# SIEVE BOOTSTRAP FOR FUNCTIONAL TIME SERIES

By Efstathios Paparoditis[*]

*University of Cyprus*

A bootstrap procedure for functional time series is proposed which exploits a general vector autoregressive representation of the time series of Fourier coefficients appearing in the Karhunen-Loève expansion of the functional process. A double sieve-type bootstrap method is developed which avoids the estimation of process operators and generates functional pseudo-time series that appropriately mimic the dependence structure of the functional time series at hand. The method uses a finite set of functional principal components to capture the essential driving parts of the infinite dimensional process and a finite order vector autoregressive process to imitate the temporal dependence structure of the corresponding vector time series of Fourier coefficients. By allowing the number of functional principal components as well as the autoregressive order used to increase to infinity (at some appropriate rate) as the sample size increases, consistency of the functional sieve bootstrap can be established. We demonstrate this by proving a basic bootstrap central limit theorem for functional finite Fourier transforms and by establishing bootstrap validity in the context of a fully functional testing problem. A novel procedure to select the number of functional principal components is introduced while simulations illustrate the good finite sample performance of the new bootstrap method proposed.

**1. Introduction.** Statistical inference for time series stemming from stationary functional processes has attracted considerable interest during the last decades and progress has been made in several directions. Estimation and testing procedures have been developed for a wide range of inference problems and for large classes of stationary functional processes; see Bosq (2000), Hörmann and Kokoszka (2012) and Horváth and Kokoszka (2012). However, the asymptotic results derived, typically depend in a complicated way on difficult to estimate, infinite dimensional characteristics of the underlying functional process. This restricts considerably the implementability of asymptotic approximations when used in practice to judge the uncertainty of estimation procedures or to calculate critical values of tests. In such situations, bootstrap methods can provide useful alternatives.

Bootstrap procedures for Hilbert space-valued time series proposed so far in the literature, are mainly attempts to adapt, to the infinite dimensional functional set-up, of bootstrap methods that have been developed for the finite dimensional (i.e., mostly univariate) time series case; cf. Lahiri (2003). Politis and Romano (1994) considered applications of the stationary bootstrap to functional, Hilbert-valued time series and showed its validity for the sample mean for functional processes satisfying certain mixing and boundeness conditions. Dehling et al. (2015) considered applications of the non-overlapping block bootstrap to U-statistics for so called near epoch dependent functional processes and Sharipov et al. (2016) to change point analysis. Franke and Nyarigue (2016) and Zhou and Politis (2016) developed some theory for different residual-based bootstrap procedures applied to a first order functional autoregressive process. Notice that the transmission of other bootstrap methods for real-valued time series to the functional set-up, like for instance of the autoregressive-sieve bootstrap, Kreiss (1988) and Kreiss et al. (2011), seems to be difficult mainly due to problems associated with the estimation (of an with sample size increasing number) of infinite dimensional autoregressive operators.

Applications of bootstrap procedures to certain inference problems in functional time series analysis have been also considered in the literature. For instance, for the construction of prediction intervals, Fernández De Castro et al. (2005) used an approach based on resampling pairs of functional observations by means of kernel-driven resampling probabilities. The same authors also apply a parametric, residual-based bootstrap approach using an estimated first order functional autoregression with i.i.d. resampling of appropriately defined functional residuals. For the same prediction problem, Hyndman and Shang (2009) applied different bootstrap approaches including bootstrapping the functional curves by randomly disturbing the forecasted scores using residuals obtained from univariate autoregressive fits. Aneiros-Perez et al. (2011) considered the nonparametric functional autoregressive models, while Mingotti et al. (2015) the case of the integrated functional autoregressive model. Apart from the lack of theoretical justification, the aforementioned bootstrap applications do not provide a general bootstrap methodology for functional time series as they are designed for and their applicability is restricted to the particular inference problem considered; see also McMurry and Politis (2011) and Shang (2016) for an overview.

In this paper a general and easy to implement bootstrap procedure for functional time series is proposed which generates bootstrap replicates $X_1^*$, $X_2^*, \ldots, X_n^*$ of a functional time series $X_1, X_2, \ldots, X_n$ and is applicable to a large class of stationary functional processes. The procedure avoids the ex-

plicit estimation of process operators and exploits some basic properties of the stochastic process of Fourier coefficients (scores) appearing in the well-known Karhunen-Loève expansion of the functional random variables. It is in particular shown, that under quite general assumptions, the stochastic process of Fourier coefficients obeys a so-called vector autoregressive representation and this representation plays a key role in developing a bootstrap procedure for the functional time series at hand. More specifically, to capture the essential driving functional parts of the underlying infinite dimensional process, the first $m$ functional principal components are used and the corresponding $m$-dimensional time series of Fourier coefficients is bootstrapped using a $p$th order vector autoregression fitted to the vector time series of sample Fourier coefficients. In this way, a $m$-dimensional pseudo-time series of Fourier coefficients is generated which imitates the temporal dependence structure of the vector time series of sample Fourier coefficients. Using the (truncated) Karhunen-Loève expansion, these pseudo-Fourier coefficients are then transformed to functional bootstrap replicates of the main driving, principal components, of the observed functional time series. Adding to these replicates an appropriately resampled functional noise, leads finally to the bootstrapped functional pseudo-time series $X_1^*, X_2^*, \ldots, X_n^*$.

In a certain sense, our bootstrap procedure works by using a finite rank (i.e., $m$-dimensional) approximation of the infinite dimensional structure of the underlying functional process and a $p$th order vector autoregressive approximation of its infinite order temporal dependence structure. To achieve consistency and to capture appropriately the entire infinite dimensional structure of the functional process, the number $m$ of functional principal components used as well as the order $p$ of the vector autoregression applied, are allowed to increase to infinity (at some appropriate rate) as the sample size $n$ increases to infinity. This double sieve property justifies the use of the term "sieve bootstrap" for the bootstrap procedure proposed.

We show that under quite general conditions, this bootstrap procedure succeeds in imitating correctly the entire infinite dimensional autocovariance structure of the underlying functional process. Notice that apart from the problem that instead of the unknown true scores, the time series of estimated scores is used, the asymptotic analysis of our bootstrap procedure faces additional challenges which are caused by the fact that vector autoregressions of increasing order and of increasing dimension are considered and that the lower bound of the corresponding spectral density matrix approaches zero as the dimension of the vector time series of scores used, increases to infinity. We demonstrate how the new bootstrap procedure proposed can be successfully applied to different inference problems in functional time series analysis. In

particular, we apply the proposed sieve bootstrap procedure to the problem of estimating the distribution of the functional Fourier transform which is fundamental in a multitude of applications and has attracted interest in the functional time series literature; see Cerovecki and Hörmann (2015) for some recent developments. In this context, a basic bootstrap central limit theorem is established which shows validity of the functional sieve bootstrap for this important class of statistics. Furthermore, we consider applications of the functional sieve bootstrap in the context of fully functional testing and to the two sample mean problem and show how this bootstrap procedure can be applied to consistently estimate the complicated distribution of the test statistic of interest under the null.

Using the time series of Fourier coefficients in the context of functional time series analysis has been considered by many authors in a variety of applications. Among others we mention Hyndman and Shang (2009) who, for functional autoregressive models and for the sake of prediction, used univariate autoregressions fitted to the scalar time series of scores. In the same context and more related to the approach proposed in this paper, a multivariate approach of prediction has been proposed by Aue et al. (2014) which works by fitting a vector autoregressive model to the multivariate time series of scores.

The paper is organized as follows. Section 2 derives some basic properties and discuss the autoregressive representations of the vector process of Fourier coefficients appearing in the Karhunen-Loève expansion of the functional process. Apart from being useful for bootstrap purposes, these properties are of interest on their own. The functional sieve bootstrap procedure proposed is described in Section 3 where some properties of the bootstrap functional pseudo-time series are also discussed. Asymptotic validity of the new bootstrap procedure applied to finite Fourier transforms and to fully functional testing is established in Section 4. Section 5 proposes some novel practical, data driven rules to choose the bootstrap parameters and presents some numerical simulations which investigate the finite sample performance of the functional sieve bootstrap. Comparisons with three variants of block bootstrap methods are also given. Technical proofs and auxiliary lemmas are deferred to Section 6.

## 2. The Process of Fourier Coefficients.

2.1. *The functional set-up.* We consider a (functional) stochastic process $\mathbf{X} = \{X_t, t \in \mathbb{Z}\}$ where for each $t$ (interpreted as time), $X_t$ is a random element of the separable Hilbert space $\mathcal{H} := L^2([0,1], \mathbb{R})$ with parametrization $\tau \to X_t(\tau) \in \mathbb{R}$ for $\tau \in [0,1]$. As usual we denote by $\langle \cdot, \cdot \rangle$ the in-

ner product in $\mathcal{H}$ and by $\|\cdot\|$ the induced norm defined for $x, y \in \mathcal{H}$ as $\langle x, y \rangle = \left( \int_{[0,1]} x(t)y(t)dt \right)^{1/2}$ and $\|x\| = \left( \langle x, x \rangle \right)^{1/2}$ respectively. Furthermore, for matrices $A$ and $B$ we denote by $\|A\|_F$ the Frobenius norm, we write $A \geq B$ or $B \leq A$ if $A - B$ is non-negative hermitian while for an operator $T$, $\|T\|$ denotes its operator norm and $\|T\|_{HS}$ its Hilbert-Schmidt norm, if $T$ is a Hilbert-Schmidt operator.

For the underlying functional process $\mathbf{X}$ it is assumed that its dependence structure satisfies the following assumption.

**Assumption 1** $\mathbf{X}$ is a purely non deterministic, $L^4$-$\mathcal{M}$ approximable process.

The general notion of $L^p - \mathcal{M}$ approximability refers to stochastic process $\mathbf{X} = \{X_t, t \in \mathbb{Z}\}$ with $X_t$ taking values in $\mathcal{H}$, $E\|X_t\|^p < \infty$, and where the random element $X_t$ admits the representation $X_t = f(\varepsilon_t, \varepsilon_{t-1}, \ldots)$. Here the $\varepsilon_t$'s are i.i.d. random elements in $\mathcal{H}$ and $f$ some measurable function $f : \mathcal{H}^\infty \to \mathcal{H}$. If for $\{\widetilde{\varepsilon}_t, t \in \mathbb{Z}\}$ an independent copy of $\{\varepsilon_t, t \in \mathbb{Z}\}$ and $X_t^{(\mathcal{M})} = f(\varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_{t-\mathcal{M}+1}, \widetilde{\varepsilon}_{t-\mathcal{M}}, \widetilde{\varepsilon}_{t-\mathcal{M}-1}, \ldots)$, the condition

$$\sum_{k=1}^\infty \left( E\|X_k - X_k^{(k)}\|^p \right)^{1/p} < \infty,$$

is satisfied, then $\mathbf{X}$ is called $L^p - \mathcal{M}$ approximable. $L^p - \mathcal{M}$ approximability is a notion of weak dependence which applies to many commonly used functional time series models, like linear functional processes, functional ARCH processes, etc.; see Hörmann and Kokoszka (2010) for more details.

Let $\mu := EX_0 \in \mathcal{H}$ be the mean of $\mathbf{X}$ which by stationary is independent of $t$ and for which we assume $\mu = 0$ for simplicity. We denote by $C_h$ the autocovariance operator $C_h : \mathcal{H} \to \mathcal{H}$ at lag $h \in \mathbb{Z}$ defined by $C_h(\cdot) = E\langle X_t - \mu, \cdot \rangle (X_{t+h} - \mu)$. Associated with the autocovariance operator is the autocovariance function $c_h : [0,1] \times [0,1] \to \mathbb{R}$ with $c_h(\tau, \nu) = E(X_t(\tau) - \mu(\tau))(X_{t+h}(\nu) - \mu(\nu))$, $\tau, \nu \in [0,1]$, that is, $C_h$ is an integral operator with kernel function $c_h$.

Assumption 1 implies that $\sum_h \|C_h\|_{HS} < \infty$ and that for every $\omega \in \mathbb{R}$ the spectral density operator

$$\mathcal{F}_\omega(x) = (2\pi)^{-1} \sum_{h \in \mathbb{Z}} C_h(x) e^{-ih\omega}, \quad x \in \mathcal{H}$$

is well defined, continuous in $\omega$, selfadjoint and trace class, Hörmann et al. (2015); see also Panaretos and Tavakoli (2013) for similar properties under different weak dependence conditions. In what follows we will strengthen

somehow the assumption on the norm summability of the autocovariance operator to the following requirement.

**Assumption 2** $\sum_h (1 + |h|)^r \|C_h\|_{HS} < \infty$ for some $r \geq 0$.

Furthermore, we will assume that the spectral density operator $\mathcal{F}_\omega$ satisfies the following condition.

**Assumption 3** For all $\omega \in [0, \pi]$, the operator $\mathcal{F}_\omega$ is of full rank, i.e., $\ker(\mathcal{F}_\omega) = 0$.

For real-valued univariate processes, $\ker(\mathcal{F}_\omega) = 0$ is equivalent to the condition that the spectral density is everywhere in $[0, \pi]$ strictly positive while for multivariate process to the non-singularity of the spectral density matrix for every frequency $\omega \in [0, \pi]$. Notice that all eigenvalues $\nu_j(\omega)$, $j = 1, 2, \ldots$ of $\mathcal{F}_\omega$ are positive and that $\sum_{j=1}^\infty \nu_j(\omega) < \infty$ by the trace class property of $\mathcal{F}_\omega$.

