A Darling-Erdős-type CUSUM-procedure for functional data II

Leonid Torgovitski∗†

Mathematical Institute, University of Cologne
Weyertal 86-90, 50931, Cologne, Germany

Abstract

This paper extends the class of the famous Darling-Erdős-type cumulative sums (CUSUM) procedures to functional time series under short range dependence conditions which are satisfied by functional analogues of many popular time series models (including functional AR and ARCH). We follow a data driven, projection-based approach where the lower-dimensional subspace is determined by (generalized) functional principal components which are eigenfunctions of the long run covariance operator. This second-order structure is generally unknown and estimation is crucial - it plays an even more important role than in the classical univariate setup because it generates the finite-dimensional subspaces. We discuss suitable estimates and demonstrate empirically that altogether the procedure performs well under moderate temporal dependence. The procedure is finally applied to publicly accessible electricity data from a German power company.

Keywords  Functional data analysis, Change-point, $L^κ$-$m$-approximable time series, Darling-Erdős, Long run variance

AMS Subject Classification  62G05, 62G10, 62G20

1 Introduction

The interest and the research activities in the change-point analysis for multivariate, high-dimensional and especially for functional data are enormous which is a consequence of the increasingly growing computational capacities. These activities are reflected by the amount (and the frequency) of published works and in particular by survey articles that appeared recently (e.g. Kokoszka (2012), Aue and Horváth (2013) and the invited discussion paper by Horváth and Rice (2014a)). One of the fundamental and most studied problems in change-point analysis is concerned with a simple abrupt change in the mean - the “at most one change” (AMOC) model. We consider this problem in the functional setup, i.e. where each observation is a curve and the mean is a deterministic function. We want to know whether the overall shapes of the curves have changed over time at some arbitrary time point or

∗E-mail: ltorgovi@math.uni-koeln.de
†Research partially supported by the Friedrich Ebert Foundation.
not. Our investigations are based on the work of Berkes et al. (2009) who studied the same problem and introduced a (differently weighted) nonparametric CUSUM procedure for detection of changes in the mean of functional observations in the i.i.d. setting. They suggested an intuitive approach which essentially relies on a multivariate CUSUM by projecting the functional time series on a finite dimensional subspace which captures the dynamics of the data in a beneficial manner in order to obtain reasonable power. The authors followed hereby the methodology of Gabrys and Kokoszka (2007) and proposed to select the subspace spanned by functional principal components (FPC’s), i.e. the eigenfunctions of the covariance operator, in view of the well known optimality properties regarding dimension reduction (cf. Ramsay and Silverman (2005)). Since then, FPC’s - which play widely known an outstanding role in functional data analysis - have been successfully incorporated into many further functional stability-testing procedures under independence as well as under dependence (cf. Horváth and Kokoszka (2012) for an overview and also Berkes et al. (2009), Hörmann and Kokoszka (2010) and Aston and Kirch (2012) in particular for the change in the mean setting). Later on it was realized that generalized FPC’s, given as eigenfunctions of the so-called long run covariance operator, are advantageous (cf. Horváth et al. (2012b, 2013) and Torgovitski (2014)).

In this article we will stick to the latter approach. We pick up and continue the work of Torgovitski (2014) (cf. also Zhou (2011)) extending the results from the $m$-dependent setting to the more challenging and realistic setting of weakly dependent time series with a focus on the framework of $L^κ$-$m$-approximability (cf., e.g. Hörmann and Kokoszka (2010) and Horváth et al. (2012b)). As in Torgovitski (2014) the procedure will be based on the dimension-reduction approach of Berkes et al. (2009). To establish asymptotics we will incorporate several steps outlined in Torgovitski (2014) and make use of recently obtained results of Berkes et al. (2011, 2013), Horváth et al. (2012b) and Aue et al. (2014).

In order to formalize the testing problem we have to introduce some notation first. We consider functional data $X(\cdot)$ as a random element on some probability space $(\Omega, \mathcal{A}, P)$ with the state space $L^2[0,1]$. Throughout, let $L^2[0,1]$ denote the space of square-integrable functions w.r.t. the Lebesgue measure on $[0,1]$ equipped with the usual inner product and the corresponding norm, denoted by $\langle \cdot, \cdot \rangle$ or $\| \cdot \|$, respectively. We also assume product measurability of $X(t) = X(t, \omega)$ w.r.t. $(t, \omega) \in [0,1] \times \Omega$. The mean of $X(\cdot)$ is defined as the unique function $\mu(\cdot)$, such that $\langle x, \mu \rangle = E\langle x, X \rangle$ holds true for all $x \in L^2[0,1]$ given that $E\|X\| < \infty$.

We assume that the observable sequence $\{X_i\}_{i \in \mathbb{Z}}$ consists of $L^2[0,1]$-valued random elements which are given by the functional “signal plus noise” model

$$X_i(t) = \mu_i(t) + Y_i(t), \quad t \in [0,1],$$

with mean functions $\mu_i(\cdot) \in L^2[0,1]$ and with innovations fulfilling the basic Assumption (M) below where the dependence structure will be specified later on.

**Assumption (M).** (i) The functional centered innovation sequence $\{Y_i\}_{i \in \mathbb{Z}}$ is strictly stationary; (ii) $E\|Y_i\|^{\nu} < \infty$ for some $\nu > 2$.

We want to test retrospectively the null hypothesis of no change in the mean, i.e.

$$H_0 : \mu_1(\cdot) = \ldots = \mu_n(\cdot)$$
against the alternative of a change in the mean

\[ H_A : \mu_1(\cdot) = \ldots = \mu_{[n\theta]}(\cdot) \neq \mu_{[n\theta]+1}(\cdot) = \ldots = \mu_n(\cdot) \]

at some unknown time point characterized by some (unknown) constant change parameter \( \theta \in (0, 1) \).

The structure of this article is as follows. In Section 2 we formulate the dependence concept of \( L^\kappa\)-\( m \)-approximability for our observations. In Section 3 we present the testing procedure together with the asymptotics under the null hypothesis and under the alternative. Section 4 is devoted to estimation of generalized FPC’s. The performance is finally demonstrated in Section 5 including an application example. All proofs are postponed to Section 6.

2 Weakly dependent time series

We consider the concept of \( L^\kappa\)-\( m \)-approximable time series which is currently of major interest in univariate, multivariate and functional settings and covers many relevant time series models (cf., e.g. amongst many others [Aue et al (2009), Hörmann and Kokoszka (2010), Aston and Kirch (2012), Horváth et al (2012b, 2013), Jirak (2012, 2013), Berkes et al (2013), Chochola et al (2013) and Hörmann and Kidziński (2014)]. Hörmann and Kokoszka (2010) and Berkes et al (2011) contain extensive discussions and comparisons to other (related) dependence concepts.

We formulate the dependence concept in general separable Hilbert spaces \( H \) but having the special cases \( H = L^2[0,1] \) and \( H = \mathbb{R}^d \) in mind. The reason for this is that we will use the notion of \( L^\kappa\)-\( m \)-approximability in both spaces because we will deal with appropriate \( \mathbb{R}^d \)-valued approximations of the original \( L^2[0,1] \) valued time series. For a moment, let \( \| \cdot \|_H \) denote a norm in the space \( H \) and recall that \( E(\|X\|_H^\kappa)^{1/\kappa} \) is the \( L^\kappa(\Omega, \mathcal{P}) \)-norm for the real-valued random variable \( \|X\|_H \). Later on we will write \( \| \cdot \| \) for the \( L^2 \) norm or \( | \cdot | \) for the Euclidean norm, respectively.

**Definition 2.1.** The \( H \)-valued sequence \( \{Y_i\}_{i \in \mathbb{Z}} \), defined on some common probability space \( (\Omega, \mathcal{A}, \mathcal{P}) \), is \( L^\kappa\)-\( m \)-approximable with rate \( \delta(m) = o(1), \delta(m) \geq 0 \) and with \( \kappa \geq 2 \) iff \( E(\|Y_0\|_H^\kappa) < \infty \) and the following conditions hold true:

1. The \( Y_i \)'s admit a Bernoulli shift representation, i.e.

\[ Y_i = f(\ldots, \varepsilon_{i+1}, \varepsilon_i, \varepsilon_{i-1}, \ldots), \tag{2.1} \]

where the innovations \( \varepsilon_i \) are i.i.d. \( S \)-valued random elements, \( S \) is some measurable space and \( f \) is a measurable mapping \( f : S^\infty \to H \).

2. The \( Y_i \)'s are approximated by \( m \)-dependent random variables in the sense that, as \( m \to \infty \),

\[ E(\|Y_0 - Y_0^{(m)}\|_H^\kappa)^{1/\kappa} = O(\delta(m)), \tag{2.2} \]

where \( Y_i^{(m)} \) are \( m \)-dependent copies of \( Y_i \) defined via

\[ Y_i^{(m)} = f(\ldots, \varepsilon_{i+M}, \varepsilon_{i+(M-1)}, \ldots, \varepsilon_i, \ldots, \varepsilon_{i-(M-1)}, \varepsilon_{i-M}, \ldots) \tag{2.3} \]

with \( M = \lfloor m/2 \rfloor \) for all integer \( m \geq 0 \) and where the family

\[ \{\varepsilon_i, \varepsilon_i^{(k,j)}, i,j,r,k \in \mathbb{Z}, k \geq 0\} \]

is i.i.d.
3. If \( \{ Y_i \} \) is function-valued with \( H = L^2[0, 1] \) then \((B_H \otimes \mathcal{A}) - B_{\mathbb{R}} \) measurability is assumed, where \( B_H \) is the Borel sigma-algebra w.r.t. \( \| \cdot \|_H \).

Typical conditions on the rate \( \delta(m) \) are summability \( \sum_{m=1}^\infty \delta(m) < \infty \), polynomial decay \( \delta(m) = \mathcal{O}(m^{-\nu}) \) for some \( \nu > 2 \) or exponential decay \( \delta(m) = \mathcal{O}(\exp(-cm)) \) with some \( c > 0 \), where the latter is the strongest condition but already satisfied for many models (e.g. covers amongst others the \( H \)-valued AR(1)). We are interested in the causal case of \( \mathcal{L}_\kappa -m \)-approximability, i.e. that \( Y_i = f(\varepsilon_i, \varepsilon_{i-1}, \ldots) \) which is a special case of (2.1). The two-sided noncausal representation (2.1) appears to be useful in the proofs, where we will deal with time-inversed \( \mathcal{L}_\kappa -m \)-approximable time series \( \{ Y_{-i} \}_{i \in \mathbb{Z}} \). Therefore, observe that \( Y_{-i} = f(\varepsilon_{-i}, \varepsilon_{-(i+1)}, \varepsilon_{-(i+2)}, \ldots) \) is noncausal but still \( \mathcal{L}_\kappa -m \)-approximable according to the two-sided Definition 2.1 above.

Remark 2.2. Note that some recent literature (e.g. Berkes et al (2013) and Horváth et al (2013)) works with a slightly modified condition (2.2) where the left-hand side of (2.2) is substituted by \( E(\| Y_0 - Y_0^{(m)} \|_H) \) with some \( \tilde{\kappa} > \kappa \).

3 The testing procedure

For the sake of generality and clarity, the testing procedure will be described in a unifying functional framework where we separate and highlight those conditions which essentially allow us to derive suitable asymptotics without a priori specifying a particular time series model or a dependence concept. This is inspired by the presentation of Aston and Kirch (2012) and Kamgaing and Kirch (2014). The conditions presented below in this section were mostly implicitly contained in Torgovitski (2014). Here, the theoretical focus will be on verification of all stated conditions for \( \mathcal{L}_\kappa -m \)-approximable time series.

