup_{N\to\infty} N E|\sigma_{\ell k} - \hat{c}_\ell \hat{d}_k \hat{\sigma}_{\ell k}|^2 \le \kappa_1 \left(\frac{1}{\alpha_k^2} + \frac{1}{(\alpha'_\ell)^2}\right),$$

where κ_1 is a constant independent of k and ℓ.

PROOF. It follow from elementary inequalities that

$$|\sigma_{\ell k} - \hat{c}_\ell \hat{d}_k \hat{\sigma}_{\ell k}|^2 \le 2T_1^2 + 2T_2^2,$$

where

$$T_1 = \frac{1}{N}\int\int \left(\sum_{i=1}^N (X_i(s)Y_i(t) - E[X_i(s)Y_i(t)])\right) u_k(s)v_\ell(t)\,dt\,ds;$$

$$T_2 = \frac{1}{N}\sum_{i=1}^N \int\int E[X_i(s)Y_i(t)][u_k(t)v_\ell(s) - \hat{d}_k\hat{u}_k(t)\hat{c}_\ell\hat{v}_\ell(s)]\,dt\,ds.$$

By the Cauchy–Schwarz inequality and (A.2) we obtain

$$T_1^2 \le \frac{1}{N^2}\int\int \left(\sum_{i=1}^N X_i(s)Y_i(t) - E[X_i(s)Y_i(t)]\right)^2 dt\,ds;$$

$$T_2^2 = 2v_2^2(X)v_2^2(Y)(\|u_k - \hat{d}_k\hat{u}_k\|^2 + \|v_\ell - \hat{c}_\ell\hat{v}_\ell\|^2).$$

Hence by similar arguments as we used for the proof of Theorem 3.1 we get $NET_1^2 = O(1)$. The proof follows now immediately from Lemma 3.2 and Theorem 3.1. $\square$

Now we are ready to verify (6.4). We have

$$\hat{\psi}_{KL}(t,s) = \sum_{k=1}^{K}\sum_{\ell=1}^{L} \hat{\lambda}_\ell^{-1}\hat{\sigma}_{\ell k}\hat{u}_k(t)\hat{v}_\ell(s).$$

The orthogonality of the sequences $\{u_k\}$ and $\{v_\ell\}$ and (6.3) imply that

$$\begin{aligned}\iint\Bigg(\sum_{k>K}\sum_{\ell>L}\lambda_\ell^{-1}\sigma_{\ell k}u_k(t)v_\ell(s)\Bigg)^2 dt\,ds &= \sum_{k>K}\sum_{\ell>L}\iint \lambda_\ell^{-2}\sigma_{\ell k}^2 u_k^2(t)v_\ell^2(s)\,dt\,ds \\ &= \sum_{k>K}\sum_{\ell>L}\lambda_\ell^{-2}\sigma_{\ell k}^2 \to 0 \qquad (L,K\to\infty).\end{aligned}$$

Therefore, letting

$$\psi_{KL}(t,s) = \sum_{k=1}^{K}\sum_{\ell=1}^{L}\lambda_\ell^{-1}\sigma_{\ell k}u_k(t)v_\ell(s),$$

(6.4) will follow once we show that

$$\iint[\psi_{KL}(t,s) - \hat{\psi}_{KL}(t,s)]^2\,dt\,ds \xrightarrow{P} 0 \qquad (N\to\infty).$$

Notice that by the Cauchy–Schwarz inequality the latter relation is implied by

$$KL\sum_{k=1}^{K}\sum_{\ell=1}^{L}\iint[\lambda_\ell^{-1}\sigma_{\ell k}u_k(t)v_\ell(s) - \hat{\lambda}_\ell^{-1}\hat{\sigma}_{\ell k}\hat{u}_k(t)\hat{v}_\ell(s)]^2\,dt\,ds \xrightarrow{P} 0 \tag{A.12}$$

$$(N\to\infty).$$

A repeated application of (A.2) and some basic algebra yield

$$\begin{aligned}&\tfrac{1}{4}[\lambda_\ell^{-1}\sigma_{\ell k}u_k(t)v_\ell(s) - \hat{\lambda}_\ell^{-1}\hat{\sigma}_{\ell k}\hat{u}_k(t)\hat{v}_\ell(s)]^2 \\ &\quad\le \lambda_\ell^{-2}|\sigma_{\ell k} - \hat{c}_\ell\hat{d}_k\hat{\sigma}_{\ell k}|^2\hat{u}_k^2(t)\hat{v}_\ell^2(s) + \hat{\sigma}_{\ell k}^2|\lambda_\ell^{-1} - \hat{\lambda}_\ell^{-1}|^2\hat{u}_k^2(t)\hat{v}_\ell^2(s) \\ &\qquad + \sigma_{\ell k}^2\lambda_\ell^{-2}|u_k(t) - \hat{d}_k\hat{u}_k(t)|^2 v_\ell^2(s) + \sigma_{\ell k}^2\lambda_\ell^{-2}|v_\ell(s) - \hat{c}_\ell\hat{v}_\ell(s)|^2\hat{u}_k^2(t).\end{aligned}$$