2.2. *Vector autoregressive representation.* Since $C_0 = \int_{-\pi}^{\pi} \mathcal{F}_\omega d\omega$, the positivity of $\mathcal{F}_\omega$ implies that the covariance operator $C_0$ has full rank, that is, its eigenvalues $\lambda_j$ satisfy $\lambda_j > 0$ for all $j \geq 1$. By the symmetry and compacteness of $C_0$, the random element $X_t$ admits the well known Karhunen-Loève representation

$$(2.1) \qquad X_t = \sum_{j=1}^\infty \langle X_t, v_j \rangle v_j, \quad t \in \mathbb{Z},$$

where $v_j$, $j = 1, 2, \ldots$, are the orthonormalized eigenfunctions that correspond to the eigenvalues $\lambda_j$, $j = 1, 2, \ldots$, of $C_0$. For $t \in \mathbb{Z}$, let $\xi_{j,t} := \langle X_t, v_j \rangle$, $j \geq 1$, and consider any subset of indices $M = \{j_1, j_2, \ldots, j_m\} \subset \mathbb{N}$ with $j_1 < j_2 < \ldots < j_m$, $m < \infty$. Later on, we will concentrate on the specific set $M = \{1, 2, \ldots, m\}$ which will be the set of the $m$ largest eigenvalues of the covariance operator $C_0$.

Consider now the $m$-dimensional process $\xi^{(M)} = \{\xi_t^{(M)} = (\xi_{j_s,t}^{(M)}, s = 1, 2, \ldots, m)^\top, t \in \mathbb{Z}\}$. Observe that $\xi^{(M)}$ is strictly stationary, purely non deterministic and has mean zero, i.e., $E(\xi_t^{(M)}) = (\langle EX_t, v_{j_s} \rangle, s \in M) = 0$. Furthermore, its autocovariance matrix function $\Gamma_{\xi^{(M)}}(h) = E(\xi_t^{(M)} \xi_{t+h}^{(M)^T})$, $h \in \mathbb{Z}$, is given by $\Gamma_{\xi^{(M)}}(h) = \left( \langle C_h(v_{j_s}), v_{l_r} \rangle \right)_{s,r=1,2,\ldots,m}$ and satisfies by Assumption 2,

$$\sum_{h=-\infty}^\infty (1 + |h|)^r \|\Gamma_{\xi^{(M)}}(h)\|_F = \sum_{h=-\infty}^\infty (1 + |h|)^r \left( \sum_{s,r=1}^m \langle C_h(v_{j_s}), v_{l_r} \rangle^2 \right)^{1/2}$$

$$(2.2) \qquad\qquad\qquad \leq \sum_{h=-\infty}^\infty (1 + |h|)^r \|C_h\|_{HS} < \infty.$$

Note that the bound on the right hand side above is independent of the set $M$ and that although by construction it holds true that $Cov(\xi_{r_1,t}^{(M)}, \xi_{r_2,t}^{(M)}) = 0$ for $r_1 \neq r_2$, the random variables $\xi_{r_1,t}$ and $\xi_{r_2,s}$ may be correlated for $t \neq s$. The summability property (2.2) implies that the $m$-dimensional vector process $\xi^{(M)}$ possesses a continuous spectral density matrix $f_{\xi^{(M)}}(\cdot)$ which is given by

$$f_{\xi^{(M)}}(\omega) = (2\pi)^{-1} \sum_{h=-\infty}^{\infty} \Gamma_{\xi^{(M)}}(h) e^{-i\omega h}, \quad \omega \in \mathbb{R}.$$

Moreover, $f_{\xi^{(M)}}$ satisfies the following boundeness conditions.

LEMMA 2.1.   *Under Assumption 1 and 3 and Assumption 2 with $r = 0$, the spectral density $f_{\xi^{(M)}}$ satisfies*

$$(2.3) \qquad \delta_M I_m \leq f_{\xi^{(M)}}(\omega) \leq c\, I_m, \quad \text{for all } \omega \in [0, \pi],$$

*where $\delta_M$ and $c$ are real numbers ($\delta_M$ depends on the set $M$), such that $0 < \delta_M \leq c < \infty$ and $I_m$ is the $m \times m$ unity matrix.*

The continuity and the boundeness properties of the spectral density matrix $f_{\xi^{(M)}}(\cdot)$ stated in Lemma 2.1, imply that the process $\xi^{(M)}$ obeys a so called vector autoregressive representation; Cheng and Pourahmadi (1983), see also Wiener and Masani (1958). That is, there exist an infinite sequence of $m \times m$-matrices $\{A_j^{(M)}, j \in \mathbb{N}\}$ and a full rank $m$-dimensional white noise process $\{e_t^{(M)}, t \in \mathbb{Z}\}$, such that $\xi_t^{(M)}$ can be expressed as

$$(2.4) \qquad \xi_t^{(M)} = \sum_{j=1}^{\infty} A_j^{(M)} \xi_{t-j}^{(M)} + e_t^{(M)}, \ t \in \mathbb{Z},$$

where the coefficients matrices satisfy $\sum_{j \in \mathbb{N}} (1 + j)^r \|A_j^{(M)}\|_F < \infty$ and $\{e_t^{(M)}, t \in \mathbb{Z}\}$ is a zero mean white noise innovation process, that is $E(e_t^{(M)}) = 0$ and $E(e_t^{(M)} e_s^{(M)\top}) = \delta_{t,s} \Sigma_e^{(M)}$, with $\delta_{t,s} = 1$ if $t = s$, $\delta_{t,s} = 0$ otherwise and $\Sigma_e^{(M)}$ a full rank $m \times m$ covariance matrix. We stress here the fact that (2.4) does not describe a model for the process of Fourier coefficients $\xi_t^{(M)}$ and should not be confused with the so-called linear, infinite order vector autoregressive (VAR($\infty$)) process driven by independent, identically distributed (i.i.d.) innovations. In fact, representation (2.4) is the autoregressive analogue of the well-known (moving average) Wold representation of $\xi_t^{(M)}$ with respect to the same white noise innovation process $\{e_t^{(M)}, t \in \mathbb{Z}\}$. This autoregressive representation is valid for any stationary and purely non

deterministic process the spectral density matrix of which is continuous and satisfies the boundness conditions (2.3); see also Cheng and Pourahmadi (1983) and Pourahmadi (2001) for details. In contrast to the Wold representation, the autoregressive representation (2.4) seems to be more appealing for statistical purposes, since it express the vector time series of Fourier coefficients $\xi_t^{(M)}$ as a function of its (in principle) observable past values $\xi_{t-j}^{(M)}$, $j = 1, 2, \ldots$.

In what follows we assume that the eigenvalues are in descending order, i.e., that $\lambda_1 > \lambda_2 > \cdots > \lambda_m > 0$ and we consider the set $M = \{1, 2, \ldots, m\}$ of the $m$ largest eigenvalues of $C_0$. The corresponding normalized eigenfunctions (principal components) are denoted by $v_j$, $j = 1, 2, \ldots, m$ and are (up to a sign) uniquely identified. Furthermore, by Parseval's identity, the quantity $\sum_{j=1}^{m} \lambda_j$ describes the variance of $X_t$ captured by the first $m$ functional principal components. To simplify notation we surpass in the following the upper index $(M)$ and write simple $\xi_t$ for $\xi_t^{(M)}$ respectively $f_\xi$ for $f_{\xi(M)}$, keeping in mind that the $j$th component $\xi_{j,t} = \langle X_t, v_j \rangle$ of $\xi_t = (\xi_{1.t}, \xi_{2,t}, \ldots, \xi_{m,t})^\top$ is obtained using the orthonormalized eigenfunction $v_j$ which corresponds to the $j$th largest eigenvalue $\lambda_j$ of $C_0$, $j = 1, 2, \ldots, m$. Furthermore, we write $A_j(m)$, $e_t(m)$, $\delta_m$ and $\Sigma_e(m)$ for $A_j^{(M)}$, $e_t^{(M)}$, $\delta_M$ and $\Sigma_e^{(M)}$, respectively.

### 3. The Functional Sieve Bootstrap Procedure.

3.1. *The bootstrap procedure.* The basic idea of our procedure is to generate pseudo-replicates $X_1^*, X_2^*, \ldots, X_n^*$ of the functional time series at hand by first bootstrapping the $m$-dimensional time series of Fourier coefficients $\xi_t = (\xi_{1,t}, \xi_{2,t}, \ldots, \xi_{m,t})^\top$, $t = 1, 2, \ldots, n$, corresponding to the first $m$ principal components. This $m$-dimensional time series of Fourier coefficients is bootstrapped using the autoregressive representation of $\xi_t$ discussed in Section 2.2. The generated $m$-dimensional pseudo-time series of Fourier coefficients is then transformed to functional principal pseudo-components by means of the truncated Karhunen-Loève expansion $\sum_{j=1}^{m} \xi_{j,t} v_j$. Adding to this an appropriately resampled functional noise leads to the functional pseudo-time series $X_1^*, X_2^*, \ldots, X_n^*$. However, since the $\xi_t$'s are not observed, we work with the time series of estimates scores. This idea is precisely described in the following functional sieve bootstrap algorithm.

*Step 1*: Select a number $m = m(n)$ of functional principal components and an autoregressive order $p = p(n)$, both finite and depending on $n$.

*Step 2*: Let

$$\widehat{\xi}_t = (\widehat{\xi}_{j,t} = \langle X_t, \widehat{v}_j \rangle, j = 1, 2, \ldots, m)^\top, \quad t = 1, 2, \ldots, n,$$

be the $m$-dimensional time series of estimated Fourier coefficients, where $\widehat{v}_j$, $j = 1, 2, \ldots, m$ are the estimated eigenfunctions corresponding to the estimated eigenvalues $\widehat{\lambda}_1 > \widehat{\lambda}_2 > \cdots > \widehat{\lambda}_m$ of the sample covariance operator $\widehat{C}_0 = n^{-1} \sum_{t=1}^{n} (X_t - \overline{X}_n) \otimes (X_t - \overline{X}_n)$, $\overline{X}_n = n^{-1} \sum_{t=1}^{n} X_t$.

*Step 3*: Let $\widehat{X}_{t,m} = \sum_{j=1}^{m} \widehat{\xi}_{j,t} \widehat{v}_j$ and define the functional residuals $\widehat{U}_{t,m} = X_t - \widehat{X}_{t,m}$, $t = 1, 2, \ldots, n$.

*Step 4*: Fit a $p$th order vector autoregressive process to the $m$-dimensional time series $\widehat{\xi}_t$, $t = 1, 2, \ldots, n$, denote by $\widehat{A}_{j,p}(m)$, $j = 1, 2, \ldots, p$, the estimates of the autoregressive matrices and by $\widehat{e}_{t,p}$ the residuals,

$$\widehat{e}_{t,p} = \widehat{\xi}_t - \sum_{j=1}^{p} \widehat{A}_{j,p}(m)\widehat{\xi}_{t-j}, \quad t = p+1, p+2, \ldots, n.$$

Different estimators $\widehat{A}_{j,p}(m)$, $j = 1, 2, \ldots, p$ can be used, but we focus in the following on Yule-Walker estimators; cf. Brockwell and Davis (1991).

*Step 5*: Generate a $m$-dimensional pseudo time series of scores $\xi_t^* = (\xi_{1,t}^*, \xi_{2,t}^*, \ldots, \xi_{m,t}^*)$, $t = 1, 2, \ldots, n$, using

$$\xi_t^* = \sum_{j=1}^{p} \widehat{A}_{j,p}(m)\xi_{t-j}^* + e_t^*,$$

where $e_t^*$, $t = 1, 2, \ldots, n$ are i.i.d. random vectors having as distribution the empirical distribution of the centered residual vectors $\widetilde{e}_{t,p} = \widehat{e}_{t,p} - \overline{\widehat{e}}_n$, $t = p+1, p+2, \ldots, n$ and $\overline{\widehat{e}}_n = (n-p)^{-1} \sum_{t=p+1}^{n} \widehat{e}_{t,p}$.

*Step 6*: Generate a pseudo-functional time series $X_1^*, X_2^*, \ldots, X_n^*$, where

$$(3.1) \qquad X_t^* = \sum_{j=1}^{m} \xi_{j,t}^* \widehat{v}_j + U_t^*, \qquad t = 1, 2, \ldots, n,$$

and $U_1^*, U_2^*, \ldots, U_n^*$ are i.i.d. random functions obtained by choosing with replacement from the set of centered functional residuals $\widehat{U}_{t,m} - \overline{\widehat{U}}_n$, $t = 1, 2, \ldots, n$ and $\overline{\widehat{U}}_n = n^{-1} \sum_{t=1}^{n} \widehat{U}_{t,m}$.

Some comments regarding the above algorithm are in order. Notice first that $X_1^*, X_2^*, \ldots, X_n^*$ are functional pseudo-random variables and that the autoregressive representation of the vector time series of Fourier coefficients is solely used as a tool to bootstrap the $m$ main functional principal components of the functional time series at hand. In fact, it is this autoregressive representation which allows the generation of the pseudo-time series of Fourier coefficients $\xi_1^*, \xi_2^*, \ldots, \xi_n^*$ in Step 4 and Step 5 in a way that imitates the dependence structure of the sample Fourier coefficients $\xi_1, \xi_2, \ldots, \xi_n$. These pseudo-Fourier coefficients are transformed to bootstrapped main principal components by means of the truncated and estimated Karhunen-Loève expansion which together with the additive functional noise $U_t^*$, lead to the new functional pseudo-observations $X_1^*, X_2^*, \ldots, X_n^*$.