The CUSUM procedure, which will be presented below, belongs to the class of FPC-based approaches and utilizes the second-order structure of the time series for an appropriate data-driven subspace selection. As mentioned in the introduction, we will assume a functional time series \( \{ X_i \} \) with functional innovations \( \{ Y_i \} \) and work with generalized FPC’s following, e.g., Horváth et al (2012b, 2013) and Torgovitski (2014). Those are eigenfunctions of the long run covariance operator \( C \) of \( \{ Y_i \}_{i \in \mathbb{Z}} \) which, given Assumption (M), can be formally defined as an integral operator

\[
(Cx)(t) = \int \zeta(t, s)x(s)ds,
\]

\( t \in [0, 1], x \in L^2[0, 1], f := \int_0^1 \) with kernel

\[
\zeta(t, s) = \sum_{r \in \mathbb{Z}} E[Y_0(t)Y_r(s)]. \tag{3.1}
\]

This operator is well-defined if

\[
\zeta \in L^2([0, 1] \times [0, 1]) \tag{3.2}
\]

holds true in which case it is symmetric Hilbert-Schmidt and positive. Hence, \( C \) can be written using the spectral decomposition as

\[
(Cx)(t) = \sum_{j=1}^\infty \lambda_j \langle x, v_j \rangle v_j(t), \tag{3.3}
\]
For testing of $H_0$ against $H_A$ we will work with the following CUSUM statistic

$$T_n(X; v, \lambda) = \max_{1 \leq k < n} w(k/n) \left( \frac{1}{\lambda_r} \sum_{r=1}^{d} \left\langle \frac{n^{1/2} \sum_{i=1}^{k} (X_i - \bar{X}_n), v_r} {\sqrt{k}} \right\rangle^2 \right)^{1/2},$$  \hspace{1cm} (3.4)

where

$$w(t) = (t(1-t))^{-1/2}$$

is a suitable weight function related to the variance of a Brownian bridge and $d \in \mathbb{N}$ is a fixed positive integer specifying the dimension. Notice, that in (3.4) $X$, $v$ or $\lambda$ represent $\{X_i\}_{i \in \mathbb{N}}$, $\{v_i\}_{i \in \mathbb{N}}$ or $\{\lambda_i\}_{i \in \mathbb{N}}$, respectively. The right-hand side of (3.4) can also be written compactly as

$$\max_{1 \leq k < n} w(k/n)|n^{-1/2} \sum_{i=1}^{k} (X_i - \bar{X}_n)|_{\Sigma}$$

to emphasize that this CUSUM is based on the multivariate projected version of (1.1), i.e. on

$$X_i = \mu_i + Y_{it},$$  \hspace{1cm} (3.5)

where the $r$-th components are the data scores $X_{i,r} = \langle X_i, v_r \rangle$, the innovation scores $Y_{i,r} = \langle Y_i, v_r \rangle$ or the projected means $\mu_{i,r} = \langle \mu_i, v_r \rangle$ for $r = 1, \ldots, d$, respectively. Here, $| \cdot |_{\Sigma} = \Sigma^{-1/2} \cdot$ is a norm, where

$$\Sigma = \text{diag}(\lambda_1, \ldots, \lambda_d)$$

is a proper standardization matrix (cf. [Torgovitski (2014)] and Remark 3.1 below).

Remark 3.1. The matrix $\Sigma$ is the long run covariance matrix of the projected time series $\{Y_i\}_{i \in \mathbb{Z}}$ which can be seen utilizing the orthonormality of the eigenfunctions:

$$\sum_{k \in \mathbb{Z}} E[Y_{0,t} Y_{k,j}] = \int \left( \int \zeta(t, s)v_j(s)ds \right) v_i(t)dt = \lambda_j \delta_{i,j}.$$  

The diagonalization of $\Sigma$ is hereby one of the main advantages of working with the long run covariance operator $C$. Also, the existence of $\Sigma$ is inherited from the existence of $C$.

In applications, especially in the functional setup, the covariance structure, e.g. in our case the covariance operator $C$, will be rarely known. Associated quantities, here the eigenstructure $\{v_i\}_{i \in \mathbb{N}}$ and $\{\lambda_i\}_{i \in \mathbb{N}}$, are therefore also usually unknown and have to be estimated. Therefore, let $\{\tilde{v}_i\}_{i \in \mathbb{N}}$ be orthonormal functions which
together with scalars \( \{ \hat{\lambda}_i \}_{i \in \mathbb{N}} \) denote some generic estimates which will be specified later on. Thus, instead of working with \( T_n = T_n(X; v, \lambda) \) we will consider
\[
\hat{T}_n = T_n(X; \hat{v}, \hat{\lambda})
\] (3.6)
having in mind that formally \( \hat{T}_n = \infty \) if \( \hat{\lambda}_r = 0 \) for some \( 1 \leq r \leq d \) under the convention \( 1/0 := \infty \).

The following conditions (L), (P1), (P2), (A1), (A2) and (B1), (B2) are the main building blocks which allow us to prove the limiting distribution of \( \hat{T}_n \) under the null hypothesis and consistency under the alternative. We proceed with conditions under the null hypothesis where the first condition (L) states the availability of a multivariate Darling-Erdős-type limit theorem for \( T_n \) which will be a cornerstone for our further considerations. (See Berkes et al (2009), Hörmann and Kokoszka (2010) and Aston and Kirch (2012) for procedures based on multivariate functional central limit theorems.)

**Assumption (L).** It holds that, as \( n \to \infty \),
\[
\lim_{n \to \infty} P \left( a(\log n) T_n(Y; v, \lambda) - b_d(\log n) \leq x \right) = \exp(-2 \exp(-x)) \quad (3.7)
\]
for all \( x \in \mathbb{R} \), where
\[
a(t) = (2 \log t)^{1/2}
\]
and \( b_d(t) = 2 \log t + (d/2) \log \log t - \log \Gamma(d/2) \) denote the well known normalizing functions.

In the i.i.d. setting (3.7) is immediately implied by Csörgő and Horváth (1997, Theorem 1.3.1) if (ii) of (M) and (3.2) holds true. For strictly stationary \( m \)-dependent sequences (L) follows analogously using strong invariance principles derived in Horváth et al (1999) (cf. Torgovitski (2014)). Verification of (L) under \( L^\kappa \)-m-approximability will be carried out further below but has now to be based on strong approximations derived recently in Aue et al (2014). Further, e.g. mixing-type, conditions which ensure (3.7) are discussed in Kamgaing and Kirch (2014).

**Remark 3.2.** It is worth noting, that strictly stationary \( m \)-dependent sequences, as considered in Horváth et al (1999) or in Torgovitski (2014), are generally not \( L^\kappa \)-m-approximable. As pointed out by Berkes et al (2011, Section 3.1), they do not necessarily possess representation (2.1).

The following conditions on maxima of weighted (backward) partial sums of the innovations together with the subsequent conditions on rates for \( \{ \hat{v}_i \}_{i \in \mathbb{N}} \) and \( \{ \hat{\lambda}_i \}_{i \in \mathbb{N}} \) will ensure a proper interplay between the functional data and the multivariate statistic \( \hat{T}_n \).

**Assumption (P1).** It holds that, as \( n \to \infty \),
\[
\max_{1 \leq i \leq d} \left| \hat{\lambda}_i - \lambda_i \right| = o_P((\log \log n)^{-1}).
\]

\[\text{Assumption (A1).} \quad \text{Under } H_0 \text{ it holds that, as } n \to \infty, \]
\[
\max_{i=1,\ldots,d} |\hat{\lambda}_i - \lambda_i| = o_P((\log \log n)^{-1}).
\]
Assumption (A2). Under $H_0$ it holds that, as $n \to \infty$,
\[
\max_{i=1, \ldots, d} |\hat{v}_i - s_i v_i| = o_P((\log \log n)^{-1/2}/g(n)),
\]
where $s_i$’s are random with $s_i \in \{1, -1\}$ and the rate function $g(n)$ is the same as in (P1).

The random $s_i$’s are typical in the functional setup and show up essentially due to the non-uniqueness of eigenfunctions but apparently do not affect the statistic (3.4). As already indicated, above assumptions are sufficient to obtain the limiting distribution of $\hat{T}_n$ which is stated in the next proposition and allows us to obtain critical values.

**Theorem 3.3.** Let Assumptions (3.2), (L), (P1), (A1), (A2) and $\lambda_d > 0$ hold true. Then under $H_0$ it holds that, as $n \to \infty$,
\[
\lim_{n \to \infty} P\left( a(\log n) \hat{T}_n - b_d(\log n) \leq x \right) = \exp(-2\exp(-x)) \quad (3.8)
\]
for all $x \in \mathbb{R}$.

**Remark 3.4.** By considering the univariate analogue of Assumption (P1) we note that $g(n) = (\log \log n)^{1/2}$ is the best possible rate (cf. e.g. Csörgő and Horváth (1997, Theorem A.4.1) for the famous Darling-Erdős asymptotics for i.i.d. random variables) and if such a rate holds true in (P1), the rates in (A1) and (A2) coincide and are both of order $o_P(\log \log n)$.

To the best of our knowledge, the law of the iterated logarithm and results of Daling-Erdős-type are not proven in the functional framework of $L^\kappa$-m-approximability so far. Nevertheless, the results shown below (cf. Propositions 3.8, 3.9 and Corollary 3.10) allow us to derive logarithmic rates $g(n)$ which are sufficient for our practical purposes.

Before proceeding with the verification of conditions (L) - (A2), we turn to the alternative and state the conditions which ensure that the procedure detects changes $\Delta(t) = \mu_n(t) - \mu_1(t)$ with $\Delta \neq 0$ in $L^2[0, 1]$ with probability tending to 1, as $n \to \infty$. Analogous to Assumption (P1), we need a bound on partial sums and conditions on estimators $\hat{\lambda}_j$ and $\hat{v}_j$.

**Assumption (P2).** The weak law of large numbers holds true, i.e. it holds that, as $n \to \infty$
\[
\| \sum_{i=1}^{n} Y_i \| = o_P(n).
\]

Estimators $\hat{\lambda}_j$ appear in the denominator of (3.6) and therefore need to be bounded.

**Assumption (B1).** Under $H_A$ it holds that, as $n \to \infty$,
\[
\hat{\lambda}_1 = o_P(n/(\log \log n)).
\]

**Assumption (B2).** Under $H_A$ it holds that, as $n \to \infty$,
\[
|\langle \Delta, \hat{v}_r \rangle | \xrightarrow{P} \xi, \quad (3.9)
\]
for some $1 \leq r \leq d$ and some $\xi > 0$. 


Remark 3.5. Condition (3.9) has an intuitive interpretation. Therefore, notice that the related condition \( \langle \Delta, v_r \rangle = 0 \) for all \( 1 \leq r \leq d \) is equivalent to
\[
\mu_1 = \ldots = \mu_n,
\]
i.e. there would be no change in the projected time series \( \{X_i\} \). Hence, (3.9) means that the change \( \Delta \) has to be “asymptotically visible” in the projected time series \( \{\hat{X}_{i,r}\} \) where \( \hat{X}_{i,r} = \langle X_i, \hat{v}_r \rangle \).

The above conditions are sufficient to state the following consistency result.

**Theorem 3.6.** Let Assumptions (3.2), (L), (P2), (B1), (B2) and \( \lambda_d > 0 \) hold true. Then under \( H_A \) it holds that, as \( n \to \infty \),
\[
(\log \log n)^{-1/2} \hat{T}_n \xrightarrow{P} \infty. \tag{3.10}
\]

Condition (B2) is verified under \( H_A \) typically via the simple relation
\[
\left| \langle \Delta, \hat{v}_r \rangle - \langle \Delta, w_r \rangle \right| = O_P( \| \hat{v}_r - c_r w_r \| ) = o_P(1) \tag{3.11}
\]
on showing that \( \| \hat{v}_r - c_r w_r \| = o_P(1) \) (where \( c_r \in \{0, 1\} \) are random) together with \( \langle \Delta, w_r \rangle \neq 0 \) for some \( 1 \leq r \leq d \). In the i.i.d. setting natural estimates \( \hat{v}_r \) are given by the empirical functional principal components (cf. Berkes et al. (2009)) and it is shown that they converge (up to signs) to eigenfunctions \( w_r \) of a contaminated operator (cf. also Aston and Kirch (2012) and Torgovitski (2014)). Therefore, however, technical conditions on the eigenstructure of the contaminated operator together with the orthogonality condition \( \langle \Delta, w_r \rangle \neq 0 \) for some \( 1 \leq r \leq d \) have to be additionally imposed. In our setup, estimates \( \hat{v}_r \) can always be chosen such that \( \langle \Delta, w_r \rangle \neq 0 \) (and therefore Assumption (B2)) is fulfilled even with \( r = 1 \) where \( w_1 = \Delta / \| \Delta \| \). This is stated in Proposition 4.3 and has been observed by Horváth et al. (2013) in a related context.

