Statistics Surveys
Vol. 17 (2023) 119–168
ISSN: 1935-7516
https://doi.org/10.1214/23-SS143

White noise testing for functional time series

Mihyun Kim

West Virginia University
e-mail: mihyun.kim@mail.wvu.edu

Piotr Kokoszka

Colorado State University
e-mail: Piotr.Kokoszka@colostate.edu

Gregory Rice

University of Waterloo
e-mail: grice@uwaterloo.ca

Abstract: We review white noise tests in the context of functional time series, and compare many of them using a custom developed `R` package `wwntests`. The tests are categorized based on whether they are conducted in the time domain or spectral domain, and whether they are valid for i.i.d. or general uncorrelated noise. We also review and extend several residual-based goodness-of-fit tests of popular models used in functional data analysis. Through numerous simulation experiments and a data application, we demonstrate the use of these tests, and are able to provide practical guidance on their implementation, benefits, and drawbacks.

Received August 2022.

Contents

1. Introduction

The objective of this paper is to review, compare, and extend white noise tests in the context of Functional Time Series (FTS). White noise tests are critical to many aspects of analyzing a time series, mainly through their use in exploring the serial dependence structure of the series and performing time series model diagnositic checks, and the literature surrounding their application to FTS has grown steadily over the last fifteen years. We refer the reader to the text books of Li (2004) and Francq and Zakoian (2010) for reviews of white noise tests for scalar and vector valued time series.

As in the analysis of univariate or multivariate time series, white noise tests for FTS can be categorized into two groups: the time domain approach based on autocovariance operators, and the spectral domain approach based on spectral density operators. The tests can further be categorized based on type of the null hypothesis considered. Either the hypothesis of i.i.d. white noise can be tested, denoted by $\mathcal{H}_0$-IID, or the hypothesis of uncorrelated white noise can be tested, denoted by $\mathcal{H}_0$-WN. Precise definitions will be given in the following. At this point, it is informative to list the tests surveyed in this paper in Table 1.1. The tests with an asterisk are compared in a simulation study conducted in Section 6.

A common technique to evaluate the goodness-of-fit of an FTS model is to apply white noise tests to model residuals. Since, even if the FTS model is well specified, model residuals typically share a common dependence on estimators of the model parameters, white noise tests applied to residuals often must be adjusted to account for this. This paper also surveys and compares via simulation available goodness-of-fit tests for FTS models. We also provide a new

contribution in this setting, which extends, both via a new theory and an implementation, the test of Kokoszka *et al.* (2017) to testing goodness-of-fit of a functional autoregressive model, the FAR(1). This is motivated by the comparatively good performance of this extension, and since it provides a time domain analog of a spectral domain test derived by Zhang (2016).

All tests we compare are implemented in the companion `R` package called `wwntests`, which is available on the Comprehensive R Archive Network (CRAN).

TABLE 1.1
White noise tests for FTS considered in this paper. The tests marked with an asterisk (*) are compared in a simulation study in Section 6.

Null type	Time domain	Spectral domain
$\mathcal{H}_0$-IID	Gabrys and Kokoszka (2007)* Horváth *et al.* (2013) Rice *et al.* (2020)	Characiejus and Rice (2020)* Hlávka *et al.* (2021)
$\mathcal{H}_0$-WN	Kokoszka *et al.* (2017)* Bücher *et al.* (2023)	Zhang (2016)* Bagchi *et al.* (2018)

The paper is organized as follows. Section 2 is dedicated to the review of the most relevant concepts related to FTS. In particular, we define the two white noise null hypotheses. Time domain tests are reviewed in Section 3, and spectral domain tests are discussed in Section 4. Section 5 contains reviews of goodness-of-fit tests for FTS models. In Section 6, we compare the performance of selected tests and also present a data example that illustrates their applicability.

The paper contains a comprehensive appendix. The information presented in it is not essential to learn about white noise testing for FTS, but it is useful. In the analysis of scalar time series, white noise tests are typically accompanied by normality tests, most commonly normal QQ-plots. Testing for periodicity and change points is also very common. Such tests in the context of FTS are reviewed in Section A. Section B is dedicated to the derivation, asymptotic justification and practical implementation of a time domain goodness-of-fit test of the FAR(1) model. Section C presents additional simulation results that provide further justification for specific conclusions reported in the main text.

2. Functional data analysis background

In this section, we review the concepts of Functional Data Analysis (FDA) that are essential to understand this paper. We also introduce the notation and terminology used in the remainder of the paper. All concepts introduced in this section are explained in much greater depth in Ramsay and Silverman (2005), Horváth and Kokoszka (2012) and Hsing and Eubank (2015), as well as in the introductory text of Kokoszka and Reimherr (2017). We begin by defining the null hypotheses.

2.1. White noise null hypotheses

Suppose $X_1, X_2, \ldots, X_N$ are a sample of functional observations. In FDA, the observations are most commonly considered as elements of the function space

$L^2 = L^2(\mathcal{T})$, where $\mathcal{T}$ is some domain, a compact interval being the most common case. From a theoretical perspective we consider the observed functions as representatives of equivalence classes of almost every equal functions in L^2, and view L^2 as a separable Hilbert space. The inner product in L^2 is defined by $\langle x, y\rangle = \int x(t)y(t)dt$, and it generates the norm $\|x\| = \sqrt{\langle x, x\rangle}$, where $\int = \int_{\mathcal{T}}$. If $X_1, X_2, \ldots, X_N$ are identically distributed functions in L^2 with the same distribution as X, then their common mean function can be defined pointwise $\mu(t) = EX(t)$, and it coincides with the Bochner and Pettis integrals in L^2 under the assumption $E\|X\| < \infty$.

We are interested in testing the following two null hypotheses:

$\mathcal{H}_0$-IID: the X_i are independent and identically distributed elements of L^2;

$\mathcal{H}_0$-WN: the X_i form a white noise sequence in L^2 according to Definition 2.1.

Definition 2.1. A sequence $\{X_i, -\infty < i < \infty\}$ of L^2 random elements is a (mean zero) white noise if

$$0 < E\|X_i\|^2 < \infty, \quad EX_i(t) = 0 \tag{2.1}$$

and for any $h \neq 0$, and any i,

$$E\left[X_i(t)X_{i+h}(s)\right] = 0, \quad \text{for almost every } t, s \in \mathcal{T}. \tag{2.2}$$

We denote $E\|X_i\|^2 = \sigma^2$.

Specific tests discussed in the following generally impose additional assumptions on the data under the null hypothesis, but the dichotomy reflected in our definitions of $\mathcal{H}_0$-IID and $\mathcal{H}_0$-WN is the central aspect, similarly as in traditional time series analysis. We note that $E\left[X_i(t)X_{i+h}(s)\right]$ is the autocovariance function or kernel of $\{X_i\}$, whose general definition with nonzero mean will be introduced in Section 2.3. Definition 2.1 is analogous to the definition of white noise for a multivariate time series: to be white noise random vectors must have zero mean and their autocovariance matrices at nonzero lags must vanish. Also, note that, under (2.1), condition (2.2) is equivalent to the condition

$$E\left[\langle X_i, x\rangle \langle X_{i+h}, y\rangle\right] = 0, \quad \text{for any } x, y \in L^2,$$

which appears in Definition 3.1 of Bosq (2000). This follows from the identity

$$E\left[\langle X_i, x\rangle \langle X_{i+h}, y\rangle\right] = \iint E\left[X_i(t)X_{i+h}(s)\right] x(t)x(s)dtds.$$

The broader null $\mathcal{H}_0$-WN contains functional processes that are uncorrelated, but possibly dependent; for example, fARCH models introduced by Hörmann *et al.* (2013) or fGARCH models of Aue *et al.* (2017) and Cerovecki *et al.* (2019). Therefore, it is important to make the distinction between $\mathcal{H}_0$-IID and $\mathcal{H}_0$-WN in the context of model diagnostic checking.

2.2. Functional principal components

Functional principal components (FPCs) are used in many procedures discussed in this paper, so it is convenient to display formulas that can be readily referred to.

FPCs are defined under the condition $E\|X\|^2 < \infty$, which we assume throughout this paper. The covariance operator of X is defined by

$$C(x) := E[\langle X - \mu, x\rangle (X - \mu)], \quad x \in L^2, \tag{2.3}$$

where $\mu = EX$. The eigenfunctions of C, denoted by $v_j, j \geq 1$, are the FPCs, i.e. $C(v_j) = \lambda_j v_j$. The FPCs lead to the Karhunen–Loéve expansion

$$X_i(t) = \mu(t) + \sum_{j=1}^{\infty} \xi_{ij} v_j(t), \quad \xi_{ij} = \langle X_i - \mu, v_j\rangle . \tag{2.4}$$

The *scores* ξ_{ij} satisfy $E\xi_{ij} = 0$, $E\xi_{ij}^2 = \lambda_j$, $E[\xi_{ij}\xi_{ij'}] = 0$ for $j \neq j'$. Expansion (2.4) is not directly accessible because μ and the v_j are unknown population parameters. The mean function μ is most commonly estimated by the average $\bar{X}_N = N^{-1}\sum_{i=1}^{N} X_i$. The FPCs v_j and the eigenvalues λ_j are estimated by $\hat{v}_j$ and $\hat{\lambda}_j$ defined as the solutions to the equations

$$\widehat{C}(\hat{v}_j)(t) = \hat{\lambda}_j \hat{v}_j(t), \quad t \in \mathcal{T}, \quad 1 \leq j \leq N, \tag{2.5}$$

where $\widehat{C}$ is the sample covariance operator defined by

$$\widehat{C}(x)(t) = \frac{1}{N}\sum_{i=1}^{N} \left\langle X_i - \bar{X}_N, x\right\rangle (X_i - \bar{X}_N)(t), \quad x \in L^2.$$

Each curve X_i can then be approximated by a linear combination of a finite set of the estimated FPCs $\hat{v}_j$, i.e. $X_i(t) \approx \bar{X}_N + \sum_{j=1}^{p} \hat{\xi}_{ij}\hat{v}_j(t)$, where the $\hat{\xi}_{ij} = \left\langle X_i - \bar{X}_N, \hat{v}_j\right\rangle$ are the sample scores. Each $\hat{\xi}_{ij}$ quantifies the contribution of the curve $\hat{v}_j$ to the shape of the curve X_i. Thus, the vector of the sample scores, $[\hat{\xi}_{i1}, \hat{\xi}_{i2}, \ldots, \hat{\xi}_{ip}]^\top$, encodes the shape of X_i to a good approximation.

To illustrate, Figure 2.1 displays the first three sample FPCs, $\hat{v}_1, \hat{v}_2, \hat{v}_3$, for intraday return curves R_i for Walmart stock from July 05, 2006 to Dec 30, 2011. These curves are described in detail in Section 1.4 of Kokoszka and Reimherr (2017) and several papers. The curves R_i show how a return on an investment changes throughout a trading day, and its two examples are shown in Figure 2.2. The first estimated FPC $\hat{v}_1$ can be identified with a monotonic trend throughout the day. If the score corresponding to it is large, trading in this stock on a given day was dominated by a systematic increase (or decline if the score is negative) in the price of the stock. Notice the gradually decreasing slope of $\hat{v}_1$, which reflects the well-known fact that the most intense trading takes place after the opening of the trading floor. The second FPC, $\hat{v}_2$, has a large score if there is a significant reversal in investor sentiment during a given trading day. These observations are illustrated in Figure 2.2.