The estimated eigenfunctions $\widehat{v}_j$ used in Step 2 may point in an opposite direction than the eigenfunctions $v_j$. In asymptotic derivations this is commonly taking care off by considering the sign corrected estimator $\widehat{s}_j \widehat{v}_j$, where the (unobserved) random variable $\widehat{s}_j$ is given by $\widehat{s}_j = \text{sign}(\langle \widehat{v}_j, v_j \rangle)$. However, since in our setting adding this sign correction will not affect the asymptotic results derived, we assume for simplicity throughout this paper, that $\widehat{s}_j = 1$, for $j = 1, 2, \ldots, m$.

REMARK 3.1.    To simplify notation we have assumed that the mean of $\mathbf{X}$ is zero. If $EX_t = \mu \neq 0$ then the sieve bootstrap algorithm can be appropriately modified by defining the pseudo-random element $X_t^*$ in *Step 6* as $X_t^* = \overline{X}_n + \sum_{j=1}^m \xi_{j,t}^* \widehat{v}_j + U_t^*$, $t = 1, 2, \ldots, n$. Notice that since under Assumption 1, $\|\overline{X}_n - \mu\| = O_P(n^{-1/2})$, see Hörmann and Kokoszka (2012), the asymptotic results derived in this paper are not affected, i.e., $EX_t = 0$ is not a stringent assumption.

REMARK 3.2.    Modifications of the above basic bootstrap algorithm are possible which concern the resampling schemes used to generate the vector of pseudo-innovations $e_t^*$ and/or the bootstrap functional noise $U_t^*$. To elaborate, and as we will see in the sequel, for general stationary processes satisfying Assumption 1, the applied i.i.d. resampling used to generate the pseudo-innovations $e_t^*$ in *Step 5*, suffices in order to capture the entire, infinite dimensional second order structure of the underlying functional process $\mathbf{X}$. However, a modification of this i.i.d. resampling scheme may be needed if higher order dependence characteristics of the underlying functional process beyond those of order two, should also be correctly mimicked by the functional pseudo-time series $X_1^*, X_2^*, \ldots, X_n^*$. In such a case, the i.i.d. resampling used to generate the $e_t^*$'s in *Step 5* can be replaced by other resampling schemes (i.e., block bootstrap schemes) that are able to capture higher order

dependence characteristics of the white noise process $\{e_t, t \in \mathbb{Z}\}$ appearing in (2.4).

3.2. *Some properties of the bootstrap functional process.* As usual, all considerations made regarding the bootstrap procedure are made conditionally on the observed functional time series $X_1, X_2, \ldots, X_n$. The generation mechanism of the pseudo-time series $X_1^*, X_2^*, \ldots, X_n^*$, enables us to consider the bootstrap functional process $\mathbf{X}^* = \{X_t^*, t \in \mathbb{Z}\}$, where for $t \in \mathbb{Z}$, $X_t^* = \sum_{j=1}^{m} \mathbf{1}_j^\top \xi_t^* \widehat{v}_j + U_t^*$, with $\{\xi_t^* = (\xi_{1,t}^*, \ldots, \xi_{m,t}^*)^\top, t \in \mathbb{Z}\}$ generated as $\xi_t^* = \sum_{j=1}^{p} \widehat{A}_{j,p}(m)\xi_{t-j}^* + e_t^*$ and the $U_t^*$'s are i.i.d. functional random variable taking values in the set $\{\widehat{U}_{t,m} - \overline{\widehat{U}}_n, t = 1, 2, \ldots, n\}$ with probability $1/n$. In the above notation $\mathbf{1}_j$ is the $m$-dimensional vector $\mathbf{1}_j = (0, \ldots, 0, 1, 0, \ldots, 0)^\top$, where the unity appears in the $j$th position.

It is easy to see that $\mathbf{X}^*$ is a strictly stationary functional process with mean function $E^* X_t^* = 0$ and autocovariance operator $C_h^* : \mathcal{H} \to \mathcal{H}$ given, for $h \in \mathbb{Z}$, by

$$C_h^*(\cdot) = \sum_{j_1=1}^{m} \sum_{j_2=1}^{m} \mathbf{1}_{j_1}' \mathbf{\Gamma}_h^* \mathbf{1}_{j_2} \langle \widehat{v}_{j_1}, \cdot \rangle \widehat{v}_{j_2} + I(h=0) E^* \langle U_t^*, \cdot \rangle U_t^*,$$

where $\mathbf{\Gamma}_h^* = E^*(\xi_t^* \xi_{t+h}^{*T})$ is the $m \times m$ autocovariance matrix at lag $h$ of $\{\xi_t^*, t \in \mathbb{Z}\}$. $C_h^*$ is a Hilbert-Schmidt operator since it is, for $h \neq 0$, a finite rank operator while for $h = 0$ it is the sum of a finite rank operator and of the (Hilbert-Schmidt) empirical covariance operator of the functional pseudo-innovations $C_U^* = E^* \langle U_t^*, \cdot \rangle U_t^* = n^{-1} \sum_{t=1}^{n} \langle \widehat{U}_{t,m} - \overline{\widehat{U}}_n, \cdot \rangle (\widehat{U}_{t,m} - \overline{\widehat{U}}_n)$.

If the (estimated) vector autoregressive process used to generate the time series of pseudo-scores $\xi_t^*$ is stable, then the dependence structure of the bootstrap process $\mathbf{X}^*$ can be precisely described. This is stated in the following proposition. Notice that the required stability condition of the estimated autoregressive polynomial is fulfilled, if for instance, $\widehat{A}_{j,p}$, $j = 1, 2, \ldots, p$, are the Yule-Walker estimators; cf. Brockwell and Davis (1991), Ch. 11.4.

PROPOSITION 3.1. *If $p, m \in \mathbb{N}$ is such that the estimator $\widehat{A}_{j,p}$, $j = 1, 2, \ldots, p$, used in Step 4 of the functional sieve bootstrap algorithm is well defined and satisfies $det(\widehat{A}_{p,m}(z)) \neq 0$ for all $|z| \leq 1$, where $\widehat{A}_{p,m}(z) = I_m - \sum_{j=1}^{p} \widehat{A}_{j,p}(m)z^j$, $z \in \mathbb{C}$, then, conditionally on $X_1, X_2, \ldots, X_n$, the bootstrap process $\mathbf{X}^*$ is $L^2 - \mathcal{M}$ approximable.*

The $L^2 - \mathcal{M}$ approximability of $\mathbf{X}^*$ implies that $\sum_h \|C_h^*\|_{HS} < \infty$, see

Hörmann et al. (2015), which can be also easily verified since

$$\sum_{h\in\mathbb{Z}} \|C_h^*\|_{HS} \le \sum_{h\in\mathbb{Z}} \|\Gamma_h^*\|_F + I(h=0)\|C_U^*\|_{HS} = O_P(1).$$

Furthermore, and because of the $L^2 - \mathcal{M}$ approximability property, the bootstrap process $\mathbf{X}^*$ possesses for every $\omega \in \mathbb{R}$ a spectral density operator $\mathcal{F}_{\omega,m}^*$ defined by

$$(3.2) \qquad \mathcal{F}_{\omega,m}^*(x) = (2\pi)^{-1}\sum_{h\in\mathbb{Z}} C_h^*(x)e^{-ih\omega}, \quad x \in \mathcal{H}.$$

$C_h^*$ and $\mathcal{F}_{\omega,m}^*$ are essentially finite rank approximations of the corresponding population operators $C_h$ and $\mathcal{F}_\omega$ respectively. Thus and in order for the bootstrap process $\mathbf{X}^*$ to capture the infinite dimensional structure of the underlying functional process and the infinite order dependence structure of the vector time series generating the scores, the dimension $m$ as well as the autoregressive order $p$, used in the functional sieve bootstrap algorithm, have to increase to infinity (at some appropriate rate) as the sample size $n$ increases to infinity. This rate should take into account the fact that the true scores and eigenfunctions appearing in the Karhunen-Loève expansion are not observed and, therefore, sample estimates are used instead. Furthermore, the lower bound $\delta_m$ of the spectral density matrix of the scores $f_\xi$, approaches zero as the sample size $n$ increases to infinity. This is due to the fact that the eigenvalues $\nu_j(\omega)$ of the spectral density operator $\mathcal{F}_\omega$ converge to zero as $j \to \infty$. These facts make the asymptotic analysis quite involved and impose several restrictions regarding the behavior of $m$ and $p$ with respect to the sample size $n$ which are summarized in the following assumption.

**Assumption 4** The sequences $m = m(n)$ and $p = p(n)$ satisfy $m \to \infty$ and $p \to \infty$ as $n \to \infty$ such that,

(i)  $m^{3/2} = O(p^{1/2})$

(ii)  $\dfrac{p^7}{n^{1/2}\lambda_m^2}\sqrt{\sum_{j=1}^{m} \dfrac{1}{\alpha_j^2}} \to 0$, where $\alpha_1 = \lambda_1 - \lambda_2$ and $\alpha_j = \min\{\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1}\}$ for $j = 2, 3, ..., m$.

(iii)  $\delta_m^{-1}\sum_{j=p+1}^{\infty} j^r\|A_j(m)\|_F \to 0$ for some $r \ge 0$, where $\delta_m$ is the lower bound of the spectral density matrix $f_\xi$ given in (2.3).

(iv)  $m^4 p^2\|\widetilde{A}_{p,m} - A_{p,m}\|_F = O_P(1)$, where $\widetilde{A}_{p,m} = (\widetilde{A}_{1,p}(m), \ldots, \widetilde{A}_{p,p}(m))$, and $A_{p,m} = (A_{1,p}(m), \ldots, A_{p,p}(m))$. Here, $\widetilde{A}_{j,p}$, $j = 1, 2, \ldots, p$ denotes the same estimator as $\widehat{A}_{j,p}$, $j = 1, 2, \ldots, p$, based on the true

vector time series of scores $\xi_1, \xi_2, \ldots, \xi_n$ instead of their estimates $\widehat{\xi}_1, \widehat{\xi}_2, \ldots, \widehat{\xi}_n$ and $A_{j,p}(m)$, $j = 1, 2, \ldots, m$ are the coefficient matrices of the best (in the mean square sense) linear predictor of $\xi_t$ based on $\xi_{t-j}$, $j = 1, 2, \ldots, p$.

Assumption 4(i) restricts the rate with which the dimension $m$ is allowed to increase to infinity compared with that of $p$. Assumption 4(ii) is imposed in order to control the error made by the fact that the bootstrap procedure is based on estimated scores and eigenfunctions instead on the unobserved true quantities in a context where the dimension $m$ and the autoregressive order $p$, both, increase to infinity and the lower bound of the spectral density matrix of the $m$-dimensional vector of scores approaches zero as $m$ increases to infinity. Part (iii) relates the rate of increase of the autoregressive order $p$ to the lower bound of the spectral density matrix $f_\xi$ and the decay of the norm of the autoregressive matrices to zero. Part (iv) is essentially a requirement on the rate at which $m$ and $p$ are allowed to increase to infinity taking into account the convergence rate of the estimator $\widetilde{A}_{j,p}$, $j = 1, 2, \ldots, p$ based on the true scores. For instance, calculations similar to that in the proof of Lemma 6.3 yield for the Yule-Walker estimator that $\|\widetilde{A}_{p,m} - A_{p,m}\|_F = O_P(mpn^{-1/2}(\sqrt{m}\lambda_m^{-1} + p)^2)$ which, taking into account Assumption 4(i), implies that Assumption 4(iv) is satisfied if $m, p \to \infty$ slowly enough with $n$ such that $mp^6 = O(\sqrt{n}\lambda_m^2)$ and $p\lambda_m^2 = O(m^2)$. Notice that, for real valued-random variables, such assumptions relating the rate of increase of the autoregressive parameters to the convergence rate of the estimators used, are common in the autoregressive-sieve bootstrap literature; see Kreiss et al. (2011) and Meyer and Kreiss (2015). However, the situation here is much more involved since in our context, not only the order $p$ but also the dimension $m$ of the vector autoregression has to increase to infinite with the sample size by taking into account the fact that $\lambda_m$ converges to zero as $m$ increases to infinity.

The following lemma illustrates the rate conditions imposed in Assumption 4 by considering two particular examples of the behavior of the difference $\lambda_j - \lambda_{j+1}$ which is related to the rate of decrease of the eigenvalues $\lambda_j$. According to this lemma, $p$ may increase to infinity as $n^a$ for some $a > 0$ while the rate of increase of $m$ depends on the rate of decrease of $\lambda_j - \lambda_{j+1}$ respectively of the eigenvalues $\lambda_j$, $j = 1, 2, \ldots$. If these differences decrease with a geometric rate, then $m$ may increase at most logarithmically in the sample size $n$, while if the same differences decrease with a polynomial rate, then $m$ may increase to infinity faster, like $n^\zeta$ for some appropriate $\zeta > 0$.