We turn to the verification of the conditions stated in Assumptions (L), (P1) and (P2) in case of \( L^\kappa \)-\( m \)-approximability as described in Section 2. Conditions of Assumptions (A1) and (A2) as well as (B1) and (B2) concerning estimation will be verified separately in the next section.

**Theorem 3.7.** Let \( \{Y_i\}_{i \in \mathbb{Z}} \) be causal and \( L^\kappa \)-\( m \)-approximable with \( \kappa > 2 \), a rate \( \delta(m) = m^{-\gamma} \) for some \( \gamma > 2 \). Then Assumptions (3.2), (L) and (P2) are fulfilled.

The next proposition applies the famous work of Móricz (1976) which, in combination with a result of Tomács and Libor (2006), will allow us to establish condition (P1).

**Proposition 3.8.** Assume that for some constant \( \kappa > 2 \), it holds that, as \( n \to \infty \),
\[
\left[ E \left\| \sum_{i=1}^{n} Y_i \right\|^\kappa \right]^{1/\kappa} = \mathcal{O}(n^{1/2}). \tag{3.12}
\]
Then, it holds that, as \( n \to \infty \),
\[
\max_{1 \leq k \leq n} k^{-1/2} \left\| \sum_{i=1}^{k} Y_i \right\| = \mathcal{O}_P((\log n)^{1/\kappa}).
\]
Now, in order to verify Assumption (P1), it is sufficient to show \((3.12)\) for \(\{Y_t\}_{t \in \mathbb{Z}}\) and \(\{Y_{-t}\}_{t \in \mathbb{Z}}\). For \(L^\kappa\)-\(m\)-approximable and causal time series, \((3.12)\) follows from Berkes et al. (2013) Theorem 3.3 given that \(\kappa \in (2,3)\). We verify \((3.12)\) for \(L^\kappa\)-\(m\)-approximable noncausal time series based on Berkes et al. (2011) Proposition 4.

**Proposition 3.9.** Let \(\{Y_t\}_{t \in \mathbb{Z}}\) be \(L^\kappa\)-\(m\)-approximable (not necessarily causal) with \(\kappa \in [2,3)\) and \(\sum_{m=1}^{\infty} \delta(m) < \infty\). Then, as \(n \to \infty\), it holds that

\[
\left[ E\|\sum_{i=1}^{n} Y_i\|^\kappa \right]^{1/\kappa} = O(n^{1/2}).
\]

A combination of Propositions 3.8 and 3.9 immediately yields the following:

**Corollary 3.10.** Let \(\{Y_t\}_{t \in \mathbb{Z}}\) be \(L^\kappa\)-\(m\)-approximable with \(\kappa > 2\) and causal with \(\sum_{m=1}^{\infty} \delta(m) < \infty\), then condition (P1) is fulfilled with rate \(g(n) = (\log n)^{1/2}\).

Altogether, Theorem 3.7 and Corollary 3.10 verify conditions (3.2), (L), (P1) and (P2) under \(L^\kappa\)-\(m\)-approximability with \(\delta(m) = O(m^{-\gamma})\) for \(\gamma > 2\). The remaining Assumptions (A1), (A2), (B1) and (B2), which ensure the validity of (3.8) and (3.10), in view of corresponding Theorems 3.3 and 3.6 can be verified using suitable estimates which will be shown in the next section.

### 4 Estimation of the eigenstructure

In this section we discuss suitable estimates \(\{\hat{v}_i\}\) and \(\{\hat{\lambda}_i\}\) for \(\{v_i\}\) and \(\{\lambda_i\}\) which is, as pointed out by Horváth et al. (2012b), an intricate problem. One possibility to obtain such estimates, suggested by Horváth et al. (2012b), is to consider the eigenstructure of Bartlett-type estimators \(\hat{C}\) for \(C\) of the following general form

\[
(\hat{C}x)(t) = \int \hat{\zeta}(t,s)x(s)ds,
\]

where \(x \in L^2[0,1]\), \(t \in [0,1]\). The kernel \(\hat{\zeta}\) is given by

\[
\hat{\zeta}(t,s) = \sum_{r=-n}^{n} K(r/h_n)\hat{\zeta}_r(t,s), \quad t, s \in [0,1],
\]

(4.1)

with covariance estimators

\[
\hat{\zeta}_r(t,s) = \begin{cases} \frac{1}{n} \sum_{i=1}^{n-r} (X_i(t) - \bar{X}_n(t)) (X_{i+r}(s) - \bar{X}_n(s)), & r \geq 0, \\ \hat{\zeta}_{-r}(s,t), & r < 0, \end{cases}
\]

a symmetric, bounded and compactly supported kernel function \(K(x)\) with \(K(0) = 1\) and a bandwidth \(h_n \to \infty\) fulfilling \(h_n = o(n)\). (Notice that, due to the compact support of \(K(x)\), the summation in (4.1) is only up to \(ch_n\) for some \(c > 0\).) These estimates were explored by Horváth et al. (2012b) under \(L^\kappa\)-\(m\)-approximability in the context of a related two-sample problem.

We restrict ourselves now to the framework of \(L^\kappa\)-\(m\)-approximable time series. For the sake of clarity we consider exponential decay rates \(\delta(m) = \exp(-cm), c > 0\). This is not very restrictive and already covers important time series models, in particular the functional AR\((p)\) time series, and will be sufficient for our purposes. (The arguments of the proof can be adapted straightforward to polynomial rates
\炮(m) but then expression (4.2) below and the restrictions on the bandwidth in the
next Theorem 4.1 become nontransparent. Therefore this case is omitted here.)

Notice that \( \hat{\mathcal{C}} \) is symmetric Hilbert-Schmidt, hence has a spectral decomposition

\[
(\hat{\mathcal{C}}x)(t) = \sum_{j=1}^{\infty} \hat{\lambda}_j(x, \hat{v}_j) \hat{v}_j(t)
\]

with real eigenvalues \( \hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \) and corresponding orthonormal eigenfunctions \( \hat{v}_1(t), \hat{v}_2(t), \ldots \). Here, we use the same notation for the eigenstructure as used for
generic estimates before but this should not lead to any confusion. The follow-
ing Theorem 4.1 is an extension of Theorem 2 of Horváth et al (2012b), where
consistency has been shown (under weaker assumptions). We consider the case
of \( L^4\)-m-approximable time series (i.e. \( \kappa = 4 \)) which allows us to work with the
variances of the estimates.

**Theorem 4.1.** Let \( \{Y_i\}_{i \in \mathbb{Z}} \) be \( L^4\)-m-approximable (not necessarily causal) with \( \delta(m) = \exp(-cm), c > 0 \). Assume that, as \( x \to 0 \),

\[
|K(x) - 1| = o(x^\rho)
\]

for some \( \rho \geq 1 \) and that \( h_n = \mathcal{O}(n^{1/\gamma}), \gamma > 2 \). Then under \( H_0 \), it holds that, as \( n \to \infty \),

\[
\|\hat{\zeta} - \zeta\| = \mathcal{O}_P \left( (\log n)^{1/2} \left( \frac{h_n}{n^{1/2}} \right) + \left( \frac{\log n}{h_n} \right)^\rho \right).
\]

(4.2)

As one would expect, the smoothness of the kernel \( K(x) \) at \( x = 0 \) effects (i.e.
 improves) the rate of convergence in (4.2). Typical values are \( \rho = 1 \) for the Bartlett
kernel \( K(x) = (1 - |x|)1_{[-1,1]}(x) \), \( \rho = 2 \) for the Parzen kernel and arbitrary large
\( \rho \) for the flat-top kernels. The main implication of Theorem 4.1 for us is that
\( \|\hat{\zeta} - \zeta\| = \mathcal{O}_P(n^{-\varepsilon}) \) holds true for some \( \varepsilon > 0 \). This allows the verification of (A1)
and (A2) via Corollary 3.10 together with Lemmas 2.2 and 2.3 of Horváth and
Kokoszka (2012). For (A2) we have to assume additionally, as common in the
functional setup, that the first \( d \) eigenvalues of \( \mathcal{C} \) are simple, i.e. that

\[
\lambda_1 > \lambda_2 > \ldots > \lambda_d > \lambda_{d+1}
\]

holds true. Hence, decomposition (3.3) is unique up to signs.

**Proposition 4.2.** Under \( H_0 \) and the assumptions of Theorem 4.1 the Assumption
(A1) holds true. If additionally (4.3) is assumed, then Assumption (A2) also holds
true.

We conclude this section by an observation, which follows in view of (3.11) and
due to the Lemma B.2 of Horváth et al (2013) (cf. also (3.5) and (3.6) therein).

**Proposition 4.3.** Let \( \{Y_i\}_{i \in \mathbb{Z}} \) be \( L^4\)-m-approximable and causal. Under \( H_A \) and
the assumptions of Theorem 4.1 the Assumptions (B1) and (B2) hold true.

As already mentioned in Remark 2.2 the dependence condition in Horváth et al
(2013, disp. (2.4)) is slightly different. However, Lemma B.2 can be restated under
our conditions due to Jirak (2013 Theorem 1.2).
5 Simulations

5.1 Monte Carlo simulation

We proceed with a Monte Carlo simulation of the finite sample behavior. In order to describe our setting and implementation details we recall that $X_i(t) = \mu_i(t) + Y_i(t)$.

Simulation setup: The signal $\mu_i$ is set to $\mu_i(t) \equiv 0$ for $i = 1, \ldots, n$ under the null hypothesis and

$$
\mu_i(t) = \begin{cases} 
0, & i = 1, \ldots, \lfloor n/2 \rfloor, \\
\sin(t), & i = \lfloor n/2 \rfloor + 1, \ldots, n 
\end{cases}
$$

(5.1)

under the alternative. The innovations follow the formal functional AR(1) model

$$
Y_i(t) = \int \Psi(t, s) Y_{i-1}(s) ds + \varepsilon_i(t),
$$

(5.2)

for $t \in [0, 1]$ and $i \in \mathbb{Z}$, where the shocks $\{\varepsilon_i\}$ are assumed to be Gaussian. Under the assumption of $\|\Psi\| < 1$ this equation is known to have a unique $L^\kappa$-approximable solution where $\kappa \geq 2$ is arbitrary (due to Gaussianity of $\varepsilon_i$'s) and where the decay rate $\delta(m)$ is exponential (cf. Horváth and Kokoszka (2010)). We will analyze the performance using different kernels

$$
\Psi_G(t, s) = C_G \exp((t^2 + s^2)/2), \quad \Psi_W(t, s) = C_W \min(t, s)
$$

normalized by constants $C_G, C_W \geq 0$, such that $\|\Psi\| = \psi$ for a prescribed value $\psi \in [0, 1)$. These kernels are common benchmarks and $\Psi_G$ and $\Psi_W$ are usually referred to as Gaussian or Wiener kernels, respectively (cf. Horváth and Kokoszka (2012)).

Implementation details: We have implemented the procedure in R using the “fda”-package. The shocks $\varepsilon_i(\cdot)$ are generated as paths of Brownian bridges on $[0, 1]$ and are represented as functional objects via the fda-function Data2fd(\ldots) by using a B-Spline basis of 25 functions. The same basis is also used to represent the kernel $\Psi$ and the innovations $Y_i(\cdot)$. More precise, the bivariate function $\Psi(t, s)$ is discretized on an equidistant grid $0 = t_1 < t_2 < \ldots < t_T = 1$ and for each $k = 1, \ldots, T$ the univariate function $\Psi(t_k, \cdot)$ is then represented as a functional object. Next,

$$
I_i(t_k) = \int \Psi(t_k, s) Y_{i-1}(s) ds
$$

are computed for all $k = 1, \ldots, T$ and $I_i(\cdot)$ itself is represented as a functional object with domain $[0, 1]$, again, using the same B-Spline basis as in the previous steps. Having computed $I_i$ we are in the position to add up $Y_i(\cdot) = I_i(\cdot) + \varepsilon_i(\cdot)$. The correct representation of the dependence structure is ensured by using a (so-called) burn-in period of length $N_B = 100$, i.e. $X_1, \ldots, X_n$ are generated iteratively according to (5.1) and (5.2) beginning with $Y_{-N_B+1} := \varepsilon_{-N_B+1}$ where, finally, the first $N_B$ observations $X_{-N_B+1}, \ldots, X_0$ are discarded.