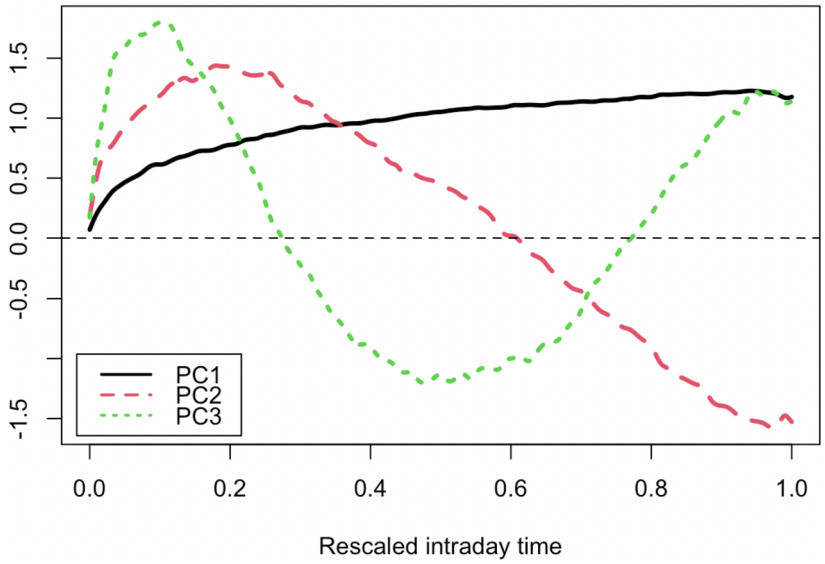

FIG 2.1. *The first three sample FPCs of intraday returns on Walmart stock based on sample of 1,378 curves.*

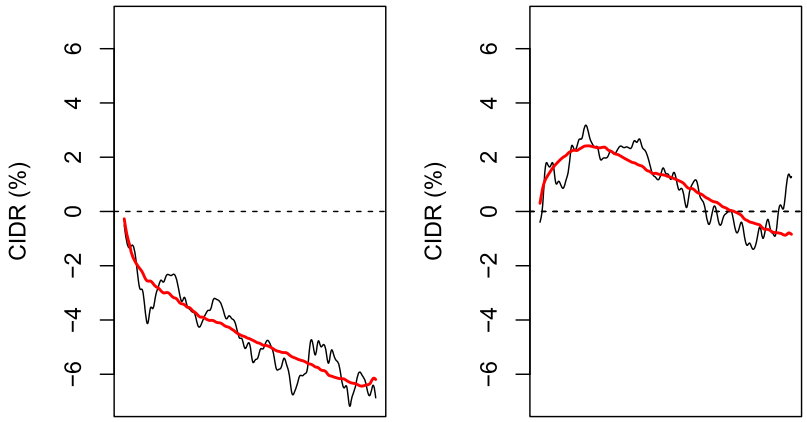

FIG 2.2. *Walmart intraday cumulative return curves on two trading days and their approximations by* $\sum_{i=1}^{3} \hat{\xi}_{ij}\hat{v}_j(t)$. *In the left panel,* $\hat{\xi}_1 = -4.7, \hat{\xi}_2 = 0.4, \hat{\xi}_3 = -0.1$, *observed on October 7, 2008. In the right panel,* $\hat{\xi}_1 = 0.8, \hat{\xi}_2 = 1.2, \hat{\xi}_3 = 0.1$, *observed on November 18, 2008.*

2.3. Autocovariance operators and spectral analysis

An integral operator Ψ with kernel $\psi(\cdot,\cdot)$ is defined by

$$\Psi(x)(t) = \int \psi(t,s)x(s)ds, \quad x \in L^2.$$

The Hilbert-Schmidt norm of Ψ is defined by

$$\|\Psi\|_{\mathcal{S}} = \left\{ \iint |\psi(t,s)|^2 dt ds \right\}^{1/2} \tag{2.6}$$

and the operator is called Hilbert–Schmidt, denoted $\Psi \in \mathcal{S}$, if it is finite.

Suppose that $\{X_i\}$ is a second-order stationary time series taking values in L^2, with the mean function $EX_i(t) = \mu(t)$. The autocovariance operator at lag h, $h \geq 0$, is defined by

$$\Gamma_h(x)(t) := \int \gamma_h(t,s)x(s)ds, \quad x \in L^2, \tag{2.7}$$

where $\gamma_h(t,s)$ is the autocovariance function defined by

$$\gamma_h(t,s) := E\left[(X_i(t) - \mu(t))(X_{i+h}(s) - \mu(s))\right], \quad t,s \in \mathcal{T}. \tag{2.8}$$

At lag zero, $\Gamma_0 = C$, where C is the covariance operator in (2.3). The functions $\gamma_h(\cdot,\cdot)$, $h \geq 0$, characterize the serial correlation in the series $\{X_i\}$. Given functional observations, $X_1, \ldots, X_N$, γ_h can be estimated using its sample counterpart

$$\hat{\gamma}_{N,h}(t,s) := \frac{1}{N} \sum_{i=1}^{N-h} (X_i(t) - \bar{X}_N(t))(X_{i+h}(s) - \bar{X}_N(s)), \quad 0 \leq h < N. \tag{2.9}$$

The operator Γ_h may then be estimated with $\hat{\Gamma}_h(x) = \int \hat{\gamma}_{N,h}(t,s)x(s)ds$. A simple graphical summary of the serial dependence in the series can be obtained by plotting

$$\hat{r}_h = \frac{\|\hat{\gamma}_{N,h}\|}{\int \hat{\gamma}_{N,0}(t,t)dt} \tag{2.10}$$

as a function of h, which we refer to as the *functional autocorrelation function* (fACF), and was studied in Mestre *et al.* (2021) and Kokoszka *et al.* (2017).

Panaretos and Tavakoli (2013b) introduced a frequency domain framework for describing the second–order structure of functional sequences. The spectral density function of $\{X_i\}$ is defined by

$$f_\omega(t,s) := \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} e^{-\mathbf{i}h\omega} \gamma_h(t,s), \quad t,s \in \mathcal{T}, \quad \omega \in [-\pi, \pi], \tag{2.11}$$

where $\mathbf{i} = \sqrt{-1}$ is the imaginary unit. If $\sum_{h \in \mathbb{Z}} \|\gamma_h\| < \infty$, the series in (2.11) is convergent in L^2. The spectral density operator is the integral operator with the kernel f_ω. Under the assumption $\sum_{h \in \mathbb{Z}} \|\Gamma_h\|_{\mathcal{S}} < \infty$, it is defined by

$$\mathcal{F}_\omega(x)(t) := \int f_\omega(t,s)x(s)ds = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} e^{-\mathbf{i}h\omega} \Gamma_h(x)(t), \quad x \in L^2. \tag{2.12}$$

A natural way of approximating the spectral density function f_ω is to use the functional periodogram

$$P_\omega(t,s) := \frac{1}{2\pi} \sum_{|h|<N} e^{-\mathbf{i}h\omega} \hat{\gamma}_{N,h}(t,s). \tag{2.13}$$

Note that the periodogram is not a consistent estimator of the spectral density even in the scalar case, see e.g., Section 10.3 of Brockwell and Davis (1991). To estimate f_ω, modifications of the periodogram, such as kernel lag–window type estimators, are typically used. They will be introduced in Section 4. A basically complete theory of the estimation of the spectral density operator under optimal conditions is developed in van Delft (2020).

3. Ljung-Box type tests

The goal of this section is to review time domain based white noise tests for FTS, including Gabrys and Kokoszka (2007), Horváth *et al.* (2013), Kokoszka *et al.* (2017), and Rice *et al.* (2020). These tests build on the ideas introduced by Box and Pierce (1970) and Ljung and Box (1978), both of which are developed in the context of traditional scalar time series. Therefore, to better understand the tests for FTS, we begin by describing the main idea of procedures in Box and Pierce (1970) and Ljung and Box (1978). Their tests are developed to check the adequacy of time series models by examining the residuals. The test statistics are based on the sum of the squares of sample autocorrelation functions (sample ACFs).

Suppose that $X_1, X_2, \ldots, X_N$ are real–valued observations, and denote their sample ACFs by $\hat{\rho}(h)$, $0 \leq h < N$. The test statistic considered by Box and Pierce (1970) is

$$\widehat{Q}_{BP} = N \sum_{h=1}^{H} \hat{\rho}^2(h), \tag{3.1}$$

where H is the number of lags. The statistic $\widehat{Q}_{BP}$ quantifies the serial correlation in the sequence up to lag H. Note that if the X_i are white noise, $\hat{\rho}(h)$ should be close to 0 for all $1 \leq h < N$. Thus, a large value of $\widehat{Q}_{BP}$ suggests that the observations might not be white noise. More specifically, it is known that the limiting distribution of the sample ACFs is multivariate normal with mean zero and identity covariance matrix, if the X_i are i.i.d. with finite fourth moment, see e.g., Theorem 7.2.2 in Brockwell and Davis (1991). Using this result, Box and Pierce (1970) note that under the i.i.d. assumption the limiting distribution of $\widehat{Q}_{BP}$ is chi–square with H degrees of freedom, χ^2_H. The i.i.d. assumption can thus be rejected at significance level α if $\widehat{Q}_{BP} > \chi^2_{1-\alpha}(H)$, where $\chi^2_{1-\alpha}(H)$ is the $(1-\alpha)$th percentile of χ^2_H. For diagnostic checking of time series models, Box and Pierce (1970) propose to use $\widehat{Q}_{BP}$ calculated from residuals. Denote the sample ACFs of the residuals from a model by $\tilde{\rho}(h)$. Since the residuals form a sequence of dependent random variables, the asymptotic

behavior of $\widehat{Q}_{BP}$ constructed by $\tilde{\rho}(h)$ is not trivial. To tackle this, Box and Pierce (1970) first establish the relationship between the residual autocorrelations $\widetilde{\rho} := [\tilde{\rho}(1), \ldots, \tilde{\rho}(H)]^\top$ and the population white noise autocorrelations $\widehat{\rho} := [\hat{\rho}(1), \ldots, \hat{\rho}(H)]^\top$; $\widetilde{\rho}$ is approximated by $(I_H - D)\widehat{\rho}$, where $I_H \in M_{H\times H}$ is the identity matrix, and $D \in M_{H\times H}$ is a matrix such that $I_H - D$ is an idempotent matrix of rank $H-d$, where d is the number of parameters to be estimated. Box and Pierce (1970) show that D depends on the model parameters, and its element d_{ij} goes to zero as i and/or j increase. From this, they argue that for sufficiently large H, the asymptotic behavior of $\widetilde{\rho}$ is close to that of $\widehat{\rho}$, so $\widehat{Q}_{BP}$ constructed from $\widetilde{\rho}$ follows a χ^2_{H-d} distribution. Box and Pierce (1970) justify their argument by means of a simulation study, but no rigorous derivation is presented.

Ljung and Box (1978) consider a statistic that modifies $\widehat{Q}_{BP}$:

$$\widehat{Q}_{LB} = N(N+2)\sum_{h=1}^{H} \hat{\rho}^2(h)/(N-h).$$

They compare the first and second moments of $\widehat{Q}_{LB}$ and $\widehat{Q}_{BP}$ and argue that $\widehat{Q}_{LB}$ follows the χ^2_H distribution more closely than $\widehat{Q}_{BP}$. For $\widehat{Q}_{LB}$ calculated from the residual autocorrelations $\widetilde{\rho}$, a similar argument is made; the asymptotic distribution of $\widehat{Q}_{LB}$ from $\widetilde{\rho}$ is closer to χ^2_{H-d} than $\widehat{Q}_{BP}$ computed from $\widetilde{\rho}$. This is supported by simulation results, but again, a rigorous mathematical justification is absent. The precise asymptotic distribution of $\widehat{Q}_{LB}$ when computed from ARMA model residuals is calculated in (McLeod, 1978). The Box–Ljung–Pierce approach was extended to multivariate time series settings, see Chitturi (1976), Hosking (1980) and Li and McLeod (1981).

3.1. Tests of $\mathcal{H}_0$-IID

Gabrys and Kokoszka (2007) introduce a test of independence and identical distribution for functional observations, which follows the Box–Ljung–Pierce paradigm. The procedure is based on the sum of the sample autocorrelation matrices computed from projections of FTS.

Suppose that $X_1, X_2, \ldots, X_N$ are a sample of functional observations in L^2. Gabrys and Kokoszka (2007) consider the null hypothesis $\mathcal{H}_0$-IID against $\mathcal{H}_A$: $\mathcal{H}_0$-IID does not hold. To test $\mathcal{H}_0$-IID, the key idea of constructing a single test statistic from infinite–dimensional functions is to obtain finite–dimensional projections of the observed functions. For dimension reduction, the FPCs introduced in Section 2.2 are employed. Recall that the unknown population scores $\xi_{ij} = \langle X_i - \mu, v_j\rangle$ in (2.4) are estimated by the sample scores $\hat{\xi}_{ij} = \langle X_i - \bar{X}_N, \hat{v}_j\rangle$, where the $\hat{v}_j$ are estimators of the FPCs v_j. We introduce the following score vectors:

$$\mathbf{Y}_i = [\xi_{i1}, \ldots, \xi_{ip}]^\top, \qquad \widehat{\mathbf{Y}}_i = [\hat{\xi}_{i1}, \ldots, \hat{\xi}_{ip}]^\top, \qquad i = 1, 2, \ldots, N.$$

Under $\mathcal{H}_0$-IID, the projections $\mathbf{Y}_i$ are i.i.d. multivariate random vectors with the mean zero vector and covariance matrix $\mathbf{V} = [v(k,l)]_{k,l=1,\ldots,p}$, where $v(k,l) = E[\xi_{ik}\xi_{il}]$. Thus, testing $\mathcal{H}_0$-IID now turns into testing i.i.d. assumption for the $\mathbf{Y}_i$.

Motivated by Chitturi (1976), Gabrys and Kokoszka (2007) define their test statistic as follows. Let $\mathbf{C}_h = [c_h(k,l)]_{k,l=1,\ldots,p}$ be the sample autocovariance matrix where

$$c_h(k,l) = \frac{1}{N}\sum_{i=1}^{N-h} \xi_{ik}\xi_{i+h,l}.$$

Denote by $\rho_{f,h}(k,l)$ the entries of matrix $\mathbf{C}_0^{-1}\mathbf{C}_h$ and by $\rho_{b,h}(k,l)$ the entries of matrix $\mathbf{C}_h\mathbf{C}_0^{-1}$. Then, they consider the random variable

$$\mathrm{GK}_{N,H} = \mathrm{GK}_{N,H}(p) = N\sum_{h=1}^{H}\sum_{k,l=1}^{p} \rho_{f,h}(k,l)\rho_{b,h}(k,l)$$

which is based on the population scores $\mathbf{Y}_i$. Note that $\mathrm{GK}_{N,H}$ cannot be computed because the FPCs v_j are unknown, so the ξ_{ij} cannot be computed from the data. Instead, they consider $\mathrm{GK}_{N,H}$ that is computed from the sample scores $\widehat{\mathbf{Y}}_i$. It can be shown that the distance between a population score and its approximation is asymptotically negligible under the existence of the fourth moment $E\|X_i\|^4 < \infty$. From this, Gabrys and Kokoszka (2007) establish that the test statistic $\mathrm{GK}_{N,H}$ computed from $\widehat{\mathbf{Y}}_i$ asymptotically follows a $\chi^2_{p^2H}$ distribution under $\mathcal{H}_0$-IID. They also consider one alternative, an FAR(1) model, see Section 5, which is one of the extensively used stationary models for FTS, and show consistency against the model under mild conditions. Zamani *et al.* (2019) put forward two methods to improve the finite sample properties of this test.