LEMMA 3.1. *Assume that $\widetilde{A}_{p,m}$ are the Yule-Walker estimators of $A_{p,m}$.*

(i)   If $\lambda_j - \lambda_{j+1} \geq C_\lambda \rho^j$ for $j = 1, 2, \ldots$, $\rho \in (0, 1)$ and $C_\lambda > 0$, then Assumption 4(i), (ii) and (iv) is satisfied if

$$p = O(n^a) \;\; and \;\; m \leq \Big(\frac{1}{6\log(\rho^{-1})}(1 - 14a) - \delta\Big)log(n),$$

for $a \in (0, 1/14)$ and some $\delta > 0$.

(ii)   If $\lambda_j - \lambda_{j+1} \geq C_\lambda j^{-\theta}$ for $j = 1, 2, \ldots$ and for some $\theta > 1$ and $C_\lambda > 0$, then Assumption 4(i), (ii) and (iv) is satisfied if

$$p = O(n^a) \quad and \;\; m = O(n^\zeta),$$

for $a \in (0, 1/14)$ and $\zeta \in [\zeta_{\min}, \zeta_{\max}]$, where $\zeta_{\min} = a/(2 + 2\theta)$ and $\zeta_{\max} = \min\{(1 - 14a)/(1 + 6\theta) - \delta, a/3\}$ for some $\delta > 0$.

Under the condition that $m$ and $p$ increase to infinity at an appropriate rate with $n$ such that Assumption 4 is satisfied, the following proposition can be established which shows that the spectral density operator $\mathcal{F}^*_{\omega,m}$ of the bootstrap process $\mathbf{X}^*$ converges, in Hilbert-Schmidt norm, to the spectral density operator $\mathcal{F}_\omega$ of the underlying functional process $\mathbf{X}$.

PROPOSITION 3.2.   *Under Assumptions 1 and 3 and Assumption 2 and 4 with $r = 2$, we have, that, as $n \to \infty$,*

$$\sup_{\omega \in [0,\pi]} \|\mathcal{F}^*_{\omega,m} - \mathcal{F}_\omega\|_{HS} \;\rightarrow\; 0,$$

*in probability.*

From the above proposition and the inversion formulae of Fourier transforms, we immediately get for the covariance operators $C^*_h$ and $C_h$ of the bootstrap process $\mathbf{X}^*$ and of the underlying process $\mathbf{X}$, that $\sup_{h \in \mathbb{Z}} \|C^*_h - C_h\|_{HS} \to 0$, in probability, as $n \to \infty$. Thus the bootstrap process $\mathbf{X}^*$, imitates asymptotically correct the entire infinite dimensional autocovariance structure of the functional process $\mathbf{X}$. This allows for the use of the bootstrap functional time $X^*_1, X^*_2, \ldots, X^*_n$ to approximate the distribution of statistics based on the functional time series $X_1, X_2, \ldots, X_n$. Some examples of such statistics are discussed in the next section.

So far we have assumed that the covariance operator $C_0$ has full rank, i.e., that its eigenvalues $\lambda_j$ are distinct which implies that, for consistency and in order to capture the entire infinite dimensional dependence structure of the underlying functional process $\mathbf{X}$, the number $m$ of principal components included, has to increase to infinity with the sample size $n$.

The situation is much simpler if we assume that $m_0 \in \mathbb{N}$ exists such that $\lambda_{m_0} > 0$ and $\lambda_j = 0$ for all $j > m_0$. In this case only the finite number of $m_0$ score time series are needed to describe the entire dependence structure of $\mathbf{X}$. We are then essentially in the finite dimensional case with the $m_0$-dimensional score process $\{\xi_t = (\langle X_t, v_j \rangle, j = 1, \ldots, m_0)^\top, t \in \mathbb{Z}\}$, possessing a spectral density matrix which is bounded from bellow by a positive constant independent of the sample size $n$. Furthermore, as in the proof of Lemma 6.3 and, because in this case $\sum_{j=1}^{m_0} \|\widehat{v}_j - v_j\|^2 = O_P(n^{-1/2})$, we get that $\|\widehat{A}_{p,m_0} - \widetilde{A}_{p,m_0}\|_F = O_P(p^4/\sqrt{n})$. Standard arguments applied in the case of the (finite dimensional) vector autoregressive-sieve bootstrap can then be used (see for instance Meyer and Kreiss (2015)), to show that under less restrictive conditions that those stated in Assumption 4, $\sup_{\omega \in [0,\pi]} \|\mathcal{F}^*_{\omega,m_0} - \mathcal{F}_\omega\|_{HS} \xrightarrow{P} 0$, in probability.

**4. Bootstrap Validity.** In this section we investigate the validity of the functional sieve bootstrap applied in order to approximate the distribution of some statistic $T_n = T(X_1, X_2, \ldots, X_n)$ of interest, when the bootstrap analogue $T_n^* = T(X_1^*, X_2^*, \ldots, X_n^*)$ is used. Notice that establishing validity of a bootstrap procedure for time series heavily depends on two issues; see also Kreiss and Paparoditis (2011). On the dependence structure of the underlying process which affects the distribution of the statistic of interest and on the capability of the bootstrap procedure used to mimic appropriately this dependence structure. Furthermore, since proving bootstrap validity is a case by case matter, we demonstrate in the following applications of the functional sieve bootstrap procedure proposed to some statistics that have recently attracted considerable interest in the functional time series literature.

4.1. *Functional finite Fourier transform.* Consider the distribution of the functional Fourier transform

(4.1) $$S_n(\omega) = \sum_{t=1}^{n} X_t e^{-it\omega}, \quad \omega \in [-\pi, \pi].$$

Notice that the sample mean $\overline{X}_n = n^{-1} S_n(0)$ is just a special case of (4.1). In order to elaborate on the limiting distribution of $S_n(\omega)$ we first fix some notation. We say that a random element $Z \in \mathcal{H}_\mathbb{C} := \mathcal{H} + i\mathcal{H}$, follows a circularly-symmetric complex Gaussian distribution with mean zero and covariance $\mathcal{G}$, we write $Z \sim CN(0, \mathcal{G})$, if

$$\begin{pmatrix} Re(Z) \\ Im(Z) \end{pmatrix} \sim N_{\mathcal{H} \times \mathcal{H}}\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \frac{1}{2} \begin{pmatrix} Re(\mathcal{G}) & -Im(\mathcal{G}) \\ Im(\mathcal{G}) & Re(\mathcal{G}) \end{pmatrix} \right);$$

see also Cerovecki and Hörmann (2017) for a general discussion of the complex Gaussian distribution.

Under a range of different weak dependence assumptions on the functional process $\mathbf{X}$, it has been shown that

$$(4.2) \qquad\qquad n^{-1/2} S_n(\omega) \Rightarrow CN(0, 2\pi \mathcal{F}_\omega)$$

as $n \to \infty$, where $\Rightarrow$ denotes weak convergence on $\mathcal{H}_\mathbb{C}$. For $\omega = 0$, such a limiting behavior has been established for linear functional processes by Merlevède et al. (1997) and for $L^p - \mathcal{M}$ approximable processes by Horvàth et al. (2013). Panaretros and Tavakoli (2013) derived the above limiting distribution of $n^{-1/2} S_n(\omega)$ for $\omega \in [0, \pi]$, under a summability condition of the functional cumulants, while more general results for the same statistic and under weaker conditions, have been recently obtained by Cerovecki and Hörmann (2017).

We propose to use the bootstrap statistic $n^{-1/2} S_n^*(\omega) = n^{-1/2} \sum_{t=1}^n X_t^* e^{-it\omega}$ in order to approximate the distribution of the statistic $n^{-1/2} S_n(\omega)$. The following theorem establishes asymptotic validity of this functional sieve bootstrap proposal for the class of functional Fourier transforms considered. In this theorem, $d$ is any metric metrizing weak convergence on $\mathcal{H}_\mathbb{C}$.

THEOREM 4.1.  *Suppose that for $\omega \in [0, \pi]$, the sequence $\{n^{-1/2} S_n(\omega), n \in \mathbb{N}\}$ in $\mathcal{H}_\mathbb{C}$ satisfies (4.2). Suppose further that Assumptions 1 and 3 and Assumption 2 and 4 with $r = 2$ are satisfied. Then, as $n \to \infty$,*

  (i)  $d\big(\mathcal{L}(n^{-1/2} S_n(\omega)), \mathcal{L}(n^{-1/2} S_n^*(\omega) | X_1, X_2, \dots, X_n)\big) \to 0$, *and*

  (ii)  $\|n^{-1} E^* S_n^*(\omega) \otimes S_n^*(\omega) - n^{-1} E S_n(\omega) \otimes S_n(\omega)\|_{HS} \xrightarrow{P} 0$,

*in probability.*

REMARK 4.1.  Notice that as a special case of the above theorem we get that, under the assumptions made, and as $n \to \infty$, $\sqrt{n} \overline{X}_n^* \Rightarrow N\big(0, \sum_{h \in Z} C_h\big)$, in probability and $n E^* \overline{X}_n^* \otimes \overline{X}_n^* \xrightarrow{P} 2\pi \mathcal{F}_0$, which provides one of the first instances of a central limit theorem for the bootstrap for functional time series under the weak dependence conditions stated in Assumption 1.

4.2. *Fully functional testing.*  In a variety of functional testing situations one is faced with the problem that the limiting distribution under the null of a fully functional test statistic, depends, in a complicated way, on difficult to estimate characteristics of the underlying functional process. This makes

the practical implementation of asymptotic results derived in order to calculate critical values of tests a difficult task. To overcome this problem, a common approach in the literature is to consider tests based on finite dimensional projections. However, such tests have non-degenerated power only for alternatives which are not orthogonal to the space captured by the particular projections considered; see Horváth et. al (2013) and Horváth et al. (2014) for examples. Using as an example the two sample mean problem, we demonstrate in the following how the sieve bootstrap procedure proposed in this paper, can be successfully applied to approximate the null distribution of a fully functional test.

Let $\mathbf{X} = \{X_t, t \in \mathbb{Z}\}$ and $\mathbf{Y} = \{Y_t, t \in \mathbb{Z}\}$ be two independent, strictly stationary functional processes with mean functions $\mu_X = EX_t$ and $\mu_Y = EY_t$ respectively and consider the testing problem $H_0 : \mu_X = \mu_Y$ against the alternative $H_1 : \mu_X \neq \mu_Y$. Given two time series $X_1, X_2, \ldots, X_{n_1}$ and $Y_1, Y_2, \ldots, Y_{n_2}$ stemming from $\mathbf{X}$ and $\mathbf{Y}$ respectively, a natural test statistic for these hypotheses is given by

$$U_{n_1,n_2} = \frac{n_1 n_2}{n_1 + n_2} \|\overline{X}_{n_1} - \overline{Y}_{n_2}\|^2,$$

where $\overline{X}_{n_1} = n_1^{-1} \sum_{t=1}^{n_1} X_t$ and $\overline{Y}_{n_2} = n_2^{-1} \sum_{t=1}^{n_2} Y_t$. If both processes satisfy Assumption 1 and $n_1, n_2 \to \infty$ such that $n_1/(n_1 + n_2) \to \theta \in (0,1)$, it has been shown in Horváth et al. (2013), that $U_{n_1,n_2} \xrightarrow{d} \int_0^1 \Gamma^2(\tau)d\tau$, where $\{\Gamma(\tau), \tau \in [0,1]\}$ is a mean zero Gaussian process with covariance function $E(\Gamma(\tau_1)\Gamma(\tau_2)) = (1 - \theta)c_X(\tau_1, \tau_2) + \theta c_Y(\tau_1, \tau_2)$ for $\tau_1, \tau_2 \in [0,1]$ and $c_X(\tau_1, \tau_2) = Cov(X_0(\tau_1), X_0(\tau_2)) + \sum_{h \geq 1} Cov(X_0(\tau_1), X_h(\tau_2)) + \sum_{h \geq 1} Cov(X_0(\tau_2), X_h(\tau_1))$ and $c_Y(\tau_1, \tau_2) = Cov(Y_0(\tau_1), Y_0(\tau_2)) + \sum_{h \geq 1} Cov(Y_0(\tau_1), Y_h(\tau_2)) + \sum_{h \geq 1} Cov(Y_0(\tau_2), Y_h(\tau_1))$. Notice that the kernel functions $c_X$ and $c_Y$ are unknown, which makes the calculation of critical values of the test $U_{n_1,n_2}$ a difficult task.