The computation of the long run covariance estimator is carried out following Horváth et al (2012a) by using 25 Fourier basis functions. For simplicity we make use of a plain kernel $K(x) = 1[-1,1](x)$, which yields satisfactory results. The overall picture remains comparable if one chooses e.g. a flat top kernel as in Horváth et al (2012b) disp. (4.1)) or a Bartlett kernel, instead.
Critical values: The convergence in (3.8) is rather slow. For that reason, we follow the idea investigated by Csörgő and Horváth (1997), also successfully applied in a functional setting by Torgovitski (2014), by using quantiles of
\[
V_n = \sup_{t \in I_n} \frac{\|B^d(t)\|}{(1 - t)^{1/2}},
\]
(5.3)
where \(\{B^d(t), 0 \leq t \leq 1\}\) is a \(d\)-dimensional process with components given by independent standard Brownian bridges \(\{B_i(t), 0 \leq t \leq 1\}\) and e.g. \(I_n = [h_n, 1 - h_n]\) with \(h_n = (\log n)^{3/2}/n\). Asymptotic correctness of this choice follows from (3.8) (cf. Csörgő and Horváth (1997, Corollary 1.3.1) and the proof of Torgovitski (2014, Corollary 4.3)). An essential point is that quantiles of (5.3) can be computed using the expansion
\[
P(V_n \geq x) = \frac{x^d}{2^{d/2} \Gamma(d/2)} \left\{ \left( 1 - \frac{1 - h_n^2}{x^2} \right) \log \frac{1 - h_n^2}{h_n^2} + \frac{4}{x^2} + O(x^{-4}) \right\}.
\]
(5.4)
This representation is well known as Vostrikova’s tail approximation (see Vostrikova (1981, disp. (18)) and also Csörgő and Horváth (1997, disp. (1.3.26))).

Dimension and bandwidth selection: Parameters \(d\) and \(h = h_n\) remain to be specified where especially the selection of \(h\) is known to be a complex practical problem. For example \(d\) can be chosen according to the generalized CPV-Criterion (cf. Section 4.1 of Horváth et al (2013)) and \(h\) could be specified (in appropriate cases) guided by rules from scalar time series as demonstrated by Hörmann and Kokoszka (2010). However, both issues are not the focus of our research and therefore an overview for various parameters is presented in the tables below.

The behavior under \(H_0\) or under \(H_A\), respectively, is demonstrated in Tables 1 - 4 (based on 1000 repetitions). For moderate dependence and moderate sample sizes the procedure performs very well (comparable for both kernels \(\Psi_W, \Psi_G\)) and shows overall robustness with respect to the selection of various dimensions \(d\) and bandwidths \(h\). With increasing sample size the “bias” due to the dependencies fades out which is in accordance with the nonparametric nature of the procedure.

5.2 Application to Load Profiles
As a real-life example we take a closer look at electricity consumption data. This is inspired by the analysis of electricity data of Horváth and Rice (2014b). We consider load profiles for the low voltage electricity network of the German electricity distribution company E.ON Mitte AG (now “EnergieNetz Mitte”) for 2012. This dataset is publicly accessible through www.energienetz-mitte.de.

Data description: The load profiles are based on quarter-hourly measurements (in kW), i.e. each of 366 days consists of 96 highly correlated observations. We split the original time series into segments corresponding to days and view each daily record as a curve, that is treat it as functional data (cf. Figure 1).
| n   | ||Ψ_G|| | d | h = 1 | h = 2 |
|-----|-------|---|------|------|
| 50  |       |   | 1    | 2    | 3    | 4    | 5    | 1    | 2    | 3    | 4    | 5    |
|     | 0.1   |   | 7.2  | 6.9  | 5.1  | 3.4  | 3.9  | 9.9  | 6.7  | 4.9  | 5.1  | 5.2  |
|     | 0.2   |   | 8.7  | 6.2  | 4.3  | 3.6  | 2.6  | 8.2  | 6.5  | 5.9  | 6.2  | 5.8  |
|     | 0.4   |   | 8.7  | 5.3  | 3.4  | 3.9  | 2.3  | 7    | 5.8  | 5.3  | 5.3  | 5.7  |
|     | 0.6   |   | 8.9  | 5.2  | 3.2  | 2.8  | 2.3  | 4.7  | 2.3  | 3.1  | 3.2  | 4    |
|     | 0.8   |   | 11.8 | 7.9  | 4.7  | 2.9  | 2.9  | 6.4  | 3.1  | 3    | 3.1  | 3.8  |
| 100 |       |   | 11.5 | 10.1 | 9.7  | 8.6  | 6.5  | 9.8  | 9.4  | 7.9  | 6.1  | 5.9  |
|     | 0.2   |   | 7.8  | 7    | 5.5  | 5.4  | 4.6  | 9    | 7.3  | 6    | 5.1  | 4.8  |
|     | 0.4   |   | 9.3  | 6.8  | 5.6  | 5.9  | 4.9  | 6.7  | 4.6  | 5.1  | 3.8  | 3.3  |
|     | 0.6   |   | 14.6 | 9.3  | 8.8  | 6.1  | 4.8  | 6.7  | 4.9  | 3.9  | 4.5  | 4    |
|     | 0.8   |   | 16.1 | 11.1 | 8.2  | 7    | 5.2  | 7.1  | 6    | 4.6  | 4.1  | 3.5  |
| 300 |       |   | 10.8 | 10.5 | 12.1 | 10.8 | 9.6  | 11.1 | 9.3  | 7.6  | 6.9  | 7.8  |
|     | 0.2   |   | 9.6  | 10.1 | 8.9  | 8.5  | 8.2  | 9    | 9.8  | 8    | 7.7  | 7    |
|     | 0.4   |   | 10.9 | 10   | 9.5  | 9.6  | 9.3  | 9.7  | 10.9 | 8.5  | 7.8  | 8.5  |
|     | 0.6   |   | 14.4 | 12   | 9.1  | 8.4  | 8.2  | 7.4  | 6.9  | 6.4  | 6.5  | 6.7  |
|     | 0.8   |   | 20.2 | 15.2 | 12.9 | 12.8 | 9.8  | 10.1 | 8.8  | 7.9  | 8.1  | 7.4  |
| 500 |       |   | 10.8 | 9.3  | 9.4  | 9    | 8.9  | 10.1 | 8.8  | 8.3  | 7.9  | 9.1  |
|     | 0.2   |   | 9.7  | 9.1  | 9    | 9.7  | 9.6  | 8.1  | 8.3  | 6.8  | 6.4  | 7.1  |
|     | 0.4   |   | 11.7 | 11.2 | 10.8 | 10.2 | 10.1 | 9.5  | 9.5  | 9.9  | 9    | 8.4  |
|     | 0.6   |   | 15.5 | 12.7 | 11.1 | 11.3 | 10.2 | 10.3 | 9.2  | 8.3  | 7.8  | 7.8  |
|     | 0.8   |   | 20.6 | 18.1 | 15.6 | 14.1 | 13   | 12.4 | 9.1  | 10.3 | 9.1  | 8.3  |
| 300 |       |   | 10.8 | 9.3  | 9.4  | 9    | 8.9  | 10.1 | 8.8  | 8.3  | 7.9  | 9.1  |
|     | 0.2   |   | 9.7  | 9.1  | 9    | 9.7  | 9.6  | 8.1  | 8.3  | 6.8  | 6.4  | 7.1  |
|     | 0.4   |   | 11.7 | 11.2 | 10.8 | 10.2 | 10.1 | 9.5  | 9.5  | 9.9  | 9    | 8.4  |
|     | 0.6   |   | 15.5 | 12.7 | 11.1 | 11.3 | 10.2 | 10.3 | 9.2  | 8.3  | 7.8  | 7.8  |
|     | 0.8   |   | 20.6 | 18.1 | 15.6 | 14.1 | 13   | 12.4 | 9.1  | 10.3 | 9.1  | 8.3  |
| 500 |       |   | 10.8 | 9.3  | 9.4  | 9    | 8.9  | 10.1 | 8.8  | 8.3  | 7.9  | 9.1  |
|     | 0.2   |   | 9.7  | 9.1  | 9    | 9.7  | 9.6  | 8.1  | 8.3  | 6.8  | 6.4  | 7.1  |
|     | 0.4   |   | 11.7 | 11.2 | 10.8 | 10.2 | 10.1 | 9.5  | 9.5  | 9.9  | 9    | 8.4  |
|     | 0.6   |   | 15.5 | 12.7 | 11.1 | 11.3 | 10.2 | 10.3 | 9.2  | 8.3  | 7.8  | 7.8  |
|     | 0.8   |   | 20.6 | 18.1 | 15.6 | 14.1 | 13   | 12.4 | 9.1  | 10.3 | 9.1  | 8.3  |

Table 1: Empirical sizes for functional AR(1) with Gaussian kernel
Ψ_G; nominal level of 10%
Table 2: Empirical sizes for functional AR(1) with Wiener kernel $\Psi_W$; nominal level of 10%  

| $n$ | $\|\Psi_W\|$ | $d$ | $h = 1$ | $h = 2$ | $h = 3$ | $h = 4$ | $h = 1$ | $h = 2$ | $h = 3$ | $h = 4$ |
|-----|----------------|-----|---------|---------|---------|---------|---------|---------|---------|---------|
| 50  | 0.1            | 9.6 | 7.6     | 4.6     | 3.9     | 2.9     | 10.1    | 8       | 6.8     | 6.4     | 5.9     |
|     | 0.2            | 9.8 | 5.9     | 4.6     | 2.8     | 2.8     | 8       | 4.4     | 4       | 4.8     | 5.1     |
|     | 0.4            | 9.9 | 5.7     | 3.8     | 2.9     | 2.3     | 5.2     | 3.5     | 2.8     | 2.4     | 3.1     |
|     | 0.6            | 11.2| 5.2     | 2.7     | 2.1     | 1.8     | 4.8     | 3.3     | 2.5     | 2.7     | 3.2     |
|     | 0.8            | 20.6| 11.1    | 4.8     | 3       | 2.5     | 3.8     | 1.9     | 1.4     | 2.1     | 2.1     |
| 100 | 0.1            | 8.6 | 8.1     | 6.5     | 6.7     | 6.5     | 9       | 8.7     | 6.6     | 7.1     | 6       |
|     | 0.2            | 8.8 | 8       | 7.9     | 6       | 5.8     | 8.5     | 6.7     | 6.1     | 5.1     | 4.9     |
|     | 0.4            | 10.6| 8.6     | 6.1     | 5.1     | 5       | 5.7     | 7.1     | 4.5     | 4.4     | 2.9     |
|     | 0.6            | 13.9| 10.3    | 7.4     | 4.8     | 3.9     | 6       | 4.6     | 3.2     | 2.8     | 2.7     |
|     | 0.8            | 28.6| 19.6    | 13.6    | 9.7     | 7.5     | 11.5    | 7       | 6.1     | 4.1     | 3.3     |
| 300 | 0.1            | 10.5| 9.4     | 9.1     | 8.7     | 7.3     | 10.5    | 9.9     | 9       | 6.8     | 7       |
|     | 0.2            | 9.6 | 10.3    | 10      | 8.5     | 8.9     | 9.2     | 8.5     | 8.3     | 7.6     | 7.2     |
|     | 0.4            | 10.3| 9.8     | 10.5    | 8.5     | 6.9     | 7       | 6.9     | 7.3     | 5.7     | 5.7     |
|     | 0.6            | 19.1| 15.5    | 12.7    | 9.9     | 8.5     | 8.9     | 8.2     | 7.7     | 6.4     | 6       |
|     | 0.8            | 33.4| 25.5    | 20.7    | 18.6    | 15.4    | 15.7    | 12.5    | 10.5    | 9.7     | 7.6     |
| 500 | 0.1            | 12.5| 10.2    | 9       | 8.6     | 8.6     | 10.8    | 10.5    | 10.3    | 9.3     | 9.4     |
|     | 0.2            | 10.8| 10      | 8.9     | 10.1    | 8.4     | 9.5     | 8.8     | 8.3     | 7.4     | 6.9     |
|     | 0.4            | 13.7| 13.3    | 11.3    | 9.9     | 8.9     | 9.3     | 8.5     | 9.3     | 8.4     | 7.9     |
|     | 0.6            | 18.2| 14.3    | 12.8    | 11.7    | 10.9    | 11.4    | 9.5     | 7.9     | 7.5     | 8.6     |
|     | 0.8            | 37.9| 29      | 22.4    | 19.1    | 16.7    | 17.9    | 12.7    | 9.9     | 9.9     | 8.7     |
| 100 | 0.2            | 7.4 | 8       | 7.9     | 6.5     | 6.2     | 9.2     | 9       | 7.5     | 6.8     | 6.8     |
|     | 0.4            | 7.3 | 8.1     | 5.7     | 5.8     | 5       | 7.2     | 4.9     | 5.3     | 5.8     | 6.3     |
|     | 0.6            | 5.5 | 4.3     | 4.1     | 4.1     | 4       | 5.1     | 3.4     | 3.3     | 4.2     | 4.8     |
|     | 0.8            | 5.8 | 3.9     | 3.1     | 3.3     | 3.2     | 2.8     | 3.3     | 2.9     | 4.3     | 4.7     |
| 300 | 0.2            | 9.3 | 8.4     | 7.4     | 6.5     | 6.2     | 9.2     | 9       | 7.5     | 6.8     | 6.8     |
|     | 0.4            | 7.3 | 1.1     | 7.9     | 6.8     | 6.9     | 10.1    | 9.1     | 8.2     | 7.1     | 6.4     |
|     | 0.6            | 8.9 | 7       | 6.6     | 6.6     | 5.7     | 6.2     | 5.4     | 5.7     | 4.8     | 5.8     |
|     | 0.8            | 9.7 | 7.7     | 7.1     | 7.1     | 6.1     | 6.2     | 7.3     | 4.7     | 6       | 5.5     |
| 500 | 0.2            | 9.7 | 8.4     | 7.7     | 8.4     | 8       | 9.6     | 8.6     | 9.7     | 8.5     | 7.5     |
|     | 0.4            | 9.8 | 8.9     | 7.2     | 7.1     | 6.5     | 6.3     | 7.9     | 7.6     | 8.7     | 8       |
|     | 0.6            | 9.1 | 8.1     | 6.3     | 5.7     | 5.7     | 6.7     | 6.9     | 8.8     | 7.4     | 6.4     |
|     | 0.8            | 10.7| 8.8     | 7.3     | 7.9     | 7.4     | 8.5     | 7.8     | 7.3     | 7       | 7.6     |
Table 3: Empirical power for functional AR(1) with Gaussian kernel $\Psi_G$; $\mu_1 \equiv 0$ and $\mu_n(t) = \sin(t)$; nominal level of 10%