Hence

$$\begin{aligned}&\tfrac{1}{4}\iint[\lambda_\ell^{-1}\sigma_{\ell k}u_k(t)v_\ell(s) - \hat{\lambda}_\ell^{-1}\hat{\sigma}_{\ell k}\hat{u}_k(t)\hat{v}_\ell(s)]^2\,dt\,ds \\ &\quad\le \lambda_\ell^{-2}|\sigma_{\ell k} - \hat{c}_\ell\hat{d}_k\hat{\sigma}_{\ell k}|^2 + \hat{\sigma}_{\ell k}^2|\lambda_\ell^{-1} - \hat{\lambda}_\ell^{-1}|^2 \\ &\qquad + \sigma_{\ell k}^2\lambda_\ell^{-2}(\|u_k - \hat{d}_k\hat{u}_k\|^2 + \|v_\ell - \hat{c}_\ell\hat{v}_\ell\|^2).\end{aligned}$$

Thus in order to get (A.12) we will show that

$$KL\sum_{k=1}^{K}\sum_{\ell=1}^{L}\lambda_\ell^{-2}|\sigma_{\ell k}-\hat{c}_\ell\hat{d}_k\hat{\sigma}_{\ell k}|^2\xrightarrow{P}0; \tag{A.13}$$

$$KL\sum_{k=1}^{K}\sum_{\ell=1}^{L}\hat{\sigma}_{\ell k}^2|\lambda_\ell^{-1}-\hat{\lambda}_\ell^{-1}|^2\xrightarrow{P}0; \tag{A.14}$$

$$KL\sum_{k=1}^{K}\sum_{\ell=1}^{L}\sigma_{\ell k}^2\lambda_\ell^{-2}(\|u_k-\hat{d}_k\hat{u}_k\|^2+\|v_\ell-\hat{c}_\ell\hat{v}_\ell\|^2)\xrightarrow{P}0. \tag{A.15}$$

We start with (A.13). By Lemma A.1 and Assumption 6.1 we have

$$E\left(KL\sum_{k=1}^{K}\sum_{\ell=1}^{L}\lambda_\ell^{-2}|\sigma_{\ell k}-\hat{c}_\ell\hat{d}_k\hat{\sigma}_{\ell k}|^2\right)\to 0\qquad(N\to\infty).$$

Next we prove relation (A.14). In order to shorten the proof we replace $\hat{\sigma}_{\ell k}$ by $\sigma_{\ell k}$. Otherwise we would need a further intermediate step, requiring similar arguments which follow. Now for any $0<\varepsilon<1$ we have

$$\begin{aligned}
&P\left(KL\sum_{k=1}^{K}\sum_{\ell=1}^{L}\sigma_{\ell k}^2|\lambda_\ell^{-1}-\hat{\lambda}_\ell^{-1}|^2>\varepsilon\right)\\
&\quad=P\left(KL\sum_{k=1}^{K}\sum_{\ell=1}^{L}\sigma_{\ell k}^2\lambda_\ell^{-2}\left|\frac{\hat{\lambda}_\ell-\lambda_\ell}{\hat{\lambda}_\ell}\right|^2>\varepsilon\right)\\
&\quad\le P\left(\max_{1\le\ell\le L}\left|\frac{\hat{\lambda}_\ell-\lambda_\ell}{\hat{\lambda}_\ell}\right|^2>\frac{\varepsilon}{\Psi KL}\right)\\
&\quad\le\sum_{\ell=1}^{L}P\left(\left|\frac{\hat{\lambda}_\ell-\lambda_\ell}{\hat{\lambda}_\ell}\right|^2>\frac{\varepsilon}{\Psi KL}\cap|\lambda_\ell-\hat{\lambda}_\ell|<\varepsilon\lambda_\ell\right)\\
&\qquad+\sum_{\ell=1}^{L}P\left(\left|\frac{\hat{\lambda}_\ell-\lambda_\ell}{\hat{\lambda}_\ell}\right|^2>\frac{\varepsilon}{\Psi KL}\cap|\lambda_\ell-\hat{\lambda}_\ell|\ge\varepsilon\lambda_\ell\right)\\
&\quad\le\sum_{\ell=1}^{L}\left[P\left(|\hat{\lambda}_\ell-\lambda_\ell|^2>\frac{\varepsilon}{\Psi KL}\lambda_\ell(1-\varepsilon)\right)+P(|\lambda_\ell-\hat{\lambda}_\ell|^2\ge\varepsilon^2\lambda_\ell^2)\right]\\
&\quad\le\kappa_2\left(\frac{KL^2}{\epsilon N\lambda_L}+\frac{1}{\varepsilon N\lambda_L^2}\right),
\end{aligned}$$