Since the functional sieve bootstrap procedure proposed satisfactory imitates the autocovariance structure of the underlying processes, it can be successfully applied to estimate the critical values of the test $U_{n_1,n_2}$. To elaborate, the goal is to generate two independent functional pseudo-time series $X_1^*, X_2^*, \ldots, X_{n_1}^*$ and $Y_1^*, Y_2^*, \ldots, Y_{n_2}^*$, that mimic the autocovariance structure of the processes $\mathbf{X}$ and $\mathbf{Y}$ respectively and satisfy, at the same time, the null hypothesis of interest. For this let $X_t^*$ and $Y_t^*$ be generated by means of equation (3.1) of the functional sieve bootstrap algorithm, where for the generation of the $X_t^*$'s the sample scores $\widehat{\xi}_t^{(X)} = (\widehat{\xi}_{j,t}^{(X)} = \langle X_t, \widehat{v}_j^{(X)} \rangle, j = 1, 2, \ldots, m_1)^\top, t = 1, 2, \ldots, n_1$ and for the generation of the $Y_t^*$'s, the sample scores $\widehat{\xi}_t^{(Y)} = (\widehat{\xi}_{j,t}^{(Y)} = \langle Y_t, \widehat{v}_j^{(Y)} \rangle, j = 1, 2, \ldots, m_2)^\top, t = 1, 2, \ldots, n_2$

are used in *Step 1* of this algorithm. Here $\widehat{v}_j^{(X)}$, $j = 1, \ldots, m_1$ and $\widehat{v}_j^{(Y)}$, $j = 1, \ldots, m_2$, denote the orthonormalized eigenfunctions of the $m_1$ respectively $m_2$ largest eigenvalues of the sample covariance operators $\widehat{C}_0^{(X)} = n_1^{-1} \sum_{t=1}^{n_1} (X_t - \overline{X}_{n_1}) \otimes (X_t - \overline{X}_{n_1})$ and $\widehat{C}_0^{(Y)} = n_2^{-1} \sum_{t=1}^{n_2} (Y_t - \overline{Y}_{n_2}) \otimes (Y_t - \overline{Y}_{n_2})$ respectively. Notice that generation of $X_t^*$ and $Y_t^*$ by using (3.1) ensures that $E^* X_t^* = E^* Y_t^* = 0$, that is the generated functional pseudo-time series $X_1^*, X_2^*, \ldots, X_{n_1}^*$ and $Y_1^*, Y_2^*, \ldots, Y_{n_2}^*$ satisfy the null hypothesis $H_0$. Now, let $\overline{X}_{n_1}^* = n_1^{-1} \sum_{t=1}^{n_1} X_t^*$ and $\overline{Y}_{n_2}^* = n_2^{-1} \sum_{t=1}^{n_2} Y_t^*$ and define the bootstrap analogue of $U_{n_1, n_2}$ as

$$U_{n_1, n_2}^* = \frac{n_1 n_2}{n_1 + n_2} \| \overline{X}_{n_1}^* - \overline{Y}_{n_2}^* \|^2.$$

The following theorem establishes validity of the sieve bootstrap applied to the functional testing problem considered.

THEOREM 4.2.    *Let the conditions of Theorem 4.1 be satisfied and assume that $n_1, n_2 \to \infty$ such that $n_1/(n_1 + n_2) \to \theta \in (0, 1)$. Then,*

$$\sup_{x \in \mathbb{R}} \left| P(U_{n_1, n_2} \leq x) - P(U_{n_1, n_2}^* \leq x | \mathbf{X}_{n_1}, \mathbf{Y}_{n_2}) \right| \to 0,$$

*in probability, where $P(U_{n_1, n_2}^* \leq \cdot | \mathbf{X}_{n_1}, \mathbf{Y}_{n_2})$ denotes the distribution function of $U_{n_1, n_2}^*$ conditional on $\mathbf{X}_{n_1} = (X_1, X_2, \ldots, X_{n_1})$ and $\mathbf{Y}_{n_2} = (Y_1, Y_2, \ldots, Y_{n_2})$.*

## 5. Choice of Parameters and Numerical Results.

5.1. *Choice of the sieve bootstrap parameters.* Implementation of the functional sieve bootstrap requires the choice of two tuning parameters: the order $p$ and the dimension $m$. By choosing these parameters, the problem of overfitting caused by selecting a large dimension and/or a high order vector autoregressive model, should be seriously taken into account.

Several approaches for selecting the number of principal components in functional data analysis have been proposed in the literature; see among others Yao et al. (2005) and Li et al. (2013) for the use of information type criteria. For our purpose, one useful and simple criterion for selecting the dimension $m$ is based on the ratio of the total variance explained by the number $m$ of principal components included, to the variance of $X_t$. According to this rule, $m$ is selected as the smallest positive integer for which the empirical variance ratio $(VR_n)$ satisfies $VR_n(m) = \sum_{j=1}^{m} \widehat{\lambda}_j / \sum_{j=1}^{n} \widehat{\lambda}_j \geq Q$, with $Q$ a predetermined value and $Q = 0.80$ or $Q = 0.85$ two common

choices; cf. Hórvath and Kokoszka (2012). One drawback of the VR-rule applied to functional time series, is that this criterion does not take into account dependence.

To overcome this drawback we introduce in the following a generalized variance ratio criterion. Measuring the total variability of the underlying functional process by the quantity $\int_{(-\pi,\pi]} \|\mathcal{F}_\omega\|_{HS}^2 d\omega$, yields by straightforward calculations and evaluating the Hilbert-Schmidt norm using the orthonormal basis $\{v_j, j = 1, 2, \ldots\}$, the expression

$$\int_{(-\pi,\pi]} \|\mathcal{F}_\omega\|_{HS}^2 d\omega = \sum_{l=1}^{\infty} \sum_{r=1}^{\infty} \int_{(-\pi,\pi]} \left|f_{\xi_l,\xi_r}(\omega)\right|^2 d\omega,$$

where $f_{\xi_l,\xi_r}$ denotes the cross spectral density of the score processes $\{\xi_{l,t}\}$ and $\{\xi_{r,t}\}$. Define next a functional process $\mathbf{X}_m^+ = \{X_t^+, \in \mathbb{Z}\}$, where $X_t^+ = X_{t,m}^+ + U_{t,m}^+$, $X_{t,m}^+ = \sum_{j=1}^{m} \xi_{j,t} v_j$, $U_{t,m}^+ = \sum_{j=m+1}^{\infty} \zeta_{j,t} v_j$ and $\{\zeta_{j,t}, t \in \mathbb{Z}\}, j = m+1, m+2, \ldots$, are independent, i.i.d. processes which are independent from $X_{t,m}^+$ and have mean zero and $Var(\zeta_{j,t}) = \lambda_j$. Observe that for any $m$ fixed and ignoring estimation errors, it is the dependence structure of $\mathbf{X}_m^+$ which is essentially mimicked by the functional sieve bootstrap process $\mathbf{X}^*$. This is so since in the bootstrap world, $U_{t,m} = X_t - \sum_{j=1}^{m} \xi_{j,t} v_j$ is treated as an i.i.d. process and the (possible) correlation between the processes $\{X_{t,m} = \sum_{j=1}^{m} \xi_{j,t} v_j\}$ and $\{U_{t,m}\}$ is ignored. Let $\mathcal{F}_{\omega,m}^+$ be the spectral density operator of $\mathbf{X}_m^+$. Using the same measure of total variability as for the process $\mathbf{X}$, we get

$$\int_{(-\pi,\pi]} \|\mathcal{F}_{\omega,m}^+\|_{HS}^2 d\omega = \sum_{l=1}^{m} \sum_{r=1}^{m} \int_{(-\pi,\pi]} \left|f_{\xi_l,\xi_r}(\omega)\right|^2 d\omega + (2\pi)^{-1} \sum_{l=m+1}^{\infty} \lambda_l^2.$$

Notice that the term $(2\pi)^{-1} \sum_{l=m+1}^{\infty} \lambda_l^2$ is due to integrating the squared Hilbert-Schmidt norm of the spectral density operator of the process $\{U_{t,m}^+\}$. This process is included in the definition of $\mathbf{X}_m^+$ because of the functional i.i.d. innovations $U_t^*$ used in *Step 6* of the sieve bootstrap algorithm to generate the $X_t^*$'s.

The ratio

$$GVR(m) = \int_{(-\pi,\pi]} \|\mathcal{F}_{\omega,m}^+\|_{HS}^2 d\omega \Big/ \int_{(-\pi,\pi]} \|\mathcal{F}_\omega\|_{HS}^2 d\omega,$$

can then be considered as the proportion of total variability of the process $\mathbf{X}$ captured by that of the process $\mathbf{X}_m^+$. Recall that $\left|f_{\xi_l,\xi_r}(\omega)\right|^2 = \kappa_{l,r}^2(\omega) f_{\xi_l,\xi_l}(\omega) f_{\xi_r,\xi_r}(\omega)$ with $\kappa_{l,r}$ the squared coherency between the score

processes $\{\xi_{l,t}\}$ and $\{\xi_{r,t}\}$. That is, $GVR$ explicitly takes into account the entire autocovariance structure of the processes $\mathbf{X}$ and $\mathbf{X}_m^+$. $GVR(m)$ can then be interpreted as a measure of the los on information on the dependence structure of $\mathbf{X}$ caused by the functional sieve bootstrap procedure based on $m$ principal components. Note that if $\mathbf{X}$ is a white noise process, then $GRV(m) = 1$ for every value of $m$. In this case we set $m = 0$ as the most parsimonious choice, i.e., no vector autoregression is fitted, which implies that the functional sieve bootstrap (correctly) reduces to an i.i.d. bootstrap.

Now, observe that $\lambda_j$, $\int_{(-\pi,\pi]} \left| f_{\xi_l,\xi_r}(\omega) \right|^2 d\omega$ and $\int_{(-\pi,\pi]} \|\mathcal{F}\omega\|_{HS}^2 d\omega$ can be consistently estimated by $\widehat{\lambda}_j$, $2\pi n^{-1} \sum_{j \in F_n} |I_{\xi_l,\xi_r}(\omega_j)|^2$ and $2\pi n^{-1} \sum_{j \in F_n} \|I_{n,\omega_j}\|_{HS}^2$, respectively, where $I_{\xi_l,\xi_r}(\omega) = J_{\xi_l}(\omega) J_{\xi_r}(-\omega)$ and $J_{\xi_s}(\omega) = (2\pi n)^{-1/2} \sum_{t=1}^n \xi_{s,t} \, e^{-i\omega t}$ for any $s \geq 1$. Furthermore, $I_{n,\omega}$ is the periodogram operator with kernel $I_{n,\omega}(\tau_1, \tau_2) = J_{n,\omega}(\tau_1)\overline{J}_{n,\omega}(\tau_2)$, $J_{n,\omega}(\tau) = (2\pi n)^{-1/2} \sum_{t=1}^n X_t(\tau) e^{-i\omega t}$, $\omega_j = 2\pi j/n$, $F_n = \{-N, \ldots, -1, 1, \ldots, N\}$ and $N = [n/2]$. This suggests to select the dimension $m$ as the smallest positive integer for which the empirical generalized variance ratio $(GVR_n)$ satisfies

$$GVR_n(m) = \frac{\sum_{l=1}^m \sum_{r=1}^m \frac{2\pi}{n} \sum_{j \in F_n} \left| \widehat{I}_{\xi_l,\xi_r}(\omega_j) \right|^2 + \frac{1}{2\pi} \sum_{l=m+1}^n \widehat{\lambda}_l^2}{\frac{2\pi}{n} \sum_{j \in F_n} \left\| I_{n,\omega_j} \right\|_{HS}^2} \geq Q.$$

Here $\widehat{I}_{\xi_l,\xi_r}(\omega) = \widehat{J}_{\xi_l}(\omega)\widehat{J}_{\xi_r}(-\omega)$ with $\widehat{J}_{\xi_s}(\omega) = (2\pi n)^{-1/2} \sum_{t=1}^n \widehat{\xi}_{s,t} e^{-i\omega t}$ the finite Fourier transform of the time series of estimated scores.

REMARK 5.1. $GVR_n$ has been developed for the functional sieve bootstrap situation considered in this paper. However, a simple modification of this criterion leads to an alternative to the $VR_n$ rule which is appropriate for dependent functional data and which is of interest on its own. In particular, ignoring the second term of the nominator of $GVR_n$, the following dependent variance ratio $(DVR_n)$ criterion is obtained,

$$DVR_n(m) = \sum_{l=1}^m \sum_{r=1}^m \sum_{j \in F_n} \left| \widehat{I}_{\xi_l,\xi_r}(\omega_j) \right|^2 \bigg/ \sum_{j \in F_n} \left\| I_{n,\omega_j} \right\|_{HS}^2.$$

$DVR_n$ delivers an empirical measure of the lost on information on the dependence structure of $\mathbf{X}$ associated with the use of the $m$-dimensional space and can be therefore, used as a simple criterion to select the number $m$ of principal components in a functional time series setting. Notice that if the

Hilbert-Schmidt norm in GVR is replaced by the trace norm of the spectral density operators involved and the additional term $(2\pi)^{-1} \sum_{l=m+1}^{\infty} \lambda_l^2$ is ignored, then the corresponding $DVR(m)$ ratio given by $DVR(m) = \sum_{l=1}^{m} \sum_{r=1}^{m} \int_{-\pi}^{\pi} |f_{\xi_l,\xi_r}(\omega)|^2 d\omega \Big/ \sum_{l=1}^{\infty} \sum_{r=1}^{\infty} \int_{-\pi}^{\pi} |f_{\xi_l,\xi_r}(\omega)|^2 d\omega$, reduces to the $VR(m) = \sum_{l=1}^{m} \lambda_l \Big/ \sum_{l=1}^{\infty} \lambda_l$ ratio.