| $d$ | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
|-----|---|---|---|---|---|---|---|---|---|---|
| $\parallel \Psi_G \parallel$ | 50 | 0.1 | 99.4 | 99.9 | 99.9 | 99.1 | 52.4 | 95.2 | 40.8 | 3.3 | 4 | 4 |
| | 0.2 | 99.3 | 99.8 | 99.8 | 98.6 | 52.1 | 91.8 | 37.5 | 2 | 2.2 | 3 |
| | 0.4 | 94.8 | 99.7 | 99.1 | 96.2 | 45 | 83 | 28.5 | 1.5 | 1.8 | 2.8 |
| | 0.6 | 89.3 | 98.7 | 98.9 | 94.7 | 40.6 | 68.5 | 23 | 1.8 | 1.7 | 2.2 |
| | 0.8 | 81.1 | 97.6 | 96.1 | 91.2 | 38.7 | 54 | 18.9 | 0.6 | 0.7 | 1.7 |
| | 100 | 0.1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.2 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.4 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.6 | 99.7 | 100 | 100 | 100 | 100 | 99.5 | 100 | 100 | 100 | 100 |
| | 0.8 | 98.7 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 300 | 0.1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.2 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.4 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.6 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.8 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 500 | 0.1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.2 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.4 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.6 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.8 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 300 | 0.1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.2 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.4 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.6 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.8 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |

$h = 3$

| 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|---|---|---|
| 50 | 0.1 | 45.3 | 1.9 | 2.8 | 4.2 | 5.3 | 1.1 | 2.2 | 2.7 | 4.5 | 6.3 |
| | 0.2 | 37.4 | 2.1 | 2.7 | 3.7 | 4.3 | 1.2 | 2.3 | 4 | 4.4 | 6.5 |
| | 0.4 | 28.2 | 1.9 | 2.8 | 3.8 | 5.3 | 1.1 | 2.7 | 3.9 | 4.2 | 7.2 |
| | 0.6 | 15.8 | 1 | 2 | 3 | 3.7 | 0.4 | 1.6 | 2.4 | 4.3 | 5.5 |
| | 0.8 | 9.7 | 0.7 | 1 | 2 | 3.4 | 1 | 1.2 | 2.9 | 4.3 | 5.8 |
| 100 | 0.1 | 100 | 100 | 100 | 13.6 | 3.8 | 99.9 | 93.2 | 2.2 | 3 | 2.8 |
| | 0.2 | 100 | 100 | 99.1 | 12.7 | 4.6 | 99.9 | 91.6 | 2.3 | 3 | 3.7 |
| | 0.4 | 99.9 | 100 | 98.9 | 9.3 | 2.9 | 99.3 | 86.5 | 2.5 | 3.4 | 3.9 |
| | 0.6 | 99 | 100 | 97.6 | 8.9 | 1.9 | 93.8 | 81 | 2.8 | 3.9 | 4.1 |
| | 0.8 | 91.5 | 99.7 | 95.8 | 8.9 | 2.6 | 82.1 | 72.3 | 0.6 | 2.2 | 2.2 |
| 300 | 0.1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.2 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.4 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.6 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.8 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 500 | 0.1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.2 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.4 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.6 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | 0.8 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
Table 4: Empirical power for functional AR(1) with Wiener kernel $\Psi_W$; $\mu_1 \equiv 0$ and $\mu_n(t) = \sin(t)$; nominal level of 10%

| $n$ | $\|\Psi_W\|$ | $d$ | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
|-----|----------------|-----|---|---|---|---|---|---|---|---|---|---|
|     |                |     | $h = 1$ |   |   |   |   | $h = 2$ |   |   |   |   |
| 50  | 0.1            | 99  | 99.9 | 99.7 | 98.2 | 54.3 | 93.3 | 36.6 | 2.2 | 3.5 | 4.1 |
|     | 0.2            | 97.9 | 99.7 | 99.3 | 97.2 | 53.1 | 91.3 | 31.4 | 2.3 | 3.1 | 3.4 |
|     | 0.4            | 93  | 98.8 | 99  | 95.4 | 43.4 | 73.7 | 22.4 | 2   | 2.7 | 2.7 |
|     | 0.6            | 81.1 | 93.4 | 94.7 | 90.9 | 36.4 | 57.1 | 12.9 | 1.4 | 1.2 | 1.5 |
|     | 0.8            | 73.2 | 81.9 | 89.1 | 84.6 | 40.3 | 37.3 | 9.5  | 1.5 | 1.3 | 1.6 |
| 100 | 0.1            | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
|     | 0.2            | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
|     | 0.4            | 99.9 | 100 | 100 | 100 | 100 | 100 | 99.6 | 100 | 100 | 100 | 100 |
|     | 0.6            | 98.3 | 100 | 100 | 100 | 100 | 100 | 96.9 | 100 | 100 | 100 | 100 |
|     | 0.8            | 93.4 | 99.7 | 100 | 100 | 100 | 100 | 85.2 | 98 | 99.9 | 100 | 100 |
| 300 | 0.1            | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
|     | 0.2            | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
|     | 0.4            | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
|     | 0.6            | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
|     | 0.8            | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 500 | 0.1            | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
|     | 0.2            | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
|     | 0.4            | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
|     | 0.6            | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
|     | 0.8            | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
|     |                |     | $h = 3$ |   |   |   |   | $h = 4$ |   |   |   |   |
| 50  | 0.1            | 40.1 | 1.7 | 2.6 | 3.1 | 4.6 | 1.2 | 2.1 | 4.1 | 4.6 | 7  |
|     | 0.2            | 31.6 | 1.5 | 2.2 | 3.2 | 4.4 | 0.8 | 1.7 | 2.1 | 3.2 | 6  |
|     | 0.4            | 21  | 2.1 | 2.4 | 2.8 | 3.8 | 0.8 | 2.1 | 3.6 | 5  | 7.8 |
|     | 0.6            | 10.5 | 1.5 | 2.2 | 2.2 | 3.3 | 0.5 | 1.6 | 1.6 | 2.3 | 5  |
|     | 0.8            | 2.8  | 0.5 | 1.1 | 1.6 | 2.5 | 0.9 | 1   | 1.4 | 1.6 | 3  |
| 100 | 0.1            | 100 | 100 | 99.8 | 10.5 | 3.5 | 99.9 | 92.5 | 3  | 3.6 | 4.8 |
|     | 0.2            | 99.9 | 100 | 99.4 | 10.5 | 3.5 | 99.7 | 90  | 2.5 | 3.4 | 3.3 |
|     | 0.4            | 99.3 | 99.7 | 97.7 | 10.9 | 3.2 | 97.2 | 76.7 | 2.9 | 2.6 | 2.8 |
|     | 0.6            | 93  | 98.1 | 95.4 | 9  | 2.9 | 81.4 | 56.3 | 1.7 | 1.8 | 2.9 |
|     | 0.8            | 69.3 | 92.4 | 88 | 9.2 | 2.4 | 50.1 | 46.3 | 1.6 | 1.7 | 2  |
| 300 | 0.1            | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
|     | 0.2            | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
|     | 0.4            | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
|     | 0.6            | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
|     | 0.8            | 99.9 | 100 | 100 | 100 | 100 | 100 | 99.8 | 100 | 100 | 100 | 100 |
| 500 | 0.1            | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
|     | 0.2            | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
|     | 0.4            | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
|     | 0.6            | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
|     | 0.8            | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
In order to apply the testing procedure we proceed as described in Section 5 and represent the discrete daily records as functional objects in rescaled time $t \in [0,1]$ using 25 $B$-Spline basis-functions, hereby smoothing the data. In the next step, in view of the obviously different stochastic pattern, we remove all curves which correspond to weekends. The remaining dataset consists of 261 curves $X_1(t), \ldots, X_{261}(t)$ corresponding to workdays (cf. Figure 2). Now, to gain stationarity, all curves are log-transformed via

$$\tilde{X}_i(t) = \log(X_i(t)/X_i(0)).$$

For a discussion on this transformation we refer to Horváth and Rice (2014b). In the third step, we discard the observations $\tilde{X}_1(t), \ldots, \tilde{X}_{100}$ (i.e. the data before May 21st) which show a somewhat erratic behavior.
The remaining observations $\tilde{X}_{101}(t), \ldots, \tilde{X}_{261}(t)$ exhibit an obvious large abrupt change in the mean at $\tilde{X}_{216}$, i.e. at line “d” in Figure 2 and a quite stationary behavior before and after the jump.