Horváth *et al.* (2013) develop a test procedure to test $\mathcal{H}_0$-IID, which does not involve projections. The test is based on the sample autocovariance functions obtained directly from functional observations. To display the test statistic, we first recall the sample autocovariance functions $\hat{\gamma}_{N,h}$ in (2.9). Horváth *et al.* (2013) considered the following test statistic

$$\widehat{V}_{N,H} = N\sum_{h=1}^{H(N)} \|\hat{\gamma}_{N,h}\|^2 = N\sum_{h=1}^{H(N)} \iint \hat{\gamma}^2_{N,h}(t,s)\, dtds. \tag{3.2}$$

The form of $\widehat{V}_{N,H}$ is an extension of $\widehat{Q}_{BP}$ and $\widehat{Q}_{LB}$ to a FTS setting; it is the sum of the L^2 norms of $\hat{\gamma}_{N,h}$ that quantifies the serial correlation in a sequence of functional observations.

To get a standard normal approximation, it is assumed in Horváth *et al.* (2013) that the number of lags $H(N)$ goes to infinity with the sample size N, in such a way that $H(N) = O((\log N)^\alpha)$ for some $\alpha > 0$. With this assumption and the condition $E\|X_i\|^4 < \infty$, Horváth *et al.* (2013) establish that the limit distribution of the scaled $\widehat{V}_{N,H}$ under $\mathcal{H}_0$-IID is the standard normal distribu-

tion;

$$\frac{1}{\sqrt{2H(N)\hat{\sigma}_N^2}}\left(\widehat{V}_{N,H} - H(N)\hat{\mu}_N\right) \xrightarrow{d} N(0,1),$$

where

$$\hat{\mu}_N = \left(\int \hat{\gamma}_{N,0}(t,t)\ dt\right)^2, \quad \hat{\sigma}_N^2 = \left(\iint \hat{\gamma}_{N,0}^2(t,s)\ dtds\right)^2.$$

For alternative, a broad class that contains FAR(1) and functional linear process models is considered:

$\mathcal{H}_A$: the X_i are stationary and ergodic sequence such that for some $h \geq 1$,

$$\iint \gamma_h^2(t,s)\ dtds > 0,$$

where γ_h is defined in (2.8). Under $\mathcal{H}_A$, the consistency of the test procedure is shown with the assumptions $H \to \infty$, $N/H^{3/2} \to \infty$, and $E\|X_i\|^2 < \infty$.

3.2. Tests of $\mathcal{H}_0$-WN

In this section, we review time domain tests of $\mathcal{H}_0$-WN. Functional GARCH–type models are included in $\mathcal{H}_0$-WN.

Kokoszka *et al.* (2017) consider the statistic $\widehat{V}_{N,H}$ in (3.2) and develop a testing procedure based on it that is applicable to test $\mathcal{H}_0$-WN. To introduce a distinction from $\widehat{V}_{N,H}$ developed for testing $\mathcal{H}_0$-IID, we define in the context of $\mathcal{H}_0$-WN testing

$$\mathrm{KRS}_{N,H} = N\sum_{h=1}^{H} \|\hat{\gamma}_{N,h}\|^2,$$

in which H is fixed and does not depend on N. The asymptotic distribution of $\mathrm{KRS}_{N,H}$ is obtained assuming $\mathcal{H}_0$-WN, which results in a more complex limiting distribution.

The null distribution of $\mathrm{KRS}_{N,H}$ is established under the assumption of L^4-m-approximability, see Definition B.1 in the appendix, and some first, second, and fourth order moment characteristics matching those of functional GARCH–type models. These conditions define a broad class of nonlinear FTS allowing for conditional heteroscedasticity. Kokoszka *et al.* (2017) show that the asymptotic distribution of $\mathrm{KRS}_{N,H}$ is a weighted sum of independent χ_1^2 random variables. It is non-pivotal, i.e., the distribution depends on the covariance structure of a process. To approximate it, the Welch–Satterthwaite approximation method is employed, which entails approximating this distribution with a scaled χ^2 distribution of the form $R = \beta\chi_\nu^2$, where β and ν are estimated so that the first two moments of R approximately match those of the limiting distribution. This is described in detail in Section 2.4 of Kokoszka *et al.* (2017). They also verify that the test statistic is consistent under the alternative of $\gamma_h \neq 0$, for some $0 < h \leq H$.

In the setting of locally stationary functional time series introduced in van Delft and Eichler (2018), Bücher *et al.* (2023) developed a general white noise test based on $\max_{h\in\{1,\ldots,H\}} N\|\hat{\gamma}_{N,h}\|^2$.

4. Spectral domain tests

In this section, we discuss spectral domain white noise tests for FTS. As in Section 3, we first review main concepts of frequency domain tests in the context of univariate time series and then extend them to the FTS context.

Spectral analysis provides an alternative way of characterizing the structure of a random process. A key tool of spectral analysis is the spectral density. Suppose that $\{X_i\}$ is a stationary, real–valued time series with its ACFs $\gamma(h)$ satisfying $\sum_{h=-\infty}^{\infty}|\gamma(h)|<\infty$. Then, the spectral density of $\{X_i\}$ is defined by

$$f(\omega)=\frac{1}{2\pi}\sum_{|h|<\infty} e^{-\mathbf{i}h\omega}\gamma(h), \qquad \omega\in[-\pi,\pi].$$

Note that the spectral density of a white noise does not depend on the frequency ω. Therefore, if a process is not white noise, its spectral density will deviate from a constant function on the frequency domain. Spectral domain tests are based on the spectral density of a process. In the context of scalar time series, spectral domain tests are developed to test various hypothesis, see e.g., Grenander and Rosenblatt (1953, 2008), Durbin and Brown (1967), Bartlett (1978), Durlauf (1991), and Hong (1996). We briefly review the approach of Hong (1996), which is extended to FTS by Characiejus and Rice (2020) and Bagchi *et al.* (2018).

Hong (1996) proposes goodness-of-fit tests for linear dynamic regression models. His test statistics are derived from the distance between a kernel–based spectral density estimator and the constant spectral density. To measure the distance, Hong (1996) considers three types of measures: the L^2 norm, the Hellinger metric, and the Kullback–Leibler information criterion. We describe only the L^2 norm statistic here since the statistics from the other two measures are obtained in a similar manner. Suppose that $X_1, X_2, \ldots, X_N$ are real–valued observations. Denote by $\hat{\gamma}(h)$, $0\le h<N$, the sample autocovariance functions of the residuals from fitting a linear dynamic regression model to the X_i. To estimate the spectral density of the errors, Hong (1996) employs a kernel lag–window estimator $\hat{f}_N$ defined by

$$\hat{f}_N(\omega)=\frac{1}{N}\sum_{|h|<N} K\left(\frac{h}{B_N}\right)\hat{\gamma}(h)e^{-\mathbf{i}h\omega}, \quad \omega\in[-\pi,\pi],$$

where B_N is the bandwidth and $K(\cdot)$ is a kernel function. The L^2–distance between the estimated spectral density of the residuals and the white noise spectral density is defined by

$$D(\hat{f}_N;f_0)=\left[2\pi\int_{-\pi}^{\pi}\left|\hat{f}_N(\omega)-\frac{1}{2\pi}\right|^2 d\omega\right]^{1/2}.$$

Hong (1996) then shows that the asymptotic distribution of the standardized $D(\hat{f}_N; f_0)$ is $N(0,1)$ under the i.i.d. assumption. The same holds for statistics derived from the other distances.

4.1. Tests of $\mathcal{H}_0$-IID

Characiejus and Rice (2020) construct test statistics based on the distance between the spectral density operator of a functional process and a white noise spectral density operator. They employ the Hilbert–Schmidt norm (2.6) to measure the distance and use a kernel lag–window type estimator to estimate the spectral density operator. The null distributions of the statistics are established under $\mathcal{H}_0$-IID.

Suppose $\{X_i\}$ is a mean-zero second-order stationary time series taking values in L^2. Recall the spectral density operator $\mathcal{F}_\omega$ in (2.12). In Characiejus and Rice (2020), the squared distance between the spectral density of a functional process and the spectral density corresponding to the white noise is defined by

$$D^2 = 2\pi \int_{-\pi}^{\pi} \left\| \mathcal{F}_\omega - (2\pi)^{-1}\Gamma_0 \right\|_{\mathcal{S}}^2 \, d\omega. \tag{4.1}$$

In this notation, the testing problem can be specified as $\mathcal{H}_0 : D^2 = 0$ vs. $\mathcal{H}_A : D^2 > 0$. However, it must be kept in mind that the behavior of the test is established under a more restrictive null hyposthesis of $\mathcal{H}_0$-IID. Characiejus and Rice (2020) work with a kernel lag–window estimator of $\mathcal{F}_\omega$ given by

$$\widehat{\mathcal{F}}_\omega = \frac{1}{2\pi} \sum_{|h|<N} K\left(\frac{h}{H_N}\right) \widehat{\Gamma}_h e^{-\mathbf{i}h\omega},$$

where H_N is the bandwidth parameter, and $K(\cdot)$ is a symmetric kernel. To construct a pivotal test statistic, the distance D^2 is approximated by

$$\widehat{D}^2 = 2\pi \int_{-\pi}^{\pi} \left\| \widehat{\mathcal{F}}_\omega - (2\pi)^{-1}\widehat{\Gamma}_0 \right\|_{\mathcal{S}}^2 d\omega = 2 \sum_{h=1}^{N-1} K^2\left(\frac{h}{H_N}\right) \left\| \widehat{\Gamma}_h \right\|_{\mathcal{S}}^2. \tag{4.2}$$

By scaling $\widehat{D}^2$, the test statistic is defined by

$$\mathrm{CR}_{N,K} = \mathrm{CR}_{N,K}(H_N) = \frac{2^{-1}N\widehat{D}^2 - \hat{\sigma}^4 C_N(K)}{\|\widehat{\Gamma}_0\|_{\mathcal{S}}^2 \sqrt{2D_N(K)}}, \tag{4.3}$$

where $\hat{\sigma}^2 = N^{-1}\sum_{i=1}^{N} \|X_i\|^2$ and

$$C_N(K) = \sum_{h=1}^{N-1} \left(1 - \frac{h}{N}\right) K^2\left(\frac{h}{H_N}\right),$$

$$D_N(K) = \sum_{h=1}^{N-2} \left(1 - \frac{h}{N}\right)\left(1 - \frac{h+1}{N}\right) K^4\left(\frac{h}{H_N}\right).$$

Under $\mathcal{H}_0$-IID, the statistic $\mathrm{CR}_{N,K}$ in (4.3) converges to the standard normal distribution. A modification analogous to the transformation proposed by Chen and Deo (2004) is also considered to get better approximations.

Hlávka *et al.* (2021) also test $\mathcal{H}_0$-IID, but they propose an approach based on the characteristic function (ch.f.). The key idea of the procedure is that independence is equivalent to the factorization of the joint ch.f.. Hlávka *et al.* (2021) study the case of discretely observed functions.