Notice that both, the VR and the GVR criterion, refer to a fixed sample size $n$ and the purpose is to select the number of principal components in a way which ensures that a desired fraction $Q$ of the variance of the process is captured by the number of principal components included in the analysis. This is important for our bootstrap proposal where the objective is to appropriately mimic the dependence structure of the functional time series at hand. However, consistency requires that $m$ increases to infinity with $n$ which is not the case if $Q$ remains fixed with $n$. At the same time and as we have seen, the rate at which $m$ has to increase to infinity should take into account the rate of decrease of the eigenvalues $\lambda_j$ respectively of the differences $\lambda_j - \lambda_{j+1}$ to zero. One way to accommodate such aspects in our practical selection of $m$, is to combine the discussed VR respectively GVR criterion with an approach for selecting $m$ proposed by Hörmann and Kidziński (2015) and which explicitly takes into account the behavior of the eigenvalues $\widehat{\lambda}_j$. To elaborate, denote by $m_{n,E}$ the number of principal components selected by the rule

$$m_{n,E} = \operatorname{argmax}\Big\{ j \geq 1 : \frac{\widehat{\lambda}_1}{\widehat{\lambda}_j} \leq \sqrt{n}/\log(n) \Big\}.$$

Notice that $m_{n,E}$ allows for the $j$-th principal component to be included in the analysis if the corresponding estimated eigenvalue $\widehat{\lambda}_j$ is big enough, i.e., if the ratio $1/\widehat{\lambda}_j$ does not exceed the threshold $\sqrt{n}/\log(n)$. The nominator $\widehat{\lambda}_1$ acts solely as a normalization to adapt for scaling; for this and for the choice the particular threshold see Hörmann and Kidziński (2015). Denote now by $m_{n,Q}$ the number of principal components selected using, the VR or the GVR criterion for some given $Q$. The practical selection of $m$ we then propose is to set this parameter equal to

$$\widehat{m}_n = \max\{m_{n,Q}, m_{n,E}\}.$$

According to this proposal, only those principal directions are included in the analysis the eigenvalues of which can be estimated with a reasonably accuracy ensuring at the same time that the number of principal components selected explains at least a desired portion of the variability of the time

series at hand. We remark that although for functional time series the GVR criterion is theoretically more appealing, for short time series of $n \le 100$ observations, we still recommend the use the VR-criterion since it leads to selections of $m$ with a smaller variability avoiding, therefore, the potential fit of vector autoregressions of large dimensions and/or of high orders which is an important issue for small samples sizes; see also Section 5.2 for details.

Once the dimension $m$ has been selected, the order $p$ of the vector autoregression fitted can be chosen using the AICC criterion; see Hurvich and Tsai (1993). This criterion is preferred because it is based on an approximately unbiased estimator of the expected Kullback-Leibler information of the fitted model and, more importantly, avoids overfitting. The order $p$ is then selected by minimizing $AICC(p) = n \log |\widehat{\Sigma}_{e,p}| + n(nm + pm^2)/(n - m(p+1) - 1)$, over a range of possible values of $p$, where $\widehat{\Sigma}_{e,p} = n^{-1} \sum_{t=p+1}^{n} \widehat{e}_{t,p} \widehat{e}_{t,p}^T$ and $\widehat{e}_{t,p}$ is defined in *Step 4* of the functional sieve bootstrap algorithm.

5.2. *Simulations.* To investigate the finite sample behavior of the functional sieve bootstrap (FSP) we have performed simulations using time series stemming from a first order functional moving average process given by

$$(5.1) \qquad\qquad X_t = \varepsilon_t + \Theta(\varepsilon_{t-1}).$$

as in Aue et al. (2015). To elaborate, $\Theta$ is specified as $\Theta = 0.8\Psi$, where $\Psi$ is a linear operator, $\Psi : \mathcal{H}_D \to \mathcal{H}_D$, $\mathcal{H}_D = \overline{sp}\{f_1, f_2, \ldots, f_D\}$, $D = 21$ and $f_j$, $j = 1, 2, \ldots, D$ are Fourier basis functions on the interval $[0, 1]$. Notice that for $x \in \mathcal{H}_D$, $x = \sum_{j=1}^{D} c_j f_j$ with $c_j = \langle x, f_j \rangle$, the operator $\Psi$ acts as $\Psi(x) = \sum_{j=1}^{D} \sum_{l=1}^{D} c_j \langle \Psi(f_j), f_l \rangle f_l = (B_\Psi c)' v$, where $c = (c_1, \ldots, c_D)'$ and $v = (f_1, \ldots, f_D)'$ and the matrix $B_\Psi$ has element in the $j$th column and $l$th row given by $\langle \Psi(f_j), f_l \rangle$. Following Aue et al. (2015) the operator $\Psi$ was chosen at random. For this a $D \times D$ matrix of independent, normal random variables with mean zero was first generated where its $(j_1, j_2)$th element has standard deviation $\sigma_{j_1, j_2} = j_1^{-1} j_2^{-1}$. This matrix was then scaled so that the resulting matrix $B_\Psi$ has induced norm equal to 1 and in every iteration of the simulation runs $B_\Psi$ was newly generated. The corresponding i.i.d. innovations $\varepsilon_t$ in (5.1) were generated as $\varepsilon_t = \sum_{j=1}^{D} Z_{t,j} f_j$, where $Z_{t,j}$ are i.i.d. Gaussian with mean zero and standard deviation equal to $j^{-1}$.

We first consider the performance of the VR and GVR criteria in selecting the number $m$ of principal components, when $Q = 0.85$. Table 1 shows the frequencies of selected dimensions $m$ over $R = 1000$ replications of the considered FMA(1) model for different sample sizes. As it is seen from this table, the VR criterion is quite stable over the different sample sizes leading to the selections $m = 3$ or $m = 4$ in almost all situations. The GVR criterion

TABLE 1

*Frequency of selected values of $m$ by the VR and the GVR criterion ($R = 1000$ replications).*

| | | m = | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|---|
| n=100 | $VR_n$ | | 0 | 0.3 | 67.1 | 32.6 | 0 | 0 | 0 |
| | $GVR_n$ | | 0 | 0.9 | 19.6 | 55.2 | 22.8 | 1.5 | 0 |
| n=200 | $VR_n$ | | 0 | 0 | 62.7 | 37.3 | 0 | 0 | 0 |
| | $GVR_n$ | | 0 | 0.1 | 10.4 | 68.7 | 20.6 | 0.2 | 0 |
| n=300 | $VR_n$ | | 0 | 0 | 66.2 | 33.8 | 0 | 0 | 0 |
| | $GVR_n$ | | 0 | 0 | 4.3 | 75.4 | 20.3 | 0 | 0 |
| n=500 | $VR_n$ | | 0 | 0 | 64.7 | 35.3 | 0 | 0 | 0 |
| | $GVR_n$ | | 0 | 0 | 0.9 | 83.0 | 16.1 | 0 | 0 |
| n=1000 | $VR_n$ | | 0 | 0 | 64.1 | 35.9 | 0 | 0 | 0 |
| | $GVR_n$ | | 0 | 0 | 0.2 | 89.3 | 10.5 | 0 | 0 |

exhibits a greater variability for small sample sizes ($n \leq 100$) and becomes more concentrated around the dimensions $m = 4$ and $m = 5$ as $n$ increases. Observe that because the GVR criterion explicitly takes into account the dependence structure of the processes involved, it selects more frequently the larger dimension $m = 4$ compared to the dimension $m = 3$ which is more frequently selected by the VR criterion. Notice further that the smaller variability of the VR rule for small sample sizes, prohibits the selection of vector autoregressions of large dimension which is particularly important in our set-up. Thus for $n \leq 100$ observations we recommend to apply the $\widehat{m}_n$ rule using the VR criterion to calculate $m_{n,Q}$ and the AICC criterion in order to select the values of $m$ and $p$.

To investigated the behavior of $\widehat{m}_n$ for the FMA(1) model considered, we use a range of sample sizes with $m_{n,Q}$ chosen according to the $VR$ ($n \leq 100$) respectively $GVR$ criterion with $Q = 0.85$. Table 1 of the supplementary material shows the results obtained over $R = 1000$ repetitions for each of the sample sizes considered. As it is seen from this table, the behavior of $\widehat{m}_n$ is dominated for small to moderate sample sizes by $m_{n,Q}$ ensuring, therefore, the desired description of the variability of the functional time series by the number $m$ of principal components selected. However, as $n$ increases the behavior of $\widehat{m}_n$ becomes dominated by $m_{n,E}$ which allows for the number of principal components selected as well as for the part of the variance explained, to increase with $n$.

We next consider the behavior of the FSB procedure in estimating the standard deviation of the sample mean $\sqrt{n}\overline{X}_n(\tau_j) = n^{-1/2} \sum_{t=1}^{n} X_t(\tau_j)$, calculated for time series of length $n = 100$ observations and for $\tau_j$, $j = 1, 2, \ldots, T$, $T = 21$, equidistant time points in the interval $[0, 1]$. The exact standard deviation of the sample mean is estimated using 20,000 replications

TABLE 2

*Averaged absolute bias (ABias), Averaged relative bias (RBias) and Averaged standard deviation (AStd) of the moving block bootstrap (MBB), the tapered block bootstrap (TBB), the stationary bootstrap (SB) and the functional sieve bootstrap (FSB) estimates of the standard deviation of the sample mean $\overline{X}_n$.*

|  | MBB | | TBB | | SB | | FSB | | |
|---|---|---|---|---|---|---|---|---|---|
|  | $b_1 = 5$ | $b_2=9$ | $b_1 = 7$ | $b_2 = 6$ | $b_1 = 5$ | $b_2 = 6$ | $(2,3)$ | $(3,3)$ | $(\widehat{m},\widehat{p})$ |
| ABias | 0.206 | 0.208 | 0.139 | 0.153 | 0.255 | 0.256 | 0.037 | 0.054 | 0.121 |
| RBias | 0.091 | 0.092 | 0.061 | 0.068 | 0.112 | 0.113 | 0.016 | 0.024 | 0.053 |
| AStd | 0.321 | 0.406 | 0.350 | 0.312 | 0.341 | 0.371 | 0.445 | 0.462 | 0.484 |

of the moving average model (5.1). All estimates presented are based on $R = 1,000$ replications and $B = 1,000$ bootstrap repetitions. Table 2 of the supplementary material shows the FSB estimates obtained using some different values of the bootstrap parameters $m$ and $p$ as well as for the values of these parameters chosen by means of the $\widehat{m}_n$ and $AICC$ rule and which are denoted by $(\widehat{m},\widehat{p})$. Note that $(m,p) = (3,3)$ is the most frequently chosen pair using this data driven selection rule. As this table shows the FSB estimates are quite good even for the short functional time series of $n = 100$ observations. These estimates also seem not to be very sensitive with respect to the different choices of the parameter $m$ used to truncate the Karhunen-Loéve expansion.

Table 2 compares the results using the FSB procedure with those of three different block bootstrap methods, the moving block bootstrap (MBB), the tapered block bootstrap (TBB) and the stationary bootstrap (SB). To asses the overall behavior of the different bootstrap estimates, we use the averaged absolute bias (ABias), $T^{-1}\sum_{j=1}^{T}|\overline{\sigma}^*(\tau_j) - \sigma(\tau_j)|$, the averaged relative bias (RBias), $T^{-1}\sum_{j=1}^{T}|\overline{\sigma}^*(\tau_j)/\sigma(\tau_j) - 1|$ and the averaged standard deviation of the bootstrap estimates (AStd), calculated as $T^{-1}\sum_{j=1}^{T}\sqrt{\widehat{Var}(\sigma^*(\tau_j))}$, where $\sigma(\tau_j)$ is the estimated exact standard deviation, $\widehat{Var}(\sigma^*(\tau_j)) = (R-1)^{-1}\sum_{r=1}^{R}(\sigma_r^*(\tau_j) - \overline{\sigma}^*(\tau_j))^2$, with $\sigma_r^*(\tau_j)$ denoting the bootstrap estimate of $\sigma(\tau_j)$ obtained in the $r$th replication, $r = 1,2,\ldots,R$, and $\overline{\sigma}^*(\tau_j) = R^{-1}\sum_{r=1}^{R}\sigma_r^*(\tau_j)$. For the three block bootstrap methods considered we report the results for two block sizes denoted by $b_1$ and $b_2$, for which the corresponding methods achieve the two lowest ABias respectively RBias values. Thus the results presented for the three block bootstrap methods in Table 2 are those having the overall lowest bias. Finally, for the FSB procedure we report the results for the values $(m,p) = (2,3)$, $(m,p) = (3,3)$ and for the values of these parameters chosen by the $\widehat{m}_n$ and $AICC$ rule denoted by $(\widehat{m},\widehat{p})$.

As it is seen from Table 2, between the three block bootstrap estimators considered, the MBB estimator seems to behave better that the SB estimator, while both estimators are outperformed by the TBB estimator. However, compared to the FSB estimates, all block bootstrap estimates are quite biased and they are clearly outperformed by the FSB estimates. This is true even for the case where the parameters of the FSB procedure are chosen data dependent, where the bias of the FSB estimates is smaller that the lowest bias achieved by the block bootstrap methods. The FSB estimates have a larger standard deviation which, however, is not surprising taking into account the fact that this bootstrap method requires the estimation of $m^2 p$ autoregressive coefficients. It is worth investigating whether the standard deviation of the FSB estimates can be reduced by using sparse methods to fit the vector autoregression involved in the bootstrap procedure.

The results of a small simulation study investigating the finite sample size and power behavior of the bootstrap based, fully functional test for the two-sample mean problem considered in Section 4.2, are presented in the supplementary material.