Data analysis: Due to the large jump, we expect the procedure to reject the null hypothesis distinctly (the slight trend in segment “c - d” of Figure 2 should not affect the performance), which is confirmed in Table 5: the procedure rejects the null hypothesis for a wide range of parameters (where the $p$-values are based on approximation (5.4)). The tests are carried out using a plain kernel $K(x) = 1_{[-1,1]}(x)$, different bandwidths $h = 0, \ldots, 4$ and various subspace dimensions $d = 1, \ldots, 6$. (We set $\hat{\zeta}(t,s) = \hat{\zeta}_0(t,s)$ for $h = 0$.) Note that $\hat{T}_n$ is largest (or vice versa the $p$-values are smallest) for $h = 0$, i.e. when the dependence structure is not taken into account. However, the results for $h > 0$ are more reliable since there is evidence of dependencies in the data described in the following. We checked for independence using the Portmanteau Test of Independence of Gabrys and Kokoszka (2007). Note that this procedure requires mean zero data. Therefore, we center the data by the sample mean and in order to minimize the influence of small trends we restrict our considerations to the segment $\tilde{X}_{115}(t), \ldots, \tilde{X}_{190}(t)$ (i.e. segment “a - c” in Figure 2). The test yields small $p$-values $< 10^{-10}$ for a wide range of parameters $d$ (number of principal components) and $\hat{H}$ (maximum lag). We obtain somewhat larger, but still small, $p$-values (cf. Table 6) if we restrict ourselves further to the segment $\tilde{X}_{115}(t), \ldots, \tilde{X}_{165}(t)$, (i.e. segment “a - b” in Figure 2).
Table 6: $p$-values from the Portmanteau Test of Independence of $\hat{\lambda}$ applied to the segment $\hat{X}_{115}(t), \ldots, \hat{X}_{165}(t)$ of the load profile dataset. $d$ represents the number of principal components and $\hat{H}$ denotes the maximum lag used for the test.

| $\hat{H}$ | $d = 1$ | $d = 2$ | $d = 3$ | $d = 4$ |
|-----------|---------|---------|---------|---------|
| 1         | 0.0011  | 0.0228  | 0.0002  | 0.0023  |
| 2         | < 0.0001 | < 0.0001| < 0.0001| < 0.0001|
| 3         | < 0.0001 | < 0.0021| < 0.0001| < 0.0001|

Remark 5.1. The observed change is in accordance with the fact that electricity consumption in the winter is higher than in the summer. However, note that the curves in Figures 2 and 3 which correspond to the winter months are below those corresponding to summer moths, due to the rescaling $X_i(t)/X_i(0)$: The electricity demand during morning and daytime in winter and summer months is comparable. However, in the winter the demand in the evening and especially at midnight, i.e. at $X_i(0)$, is much higher.

6 Proofs

Proof of Theorem 3.3. We outline the important steps, thereby following the proof of Torgovitski (2014, Theorem 4.1), which in turn is largely based on considerations of Berkes et al (2009). Going through the proofs of Lemmas 6.4 and 6.5 of Torgovitski (2014) we see that Assumptions (P1) and (A2) ensure that, as $n \to \infty$,

$$|T_n(X; \hat{\nu}, \lambda) - T_n(Y; \nu, \lambda)| = o_P((\log \log n)^{-1/2})$$  \hspace{1cm} (6.1)

and therefore, taking Assumption (L) into account

$$\lim_{n \to \infty} P\left( a(\log n) T_n(X; \hat{\nu}, \lambda) - b_d(\log n) \leq x \right) = \exp(-2\exp(-x))$$  \hspace{1cm} (6.2)

holds true. Now, Assumption (A1) implies that $\lim_{n \to \infty} P(\hat{\lambda}_d > c) = 1$, for some $c > 0$, and $\max_{i=1, \ldots, d} |\hat{\lambda}_i^{1/2} - \lambda_i^{1/2}| = o_P((\log \log n)^{-1})$ hold true. Following the arguments of Torgovitski (2014, Lemma 6.7) we see that (6.2) and Assumption (A1) imply that

$$|T_n(X; \hat{\nu}, \hat{\lambda}) - T_n(X; \hat{\nu}, \lambda)| = o_P((\log \log n)^{-1/2}).$$

The assertion follows immediately by using (6.1) and Assumption (L).
Proof of Theorem 3.6. Let $m = \lfloor n\theta \rfloor$. Using standard arguments we obtain for sufficiently large $n$, on a set where $\lambda_r > 0$, that
\[
T_n \geq c_1 w(m/n) \left( n^{-1/2} \sum_{i=1}^{m} (X_i - \bar{X}_n), \hat{\nu}_r \right) |\hat{\lambda}_r|^{-1/2} \\
\geq c_2 w(\theta) \left| \left\langle n^{-1} \sum_{i=1}^{m} (Y_i - \bar{Y}_n), \hat{\nu}_r \right\rangle - \left( \left\langle \Delta, \hat{\nu}_r \right\rangle - \xi \right) \frac{m(n-m)}{n^{3/2}n^{1/2}} \right| \\
+ \xi \frac{m(n-m)}{n^{3/2}n^{1/2}} \times (\log \log n)^{1/2} \frac{n^{1/2}}{(\log \log n)^{1/2} \lambda_r^{1/2}} \\
=: c_2 w(\theta) |A_1 + A_2 + A_3| (\log \log n)^{1/2} \left( \frac{\lambda_r \log \log n}{n} \right)^{-1/2}.
\]
We get $A_1 = o_P(1)$ in view of Assumption (P2), $||\hat{\nu}_r|| = 1$ and the Cauchy-Schwarz inequality. Further, $A_2 = o_P(1)$ holds true on account of Assumption (B2) and the third term $A_3$ converges towards a nonzero constant, again due to (B2). The assertion follows now in view of Assumption (B1).

Proof of Theorem 3.7. Property (3.2) is stated in Theorem 1 of [Horváth et al. 2012b] and (P2) follows from ergodicity and stationarity. Note that (P2) is also immediately implied by [Berkes et al. 2013, Theorem 3.3]. We proceed with the verification of Assumption (L). Going carefully through the proof of [Csörgő and Horváth 1997, Theorem 4.1.3] and taking [Schmitz 2011, Theorem 2.1.4] into account -replacing all considerations for univariate time series with multivariate analogues- we see that it suffices to show the following conditions (C1) - (C3) below (cf. also [Kamgaing and Kirch 2014, Theorem 1.2.1]). Hereby, it is crucial that $L^\kappa$-$m$-approximable time series fulfill Assumption (M) by definition.

For one thing, we need a forward and a backward approximation of the projected innovations $\{Y_t\}$ (see (3.5)) by centered multivariate Brownian motions $\{W_1(n)\}$, $\{W_2(n)\}$ with covariance matrix $\Sigma$ (cf. Remark 3.1), i.e. we need conditions
\[
\left\lvert \sum_{i=1}^{n} Y_i - W_1(n) \right\rvert = O(n^{1/2-\eta}) \quad \text{a.s.,} \quad (C1) \\
\left\lvert \sum_{i=1}^{n} Y_{-i} - W_2(n) \right\rvert = O(n^{1/2-\eta}) \quad \text{a.s.} \quad (C2)
\]
for some $\eta > 0$. (We do not require independence between $\{W_1(n)\}$ and $\{W_2(n)\}$.)

For another thing we need the asymptotic independence and the asymptotic distribution of
\[
A_n^* := a(\log n) A_n - b_d(\log n), \\
B_n^* := a(\log n) B_n - b_d(\log n)
\]
with
\[
A_n = \max_{1 \leq k \leq n/\log n} k^{-1/2} \left\lvert \sum_{i=1}^{k} Y_i \right\rvert, \\
B_n = \max_{n - n/\log n \leq k < n} (n - k)^{-1/2} \left\lvert \sum_{i=k+1}^{n} Y_i \right\rvert.
\]
i.e. that
\[
\lim_{n \to \infty} P(A_n^* \leq s, B_n^* \leq t) = \lim_{n \to \infty} P(A_n^* \leq s)P(B_n^* \leq t) = \exp(-\exp(-s))\exp(-\exp(-t))
\] (C3)
holds true for all \(s, t \in \mathbb{R}\).

These conditions (C1) - (C3) are similar to conditions A.3 (i)-(iii) of Kamgaing and Kirch (2014, Theorem 1.2.1). Condition (C1) replaces Assumption A.3 (i), condition (C2) substitutes Assumption A.3 (ii) (cf. Schmitz (2011)) and finally condition (C3) corresponds to Assumption A.3 (iii) (up to normalizing sequences \(a(\log n)\) and \(b_d(\log n)\), cf. Csörgö and Horváth (1997, Theorem 4.1.3)). Notice that the additional Assumption A.1 of Kamgaing and Kirch (2014, Theorem 1.2.1) is trivially fulfilled in our setting, due to the shape of our test statistic.

We begin by showing (C1) and (C2). It is easy to see, that the projected time series \(\{Y_i\}_{i \in \mathbb{Z}}\) remains \(L^\kappa\)-\(m\)-approximable (now in \(\mathbb{R}^d\)) with same \(\kappa > 2\) and same rate \(\delta(m)\). In particular \(E|Y_0|^\kappa < \infty\), \(\kappa > 2\), holds true and \(\Sigma\) is obviously positive definite due to \(\lambda_d > 0\) (cf. also Remark 3.1). Hence, (C1) follows immediately from Theorem A.1 of Aue et al (2014) taking Csörgö and Horváth (1997, dispersion (A.1.16)) into account. Furthermore, a careful examination of the proof of Aue et al (2014, Theorem A.1) shows that their arguments do not rely on causality and their strong approximation in Theorem A.1 could be restated for general (noncausal) \(L^\kappa\)-\(m\)-approximable multivariate time series using Definition 2.1 with \(H = \mathbb{R}^d\) and \(\delta(m) = m^{-\gamma}\) for some \(\gamma > 2\). Note that in the proof of Theorem A.1 of Aue et al (2014), coupling expressions like, e.g.,
\[
E[Y_{0,r}Y_{j,r}] = E[Y_{0,r}(Y_{j,r} - Y_{j,r}^{(j)})],
\]
are only valid under causality, since \(E[Y_{0,r}Y_{j,r}^{(j)}] = 0\) is necessary. However, relation \(E[Y_{0,r}Y_{j,r}^{(j)}] = 0\) is valid under noncausality and it is known that the above expression can be easily replaced by
\[
E[Y_{0,r}Y_{j,r}] = E[Y_{j,r}(Y_{0,r} - Y_{0,r}^{(j)})] + E[Y_{0,r}(Y_{j,r} - Y_{j,r}^{(j)})].
\] (6.3)

Now, observe that after time inversion \(\{Y_{-i}\}_{i \in \mathbb{Z}}\) remains \(L^\kappa\)-\(m\)-approximable (in the sense of Definition 2.1) with the same \(\kappa\), the same rate \(\delta(m)\) and with the same long run covariance matrix \(\Sigma\). Hence, according to previous considerations, (C2) holds true, as well.

Finally, we verify (C3). In the setting of linear processes, Csörgö and Horváth (1997, Theorem 4.1.3) have shown asymptotic independence of \(A_n\) and \(B_n\) by replacing the \(Y_i\)'s in \(B_n\) by truncated approximations (cf. Csörgö and Horváth (1997, p.308)). We adapt this approach in a straightforward manner by considering \(m\)-dependent copies \(Y(m)\) (cf. (2.3)) and defining
\[
B_n' := \max_{n-n/\log n \leq k < n} (n-k)^{-1/2} \sum_{i=k+1}^{n} Y(m)_i,
\]
where \(m_n := n - 3n/\log n\). The representation of \(Y_i\)'s as a shift of i.i.d. random variables and the construction of the \(Y(m)_i\)'s ensures that \(Z_{k,r} := Y_{k,r} - Y(m)_k\) are equally distributed for all \(k\). Hence, it holds that
\[
E|Z_{k,r}|^2 = E|Z_{0,r}|^2 = O(m_n^{-2\gamma})
\]
for all $k \in \mathbb{Z}$ and $r = 1, \ldots, d$. Furthermore,
\begin{align*}
|B_n - B'_n| &\leq \max_{n-n/\log n \leq k < n} (n-k)^{-1/2} \left| \sum_{i=k+1}^{n} (Y_i - Y_i^{(m_n)}) \right| \\
&\leq d \sum_{r=1}^{d} \left( \max_{n-n/\log n \leq k < n} (n-k)^{-1/2} \left| \sum_{i=k+1}^{n} Z_{i,r} \right| \right).
\end{align*}
An application of the Hájeck-Rényi type inequality of Kounias and Weng (1969, Theorem 2) yields that
\begin{align*}
P \left( \max_{n-n/\log n \leq k < n} (n-k)^{-1/2} \left| \sum_{i=k+1}^{n} Z_{i,r} \right| > (\log n)^{-1/2} \right) \\
&\leq \left( (\log n)^{1/2} \sum_{k=n-n/\log n}^{n-1} (n-k)^{-1/2} \left( \text{E}|Z_{k,r}|^2 \right)^{1/2} \right)^{2} \\
&= (\log n) \text{E}|Z_{0,r}|^2 \left( \sum_{k=1}^{n/\log n} k^{-1/2} \right)^{2} = \mathcal{O} \left( (\log n)^{m_n-2\nu} \frac{n^{2\nu}}{\log n} \right) = \mathcal{O}(n^{1-2\nu}),
\end{align*}
which implies
\[ B_n = B'_n + o_P(a(\log n)^{-1}). \]

Now, observe that $B'_n$ and $A_n$ are independent because the sets $\{Y_i, i \leq n/\log n\}$ and
\[ \{Y_i^{(m_n)}, i \geq n - n/\log n, m_n = n - 3n/\log n\} \]
are obviously independent for sufficiently large $n$. Finally, (C3) follows from Horváth (1993, Lemma 2.2) taking Davidson (1982, Lemma 29.5) into account.