by an application of the Markov inequality and Theorem 3.2. According to our Assumption 6.1 this also goes to zero for $N\to\infty$.

Finally we prove (A.15). By Lemma 3.2 and Theorem 3.1 we infer that

$$
\begin{aligned}
&E\left(KL\sum_{k=1}^{K}\sum_{\ell=1}^{L}\sigma_{\ell k}^2\lambda_\ell^{-2}(\|u_k-\hat{d}_k\hat{u}_k\|^2+\|v_\ell-\hat{c}_\ell\hat{v}_\ell\|^2)\right)\\
&\quad\le\kappa_3\frac{KL}{N}\sum_{k=1}^{K}\sum_{\ell=1}^{L}\sigma_{\ell k}^2\lambda_\ell^{-2}\left(\frac{1}{\alpha_k^2}+\frac{1}{\alpha_\ell'^2}\right)\\
&\quad\le 2\kappa_3\Psi\frac{KL}{N\min\{h_L,h_K'\}^2}.
\end{aligned}
$$

Assumption 6.1(ii) assures that the last term goes to zero. The proof is now complete.

Acknowledgments. We thank P. Hall, H.-G. Müller, D. Paul and J.-L. Wang for their useful comments. We are grateful to two referees for raising a number of points which helped us clarify several important issues. Finally, we thank the Associate Editor for constructive criticism and clear guidelines.

REFERENCES

[1] ANDERSON, T. W. (1994). *The Statistical Analysis of Time Series*. Wiley, New York. MR0283939

[2] ANDREWS, D. W. K. (1984). Nonstrong mixing autoregressive processes. *J. Appl. Probab.* **21** 930–934. MR0766830

[3] ANDREWS, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59** 817–858. MR1106513

[4] ANTONIADIS, A. and SAPATINAS, T. (2003). Wavelet methods for continuous time prediction using Hilbert-valued autoregressive processes. *J. Multivariate Anal.* **87** 133–158. MR2007265

[5] AUE, A., HORVÁTH, S., HÖRMANN L. and REIMHERR, M. (2009). Detecting changes in the covariance structure of multivariate time series models. *Ann. Statist.* **37** 4046–4087.

[6] BENKO, M., HÄRDLE, W. and KNEIP, A. (2009). Common functional principal components. *Ann. Statist.* **37** 1–34. MR2488343

[7] BERKES, I., GABRYS, R., HORVÁTH, L. and KOKOSZKA, P. (2009). Detecting changes in the mean of functional observations. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **71** 927–946.

[8] BERKES, I., HÖRMANN, S. and SCHAUER, J. (2009). Asymptotic results for the empirical process of stationary sequences. *Stochastic Process. Appl.* **119** 1298–1324. MR2508575

[9] BERKES, I., HORVÁTH, L., KOKOSZKA, P. and SHAO, Q.-M. (2005). Almost sure convergence of the Bartlett estimator. *Period. Math. Hungar.* **51** 11–25. MR2180630

[10] BERKES, I., HORVÁTH, L., KOKOSZKA, P. and SHAO, Q.-M. (2006). On discriminating between long-range dependence and changes in mean. *Ann. Statist.* **34** 1140–1165. MR2278354

[11] BESSE, P., CARDOT, H. and STEPHENSON, D. (2000). Autoregressive forecasting of some functional climatic variations. *Scand. J. Statist.* **27** 673–687.