## 6. Auxiliary Results and Proofs.

LEMMA 6.1. *Let Assumption 1, 2 and 3 be satisfied. Denote by $\Psi_j(m)$, $j = 1, 2, \ldots$, the coefficients matrices of the power series $A_m^{-1}(z)$, where $A_m(z) = I_m - \sum_{j=1}^{\infty} A_j(m) z^j$, $|z| \leq 1$, and let $\Sigma_e(m) = E(e_t(m) e_t^\top(m))$. Then,*

(i)   $\sum_{j=1}^{\infty} (1+j)^r \|A_j(m)\|_F = O(1)$,
(ii)  $\sum_{j=1}^{\infty} (1+j)^r \|\Psi_j(m)\|_F = O(1)$,  *and*
(iii) $0 < c_e \leq \|\Sigma_e(m)\|_F = O(1)$,

*where all bounds on the right hand side are valid uniformly in $m$.*

The following version of Baxter's inequality is very useful in our setting because it relates the approximation error of the coefficient matrices of the finite predictor and of the autoregressive-representation of the $m$-dimensional process of scores to the lower bound of the spectral density matrix $f_\xi(\cdot)$. It is an immediate consequence of Lemma 2.1 and of Theorem 3.2 in Meyer et al. (2016).

LEMMA 6.2. *Let Assumption 1, 2 and 3 be satisfied. Then there exists a constant $C > 0$ which does not depend on $m$, such that for all $0 \leq s \leq r - 1$,*

$$\sum_{j=1}^{p} (1+j)^s \|A_{j,p}(m) - A_j(m)\|_F \leq C \delta_m^{-1} \sum_{j=p+1}^{\infty} (1+j)^{s+1} \|A_j(m)\|_F,$$

*where $\delta_m$ is given in Lemma 2.1.*

The following lemma provides a useful bound between the estimated matrices of the autoregressive parameters based on the vector of scores $\xi_t$ and on the vector of their estimates $\widehat{\xi}_t$, $t = 1, 2, \ldots, n$. It deals with the case of the Yule-Walker estimators but similar bounds can be established along the same lines for other estimators, like for instance for least squares estimators.

LEMMA 6.3. *Let Assumption 1 be satisfied, let $\widehat{A}_{p,m} = (\widehat{A}_{j,p}(m), j = 1, 2, \ldots, p)$ and let $\widetilde{A}_{p,m} = (\widetilde{A}_{j,p}(m), j = 1, 2, \ldots, p)$ be the Yule-Walker estimators of $A_{j,p}(m)$, $j = 1, 2, \ldots, p$, based on the time series of true scores $\xi_1, \xi_2, \ldots, \xi_n$. Then,*

$$\big\|\widehat{A}_{p,m} - \widetilde{A}_{p,m}\big\|_F = O_P\Big(\Big(\frac{p\sqrt{m}}{\lambda_m} + p^2\Big)^2 \Big\{\frac{1}{n}\sum_{j=1}^{m}\frac{1}{\alpha_j^2}\Big\}^{1/2}\Big).$$

LEMMA 6.4. *Let Assumption 1 and 2 (with $r = 0$) be satisfied and $A_{p,m}(z) = I - \sum_{j=1}^{p} A_{j,p}(m)z^j$, $z \in \mathbb{C}$. There exists $p_m \in \mathbb{N}$ and a positive constant $C$ which does not depend on $m$ such that for $m \in \mathbb{N}$ and all $p > p_m$,*

$$\inf_{|z|\leq 1+1/p}\Big|det(A_{p,m}(z))\Big| \geq Cm^{-1/2}.$$

To state the next lemma we first fix the following notation. $\Psi_j(m)$, $\Psi_{j,p}(m)$, $\widetilde{\Psi}_{j,p}(m)$ and $\widehat{\Psi}_{j,p}(m)$ $j = 1, 2, \ldots$ denote the coefficient matrices in the power series expansions of $A_m^{-1}(z)$, $A_{p,m}^{-1}(z)$, $\widetilde{A}_{p,m}^{-1}(z)$ and $\widehat{A}_{p,m}^{-1}(z)$, respectively, $|z| \leq 1$. We set $\Psi_0(m) = \Psi_{0,p}(m) = \widetilde{\Psi}_{0,p}(m) = \widehat{\Psi}_{0,p}(m) = I_m$. Furthermore, $e_t(m) = \xi_t - \sum_{j=1}^{\infty} A_j(m)\xi_{t-j}$, $e_{t,p}(m) = \xi_t - \sum_{j=1}^{p} A_{j,p}(m)\xi_{t-j}$, $\widetilde{e}_{t,p}(m) = \xi_t - \sum_{j=1}^{p} \widetilde{A}_{j,p}(m)\xi_{t-j}$ and $\widehat{e}_{t,p}(m) = \widehat{\xi}_t - \sum_{j=1}^{p} \widehat{A}_{j,p}(m)\widehat{\xi}_{t-j}$, while $\widetilde{\Sigma}_{e,p}(m) = E^+(\widetilde{e}_{t,p}(m) - \overline{\widetilde{e}}_{n,p}(m))(\widetilde{e}_{t,p}(m) - \overline{\widetilde{e}}_{n,p}(m))^\top$ and $\widehat{\Sigma}_{e,p}(m) = E^*(\widehat{e}_{t,p}(m) - \overline{\widehat{e}}_{n,p}(m))(\widehat{e}_{t,p}(m) - \overline{\widehat{e}}_{n,p}(m))^\top$ with $\overline{\widetilde{e}}_{n,p}(m) = (n-p)^{-1}\sum_{t=p+1}^{n}\widetilde{e}_{t,p}(m)$ and $\overline{\widehat{e}}_{n,p}(m) = (n-p)^{-1}\sum_{t=p+1}^{n}\widehat{e}_{t,p}(m)$, where $E^+$ denotes expectation with respect to the measure assigning probability $(n-p)^{-1}$ to each $\widetilde{e}_{t,p}(m)$, $t = p + 1, p + 2, \ldots, n$.

LEMMA 6.5. *Let Assumptions 1 and 3 and Assumption 2 and 4 (r=2) be satisfied. Then, as $n \to \infty$,*

(i)  $\sum_{j=1}^{\infty} \|\widetilde{\Psi}_{j,p}(m) - \Psi_{j,p}(m)\|_F \xrightarrow{P} 0$,

(ii)  $\|\widetilde{\Sigma}_{e,p}(m) - \Sigma_{e,p}(m)\|_F \xrightarrow{P} 0$,

(iii) $\sum_{j=1}^{\infty} \|\widehat{\Psi}_{j,p}(m) - \Psi_{j,p}(m)\|_F \xrightarrow{P} 0,$

(iv) $\|\widehat{\Sigma}_{e,p}(m) - \Sigma_{e,p}(m)\|_F \xrightarrow{P} 0,$

(v) $\sum_{j=1}^{\infty} \|\Psi_{j,p}(m) - \Psi_j(m)\|_F \to 0,$

(vi) $\|\Sigma_{e,p}(m) - \Sigma_e(m)\|_F \to 0.$

**Proof of Lemma 2.1:** Expression (2.2) imediately leads, for all $\omega \in [0, \pi]$, to an upper bound of $f_{\xi^{(M)}}(\omega)$. To derive a lower bound, recall that $\Gamma_{\xi^{(M)}}(h) = (\langle C_h(v_{j_r}), v_{j_s} \rangle)_{r,s=1,2,\ldots,m}$ and observe that

$$f_{\xi^{(M)}}(\omega) = \left( \langle \mathcal{F}_\omega(v_{j_r}), v_{j_s} \rangle \right)_{r,s=1,2,\ldots,m}.$$

Let $\mu_j(\omega)$, $j = 1, 2, \ldots, m$, be the eigenvalues of $f_{\xi^{(M)}}(\omega)$ (including multiplicity). It suffices to show that $\min_{1 \le j \le m} \mu_j(\omega) \ge \delta_M > 0$ for all frequencies $\omega \in [0, \pi]$. For this let $c_j(\omega) = (c_{j,1}(\omega), c_{j,2}(\omega), \ldots, c_{j,m}(\omega))^\top \in \mathbb{C}^m$, $j = 1, 2, \ldots, m$, be the corresponding normalized eigenvectors. Then for every $j \in \{1, 2, \ldots, m\}$, we have

$$\mu_j(\omega) = c_j^\top(\omega) \left( \langle \mathcal{F}_\omega(v_{j_r}), v_{j_s} \rangle \right)_{r,s=1,2,\ldots,m} c_j(\omega)$$
$$= \langle \mathcal{F}_\omega(y_j(\omega)), y_j(\omega) \rangle > 0,$$

by the positivity of $\mathcal{F}_\omega$, where $y_j(\omega) = \sum_{r=1}^{m} c_{j,r}(\omega) v_{j_r} \in \overline{V}_M = \overline{sp}\{v_{j_1}, v_{j_2}, \ldots, v_{j_m}\}$ and $\|y_j\| = 1$. Because of the norm summability of the autocovariance matrix function $\Gamma_{\xi^M}(h)$, the spectral density $f_{\xi^{(M)}}(\omega)$ and consequently the eigenvalues $\mu_j(\omega)$, $j = 1, 2, \ldots, m$, are continuous functions of $\omega$. Let $\delta_M(\omega) = \min_{1 \le j \le m} \mu_j(\omega)$ and notice that $\delta_M(\omega)$ is continuous in $\omega$ and $\delta_M(\omega) > 0$ for all $\omega \in [0, \pi]$. Define $\delta_M = \min_{\omega \in [0,\pi]} \delta_M(\omega)$ which is positive by the continuity of $\delta_M(\cdot)$ in the compact interval $[0, \pi]$. Hence $\min_{1 \le j \le m} \mu_j(\omega) \ge \delta_M > 0$ for all $\omega \in [0, \pi]$. $\square$

**Proof of Proposition 3.1:** Recall the definition of $X_t^* = \sum_{j=1}^{m} \mathbf{1}_j^\top \xi_t^* \widehat{v}_j + U_t^*$ and observe that $\xi_t^* = \sum_{l=0}^{\infty} \widehat{\Psi}_{l,p}(m) e_{t-l}^*$, where $\widehat{\Psi}_{0,p}(m) = I_m$ and the power series $\widehat{\Psi}_{m,p}(z) = I_m + \sum_{l=1}^{\infty} \widehat{\Psi}_{l,p}(m) z^l = (I_m - \sum_{j=1}^{p} \widehat{A}_{j,p}(m) z^j)^{-1}$ converges for $|z| \le 1$. Write $X_t^* = \sum_{l=0}^{\infty} \sum_{j=1}^{m} \mathbf{1}_j^\top \widehat{\Psi}_{l,p}(m) e_{t-l}^* \widehat{v}_j + U_t^*$ and define $X_{t,M}^* = \sum_{l=0}^{M-1} \sum_{j=1}^{m} \mathbf{1}_j^\top \widehat{\Psi}_{l,p}(m) e_{t-l}^* \widehat{v}_j + \sum_{l=M}^{\infty} \sum_{j=1}^{m} \mathbf{1}_j^\top \widehat{\Psi}_{l,p}(m) e_{t-l,t}^* \widehat{v}_j + U_t^*$, where for each $t \in \mathbb{Z}$, $\{e_{s,t}^*, s \in \mathbb{Z}\}$ is an independent copy of $\{e_s^*, s \in \mathbb{Z}\}$. Notice that $X_M^* - X_{M,M}^* = \sum_{l=M}^{\infty} \sum_{j=1}^{m} \mathbf{1}_j^\top \widehat{\Psi}_{l,p}(m)(e_{M-l}^* - e_{M-l,M}^*) \widehat{v}_j$. By

Minkowski's inequality we have

$$\sqrt{E\|X_M^* - X_{M,M}^*\|^2} \le \sqrt{E\|\sum_{l=M}^{\infty}\sum_{j=1}^{m} \mathbf{1}_j^\top \widehat{\Psi}_{l,p}(m)e_{M-l}^*\widehat{v}_j\|^2}$$

$$\text{(6.1)} \qquad\qquad\qquad + \sqrt{E\|\sum_{l=M}^{\infty}\sum_{j=1}^{m} \mathbf{1}_j^\top \widehat{\Psi}_{l,p}(m)e_{M-l,M}^*\widehat{v}_j\|^2}.$$

Evaluating the first expectation term we get using $\|A\|_F^2 = tr(AA^\top)$ and the submultiplicative property of the Frobenius matrix norm, that

$$E\|\sum_{l=M}^{\infty}\sum_{j=1}^{m} \mathbf{1}_j^\top \widehat{\Psi}_{l,p}(m)e_{M-l}^*\widehat{v}_j\|^2 = \sum_{l=M}^{\infty} tr\big(\widehat{\Psi}_{l,p}(m)\Sigma^*(m)\widehat{\Psi}_{l,p}^\top(m)\big)$$

$$\le \|\widehat{\Sigma}_{e,p}^{1/2}(m)\|_F^2 \sum_{l=M}^{\infty} \|\widehat{\Psi}_{l,p}(m)\|_F^2,$$

where $\widehat{\Sigma}_{e,p}(m) = \widehat{\Sigma}_{e,p}^{1/2}(m)\widehat{\Sigma}_{e,p}^{1/2}(m)$. An identical expression appears for the second expectation term on the right hand side of (6.1). Applying Minkowski's inequality again we get by the exponential decay of $\|\widehat{\Psi}_{l,p}(m)\|_F$, that

$$\sum_{M=1}^{\infty} \sqrt{E\|X_M^* - X_{M,M}^*\|^2} \le 2\|\widehat{\Sigma}_{e,p}^{1/2}(m)\|_F \sum_{M=1}^{\infty}\sum_{l=M}^{\infty} \|\widehat{\Psi}_{l,p}(m)\|_F$$

$$= 2\|\widehat{\Sigma}_{e,p}^{1/2}(m)\|_F \sum_{l=1}^{\infty} l\|\widehat{\Psi}_{l,p}(m)\|_F = O_P(1).$$