**Proof of Proposition 3.8** The well known results of Móricz (1976) show that moment inequalities for partial sums yield analogous moment inequalities for maxima of partial sums. Furthermore, in Tomács and Libor (2006) it is shown that inequalities for maxima of partial sums yield inequalities for weighted maxima of partial sums and vice versa. Carefully inspecting the proofs of Móricz (1976, Theorem 1) and of Tomács and Libor (2006, Theorem 2.1) we observe that the same results can be restated in our functional setting with $\kappa > 2$, as well. Therefore, Móricz (1976, Theorem 1) together with assumption (3.12) and Markov’s inequality yield
\begin{align*}
x^{\kappa} P \left( \max_{1 \leq k \leq n} \left\| \sum_{i=1}^{k} Y_i \right\| \geq x \right) &\leq E \left[ \max_{1 \leq k \leq n} \left\| \sum_{i=1}^{k} Y_i \right\|^{\kappa} \right] \leq c_1 n^{\kappa/2} \quad (6.4)
\end{align*}
for all $x > 0$, $n \in \mathbb{N}$ and some $c_1 > 0$. Next, we use $n^{\kappa/2} = \mathcal{O} \left( \sum_{k=1}^{n} k^{\kappa/2-1} \right)$ and apply Tomács and Libor (2006, Theorem 2.1) to obtain,
\begin{align*}
x^{\kappa} P \left( \max_{1 \leq k \leq n} k^{-1/2} \left\| \sum_{i=1}^{k} Y_i \right\| \geq x \right) &\leq c_2 \sum_{k=1}^{n} k^{-1} \leq c_3 \log n,
\end{align*}
for all $x > 0$, $n \in \mathbb{N}$ and some $c_2, c_3 > 0$. The conclusion follows on setting $x = c_4 (\log n)^{1/\kappa}$ with a suitable constant $c_4 > 0$. \hfill \qed

**Remark 6.1.** In the previous proof we applied Móricz (1976, Theorem 1) with $g(F_{b,n}) = n^\alpha$, $b = 0$ and $\alpha = \kappa/2 > 1$. In case of $\kappa = 2$ (i.e. $\alpha = 1$) an additional logarithmic term would appear on the right-hand side of (6.4) (cf. Móricz (1976, Theorem 3)).
We proceed with the proof of Proposition 3.9. Berkes et al (2011) have shown the corresponding result in the univariate setting and their techniques, slightly modified, are directly applicable to the functional setting, as shown below. Here, we demonstrate that Proposition 4 of Berkes et al (2011) is extensible to 
noncausal centered $L^\kappa$-m-approximable functional time series $\{Y_t\}_{t \in \mathbb{Z}}$. We want to emphasize that another, more sophisticated extension -yet for causal centered $L^\kappa$-m-approximable time series-, has been developed by Berkes et al (2013, cf. Theorems 3.1, 3.3 and Remark 3.2).

Proof of Proposition 3.9. We want to point out that -up to the functional setting-the proof presented here is for most parts identical with Berkes et al (2011, Proposition 4) and that we stay very close to their exposition. To avoid misunderstandings and for the convenience of the reader, we restate their proof in the functional setting, emphasizing the necessary modifications. In adaption to our situation, the major difficulty stems from the relations (37) - (39) of Berkes et al (2011) which are not clear in the functional setting for all $\kappa > 2$ and therefore are substituted by (6.11) below. This is done using a result of Berkes et al (2013) which is, however, restricted to $\kappa \in (2, 3)$.

Let $S_n = \sum_{i=1}^{n} Y_i$ denote the partial sums, let $\psi_2 > 0$, $\psi_\kappa > 0$ be arbitrary constants and let

$$D_\kappa := \sum_{m=0}^{\infty} (E\|X_0 - X_0^{(m)}\|^\kappa)^{1/\kappa} < \infty,$$

where the inequality holds true in view of $\sum_{m=1}^{\infty} \delta(m) < \infty$. The cases $\kappa = 2$ and $2 < \kappa < 3$ are treated separately, where the former one can be seen as follows: Via the decomposition

$$Y_0(t)Y_j(s) = \left(Y_0(t) - Y_0^{(j)}(t)\right)Y_j(s) + Y_0^{(j)}(t)\left(Y_j(s) - Y_j^{(j)}(s)\right) + Y_0^{(j)}(t)Y_j^{(j)}(s),$$

using stationarity and $D_2 < \infty$ we obtain

$$E\|S_n\|^2 = \sum_{k=1}^{n} E(Y_k, Y_k) + 2 \sum_{1 \leq k < l \leq n} E(Y_k, Y_l) \leq nC_2$$

for some $C_2 > D_2/\psi_2$ and all $n \in \mathbb{N}$, which finishes the proof for $\kappa = 2$. For more details cf. Berkes et al (2011). For the latter case, i.e. $2 < \kappa < 3$, the idea is to show, that for any $n_0 \in \mathbb{N}$ there is some $C_\kappa(n_0)$ such that:

$$E\|S_n\|^\kappa \leq C_\kappa n^{\kappa/2} \quad \forall n \leq n_0 \quad \Rightarrow \quad E\|S_n\|^\kappa \leq C_\kappa n^{\kappa/2} \quad \forall n \leq 2n_0.$$  \hspace{1cm} (6.6)

Hence, by induction, it is possible to conclude that $C_\kappa$ does not depend on $n_0$ which then completes the proof.

Now, (6.6) can be verified by selecting an arbitrary $n_0 \in \mathbb{N}$ and choosing $C_\kappa(n_0)$ large enough, such that on the one hand

$$E\|S_n\|^\kappa \leq C_\kappa(n_0)n^{\kappa/2}$$

for all $n \leq n_0$ and on the other hand $C_\kappa^{1/\kappa}(n_0) > D_\kappa/\psi_\kappa$. Using Jensen’s inequality we have

$$E\|Y_k - Y_k^{(n-k)}\| \leq (E\|Y_k - Y_k^{(n-k)}\|^\kappa)^{1/\kappa},$$

for all $n \leq n_0$ and on the other hand $C_\kappa^{1/\kappa}(n_0) > D_\kappa/\psi_\kappa$. Using Jensen’s inequality we have

$$E\|Y_k - Y_k^{(n-k)}\|^\kappa \leq (E\|Y_k - Y_k^{(n-k)}\|^\kappa)^{1/\kappa}. $$

23
which, via basic inequalities for norms, yields that

\[ E \| S_{2n} \|^\kappa \leq \left( 2D_\kappa + \left( E \left[ \| Z_n + W_n \|^\kappa \right] \right)^{1/\kappa} \right)^\kappa \]  

(6.8)

where

\[ Z_n = \sum_{k=1}^{n} Y_k^{(n-k)}, \quad W_n = \sum_{k=1}^{n} Y_{n+k}^{(k-1)} \]

(cf. Berkes et al (2011) disp. (36))). Next, observe that

\[ E \| Z_n \|^p \leq \left( (E \| S_n \|^p)^{1/p} + D_p \right)^p \]

for \( p = 2 \) or \( p = \kappa \), respectively, and that the same holds true if we replace \( W_n \) by \( Z_n \). Hence, in view of (6.7), we arrive at

\[ E \| Z_n \|^\kappa \leq \frac{n \kappa}{2} C_\kappa (1 + \psi_\kappa)^\kappa, \quad E \| W_n \|^\kappa \leq \frac{n \kappa}{2} C_\kappa (1 + \psi_\kappa)^\kappa, \]  

(6.9)

for \( n \leq n_0 \). Furthermore, due to (6.5) it holds also that

\[ E \| Z_n \|^2 \leq nC_2(1 + \psi_2)^2, \quad E \| W_n \|^2 \leq nC_2(1 + \psi_2)^2, \]  

(6.10)

for all \( n \in \mathbb{N} \) (cf. Berkes et al (2011)). Recall that \( C_2^{1/2} > D_2/\psi_2 \) and that no restriction on \( n \) is needed here due to (6.5). Observe that \( Z_n \) and \( W_n \) are mean zero and independent. Therefore, by Berkes et al (2013, Lemma 3.1) we have for \( 2 < \kappa < 3 \) that

\[ E \| Z_n + W_n \|^\kappa \leq E \| Z_n \|^\kappa + E \| W_n \|^\kappa + E \| Z_n \|^2 (E \| W_n \|^2)^{\kappa/2 - 1} + E \| W_n \|^2 (E \| Z_n \|^2)^{\kappa/2 - 1}. \]

(6.11)

Now, combining (6.9), (6.10) and (6.11) yields

\[ E \| Z_n + W_n \|^\kappa \leq 2(1 + \psi_\kappa)^\kappa C_\kappa n^{\kappa/2} + 2 \left( (1 + \psi_2)^2 C_2 n \right)^{\kappa/2} \]  

(6.12)

for all \( n \leq n_0 \), which is a simple but significant modification of Berkes et al (2011, disp. (37)). Consequently, from (6.8) and (6.12) we obtain

\[ E \| S_{2n} \|^\kappa \leq \left\{ 2D_\kappa + \left( E \| Z_n + W_n \|^\kappa \right)^{1/\kappa} \right\}^\kappa \leq \left\{ 2D_\kappa + \left( 2(1 + \psi_\kappa)^\kappa C_\kappa n^{\kappa/2} + 2 \left( (1 + \psi_2)^2 C_2 n \right)^{\kappa/2} \right)^{1/\kappa} \right\}^\kappa \leq C_\kappa(n\Psi)^{\kappa/2} \]

where

\[ \Psi(\psi_\kappa, \psi_2, C_\kappa, C_2) := 2\psi_\kappa + \left[ 2(1 + \psi_\kappa)^\kappa + 2(1 + \psi_2)^\kappa C_2^{\kappa/2} C_\kappa^{-1} \right]^{1/\kappa} \downarrow 2^{1/\kappa} < 2 \]

as \((\psi_2, \psi_\kappa, C_\kappa^{-1}) \to 0\) (cf. Berkes et al (2011) disp. (38))). Copying the final arguments of Berkes et al (2011 Proof of Proposition 4) completes the proof.  

\[ \square \]
Next, we take a closer look at the estimation of the eigenstructure of $C$.