Suppose that a sample of curves, $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N$, is observed at points $t_{m;p} = m/p$, $m = 1, \ldots, p$. Denote the column vector of the available observations of the ith curve by

$$\mathbf{X}_i = [X_i(t_{1;p}), X_i(t_{2;p}), \ldots, X_i(t_{p;p})]^\top, \quad i = 1, 2, \ldots, N.$$

Under the assumption of stationarity of the functional observations, the empirical joint ch.f. of $\mathbf{X}_1$ and $\mathbf{X}_{1+h}$ is

$$\widehat{\Phi}_{1,N-h}(\mathbf{u}, \mathbf{v}) = \frac{1}{N-h} \sum_{j=1}^{N-h} \exp\left\{ i(\mathbf{u}^\top \mathbf{X}_j + \mathbf{v}^\top \mathbf{X}_{j+h}) \right\}, \quad \mathbf{u}, \mathbf{v} \in \mathbb{R}^p.$$

The corresponding marginal empirical characteristic functions are

$$\hat{\varphi}_{1,N-h}(\mathbf{u}) = \frac{1}{N-h} \sum_{j=1}^{N-h} \exp\left\{ i\mathbf{u}^\top \mathbf{X}_j \right\},$$
$$\hat{\varphi}_{1+h,N}(\mathbf{v}) = \frac{1}{N-h} \sum_{j=1}^{N-h} \exp\left\{ i\mathbf{v}^\top \mathbf{X}_{j+h} \right\}.$$

The difference between the joint ch.f. and the product of the marginal ch.f.s is

$$\widehat{D}_{N,h}(\mathbf{u}, \mathbf{v}) = \widehat{\Phi}_{1,N-h}(\mathbf{u}, \mathbf{v}) - \hat{\varphi}_{1,N-h}(\mathbf{u}) \hat{\varphi}_{1+h,N}(\mathbf{v}).$$

Based on the measure of dependence $\widehat{D}_{N,h}$ at lag h, Hlávka *et al.* (2021) introduced the test statistic defined by

$$\widehat{\Delta}_{N,H} = \sum_{h=1}^{H} (N-h) \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} |\widehat{D}_{N,h}(\mathbf{u}, \mathbf{v})|^2 w(\mathbf{u}) w(\mathbf{v}) d\mathbf{u} d\mathbf{v},$$

where w is a positive, integrable weight function. It is then proven that for suitable constants γ_p, ν_p (recall that p is the number of the discrete points in the interval $(0, 1]$),

$$\frac{\widehat{\Delta}_{N,H} - H\nu_p}{\sqrt{H\gamma_p}} \xrightarrow{d} N(0,1), \quad \text{as } H \to \infty.$$

The above relation also holds with γ_p, ν_p replaced by suitable estimators $\hat{\gamma}_p, \hat{\nu}_p$. An analog of $\widehat{\Delta}_{N,H}$ can be defined of continuously observed functions and an

analogous asymptotic normality holds. For a fixed H, the null distribution of $\widehat{\Delta}_{N,H}$ can be approximated by computing this statistic on all or a large number of permutations of $X_1, X_2, \ldots X_N$. Intuitively, this approach is justified because the functional observations are independent under $\mathcal{H}_0$-IID. The permutation test produces much more accurate empirical sizes.

4.2. Tests of $\mathcal{H}_0$-WN

Recall the spectral density function defined in (2.11). The white noise testing problem considered by Zhang (2016) is formulated as

$$\begin{aligned} \mathcal{H}_0: &\quad \forall\ \omega \in [-\pi, \pi], \quad f_\omega(t,s) = \frac{1}{2\pi}\gamma_0(t,s), \\ \mathcal{H}_A: &\quad \exists\ \omega \in [-\pi, \pi], \quad f_\omega(t,s) \neq \frac{1}{2\pi}\gamma_0(t,s). \end{aligned}$$

The idea of the test is as follows. Under $\mathcal{H}_0$-WN, the spectral density function does not change with ω, so its periodogram P_ω, defined in (2.13), changes with ω very little, in some random manner. Therefore, the differences

$$\int_0^{\pi\lambda} \mathrm{Re}P_\omega(t,s)d\omega - \lambda \int_0^{\pi} \mathrm{Re}P_\omega(t,s)d\omega$$

will be small for all λ, so continuous functionals based on the above process will have an asymptotic distribution under $\mathcal{H}_0$-WN. Zhang (2016) considers a Cramér-von-Mises functional that is shown to be equal to the following tests statistic

$$Z_N = \frac{N}{8\pi^2} \sum_{h=1}^{N-1} h^{-2} \iint \left\{\hat{\gamma}_h(t,s) + \hat{\gamma}_h(s,t)\right\}^2 dtds. \tag{4.4}$$

Under the assumption of L^4-m-approximability, see Hörmann and Kokoszka (2010) or Chapter 16 of Horváth and Kokoszka (2012), the statistic Z_N in (4.4) has a limiting distribution that can be expressed as an integral of the square of a Gaussian process. Like $\mathrm{KRS}_{N,H}$ introduced by Kokoszka *et al.* (2017), its asymptotic null distribution depends on the distribution and dependence (GARCH-like dependence) of the uncorrelated observations. Zhang (2016) develops a block bootstrap method for approximating the null distribution.

Under $\mathcal{H}_A$, $N^{-1}Z_N$ converges in probability to a positive limit. These results are stated precisely as Theorems 2.1 and 2.2 in Zhang (2016). The form of the limit under $\mathcal{H}_A$ shows that the test can detect alternatives such that $\iint \left\{\hat{\gamma}_h(t,s) + \hat{\gamma}_h(s,t)\right\}^2 dtds > 0$ for some $h > 0$.

Bagchi *et al.* (2018) also introduce a test of $\mathcal{H}_0$-WN that follows the approach of Hong (1996). As Characiejus and Rice (2020), they compare the spectral density operator of a functional process and the white noise density as shown in (4.1). However, the main difference is that Bagchi *et al.* (2018) estimate the distance D^2 in (4.1) by sums of periodogram functions while Characiejus and

Rice (2020) use a kernel lag–window estimator. Specifically, Bagchi *et al.* (2018) consider

$$B^2 = \frac{1}{2\pi} D^2 = \int_{-\pi}^{\pi} \left\| \mathcal{F}_\omega - (2\pi)^{-1} \Gamma_0 \right\|_{\mathcal{S}}^2 d\omega.$$

It is further shown that

$$B^2 = \iint \int_{-\pi}^{\pi} |f_\omega(t,s)|^2 d\omega dt ds - \frac{1}{2\pi} \iint \Big| \int_{-\pi}^{\pi} f_\omega(t,s) d\omega \Big|^2 dt ds.$$

To estimate B^2, they employ the periodogram P_ω defined in (2.13) and introduce the statistic

$$\widehat{B}_N^2 = 2\pi \iint \left(S_{N,2}(t,s) - S_{N,1}(t,s) \overline{S_{N,1}(t,s)} \right) dt ds,$$

where

$$S_{N,1} = \frac{1}{N} \sum_{i=1}^{\lfloor N/2 \rfloor} (P_{\omega_i} + \bar{P}_{\omega_i}), \qquad S_{N,2} = \frac{2}{N} \sum_{i=2}^{\lfloor N/2 \rfloor} (P_{\omega_i} \bar{P}_{\omega_{i-1}}),$$

with $\omega_i = 2\pi i/N$ for $i = 1, \ldots, \lfloor N/2 \rfloor$. The estimator is motivated by the unbiased approximation $E\widehat{B}_N^2 \approx B^2$, which follows from

$$ES_{N,1}(t,s) \approx \frac{1}{2\pi} \int_{-\pi}^{\pi} f_\omega(t,s) d\omega, \quad ES_{N,2}(t,s) \approx \frac{1}{2\pi} \int_{-\pi}^{\pi} |f_\omega(t,s)|^2 d\omega,$$

see Panaretos and Tavakoli (2013b). The asymptotic null distribution of $\widehat{B}_N^2$ is normal, and the asymptotic variance can be obtained using the P_{ω_i}.

5. Goodness-of-fit tests for functional time series models

In this section, we describe how some tests introduced above may be adapted to evaluate the goodness-of-fit of several popular functional time series models.

5.1. *Functional autoregressive models*

A sequence $\{X_i,\ -\infty < i < \infty\}$ of random variables in L^2 with mean μ is said to follow a *functional autoregressive process of order 1* (FAR(1)) if

$$X_i(t) - \mu(t) = \Psi(X_{i-1} - \mu)(t) + \varepsilon_i(t), \quad i \in \mathbb{Z}, \tag{5.1}$$

where $\Psi(X)(t) = \int \psi(t,s) X(s) ds$ for a kernel function $\psi \in L^2([0,1] \times [0,1])$, and $\{\varepsilon_i\}$ is a mean-zero, L^2-valued white noise sequence with covariance operator $\Gamma_{\epsilon,0}$. When $\left\| \Psi^j \right\|_{\mathcal{S}} < 1$ for some $j \in \mathbb{N}$, there is a unique stationary solution to (5.1) that takes the form of a functional linear process $X_i(t) = \mu(t) + \sum_{j=0}^{\infty} \Psi^j (\varepsilon_{i-j})(t)$. To ease notation, in the following discussion, we assume that $\mu(t) = 0$. When implementing the techniques described below, we always start with centering by the sample mean function.

The white noise tests introduced above may be used to evaluate the goodness-of-fit of model (5.1) to an observed functional time series by applying them to model residuals. In order to estimate the model and calculate the residuals, it is typical to employ dimension reduction techniques. Let

$$\tilde{C}^{-1}(\cdot) = \sum_{j=1}^{k_N} \frac{\langle \hat{v}_j, \cdot \rangle}{\hat{\lambda}_j} \hat{v}_j,$$

denote the Moore-Penrose inverse of $\widehat{C}$, where the empirical eigenvalues and eigenfunctions $(\hat{\lambda}_j, \hat{v}_j)_{1 \le j \le k_N}$ are defined in (2.5). Then a combination of functional principal component analysis and a least-squares principle suggest estimating Ψ with

$$\widehat{\Psi}_N(\cdot) = \widehat{\Gamma}_1 \tilde{C}^{-1}(\cdot) = \frac{1}{N} \sum_{i=2}^{N} \sum_{j=1}^{k_N} \frac{\langle \hat{v}_j, \cdot \rangle}{\hat{\lambda}_j} \langle X_{i-1}, \hat{v}_j \rangle X_i,$$

see e.g. Chapter 8 of Bosq (2000) for a detailed derivation of this estimator. In order for $\widehat{\Psi}_N$ to be a consistent estimator of Ψ, one must take k_N to be a suitably increasing sequence, although in practice k_N is typically selected using the total variance explained approach, cross-validation, or information criteria.

The model residuals are then defined as

$$\hat{\varepsilon}_i(t) = X_i(t) - \widehat{\Psi}_N(X_{i-1})(t), \quad i \in \{2, ..., N\}.$$

Zhang (2016) shows that if indeed the observations are drawn from a stationary series following (5.1), then the test statistic

$$Z_N^{(GF)} = \frac{N}{8\pi^2} \sum_{h=1}^{N-1} h^{-2} \iint \left\{ \hat{\gamma}_{\hat{\varepsilon},h}(t,s) + \hat{\gamma}_{\hat{\varepsilon},h}(s,t) \right\}^2 dtds,$$

where

$$\hat{\gamma}_{\hat{\varepsilon},h}(u,v) = \frac{1}{N} \sum_{t=1+h}^{N} \hat{\varepsilon}_t(u) \hat{\varepsilon}_{t-h}(v),$$

also converges in distribution to the squared norm of a Gaussian process. Importantly though, the covariance of this limiting Gaussian process does not coincide with the limiting covariance when the observations evolve simply as a stationary white noise sequence. The difference is attributable to the effect of estimating the operator Ψ. Because of the complicated form of this limiting process, a modified block bootstrap procedure to adjust for this effect was introduced in Section 3.2 of Zhang (2016) to estimate the null distribution of $Z_N^{(GF)}$, which may then be used to perform goodness-of-fit testing for the FAR model.

An analogous adjustment may be introduced to make the statistic $\mathrm{KRS}_{N,H}$ described in Section 3.2 applicable to testing the goodness-of-fit of the FAR(1)

model. We define the corresponding statistic as

$$\mathrm{KRS}_{N,H}^{(GF)} = N \sum_{h=1}^{H} \|\hat{\gamma}_{\hat{\varepsilon},h}\|^2.$$

It is shown in Section B that under the null hypothesis of FAR(1), $\mathrm{KRS}_{N,H}^{(GF)} \xrightarrow{d} V_H$, where V_H is a weighted sum of independent χ_1^2 random variables, with complicated weights reflecting the dependence structure of the FAR(1) model and the effect of the estimation of the autoregressive operator Ψ. Using a suitable Welch-Satterthwaite approximation, the distribution of V_H can be effectively approximated, as explained in Section B.2. Even though the theory underlying the convergence $\mathrm{KRS}_{N,H}^{(GF)} \xrightarrow{d} V_H$ is complex, but comparable to the corresponding theory of Zhang (2016), the advantage is that an analytic approximation to the asymptotic null distribution can be derived, eliminating the need for the bootstrap.

González-Manteiga *et al.* (2023) recently developed a goodness-of-fit test for the FAR(1) model based on evaluating the condition $E[X_i - \Phi(X_{i-1})|X_{i-1}] = 0$ using random projections. This test is akin to the goodness-of-fit test for the functional regression models developed in Cuesta-Albertos *et al.* (2019).