[12] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York. MR0233396

[13] BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd ed. Wiley, New York. MR1700749

[14] BOSQ, D. (2000). *Linear Processes in Function Spaces*. Springer, New York. MR1783138
[15] BOSQ, D. and BLANKE, D. (2007). *Inference and Prediction in Large Dimensions*. Wiley, Chichester.
[16] BRADLEY, R. C. (2007). *Introduction to Strong Mixing Conditions* **1**, **2**, **3**. Kendrick Press, Heber City, UT. MR2359339
[17] CAI, T. and HALL, P. (2006). Prediction in functional linear regression. *Ann. Statist.* **34** 2159–2179. MR2291496
[18] CARDOT, H., FERRATY, F., MAS, A. and SARDA, P. (2003). Testing hypothesis in the functional linear model. *Scand. J. Statist.* **30** 241–255. MR1965105
[19] CARDOT, H., FERRATY, F. and SARDA, P. (2003). Spline estimators for the functional linear model. *Statist. Sinica* **13** 571–591. MR1997162
[20] CHIOU, J.-M. and MÜLLER, H.-G. (2007). Diagnostics for functional regression via residual processes. *Comput. Statist. Data Anal.* **15** 4849–4863. MR2364544
[21] CHIOU, J.-M., MÜLLER, H.-G. and WANG, J.-L. (2004). Functional response models. *Statist. Sinica* **14** 675–693. MR2087968
[22] CUEVAS, A., FEBRERO, M. and FRAIMAN, R. (2002). Linear functional regression: The case of fixed design and functional response. *Canad. J. Statist.* **30** 285–300. MR1926066
[23] DAMON, J. and GUILLAS, S. (2002). The inclusion of exogenous variables in functional autoregressive ozone forecasting. *Environmetrics* **13** 759–774.
[24] DAUXOIS, J., POUSSE, A. and ROMAIN, Y. (1982). Asymptotic theory for principal component analysis of a vector random function. *J. Multivariate Anal.* **12** 136–154. MR0650934
[25] DOUKHAN, P. (1994). *Mixing: Properties and Examples*. Springer, New York. MR1312160
[26] DOUKHAN, P. and LOUHICHI, S. (1999). A new weak dependence and applications to moment inequalities. *Stochastic Process. Appl.* **84** 313–343. MR1719345
[27] ENGLE, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50** 987–1007. MR0666121
[28] GIRAITIS, L., KOKOSZKA, P. S. and LEIPUS, R. (2000). Stationary ARCH models: Dependence structure and Central Limit Theorem. *Econometric Theory* **16** 3–22. MR1749017
[29] GIRAITIS, L., KOKOSZKA, P. S., LEIPUS, R. and TEYSSIÈRE, G. (2003). Rescaled variance and related tests for long memory in volatility and levels. *J. Econometrics* **112** 265–294. MR1951145
[30] GOHBERG, I., GOLBERG, S. and KAASHOEK, M. A. (1990). *Classes of Linear Operators, vol. 1. Operator Theory: Advances and Applications* **49**. Birkhaüser, Basel.
[31] GORDIN, M. (1969). The central limit theorem for stationary processes. *Dokl. Akad. Nauk SSSR* **188** 739–741. MR0251785
[32] GRANGER, C. W. J. and ANDERSON, A. P. (1979). *An Introduction to Bilinear Time Series Models.* Vandenhoeck and Ruprecht, Göttingen.
[33] HALL, P. and HOSSEINI-NASAB, M. (2006). On properties of functional principal components. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68** 109–126. MR2212577
[34] HAMILTON, J. D. (1994). *Time Series Analysis*. Princeton Univ. Press, Princeton, NJ. MR1278033
[35] HÖRMANN, S. (2008). Augmented GARCH sequences: Dependence structure and asymptotics. *Bernoulli* **14** 543–561. MR2544101
[36] HÖRMANN, S. (2009). Berry-Esseen bounds for econometric time series. *ALEA* **6** 377–397.
[37] IBRAGIMOV, I. A. (1962). Some limit theorems for stationary processes. *Teor. Veroyatn. Primen.* **7** 361–392. MR0148125
[38] KARGIN, V. and ONATSKI, A. (2008). Curve forecasting by functional autoregression. *J. Multivariate Anal.* **99** 2508–2506. MR2463404
[39] KIEFER, J. (1959). K-sample analogues of the Kolmogorov–Smirnov and Cramér-von Mises tests. *Ann. Math. Statist.* **30** 420–447. MR0102882

[40] LI, Y. and HSING, T. (2007). On rates of convergence in functional linear regression. *J. Multivariate Anal.* **98** 1782–1804. MR2392433

[41] MALFAIT, N. and RAMSAY, J. O. (2003). The historical functional model. *Canad. J. Statist.* **31** 115–128. MR2016223

[42] MAS, A. and MENNETEAU, L. (2003). Perturbation approach applied to the asymptotic study of random operators. *Progr. Probab.* **55** 127–134. MR2033885

[43] MASLOVA, I., KOKOSZKA, P., SOJKA, J. and ZHU, L. (2009). Removal of nonconstant daily variation by means of wavelet and functional data analysis. *Journal of Geophysical Research* **114** A03202.