$\square$

**Proof of Theorem 4.1** Let

$$L_{n,m}^+ = \frac{1}{\sqrt{n}} \sum_{t=1}^{n}\sum_{j=1}^{m} \xi_{j,t}^+ v_j e^{-it\omega},$$

where $\xi_t^+ = (\xi_{1,t}^+, \xi_{2,t}^+, \ldots, \xi_{m,t}^+)^\top$, $t = 1, 2, \ldots, n$ with $\xi_t^+ = \sum_{j=1}^{p} \widetilde{A}_{j,p}(m)\xi_{t-j}^+ + e_t^+$, where $\widetilde{A}_{j,p}(m)$, $j = 1, 2, \ldots, p$ are the estimators of the autoregressive parameter matrices based on the vector time series of true scores $\xi_t$, $t = 1, 2, \ldots, n$ and $e_t^+$ are obtained by i.i.d. resampling from the centered residuals $\widehat{e}_t = \xi_t - \sum_{j=1}^{p} \widetilde{A}_{j,p}(m)\xi_{t-j}$, $t = p+1, p+2, \ldots, n$. That is, the pseudo-variable $L_{n,m}^+$ is obtained using the true eingefunctions $v_j$ and the

true scores $\xi_{j,t}$ instead of their estimates $\widehat{v}_j$ and $\widehat{\xi}_{j,t}$ respectively. Decompose $n^{-1/2}S_n^*(\omega)$ as

$$
n^{-1/2}S_n^*(\omega) = \frac{1}{\sqrt{n}}\sum_{t=1}^n\sum_{j=1}^m \xi_{j,t}^+ v_j e^{-it\omega} + \frac{1}{\sqrt{n}}\sum_{t=1}^n\sum_{j=1}^m \xi_{j,t}^*(\widehat{v}_j - v_j)e^{-it\omega}
$$
$$
+ \frac{1}{\sqrt{n}}\sum_{t=1}^n\sum_{j=1}^m (\xi_{j,t}^* - \xi_{j,t}^+)v_j e^{-it\omega} + \frac{1}{\sqrt{n}}\sum_{t=1}^n U_{t,m}^* e^{-it\omega}
$$
$$
= L_{n,m}^+ + V_{n,m}^* + D_{n,m}^* + R_{n,m}^*
$$

with an obvious notation for $L_{n,m}^+$, $V_{n,m}^*$, $D_{n,m}^*$ and $R_{n,m}^*$. Notice that the terms $V_{n,m}^*$ and $D_{n,m}^*$ are due to the fact that, in the bootstrap procedure, the unknown scores and eigenfunctions are replaced by their sample estimates, while $R_{n,m}^*$ is due to the $m$-dimensional approximation of the infinite dimensional structure of the underlying process. Assertion (i) of the theorem follows then from Lemma 6.6, 6.7, 6.8 and 6.9 and Slutsky's theorem.

Consider assertion (ii). Since

$$
n^{-1}\|E^* S_n^*(\omega) \otimes S_n^*(\omega) - E S_n(\omega) \otimes S(\omega)\|_{HS}
$$
$$
\leq \|n^{-1}E^* S_n^*(\omega) \otimes S_n^*(\omega) - 2\pi\mathcal{F}_{\omega,m}^*\|_{HS}
$$
$$
+ 2\pi\|\mathcal{F}_{\omega,m}^* - \mathcal{F}_\omega\|_{HS} + \|n^{-1}E S_n(\omega) \otimes S_n - 2\pi\mathcal{F}_\omega\|_{HS},
$$

it suffices in view of Proposition 3.2 and Theorem 2 of Cerovecki and Hörmann (2015), to show that the first term on the right hand side of the above inequality converges to zero in probability. For this we have using $n^{-1}E^* S_n^*(\omega)\otimes S_n^*(\omega) = n^{-1}\sum_{-n+1}^{n-1}(1 - |h|/n)C_h^*$, that this term is bounded by

$$
\sum_{|h|\geq n} \|C_h^*\|_{HS} + n^{-1}\sum_{h=-n+1}^{n-1} |h|\|C_h^*\|_{HS}.
$$

Now, since $\sum_{h\in\mathbb{Z}} \|C_h^*\|_{HS} = O_P(1)$ uniformly in $p$ and $m$, we get that $\sum_{|h|\geq n} \|C_h^*\|_{HS} = o_P(1)$ and by Kronecker's lemma that $n^{-1}\sum_{h=-n+1}^{n-1} |h| \|C_h^*\|_{HS} = o_P(1)$. To verify the uniform boundeness of $\sum_{h\in\mathbb{Z}} \|C_h^*\|_{HS}$, notice first that from the expression of $C_h^*$ given in Section 3.2 we get that $\sum_{h\in\mathbb{Z}} \|C_h^*\|_{HS} \leq \sum_{h\in\mathbb{Z}} \|\Gamma_h^*\|_F + \|C_U^*\|_{HS}$. The square of the second term on the right hand side of the last inequality equals $\|E^* U_t^* \otimes U_t^*\|_{HS}^2$ which converges to zero in probability, see the proof of Proposition 3.2. For the first term we have that $\sum_{h\in\mathbb{Z}} \|\Gamma_h^*\|_F \leq \big(\sum_{j=0}^\infty \|\widehat{\Psi}_{j,p}(m)\|_F\big)^2 \|\widehat{\Sigma}_{e,p}(m)\|_F = O_P(1)$ uniformly in $p$ and $m$ by Lemma 6.1 and Lemma 6.5. $\qquad\square$

LEMMA 6.6.  *Under the assumptions of Theorem 4.1 it holds true that,*
$R_{n,m}^* \xrightarrow{P} 0$, *as $n \to \infty$.*

LEMMA 6.7.  *Under the assumptions of Theorem 4.1 it holds true that,*
$D_{n,m}^* \xrightarrow{P} 0$, *as $n \to \infty$.*

LEMMA 6.8.  *Under the assumptions of Theorem 4.1 it holds true that,*
$V_{n,m}^* \xrightarrow{P} 0$, *as $n \to \infty$.*

LEMMA 6.9.  *Under the assumptions of Theorem 4.1 it holds true that,*
*for all $\omega \in [-\pi, \pi]$ and as $n \to \infty$,*

$$L_{n,m}^+(\omega) \Rightarrow NC(0, 2\pi \mathcal{F}_\omega),$$

*in probability.*

**Acknowledgements.**  The author thanks the Editor, the Associate Editor and two referees for their careful reading and thoughtful comments and questions that led to an improved version of the paper.

## SUPPLEMENTARY MATERIAL

**Online Supplement: "Sieve Bootstrap for Functional Time Series".** The online supplement contains the proofs that were omitted in this paper and additional numerical results.

**References.**

[1] ANEIROS-PÉREZ, G., CAO, R. and VILAR-FERNANÉZ, J. M. (2011). Functional methods for time series prediction: A nonparametric approach. *Journal of Forecasting*, **30**, 377-392.

[2] AUE, A., Dubart, D. N. and HÖRMANN, S. (2015). On the prediction of stationary functional time series. *Journal of the American Statistical association*, **110**, 378-392.

[3] BOSQ, D. (2000). *Linear Process in Function Spaces*. Springer, Berlin-Heidelberg-New York.

[4] BROCKWELL, P. and DAVIS, R. (11991). *Time Series: Theory an Methods*. Springer, Berlin-Heidelberg-New York.

[5] CEROVECKI, C. and HÖRMANN, S. (2017). On the CLT for discrete Fourier transforms of functional time series. *Journal of Multivariate Analysis*, **154**, 282-295.

[6] CHENG, R. and POURAHMADI, M. (1993). Baxter's inequality and convergence of finite predictors of multivariate stochastic processes. *Probability Theory and Related Fields*, **95**, 115-124.

[7] DEHLING, H., SHARIPOV, O. S. and WENDLER, M. (2015). Bootstrap for dependent Hilbert space valued random variables with application to von Mises statistics. *Journal of Multivariate Analysis*, **133**, 200-215.

[8] FERNÁNDEZ DE CASTRO, B., GUILLAS, S. and GONZÁLEZ MANTEIGA, W. (2005). Functional samples and bootstrap for predicting sulfur dioxide levels. *Technometrics*, **47**, 212-222.

[9] FRANKE, J. and NYARIGUE, E. (2016). Residual-based bootstrap for functional autoregressions. *Preprint*.

[10] HÖRMANN, S. and KOKOSZKA, P. (2010). Weakly dependent functional data. *Annals of Statistics*, **38**, 1845-1884.

[11] HÖRMANN, S. and KOKOSZKA, P. (2012). Functional time series. *Handbook of Statistics: Time Series Analysis-Methods and Applications*, 157-186.

[12] HÖRMANN, S. and KIDZIŃSKI, L. (2015). A note on estimation in Hilbertian linear models. *Scandinavian Journal of Statistics*, **42**, 43-62.

[13] HÖRMANN, S., KIDZIŃSKI, L. and HALLIN, M. (2015). Dynamic functional principal components. *Journal of the Royal Statistical Society: Series B*, **77**, 319-348.

[14] HORVÁTH, L. and KOKOSZKA, P. (2012). *Inference for Functional Data with Applications*. Springer, Berlin-Heidelberg-New York.

[15] HORVÁTH, L., KOKOSZKA, P. and REEDER, R. (2013). Estimation of the Mean of Functional Time Series and a Two Sample Problem. *Journal of the Royal Statistical Society: Series B*, **75**, 103-122.

[16] HORVÁTH, L., KOKOSZKA, P. and RICE, G. (2014). Testing Stationarity of Functional Time Series. *Journal of Econometric*, **179**, 66-82.

[17] HURVICH, C. M. and TSAI, C.-L. (1993). A Corrected Akaike Information Criterion for Vector Autoregressive Model Selection. *Journal of Time Series Analysis*, **14**, 271-279.

[18] HYNDMAN, R. J. and SHANG, H. L. (2009). Forecasting functional time series. *Journal of the Korean Statistical Society*, **38**, 199-211.

[19] KREISS, J.-P. (1988). Asymptotic Statistical Inference for a Class of Stochastic Processes. Habilationsschrift. Universität Hamburg.

[20] KREISS, J.-P., PAPARODITIS, E. and POLITIS, D. N. (2011). On the range of validity of the autoregressive sieve bootstrap. *Annals of Statistics*, **39**, 2103-2130.

[21] KREISS, J.-P. and PAPARODITIS, E. (2011). Bootstrap Methods for Dependent Data: A Review. *Journal of the Korean Statistical Society*, **40**, 357-378.

[22] LAHIRI, S. N. (2003). *Resampling Methods for Dependent Data*. Springer, Berlin-Heidelberg-New York.

[23] LI, Y., WANG, N. and CAROROL, R. J. (2013). Selecting the number of principal components in functional data. *Journal of the American Statistical Association*, **108**, 1284-1294.

[24] McMURRY, T. and POLITIS, D. N. (2011). Resampling methods for functional data. *The Oxford Handbook of Functional Data Analysis*, (F. Ferraty and Y. Romain, eds.) 189-209. Oxford University Press.

[25] MERLEVEDE, F., PELIGRAD, M. and UTEV, S. (1997). Sharp conditions for the CLT of linear processes in a Hilbert -space. *Journal of Theoretical Probability*, **10**, 681-693.

[26] MEYER, M. and KREISS, J.-P. (2015). On the vector autoregressive sieve bootstrap. *Journal of Time Series Analysis*, **36**, 377-397.

[27] MEYER, M., JENTSCH, C., and KREISS, J.-P. (2016). Baxter's inequality and sieve bootstrap for random fields. *Bernoulli*, to appear.

[28] MINGOTTI, N., LILLO, R. E. and ROMO, J. (2015). A random walk test for functional time series. UC3M Working papers, Statistics and Econometrics.

[29] PANARETOS, V. and TAVAKOLI, S. (2013). Fourier analysis of stationary time series in function spaces. *Annals of Statistics*, **41**, 568-603.

[30] POLITIS, D. N. and ROMANO, J. (1994). Limit theorems for weakly dependent Hilbert space valued random variables with applications to the stationary bootstrap. *Statistica Sinica*, **4**, 461-476.

[31] POURAHMADI, M. (2001). *Foundation of Time Series Analysis and Prediction Theory*. John-Wiley: New York.

[32] SHARIPOV, O., TEWES, J. and WENDLER, M. (2016). Sequential block bootstrap in a Hilbert space with application to change point analysis. *The Canadian Journal of Statistics*, forthcoming.

[33] YAO, F., MÜLLER, H. G. and WANG, J. L. (2005). Functional Data Analysis for Sparse Longitudinal Data. *Journal of the American Statistical Association*, **100**, 577-590.

[34] WIENER, N. and MASANI, P. (1958). The prediction theory of multivariate stochastic processes, II. *Acta Mathematica*, **99**, 93-137.

[35] SHANG, L. H. (2016). Resampling Methods for Dependent Functional Data. *Preprint*.

[36] ZHOU, T. and POLITIS, D. N. (2016). Kernel Estimation of First-Order Nonparametric Functional Autoregression Model and its Bootstrap Approximation. *Preprint*.

UNIVERISTY OF CYPRUS
DEPARTMENT OF MATHEMATICS AND STATISTICS
P.O.Box 20537
CY-1678 NICOSIA
CYPRUS
E-mail: stathisp@ucy.ac.cy