5.2. Functional linear models

A sequence of pairs of functional data $\{(X_i, Y_i),\ -\infty < i < \infty\}$ is said to follow a *functional linear model* (FLM) if

$$Y_i(t) = \mathcal{K}(X_i)(t) + \varepsilon_i(t) = \int \kappa(t,s) X_i(s) ds + \varepsilon_i(t), \tag{5.2}$$

where $\{\varepsilon_i, -\infty < i < \infty\}$ is a mean-zero, L^2-valued white noise sequence that is uncorrelated (at all lags) with the covariate series $\{X_i,\ -\infty < i < \infty\}$, and $\kappa \in L^2([0,1] \times [0,1])$ is a kernel function. When the pairs of functions are observed as a bivariate FTS, white noise tests can be applied to the model residuals in order to evaluate the plausibility of the white noise error assumption. Given a sample $(X_1, Y_1), ..., (X_N, Y_N)$, a commonly used estimator of the operator $\mathcal{K}$ is given by

$$\widehat{\mathcal{K}}(\cdot) = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{k_N} \frac{1}{\hat{\lambda}_j} \langle X_i, \hat{v}_j \rangle \langle \hat{v}_j, \cdot \rangle Y_i,$$

where we used the notation of Section 2.2. The derivation of $\widehat{\mathcal{K}}$ is given, for example, in Section 8.3 of Horváth and Kokoszka (2012). The model residuals may then be computed as

$$\hat{\varepsilon}_i(t) = Y_i(t) - \widehat{\mathcal{K}}(X_i)(t),\ i \in \{1, ..., N\}. \tag{5.3}$$

White noise tests applied to these residuals may then be based on the statistics $Z_N^{(GF)}$ and $\mathrm{KRS}_{N,H}^{(GF)}$. Interestingly, and in contrast to the FAR case, due

to the assumption that the covariates X_i and errors ε_i are uncorrelated at all lags, and under suitable additional moment and rate conditions on k_N, both these statistics have the same limiting distributions as if they were based on the unobservable errors $\{\varepsilon_i,\ -\infty < i < \infty\}$. This justifies the practice of applying each test to the residuals in (5.3) in the same way as described in Sections 3 and 4.

5.3. Tests for conditional heteroscedasticity and goodness-of-fit for functional GARCH models

A pronounced property of many scalar economic and financial time series, or model residuals, is that they are uncorrelated, but their squares or absolute values are strongly correlated. Conditionally heteroscedastic models, especially the GARCH models, have become popular tools, and extensive methodology and theory are now available, see e.g. Francq and Zakoian (2010). There is a much smaller body of research on functional conditionally heteroscedastic models. We review in this section its aspects related to the focus of this paper.

The test introduced by Rice *et al.* (2020) is concerned with testing $\mathcal{H}_0$-IID, but focuses on detecting conditional heteroscedasticity in functional observations. Conditional heteroscedasticity is defined by the conditions

$$E[X_i|\mathcal{F}_{i-1}] = 0,\ \mathrm{Cov}(X_i^2(t), X_{i+h}^2(s)) \neq 0, \quad \text{in } L^2 \text{ sense for some } h \geq 1, \quad (5.4)$$

where the filtration $\mathcal{F}_i$ is the sigma algebra generated by random functions $\{X_j, j \leq i\}$. The alternative hypotheses thus is

$$\mathcal{H}_A: \quad \text{the } X_i \text{ satisfy (5.4)}.$$

A key feature of conditional heteroscedasticity is the presence of serial correlation in the squared series, which motivates the following two test statistics:

$$\widehat{U}_{N,H} = N\sum_{h=1}^{H} \hat{\gamma}_{X^2}^2(h), \quad \mathrm{KRS}_{N,H}^{(CH)} = N\sum_{h=1}^{H} \|\hat{\gamma}_{X^2,N,h}\|^2, \qquad (5.5)$$

where $\hat{\gamma}_{X^2}(h)$ is the sample autocorrelation of the squared norms, $\|X_1\|^2, \ldots, \|X_N\|^2$, and

$$\hat{\gamma}_{X^2,N,h}(t,s) = \frac{1}{N}\sum_{i=1}^{N-h} \left(X_i^2(t) - \hat{\mu}_{X^2,N}(t)\right)\left(X_{i+h}^2(s) - \hat{\mu}_{X^2,N}(s)\right)$$

with $\hat{\mu}_{X^2,N}$ denoting the sample mean of the functions X_i^2. The statistic $\widehat{U}_{N,H}$ is indeed $\widehat{Q}_{BP}$ in (3.1) applied to the squared norms, and $\mathrm{KRS}_{N,H}^{(CH)}$ is the same as the statistic in (3.2), but derived from the squared functions X_i^2.

Rice *et al.* (2020) show that under $\mathcal{H}_0$-IID, $\widehat{U}_{N,H}$ converges in distribution to the χ_H^2 distribution, whereas $\mathrm{KRS}_{N,H}^{(CH)}$ converges to an infinite sum of χ_1^2 random variables with weights. The weights are products of certain eigenvalues that

can be estimated. Under $\mathcal{H}_A$, it is proven that the $N^{-1}\widehat{U}_{N,H}$ and $N^{-1}\mathrm{KRS}_{N,H}^{(CH)}$ converge to positive constants.

Rice *et al.* (2020) further study tests to evaluate the fit of functional GARCH models introduced in Hörmann *et al.* (2013), and further developed in Aue *et al.* (2017) and Cerovecki *et al.* (2019). These models have the form $X_i(t) = \sigma_i(t)\varepsilon_i(t)$ with i.i.d. mean zero functions ε_i satisfying $E\varepsilon_i^2(t) = 1$. The conditional variance process $\sigma_i(t)$ takes a GARCH(p, q) form:

$$\mathrm{fGARCH}(p,q): \quad \sigma_i^2(t) = \omega(t) + \sum_{\ell=1}^{p} \alpha_\ell(X_{i-\ell}^2) + \sum_{k=1}^{q} \beta_k(\sigma_{i-k}^2), \tag{5.6}$$

where ω is a non-negative function in L^2, and the α_ℓ's and β_k's, are kernel-integral operators with non-negative kernels. After estimating such a model, one can compute the residuals as $\hat{\varepsilon}_i(t) = X_i(t)/\hat{\sigma}_i(t)$. If the model is correct, these residuals should be approximately i.i.d. with $E\hat{\varepsilon}_i^2(t) \approx 1$. In case of the statistic $\mathrm{KRS}_{N,H}^{(CH)}$, the autocovariances $\hat{\gamma}_{X^2,N,h}(t,s)$ are replaced by

$$\hat{\gamma}_{\varepsilon^2,N,h}(t,s) = \frac{1}{N}\sum_{i=1}^{N-h} \left(\hat{\varepsilon}_i^2(t) - 1\right)\left(\hat{\varepsilon}_{i+h}^2(s) - 1\right).$$

The asymptotic null distribution of $\mathrm{KRS}_{N,H}^{(CH)}$ obtained from $\hat{\gamma}_{\varepsilon^2,N,h}$ depends on model parameters and the method used to estimate them, but can still be expressed as an infinite sum of independent χ_1^2 random variables. These asymptotic results are known when functional principal component analysis as well as least squares or Quasi-Maximum likelihood estimators are used to estimate the operators in (5.6).

6. Numerical comparisons

The goal of this section is to compare the white noise tests described in Sections 3 and 4 by means of a comprehensive simulation study. Some of these tests have been compared to each other in the papers cited above, but we would like to present a more comprehensive and independent comparison. In Section 6.1, we evaluate the empirical size and power of the tests, and in Section 6.2 we focus on goodness-of-fit testing for FLM and FAR(1) models. An application study where we investigate suitable models for Eurodollar futures curves is also provided in Section 6.3. The numerical work presented in this section was performed using `R`, R Development Core Team (2008). The code used to implement these tests is collected in the `R` package `wwntests`.

6.1. White noise testing

We begin by briefly describing the tests that we compare, including the choice of required tuning parameters.

$\mathrm{GK}_{N,H}$ is the test statistic introduced by Gabrys and Kokoszka (2007), and its null distribution is developed under $\mathcal{H}_0$-IID. It is based on projections of functional observations onto the estimated FPCs defined in (2.5). To choose the dimension of projections, we use the smallest p such that $\sum_{j=1}^{p} \hat{\lambda}_j / \sum_{j=1}^{N} \hat{\lambda}_j > 0.85$. We consider implementations with the total number of lags $H = 1, 5, 20$.

$\mathrm{CR}_{N,K}$ is introduced by Characiejus and Rice (2020), assuming $\mathcal{H}_0$-IID. It is constructed from the proximity of a kernel lag–window type estimator to the white noise spectral density. We use the Bartlett (Bl) and Parzen (Pz) kernels with bandwidths $H_N = N^{1/(2q+1)}$, where q is the order of the specific kernel, following Characiejus and Rice (2020).

$\mathrm{KRS}_{N,H}$ is introduced by Kokoszka *et al.* (2017), assuming $\mathcal{H}_0$-WN. It is based on the sum of the L^2 norm of sample autocovariance functions up to lag H. The total number of lags, $H = 1$, 5, 20, are considered.

$Z_N(b)$ is introduced by Zhang (2016), assuming $\mathcal{H}_0$-WN. This test statistic can be viewed as a special case of $\mathrm{CR}_{N,K}$ with a certain choice of kernel and bandwidth, but Z_N assumes a broader null than $\mathrm{CR}_{N,K}$. This test requires bootstrap procedures with block size b and the number of resamples B to approximate the null distribution of Z_N. We use $b \in \{5, 10\}$ and $B = 299$.

We thus compare four tests that reflect the dichotomies discussed in this paper, as summarized in the following diagram:

	$\mathcal{H}_0$-IID	$\mathcal{H}_0$-WN
Time	$\mathrm{GK}_{N,H}$	$\mathrm{KRS}_{N,H}$
Frequency	$\mathrm{CR}_{N,K}$	$Z_N(b)$

To investigate the empirical size and power of the tests, we consider the following data generating processes (DGPs):

- **IID-BM**: the X_i are the i.i.d. standard Brownian motions (BMs) on $[0, 1]$. We generate their trajectories as rescaled cumulative sums of independent normal variables.
- **fGARCH(1, 1)**: recall (5.6). The X_i satisfy $X_i(t) = \sigma_i(t)\varepsilon_i(t)$. The conditional variance has the form

$$\sigma_i^2(t) = \delta(t) + \alpha(X_{i-1}^2)(t) + \beta(\sigma_{i-1}^2)(t),$$

  where $\delta = 0.01$, and $\alpha, \beta : L^2 \to L^2$ are linear operators satisfying

$$(\alpha x)(t) = (\beta x)(t) = \int 12t(1-t)s(1-s)x(s)ds.$$

  The ε_i are i.i.d. Ornstein-Uhlenbeck processes given by $\varepsilon_i(t) = e^{-t/2}B_i(e^t)$, where the B_i are i.i.d. standard BMs. The particular settings for σ_i and ε_i are from Cerovecki *et al.* (2021).
- **FAR(1, S)–BM**: recall (5.1). The X_i satisfy

$$X_i(t) = \int \psi(t, s)X_{i-1}(s)ds + \varepsilon_i(t),$$

TABLE 6.1
Empirical sizes (in percent) based on 1000 replications for tests applied to series generated from the DGP's IID-BM and fGARCH(1,1).

DGP	IID-BM				fGARCH(1, 1)			
N	100		250		100		250	
Nominal level	5%	1%	5%	1%	5%	1%	5%	1%
$GK_{N,1}$	4.7	0.6	4.8	0.5	15.3	6.2	19.8	9.5
$GK_{N,5}$	3.5	0.1	4.2	1.2	16.4	8.4	25.8	13.0
$GK_{N,20}$	2.2	0.3	3.5	0.6	6.9	3.5	15.0	8.0
$CR_{N,Bl}$	6.0	1.1	6.2	1.2	18.4	7.9	27.8	15.9
$CR_{N,Pz}$	5.9	1.1	4.9	1.1	16.1	6.9	21.1	11.0
$KRS_{N,1}$	5.4	1.5	5.6	1.7	3.0	0.2	5.4	1.1
$KRS_{N,5}$	4.9	1.1	4.0	1.0	3.2	0.6	4.8	0.8
$KRS_{N,20}$	6.1	2.0	5.9	1.4	4.0	1.1	5.6	1.4
$Z_N(5)$	7.0	1.7	6.0	1.2	3.6	0.3	5.1	1.1
$Z_N(10)$	6.0	1.6	5.0	1.9	5.5	0.8	5.4	0.9

where the ε_i follow IID-BM. The Gaussian kernel $\psi(t,s) = c\exp\{-(t^2 + s^2)/2\}$ is assumed with the choice of c such that $\|\psi\| = S$. For S, we consider $S = 0.1, 0.3, 0.5, 0.7$.

- **FAR(1, S)–fGARCH**: the X_i follow the same process as FAR(1, S)–BM, except for the innovations ε_i following the fGARCH(1, 1).
- **FMA(5, S)–BM**: the X_i satisfy

$$X_i(t) = \int \psi(t,s)\varepsilon_{i-5}(s)ds + \varepsilon_i(t), \tag{6.1}$$

where the ε_i follow IID-BM and ψ is the Gaussian kernel. For the norm of the kernel, we consider $S = 0.1, 0.3, 0.5, 0.7$.

Note that IID-BM satisfies both $\mathcal{H}_0$-IID and $\mathcal{H}_0$-WN. The fGARCH(1, 1) model satisfies $\mathcal{H}_0$-WN, but not $\mathcal{H}_0$-IID. The FAR and FMA models with $\|\psi\| > 0$ violate both $\mathcal{H}_0$-WN and $\mathcal{H}_0$-IID. All DGPs simulate functional observations on a grid of 100 equally–spaced points on the unit interval $[0, 1]$, and we discard a burn–in period of the first 50 curves for all DGPs. Each process is generated by 1000 replications with the sample sizes $N = 100$ and $N = 250$.