[44] MASLOVA, I., KOKOSZKA, P., SOJKA, J. and ZHU, L. (2009). Estimation of Sq variation by means of multiresolution and principal component analyses. *Journal of Atmospheric and Solar-Terrestrial Physics*. To appear.

[45] MÜLLER, H.-G. and YAO, F. (2008). Functional additive models. *J. Amer. Statist. Assoc.* **103** 1534–1544. MR2504202

[46] MÜLLER, H.-G. and STADTMÜLLER, U. (2005). Generalized functional linear models. *Ann. Statist.* **33** 774–805. MR2163159

[47] PHILIPP, W. and STOUT, W. F. (1975). *Almost Sure Invariance Principles for Partial Sums of Weakly Dependent Random Variables*. *Mem. Amer. Math. Soc.* **161**. Amer. Math. Soc., Providence, RI.

[48] PÖTSCHER, B. and PRUCHA, I. (1997). *Dynamic Non-linear Econonometric Models. Asymptotic Theory*. Springer, Berlin.

[49] PRIESTLEY, M. (1988). *Nonlinear and Nonstationary Time Series Analysis.* Academic Press, London. MR0991969

[50] RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*. Springer, New York. MR2168993

[51] REISS, P. T. and OGDEN, R. T. (2007). Functional principal component regression and functional partial least squares. *J. Amer. Statist. Assoc.* **102** 984–996. MR2411660

[52] REISS, P. T. and OGDEN, R. T. (2009). Smoothing parameter selection for a class of semiparametric linear models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **71** 505–523.

[53] REISS, P. T. and OGDEN, R. T. (2010). Functional generalized linear models with images as predictors. *Biometrics*. To appear.

[54] ROSENBLATT, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Natl. Acad. Sci. USA* **42** 43–47. MR0074711

[55] ROSENBLATT, M. (1959). Stationary processes as shifts of functions of independent random variables. *J. Math. Mech.* **8** 665–681. MR0114249

[56] ROSENBLATT, M. (1961). Independence and dependence. In *Proc. 4th Berkeley Sympos. Math. Statist. Probab.* **2** 431–443. Univ. California Press, Berkeley, CA. MR0133863

[57] ROSENBLATT, M. (1971). *Markov Processes. Structure and Asymptotic Behavior.* Springer, New York. MR0329037

[58] STADTLOBER, E., HÖRMANN, S. and PFEILER, B. (2008). Qualiy and performance of a PM10 daily forecasting model. *Athmospheric Environment* **42** 1098–1109.

[59] STINE, R. (1997). Nonlinear time series. In *Encyclopedia of Statistical Sciences* (S. Kotz, C. Y. Read and D. L. Banks, eds.) 430–437. Wiley, New York.

[60] TONG, H. (1990). *Non-linear Time Series: A Dynamical System Approach.* Oxford Univ. Press, New York. MR1079320

[61] VOSTRIKOVA, L. J. (1981). Detection of "disorder" in multidimensional random processes. *Sov. Dokl. Math.* **24** 55–59.

[62] WU, W. (2005). Nonlinear system theory: Another look at dependence. *Proc. Natl. Acad. Sci. USA* **102** 14150–14154. MR2172215

[63] WU, W. (2007). Strong invariance principles for dependent random variables. *Ann. Probab.* **35** 2294–2320. MR2353389

[64] YAO, F. and LEE, T. (2006). Penalized spline models for functional principal component analysis. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68** 3–25. MR2212572

[65] YAO, F., MÜLLER, H.-G. and WANG., J.-L. (2005). Functional linear regression analysis for longitudinal data. *Ann. Statist.* **33** 2873–2903. MR2253106

[66] YAO, F., MÜLLER, H.-G. and WANG., J.-L. (2005). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100** 577–590. MR2160561

UNIVERSITÉ LIBRE DE BRUXELLES
CP 210, LOCAL O.9.115
BD DU TRIOMPHE
B-1050 BRUXELLES
BELGIUM
E-MAIL: shormann@ulb.ac.be

DEPARTMENT MATHEMATICS AND STATISTICS
UTAH STATE UNIVERSITY
3900 OLD MAIN HILL
LOGAN, UTAH 84322-3900
USA
E-MAIL: piotr.kokoszka@usu.edu