**Proof of Theorem 4.1.** Due to the symmetry of $K(x)$ the estimator $\hat{\zeta}$ can be rewritten as

$$\hat{\zeta} = \hat{\zeta}_0(t, s) + \sum_{i=1}^{n} K(i/h_n)\hat{\zeta}_i(t, s) + \sum_{i=1}^{n} K(i/h_n)\hat{\zeta}_i(s, t).$$

(6.13)

Note that in view of Hörmann and Kokoszka (2010, Theorem 3.1) the covariance estimation is of order

$$\int \int (\hat{\zeta}_0(t, s) - E[Y_0(t)Y_0(s)])^2dt \, ds = O_P(n^{-1}).$$

(6.14)

(Their arguments carry over to our case of noncausality in a straightforward manner using modifications similar to (6.3).) It remains to investigate the long run part of the estimate, where due to symmetry it suffices to consider the second term of (6.13). We define a centered version of the second expression of (6.13)

$$\hat{c}_{1,n}(t, s) = \sum_{i=1}^{n} K(i/h_n)\hat{\gamma}_i(t, s)$$

with $\hat{\gamma}_i(t, s) = n^{-1} \sum_{j=1}^{n-i} Y_j(t)Y_{j+i}(s)$ and take into account that the difference between the original expression and its centered counterpart is of order

$$\int \int \left( \sum_{i=1}^{n} K(i/h_n)\hat{\zeta}_i(t, s) - \hat{c}_{1,n}(t, s) \right)^2 \, dt \, ds = O_P(h_n^2/n).$$

(6.15)

(cf. Horváth et al. (2012b, proof of Theorem 2), as before, with straightforward modifications in view of noncausality). The centered version can be decomposed as follows:

$$\hat{c}_{1,n} - c_1 = \left[ \hat{c}_{1,n} - \hat{c}_{1,n}^{(m_n)} \right] + \left[ \hat{c}_{1,n}^{(m_n)} - E\hat{c}_{1,n}^{(m_n)} \right] + \left[ E\hat{c}_{1,n}^{(m_n)} - c_{1,n}^{(m_n)} \right] + \left[ c_{1,n}^{(m_n)} - c_1 \right],$$

where

$$c_1(t, s) = \sum_{i=1}^{\infty} E[Y_0(t)Y_i(s)],$$

$$c_{1,n}^{(m_n)}(t, s) = \sum_{i=1}^{m_n} E[Y_0^{(m_n)}(t)Y_i^{(m_n)}(s)],$$

$$\hat{c}_{1,n}^{(m_n)}(t, s) = \sum_{i=1}^{n} K(i/h_n)\hat{\gamma}_i^{(m_n)}(t, s),$$

$$\hat{\gamma}_i^{(m_n)}(t, s) = n^{-1} \sum_{j=1}^{n-i} Y_j^{(m_n)}(t)Y_{j+i}^{(m_n)}(s)$$

and $(m_n)$ indicates the $m_n$-dependent versions. The sequence $m_n$ needs to fulfill $m_n = o(h_n)$ and $m_n \to \infty$, as $n \to \infty$. The main extension of the proof of Horváth et al. (2012b) is the introduction of the additional term $E[c_{1,n}^{(m_n)}(t, s)]$ and that we allow for an increase in the dependency of $\hat{c}_{1,n}^{(m_n)}(t, s)$ with increasing sample size.
Due to stationarity, the values $\hat{\gamma}_i^{(m_n)}(t, s)$ depend only on $i, j$ and on the difference of $k - r$. Hence, we have

$$\sum_{i,j=1}^{c_n} \sum_{k,r=1}^{n} \delta_{k,r}^{i,j} = \sum_{i,j=1}^{c_n} \sum_{k,r=1}^{n} \left\{ \sum_{q=0}^{n} \delta_{z+q}^{i,j} + \sum_{q=0}^{n} \delta_{z+q}^{i,j} \right\}$$

$$= \sum_{i,j=1}^{c_n} \sum_{k,r=1}^{n} \left\{ \sum_{q=0}^{n} \delta_{0,q}^{i,j} + \sum_{q=0}^{n} \delta_{q,0}^{i,j} \right\}$$

$$\leq n \sum_{i,j=1}^{c_n} (m_n + i + j) \leq 3cnh_n^2 m_n$$

for some $c > 0$. From above considerations we obtain

$$E \int \int \left( \hat{\gamma}_{1,n}^{(m_n)}(t, s) - E\hat{\gamma}_{1,n}^{(m_n)}(t, s) \right)^2 dt ds$$

$$= \int \int \text{Var}[\hat{\gamma}_{1,n}^{(m_n)}(t, s)] dt ds = O(h_n^2 m_n/n).$$

(6.17)

Decomposing

$$Y_0^{(m)}(t) Y_i^{(m)}(s) - Y_0(t) Y_i(s)$$

$$= Y_0^{(m)}(t) \left( Y_i^{(m)}(s) - Y_i(s) \right) + \left( Y_0^{(m)}(t) - Y_0(t) \right) Y_i(s).$$

(6.18)
we obtain by stationarity and Cauchy-Schwarz’s inequality that
\[
\left( \iint \{E[Y_0^{(m_n)}(t)Y_i^{(m_n)}(s)]\}^2 dt ds \right)^{1/2} 
\leq 2 \left( E\|Y_0\|^2 \right)^{1/2} \left( E\|Y_0^{(m_n)} - Y_0\|^2 \right)^{1/2} + \left( \iint \{E[Y_0(t)Y_i(s)]\}^2 dt ds \right)^{1/2}.
\]
Since
\[
\sum_{i=1}^{\infty} \left\{ \iint \{E[Y_0(t)Y_i(s)]\}^2 dt ds \right\}^{1/2} < \infty
\]
we have, as \( n \to \infty \), that
\[
\sum_{i=1}^{m_n} \left( \iint \{E[Y_0^{(m_n)}(t)Y_i^{(m_n)}(s)]\}^2 dt ds \right)^{1/2} = O\left( m_n/\exp(cm_n) + 1 \right) = O(1)
\]
for some \( c > 0 \). Next, using standard arguments and stationarity we see that
\[
\left( \iint \{E[c_1^{(m_n)}(t,s) - c_1^{(m_n)}(t,s)]\}^2 dt ds \right)^{1/2} 
= \left( \iint \left\{ \sum_{i=1}^{m_n} (K(i/h_n)(n-i)/n - 1) E[Y_0^{(m_n)}(t)Y_i^{(m_n)}(s)] \right\}^2 dt ds \right)^{1/2}
\leq \left[ m_n/n + \max_{i=1,\ldots,m_n} |K(i/h_n) - 1| \right] 
\times \sum_{i=1}^{m_n} \left( \iint \{E[Y_0^{(m_n)}(t)Y_i^{(m_n)}(s)]\}^2 dt ds \right)^{1/2}
= O\left( (m_n/n) + (m_n/h_n)^p \right).
\]
Now, by Horváth et al (2012b, proof of Theorem 2) and the exponential decay of \( \delta(m) \) we observe that
\[
E\|\hat{c}_{1,n} - \hat{c}_{1,n}\| = O\left( m_n \left\{ E\|Y_0 - Y_0^{(m_n)}\|^2 \right\}^{1/2} 
+ \sum_{i=m_n+1}^{\infty} \left\{ E\|Y_0 - Y_0^{(i)}\|^2 \right\}^{1/2} \right) = O\left( m_n/\exp(cm_n) \right)
\]
for some \( c > 0 \). Using decomposition (6.18), stationarity and again the exponential
decay of $\delta(m)$ we get

$$
\|c_1^{(m)} - c_1\| \leq \left( \int \int \left( \sum_{i=1}^{m_n} E[Y_0^{(m_n)}(t)Y_i^{(m_n)}(s) - Y_0(t)Y_i(s)]^2 \, dt \, ds \right)^{1/2}
\right)
\left( \int \int \left( \sum_{i=m_n+1}^{\infty} E[Y_0(t)Y_i(s)]^2 \, dt \, ds \right)^{1/2}
\right)
= \mathcal{O}\left( m_n \left\{ E\|Y_0 - Y_0^{(m_n)}\|^2 \right\}^{1/2} \right)
\left( \int \int \left( \sum_{i=m_n+1}^{\infty} E\left[ (Y_0(t) - Y_0^{(i)}(t))Y_i(s) + Y_0^{(i)}(t)(Y_i(s) - Y_i^{(i)}(s)) \right]^2 \, dt \, ds \right)^{1/2}
\right)
\right)
= \mathcal{O}(m_n/\exp(cm_n))
\right)
\right)
(6.21)
$$

for some $c > 0$. Combining (6.13) - (6.21), we get

$$
\|\hat{\zeta} - \zeta\| = \mathcal{O}_P\left( m_n^{1/2} (h_n/n^{1/2}) + m_n/\exp(cm_n) + (m_n/n) + (m_n/h_n)\rho \right).
$$

Setting $m_n := \lfloor (\log n)/c \rfloor$ we obtain the desired rate (4.2). Note that the rate further simplifies to $\mathcal{O}_P((\log n)^{1/2}(h_n/n^{1/2}))$ if $\rho > \gamma/2$.

References

Aston J, Kirch C (2012) Detecting and estimating changes in dependent functional data. J Multivar Anal 109:204–220

Aue A, Hörmann S, Horváth L, Reimherr M (2009) Break detection in the covariance structure of multivariate time series models. Ann Stat 37(6B):4046–4087

Aue A, Hörmann S, Horváth L, Hušková M (2014) Dependent functional linear models with applications to monitoring structural change. Stat Sin doi:10.5705/ss.2012.233 (to appear). Supplement available on http://homepages.ulb.ac.be/∼shormann/ahhh_supp.pdf

Aue A, Horváth L (2013) Structural breaks in time series. J Time Ser Anal 34(1):1–16

Berkes I, Gabrys R, Horváth L, Kokoszka P (2009) Detecting changes in the mean of functional observations. J R Stat Soc Ser B Stat Methodol 71(5):927–946

Berkes I, Horváth L, Rice G (2013) Weak invariance principles for sums of dependent random functions. Stoch Proc Appl 123(2):385–403

Berkes I, Hörmann S, Schauer J (2011) Split invariance principles for stationary processes. Ann Prob 39(6):2441–2473

Csörgő M, Horváth L (1997) Limit Theorems in Change-point Analysis. Wiley, Chichester
Chochola O, Hušková M, Prašková Z, Steinebach, J (2013) Robust monitoring of CAPM portfolio betas. J Multivar Anal 115(2013):374-395.

Davidson J (1994) Stochastic Limit Theory: An Introduction for Econometricians. Oxford university press, New York.

Gabrys R, Kokoszka P (2007) Portmanteau test of independence for functional observations. JASA 102(480):1338–1348

Gombay E, Horváth L (1996) On the rate of approximations for maximum likelihood tests in change-point models. J Multivar Anal 56(1):120–152

Hörmann S, and Kidziński L (2014) A note on estimation in Hilbertian linear models. Scand J Stat doi:10.1111/sjos.12094

Hörmann S, Kokoszka P (2010) Weakly dependent functional data. Ann Stat 38(3):1845–1884

Horváth L (1993), The maximum likelihood method for testing changes in the parameters of normal observations. Ann. Stat. 21(2):671–680

Horváth L, Kokoszka P (2012) Inference for Functional Data with Applications. Springer, New York

Horváth L, Kokoszka P, Steinebach J (1999) Testing for changes in multivariate dependent observations with an application to temperature changes. J Multivar Anal 68(1):96–119

Horváth L, Kokoszka P, Reeder R (2012a) Estimation of the mean of functional time series and a two-sample problem. arXiv:1105.0019v1

Horváth L, Kokoszka P, Reeder R (2012b) Estimation of the mean of functional time series and a two-sample problem. J R Stat Soc Ser B Stat Methodol 75:103–122

Horváth L, Kokoszka P, Rice G (2013) Testing stationarity of functional time series. J Econometr 179:66–82

Horváth L, Rice G (2014a) Extensions of some classical methods in change point analysis. TEST 23(2):219–255

Horváth L, Rice G (2014b) Testing equality of means when the observations are from functional time series. Preprint

Jirak M (2012) Change-point analysis in increasing dimension. J Multivar Anal 111:136–159

Jirak M (2013) On weak invariance principles for sums of dependent random functionals. Stat Prob Lett 83(10):2291–2296

Kamgaing J T, Kirch C (2014) Detection of change points in discrete valued time series. In: Handbook of Discrete Valued Time series. Chapman Hall/CRC (to appear). Preprint available on http://www.math.kit.edu/stoch/~ckirch/page/publications/

Kokoszka P (2012). Dependent functional data. ISRN Probability and Statistics 2012:1–30
Kounias E G, Weng T (1969) An inequality and almost sure convergence. Ann Math Stat 40(3):1091–1093

Moricz F (1976) Moment inequalities and the strong laws of large numbers. Probab Theory Related Fields 35(4):299–314

Ramsay J, Silverman B (2005) Functional Data Analysis. Springer, New York

Schmitz A (2011). Limit theorems in change-point analysis for dependent data. PhD Thesis, University of Cologne

Tómacs T, Libor Z (2006) A Hájek-Rényi type inequality and its applications. Ann Math Inform 33:141–149

Torgovitski L (2014) A Darling-Erdős-type CUSUM-procedure for functional data. Metrika doi 10.1007/s00184-014-0487-7 (to appear)

Vostrikova L (1981) Detection of a “disorder” in a Wiener process. Theory Probab Appl 26:356–362

Zhou J (2011) Maximum likelihood ratio test for the stability of sequence of Gaussian random processes. Comput Stat Data Anal 55(6):2114–2127