To assess the empirical size of the tests, we consider IID–BM and fGARCH(1, 1). Table 6.1 reports the empirical rejection rates from tests of $GK_{N,H}$, $CR_{N,K}$, $KRS_{N,H}$, and $Z_N(b)$ at nominal levels of 5% and 1%. To investigate the empirical power, we apply those tests to data generated from FAR(1, S)–BM, FAR(1, S)–fGARCH, and FMA(5, S)–BM. The empirical power as a function of the level of serial dependence, measured by S, of the kernel is presented in Figure 6.1. Since the rejection rates from FAR(1, S)–BM are almost identical over the tests and uniformly higher than the rates from FAR(1, S)–fGARCH, we only report the results from FAR(1, S)–fGARCH and FMA(5, S)–BM. The conclusions from Table 6.1 and Figure 6.1 can be summarized as follows:

1. All tests perform well under IID–BM. In terms of power, all methods exhibit increasing power as a function of the strength of the serial dependence

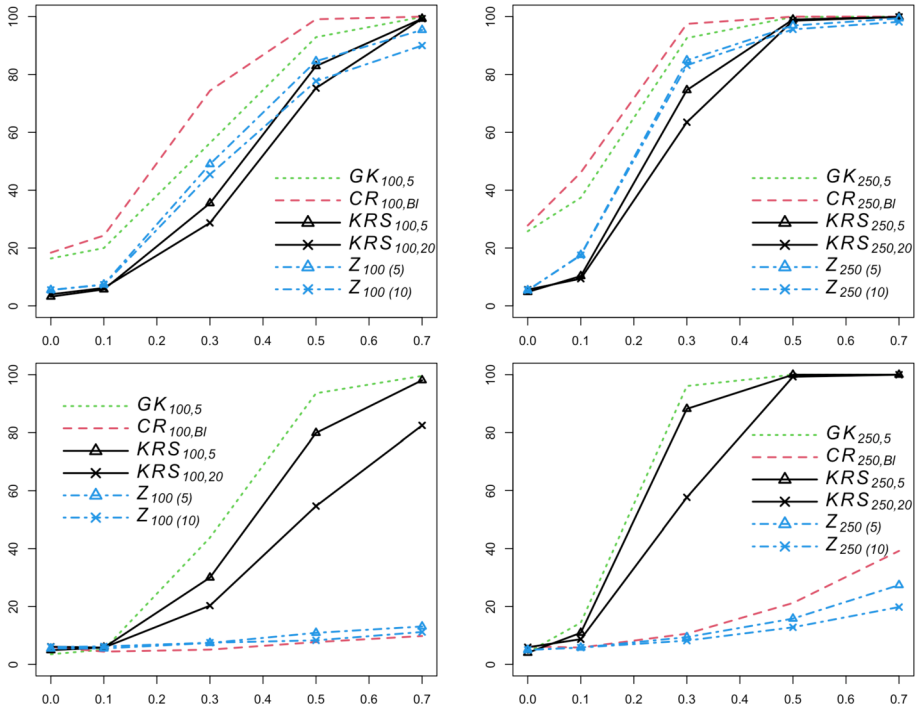

FIG 6.1. *Empirical power in percent (y-axis) for increasing values of the norm S (proxy of correlation) (x-axis), for FAR(1, S)-fGARCH (upper plots) and FMA(5, S)-BM (lower plots);* $N = 100$ *(left),* $N = 250$ *(right); significance level = 5%.*

in the series, measured by S, as well as with the sample size N.

2. The tests based on $\mathrm{GK}_{N,H}$ and $\mathrm{CR}_{N,K}$ achieve good size when the errors are i.i.d., but they over-reject severely under the fGARCH errors.
3. The tests based on $Z_N(b)$ and $\mathrm{KRS}_{N,H}$ exhibited close to nominal empirical size for i.i.d. as well as weak white noise sequences.
4. The $\mathrm{CR}_{N,K}$ and $Z_N(b)$ tests share a common feature in that decreasing weights are applied to the norms of the autocovariance kernels when calculating the test statistics. This appears to manifest itself in that these tests have strong power against autocorrelation occurring in the sequence at small lags, e.g. the FAR(1, S)–fGARCH case (top plots in Figure 6.1), but they do not consistently detect serial correlation that occurs only at longer lags, e.g. the FMA(5, S)–BM case (bottom plots in Figure 6.1). We see here that although in practice each of these statistics are consistent against autocorrelation occurring in the FTS at any lag, the weights employed in computing $\mathrm{CR}_{N,K}$ and $Z_N(b)$ decrease quickly enough so that by lag 5 even rather strong autocorrelation in the series is not reliably detected.
5. In terms of computational time, the tests based on $\mathrm{GK}_{N,H}$, $\mathrm{CR}_{N,K}$ and

$\mathrm{KRS}_{N,H}$, which use explicit approximations of their null limiting distributions, run quickly, with the computational bottleneck lying in computing and integrating high-dimensional covariance kernels. Implementing the test based on $Z_N(b)$ takes somewhat more time, since it relies on a block bootstrap. For example, computing the p-values of each test on a single sample of size $N = 250$ took 3.72s for $\mathrm{GK}_{N,5}$, 1.16s for $\mathrm{CR}_{N,Bl}$, 4.92s for $\mathrm{KRS}_{N,5}$, and 32.32s for $Z_N(10)$, using a 2.3 GHz Intel dual-core processor running a Windows 64 bit implementation of `R`.

In conclusion, our numerical comparisons suggest that the test based on $\mathrm{KRS}_{N,H}$ should be used as a default and benchmark. It demonstrated nearly nominal empirical size in all examples considered, both for i.i.d. and weak white noise sequences, competitive power, and is easy and quick to compute. Moreover, the fact that it gives equal weight to each autocovariance considered makes it somewhat more transparent and easier to interpret. Although $\mathrm{KRS}_{N,H}$ comes with a theoretical drawback that the statistic is not asymptotically powerful against autocorrelation occurring in the FTS at any lag, we have seen in practice that the spectral based tests, which do have this property, are often not powerful with typical choices of the tuning parameters to detect strong autocorrelation occurring at modest lags with even large sample sizes. We think a reasonable default choice for H is to take $H = 10$ or 20, since the approximation methods used to estimate the null distribution of $\mathrm{KRS}_{N,H}$ appear to work well for these choices, among the sample sizes we have considered ($N \geq 100$).

6.2. Simulation study of goodness-of-fit tests

We now turn our attention to the application of the tests based on $\mathrm{KRS}_{N,H}$ and $Z_N(b)$ to evaluate the goodness-of-fit of FLM and FAR(1) models, as discussed in Section 5. We focus only on these two tests because they work well when applied directly to functional observations, we want to consider non-i.i.d. errors, and finally keep the numerical study reasonably compact. The tests are constructed as follows: if the residuals satisfy $\mathcal{H}_0$-WN, the model is declared to fit well.

First, we assessed the empirical size and power of the goodness-of-fit tests for the FAR(1) model by considering the following DGPs:

- **FAR(1)–BM**: recall (5.1). The X_i follow

$$X_i(t) = \int \psi(t,s)X_{i-1}(s)ds + \varepsilon_i(t),$$

  where ε_i follows IID-BM DGP in Section 6.1, and ψ is taken to be the Gaussian kernel $\psi(t,s) = c\exp\{-(t^2+s^2)/2\}$, or the Wiener kernel $\psi(t,s) = c\min(t,s)$, with the choice of c such that $\|\psi\| = S$, with $S = 0.5, 0.8$.
- **FAR(1)–fGARCH**: the data are generated from the same process as FAR(1)–BM, except that the innovations ε_i follow the fGARCH(1,1) of Section 6.1.

- **FAR(2, S)**: the X_i follow

$$X_i(t) = \int \psi_1(t,s)X_{i-1}(s)ds + \int \psi_2(t,s)X_{i-2}(s)ds + \varepsilon_i(t),$$

  where ψ_1 and ψ_2 are the Gaussian kernel with $\|\psi_1\| = 0.5$ and $\|\psi_2\| = S$, and $S = 0.1, 0.3, 0.5, 0.7$. The ε_i follow IID–BM.
- **FAR(1)–FMA(5, S)**: the X_i satisfy

$$X_i(t) = \int \psi_1(t,s)X_{i-1}(s)ds + \int \psi_2(t,s)\varepsilon_{i-5}(s)ds + \varepsilon_i(t),$$

  where ψ_1 and ψ_2 are the Gaussian kernel with $\|\psi_1\| = 0.5$ and $\|\psi_2\| = S$, and $S = 0.1, 0.3, 0.5, 0.7$. The ε_i follow IID–BM.

We note both the DGPs FAR(1)–BM and FAR(1)–fGARCH are FAR models as in (5.1), whereas the DGPs FAR(2, S) and FAR(1)–FMA(5, S) are not FAR(1) processes. In implementing the estimation of the FAR(1) kernel, the dimension reduction parameter k_N is chosen to be the smallest number that satisfies $\sum_{j=1}^{k_N} \hat{\lambda}_j / \sum_{j=1}^{N} \hat{\lambda}_j > 0.90$.

TABLE 6.2
Empirical sizes (in percent) based on 1000 replications. The tests $\mathrm{KRS}_{N,H}^{(GF)}$, $Z_N^{(GF)}(b)$ *are applied to evaluate the goodness-of-fit of an FAR(1) model with the Wiener kernel.*

	DGP	FAR(1)–BM				FAR(1)–fGARCH			
S	N	100		250		100		250	
	Nominal level	5%	1%	5%	1%	5%	1%	5%	1%
0.5	$\mathrm{KRS}_{N,1}$	5.4	1.9	5.0	1.9	5.8	2.0	6.7	1.5
	$\mathrm{KRS}_{N,5}$	4.9	1.1	4.6	1.7	3.0	0.7	4.8	0.8
	$KRS_{N,20}$	6.6	1.0	5.6	1.3	5.3	1.4	4.9	1.3
	$Z_N(b=10)$	5.2	1.3	6.4	1.2	4.7	0.9	4.1	0.3
0.8	$\mathrm{KRS}_{N,1}$	5.5	1.6	5.9	1.3	6.4	1.3	6.3	1.2
	$\mathrm{KRS}_{N,5}$	4.8	0.9	5.4	1.7	3.6	0.7	4.7	1.0
	$\mathrm{KRS}_{N,20}$	5.6	1.7	5.8	1.2	4.7	1.8	5.2	1.4
	$Z_N(b=10)$	6.1	2.0	5.4	1.4	4.5	0.7	3.4	0.8

Empirical sizes are given in Table 6.2 for the Wiener kernel, and sizes for the Gaussian kernel are given in Table C.1 of Appendix C. Power curves as a function of S are displayed in Figure 6.2. We see that both tests have well-controlled size; the empirical power is increasing with N. Similarly as reported in Section 6.1, $\mathrm{KRS}_{N,5}^{(GF)}$ performs similarly well for both alternatives, whereas $Z_N^{(GF)}(10)$ performs worse for the FAR(1)–FMA(5, S) process.

To assess goodness-of-fit of the functional linear model, we consider the following examples:

- **FLM–BM**: recall (5.2). The Y_i is defined by

$$Y_i(t) = \int \kappa(t,s)X_i(s)ds + \varepsilon_i(t),$$

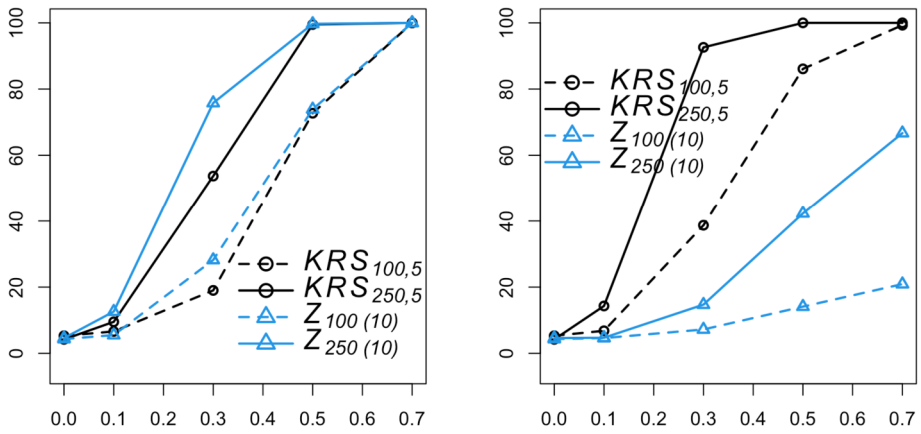

FIG 6.2. *Empirical power in percent (y-axis) for increasing values of the norm S, of the Gaussian kernel (x-axis) for FAR(2, S) (left) and FAR(1)–FMA(5, S) (right); significance level = 5%; N = 100, 250.*

TABLE 6.3
Empirical sizes (in percent) based on 1000 replications. The tests $\mathrm{KRS}_{N,H}^{(GF)}$, $Z_N^{(GF)}(b)$ *are applied to evaluate the goodness-of-fit of an FLM model with the Wiener kernel.*

	DGP	FLM–BM				FLM–fGARCH			
S	N	100		250		100		250	
	Nominal level	5%	1%	5%	1%	5%	1%	5%	1%
0.5	$\mathrm{KRS}_{N,1}$	4.4	0.7	5.0	1.1	4.9	0.5	3.8	0.3
	$\mathrm{KRS}_{N,5}$	4.7	1.2	4.9	1.3	3.9	0.4	2.9	0.6
	$\mathrm{KRS}_{N,20}$	5.9	1.4	5.6	1.6	5.9	1.5	3.7	1.1
	$Z_N(10)$	5.6	1.7	5.2	1.3	6.2	1.7	4.1	0.5
0.8	$\mathrm{KRS}_{N,1}$	5.0	1.5	5.2	1.7	5.3	0.9	5.2	1.3
	$\mathrm{KRS}_{N,5}$	4.9	1.2	4.6	1.6	4.1	1.2	3.7	1.0
	$\mathrm{KRS}_{N,20}$	7.3	2.7	6.7	2.1	5.2	1.8	5.2	1.7
	$Z_N(10)$	7.0	2.4	6.3	1.5	7.2	1.2	4.5	1.2

where the pairs (X_i, ε_i) are i.i.d., and the ε_i are independent of the X_i. The X_i are generated as $X_i(t) = it$, and the ε_i are generated from the IID-BM DGP of Section 6.1. For the regression kernel κ, we consider the Gaussian kernel $\kappa(t,s) = c\exp\{-(t^2+s^2)/2\}$ and the Wiener kernel $\kappa(t,s) = c\min(t,s)$ with the choice of c such that $\|\kappa\| = S$, with $S = 0.5, 0.8$.

- **FLM–fGARCH**: the data are generated from the same process as FLM–BM, except that the innovations ε_i follow the fGARCH(1, 1) of Section 6.1.
- **FLM–FAR(1, S)**: the data are generated from (5.2) with $X_i(t) = it$ and the Gaussian kernel such that $\|\kappa\| = 0.5$. The ε_i follow the FAR(1, S)–BM DGP of Section 6.1, with S = 0.1, 0.3, 0.5, 0.7.

Table 6.3 reports empirical sizes of the goodness-of-fit tests based on $\mathrm{KRS}_{N,H}^{(GF)}$

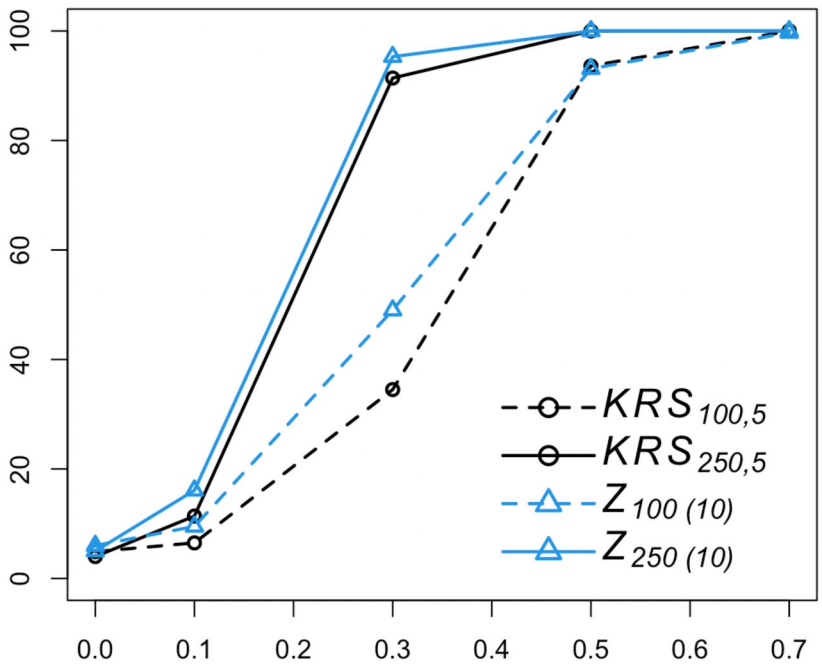

FIG 6.3. *Empirical power in percent (y-axis) for increasing values of the norm S of the Gaussian kernel (x-axis) for FLM–FAR(1, S); significance level = 5%; N = 100, 250.*

and $Z_N^{(GF)}(b)$ for FLM–BM and FLM–fGARCH generated with the Wiener kernel. The sizes for the Gaussian kernel are presented in Table C.2 of Appendix C. Both tests exhibit accurate sizes, even for $N = 100$, and for both values of S. To investigate empirical power, we examine the rejection rates of the residuals obtained by fitting an FLM to data generated from FLM–FAR(1, S). The result is reported in Figure 6.3. We observed similar performance in this case as when these tests were applied to FAR models.

6.3. Application to Eurodollar futures contract curves

In this section, we illustrate the practical application of several of the tests discussed in previous sections by considering daily Eurodollar futures curves. A Eurodollar futures contract represents an obligation to deliver 1,000,000 USD to a bank outside of the United States at a specified time. The Eurodollar futures curves consist of daily settlement prices of such a contract, which are available at monthly delivery dates for the first six months, and quarterly delivery dates for up to 10 years into the future. The sample considered in this study consists of 10 years of daily Eurodollar futures curves taken from 1994 to 2003. The curves are preprocessed using cubic-splines, following Kargin and Onatski (2008), so that the raw data can be transformed into smooth curves on a grid of 114

equally–spaced points. Denote by $X_i(t)$ the smoothed futures curve on day $i = 1, \ldots 2486$, where $t = 1, \ldots, 114$. We divide the curves into yearly series, so that we consider 10 yearly samples of functional time series, each of which consists of approximately 250 curves. A plot of the curves X_i from 2003 is displayed on the upper left of Figure 6.4.

Analyses of Eurodollar futures curves have been conducted in several papers. Kargin and Onatski (2008) argued that the curves can be modeled and forecasted using an FAR(1) model, which is also supported by the model identification procedures introduced by Mestre *et al.* (2021). Another approach to the analysis of the Eurodollar future curves is to assume that such curves follow random walk models. A study in this direction is done by Characiejus and Rice (2020), where frequency domain tests are applied to the first differences $Y_i = X_i - X_{i-1}$. The tests supported the claim that the Y_i evolve as a white noise. In this section, we explore both approaches using the white noise and goodness-of-fit tests discussed throughout the paper. All results of these tests are displayed in terms of p-values in Table 6.4.

We first examined whether the curves X_i appeared to exhibit autocorrelation. For this, the tests based on $\mathrm{GK}_{N,H}$, $\mathrm{CR}_{N,K}$, $\mathrm{KRS}_{N,H}$, and $Z_N(b)$ discussed in Section 6.1 were employed. All tests reject the i.i.d. and white noise hypotheses; the p–values from the tests are all zero in each year (this is omitted from Table 6.4 to save space). As can be seen in the upper left hand panel of Figure 6.4, the raw series exhibit strong (stochastic) trends, which leads to large in norm autocovariance kernel estimators; see the fACF plot in the upper right panel of Figure 6.4.

In order to model the evident serial dependence in the series, Kargin and Onatski (2008) suggest fitting an FAR(1) model. In order to evaluate whether the residuals of such a model fit to this data are plausibly white noise, we applied the goodness-of-fit tests $\mathrm{KRS}^{(GF)}_{N,H}$, $H \in \{1, 5, 20\}$, and $Z^{(GF)}_N(b)$ discussed in Section 6.2. Each of these tests provided strong evidence that the residuals of an FAR(1) model fit still contain substantial autocorrelation, suggesting that the FAR models do not fit this data well; see the top panel of Table 6.4.

Given the strong and fluctuating trends evident in the raw series, we also explored whether the series instead appear to evolve as a random walk with potentially conditionally heteroscedastic functional innovations. Let $Y_i(t) = X_i(t) - X_{i-1}(t)$ be the first differenced Eurodollar futures curves. The second panel of Table 6.4 reports the p–values from the tests $\mathrm{GK}_{N,H}$, $\mathrm{CR}_{N,K}$, $\mathrm{KRS}_{N,H}$, and $Z_N(b)$ in Section 6.1 applied to the Y_i each year. The p-values observed here appeared more uniformly distributed, suggesting that the white noise hypothesis is more plausible for these curves. Evidently though there are some years where strong autocorrelation is observed in the differenced curves that cannot be accounted for by allowing for conditional heteroscedasticity in the series, as seen by the results of the $\mathrm{KRS}_{N,H}$ and $Z_N(b)$ tests.

By applying the same tests to the squared differences, Y_i^2, see the third panel of Table 6.4, we observed that for most series the p-values of the same white noise tests decreased, which suggests the presence of some conditional

TABLE 6.4
The p–values of suitable tests applied to curves derived from the Eurodollar curves X_i. The curves and/or the tests are specified in the top row of each panel.

Year	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
	goodness-of-fit of FAR(1) applied to X_i									
$\mathrm{KRS}^{(GF)}_{N,1}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
$\mathrm{KRS}^{(GF)}_{N,5}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
$\mathrm{KRS}^{(GF)}_{N,20}$	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
$Z^{(GF)}_N(10)$	**0.01**	**0.00**	**0.01**	**0.00**	**0.00**	**0.00**	**0.00**	**0.02**	**0.00**	**0.00**
	first difference curves $Y_i = X_i - X_{i-1}$									
$\mathrm{GK}_{N,1}$	0.76	0.13	0.25	0.40	**0.00**	**0.02**	**0.02**	0.08	0.18	0.43
$\mathrm{GK}_{N,5}$	0.49	**0.02**	0.50	0.57	**0.00**	**0.00**	0.10	0.19	0.27	0.95
$\mathrm{GK}_{N,20}$	0.42	**0.03**	0.58	0.25	**0.00**	0.08	0.17	0.14	0.53	0.99
$\mathrm{CR}_{N,Bl}$	0.69	0.08	0.48	0.53	**0.00**	**0.00**	**0.02**	0.07	0.15	0.67
$\mathrm{CR}_{N,Pz}$	0.71	0.11	0.28	0.44	**0.00**	**0.01**	**0.01**	0.06	0.15	0.33
$\mathrm{KRS}_{N,1}$	0.74	0.07	0.23	0.31	0.06	**0.02**	**0.03**	0.14	0.17	0.28
$\mathrm{KRS}_{N,5}$	0.61	**0.03**	0.50	0.52	**0.01**	**0.01**	0.19	0.25	0.22	0.91
$\mathrm{KRS}_{N,20}$	0.63	**0.05**	0.62	0.17	**0.00**	0.12	0.22	0.18	0.52	0.99
$Z_N(10)$	0.61	0.08	0.10	0.25	0.12	**0.02**	0.06	0.11	0.25	0.40
	squared first difference curves Y_i^2									
$\mathrm{GK}_{N,1}$	0.11	0.55	0.75	0.14	**0.00**	0.44	**0.04**	**0.00**	0.59	0.33
$\mathrm{GK}_{N,5}$	**0.05**	0.07	0.42	0.32	**0.00**	0.22	0.10	**0.00**	**0.03**	0.11
$\mathrm{GK}_{N,20}$	**0.02**	0.12	**0.01**	0.78	**0.00**	0.43	0.16	**0.00**	**0.01**	**0.01**
$\mathrm{CR}_{N,Bl}$	**0.01**	0.08	0.94	0.31	**0.00**	0.51	0.12	**0.00**	0.28	0.40
$\mathrm{CR}_{N,Pz}$	0.10	0.64	0.83	0.18	**0.00**	0.53	0.06	**0.00**	0.39	0.43
$\mathrm{KRS}^{(CH)}_{N,1}$	0.11	0.68	0.83	0.18	**0.00**	0.52	0.06	**0.00**	0.39	0.43
$\mathrm{KRS}^{(CH)}_{N,5}$	**0.03**	**0.05**	0.45	0.38	**0.00**	0.20	0.14	**0.00**	**0.03**	0.09
$\mathrm{KRS}^{(CH)}_{N,20}$	**0.01**	0.12	**0.00**	0.85	**0.00**	0.38	0.31	**0.00**	**0.02**	**0.00**
	goodness-of-fit of fGARCH(1, 1) model applied to Y_i									
$\mathrm{KRS}^{(GF)}_{N,1}$	0.17	0.05	1.00	1.00	1.00	0.50	0.54	0.50	**0.01**	0.98
$\mathrm{KRS}^{(GF)}_{N,5}$	0.30	1.00	0.29	1.00	1.00	1.00	**0.00**	0.26	**0.02**	0.60
$\mathrm{KRS}^{(GF)}_{N,10}$	0.60	0.82	0.93	0.83	0.99	0.35	**0.00**	1.00	0.38	**0.00**

heteroscedasticity in the differenced series. We therefore fit an fGARCH(1, 1) model to the differenced curves Y_i in each year, and checked the goodness-of-fit of these models using the statistic $\mathrm{KRS}_{N,H}$. The results of this are presented in the last panel of Table 6.4. These generally suggested that the residuals of the fGARCH(1, 1) model applied to Y_i behave like white noise, supporting the adequacy of the model.

In conclusion, this analysis suggests that 1) FAR(1) models do not fit the Eurodollor futures curves well, and 2) a functional random walk model,

$$X_i(t) = \mu(t) + \sum_{j=0}^{i} \varepsilon_j(t),$$

where the innovations ε_j evolve as a functional GARCH process cannot be rejected as an adequate model.

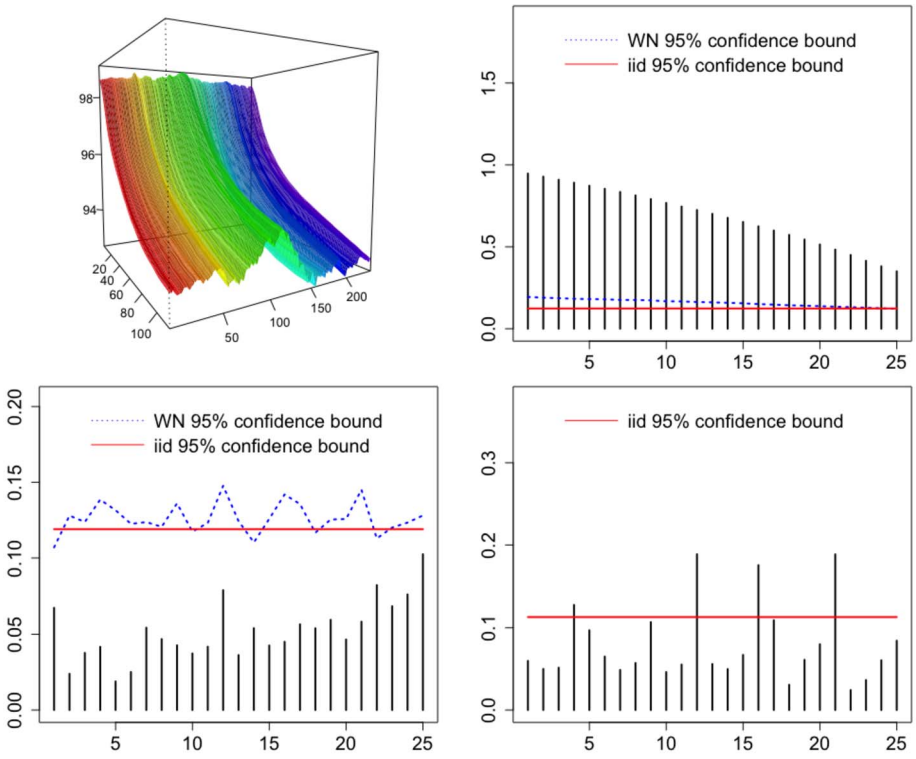

FIG 6.4. *3D Rainbow plot of the daily Eurodollar futures curves X_i in 2003, with earlier curves in red and later curves in violet (upper left). The fACF plots (see (2.10)) of the raw X_i (upper right), of the differenced curves $Y_i = X_i - X_{i-1}$ (lower left), and of the squared differenced curves Y_i^2 (lower right). The blue dotted (red solid) line is the white noise (i.i.d.) 95% critical bound for the fACF.*

7. Directions for future work

It remains an open problem how to extend several existing spectral based tests, such as in Characiejus and Rice (2020), to general, uncorrelated functional data, as well as to applications in goodness-of-fit in FTS models. An important reference that tackles this problem for univariate time series is Shao (2011), and the results in van Delft (2020) appear applicable to establish the limiting distribution of some test statistics in this class. A common problem encountered in applying white noise tests to FTS in practice is their sensitivity to outlying curves. This introduces a much larger, and largely unexplored to the knowledge of the authors, problem of developing robust methods for measuring serial dependence and performing model diagnostic checking for FTS. Recent preprints in this direction include Xin and Shang (2023) and Yeh *et al.* (2023). An emerging area in recent years is, potentially high-dimensional, vector valued FTS. White noise tests as well as diagnostic model checks for such data have not been developed

to our knowledge. See Chang *et al.* (2023), Guo and Qiao (2023), and Yuan *et al.* (2019) for some work in this direction.

Appendix A: Other diagnostics

In the analysis of scalar time series and regression, many diagnostics beyond white noise tests are applied. Perhaps the most prominent are normal QQ-plots of residuals because normality of the errors is needed to justify various F and t tests as well as prediction confidence bands. In section A.1, we review normality tests in the context of functional data. In Section A.2, we briefly discuss, chiefly by providing suitable references, other diagnostic procedures, like change-point and periodicity tests. Obviously, a FTS that exhibits such characteristics cannot be white noise. We also briefly mention a few goodness-of-fit tests beyond those investigated in Section 6.2.

A.1. Tests of normality

Verification of the normality of residuals of ARIMA and other models for scalar time series is an integral part of diagnostic checks, see e.g. Section 3.7 of Shumway and Stoffer (2017). Normality is chiefly needed to validate prediction confidence intervals produced by various methods. Standard outputs include normal QQ-plots and P-values of several normality tests applicable to i.i.d. scalar observations.

Tests of normality of functional data are studied in Górecki *et al.* (2018) and Górecki *et al.* (2020). Górecki *et al.* (2018) propose several normality tests applicable to functional time series, and the observed functions need not be independent. The starting point of these tests is the most commonly used Jarque-Bera test for i.i.d. scalar observations, Jarque and Bera (1980), and its extension to regression residuals, Jarque and Bera (1987). The idea is that standardized normal observations have approximately skewness 0 and fourth moment 3. Denoting by $\hat{\tau}$ the sample skewness and by $\hat{\kappa}$ the sample fourth moment, a direct by lengthy calculation shows that

$$\mathrm{JB}_N := N\left(\frac{\hat{\tau}^2}{6} + \frac{(\hat{\kappa}-3)^2}{24}\right) \xrightarrow{d} \chi^2(2).$$

The standard Jarque-Bera test uses the above convergence and the quantiles of the limit chi-square distribution with two degrees of freedom.

In the functional setting, the null hypothesis is that the realization $X_1, X_2, \ldots, X_N$ is generated by a functional Gaussian process. By definition, $\{X_j\}$, $X_j \in L^2$, is Gaussian if for any d, any indexes $i_1, i_2, \ldots, i_d$ and any functions $\phi_1, \phi_2, \ldots, \phi_d \in L^2$, the vector

$$\left[\langle X_{i_1}, \phi_1\rangle, \langle X_{i_2}, \phi_2\rangle, \ldots, \langle X_{i_d}, \phi_d\rangle\right]^\top$$

is normal in $\mathbb{R}^d$. Thus normality implies that the population scores ξ_{ij} in (2.4) must be normal. If the X_j are independent, the sets of scores ξ_{ij} and $\xi_{i'j'}$ are

independent if $i \neq i'$. For a fixed i, the ξ_{ij} are uncorrelated across j, hence independent under normality. Combining these observations, it is easy to show that for any p,

$$\mathrm{JB}_N^{(p)} := N \sum_{j=1}^{p} \left(\frac{\hat{\tau}_j^2}{6} + \frac{(\hat{\kappa}_j - 3)^2}{24} \right) \xrightarrow{d} \chi^2(2p), \tag{A.1}$$

where $\hat{\tau}_j$ and $\hat{\kappa}_j$ are, respectively, the sample skewness and kurtosis of the pseudo-observations

$\xi_{1,j}, \xi_{2,j}, \ldots, \xi_{N,j}$. Convergence (A.1) leads to a simple and effective test to validate if a sequence of i.i.d. functions is normal.

More complex test statistics are needed if the X_i are not white noise. According to the simulations in Górecki *et al.* (2018), if the X_i exhibit temporal dependence, the best test is based on their statistics (2.13) that combines the methodology of Lobato and Velasco (2004) and Hörmann *et al.* (2015). The *dynamic* functional principal components of Hörmann *et al.* (2015) and Panaretos and Tavakoli (2013a) decorrelate functional time series more completely than the traditional FPCs introduced in Section 2.2, and so lead to good normality tests. This decomposition was extended by Kuenzer *et al.* (2021) to a spatial setting. The functional data have the form $X(\mathbf{s}_i)$, where the $\mathbf{s}_i$ are locations on a regular spatial grid. For example, $X(\mathbf{s}_i; t)$ can be Sea Surface Temperature at location $\mathbf{s}_i$ in month t. Using this extension, Hörmann *et al.* (2021) developed a normality test for functional data with spatial dependence.

Normality tests that go beyond the skewness-kurtosis paradigm of Jarque and Bera are considered in Górecki *et al.* (2020). The authors consider many multivariate normality tests and apply them to the scores computed from functional samples or from the residuals of a functional linear model. In the first scenario, the tests are applied under the assumption that the functions in a sample are i.i.d., in the second scenario under the assumption that the linear model is valid and has i.i.d. errors. It turns out that the best tests are actually those based on skewness and kurtosis, which have to be suitably defined in the multivariate setting. Górecki *et al.* (2020) examine the normality of many data sets extensively used in FDA research. To illustrate, the well–known male/female growth curves considered in Ramsay and Silverman (2005) can be assumed normal in L^2, but the tecator curves (fat, water and protein content of meat samples) are definitely not normally distributed. Another interesting finding is that fractional anisotropy tract profiles are normal in the control group, but not normal in the MS group.

A.2. Change-point, periodicity and other goodness-of-fit tests

From the point of view of the theory of time series, the most obvious departure from stationarity, and hence from the white noise assumption, is a change point. If data structures of any form are sequentially collected, we say that there is a change point if starting from a certain point the random structure of the

data changes in some way. Thus, there is a change point $k^\star$ in the sequence $X_1, X_2, \ldots, X_N$ of functions if the distribution in L^2 of $X_1, \ldots, X_{k^\star}$ is different from the distribution of $X_{k^\star+1}, \ldots, X_N$. In change point analysis, the change point $k^\star$ is unknown and at least some characteristics of the distribution before and after the change point are unknown. In the context of FDA, the simplest form of change is the change in the mean function. Under the null hypothesis, $EX_i(t) = \mu(t)$ for some unknown function $\mu \in L^2$. Under the alternative hypothesis, $EX_i(t) = \mu_1(t)$ for $i \leq k^\star$ and $EX_i(t) = \mu_2(t)$ for $i > k^\star$. The functions μ_1 and μ_2 are unknown. In an analogous way, one can define a change point in the covariance function or other aspects of the distribution. An introduction to such tests in given in Chapter 6 of Horváth and Kokoszka (2012), with a change point detection in the FAR(1) model studied in Chapter 14. A comprehensive review of change point analysis that contains recent results related to functional data is given in Horváth and Rice (2014). A recent work that also discusses references to research on multiple change points is Chiou *et al.* (2019). Other tests of stationarity for FTS are studied in Horváth *et al.* (2014) and Aue and van Delft (2020).

Periodicity is another obvious departure from the white noise assumption, and also from stationarity. In the context of FTS, Hörmann *et al.* (2018) provide several tests of periodicity assuming that the potential period is known. For example, daily pollution curves will exhibit weekly and annual periodicity. If some method is applied to remove periodicity, one should test if the deseasonalized FTS is still periodic. Tests of Hörmann *et al.* (2018) would be useful in such contexts. Such problems have been investigated for a long time in the context of scalar time series, see McElroy and Roy (2022) for a review. Cerovecki *et al.* (2021) solve a different problem. Given a FTS, how can we test if it has an unknown period? In the analysis of scalar time series, such tests are based on the maximum of the periodogram. Cerovecki *et al.* (2021) develop analogous tests and the requisite theory in the functional and multivariate contexts.

We now discuss several papers related to the goodness-of-fit tests discussed in Section 6.2. Gabrys *et al.* (2010) derive two types of tests for the correlation of errors ε_i in the FLM (5.2). Kokoszka *et al.* (2008) develop a simple test for the significance of the kernel $\psi(\cdot, \cdot)$ in (5.2), which is also discussed in Chapter 9 of Horváth and Kokoszka (2012). A general goodness-of-fit test of the FLM (5.2) is developed in García-Portugués *et al.* (2021), and other inferential procedures in this direction can be also found in Lee *et al.* (2020) and Patilea and Sánchez-Sellero (2020). Horváth *et al.* (2022) study goodness-of-fit testing in a general functional factor model of the form

$$X_i(t) = \sum_{\ell=1}^{K} b_{i,\ell,0} f^{(N)}\ell(t; \lambda) + \epsilon_i(t),$$

where the functions $f^{(N)}\ell$ are known up to an unknown parameter vector λ. Such models are routinely used in finance.

Appendix B: A time domain test of goodness-of-fit of functional AR(1) model

B.1. Assumptions and asymptotic theory

We consider functional observations $X_1, ..., X_N$ drawn from a stationary FAR(1) model

$$X_i(t) - \mu(t) = \Psi(X_{i-1} - \mu)(t) + \varepsilon_i(t), \quad i \in \mathbb{Z},$$

where $\Psi(X)(t) = \int \psi(t,s)X(s)ds$ for a kernel function $\psi \in L^2([0,1] \times [0,1])$, and $\{\varepsilon_i, -\infty < i < \infty\}$ a mean-zero, L^2-valued white noise sequence with covariance operator $\Gamma_{\epsilon,0}$. We assume going forward without loss of generality that $\mu = 0$. For a non-decreasing integer sequence $k_N \to \infty$, put

$$\tilde{C}^{-1}(\cdot) = \sum_{j=1}^{k_N} \frac{\langle \hat{v}_j, \cdot \rangle}{\hat{\lambda}_j} \hat{v}_j$$

and

$$\widehat{\Psi}_N(\cdot) = \widehat{\Gamma}_1 \tilde{C}^{-1}(\cdot) = \frac{1}{N} \sum_{i=2}^{N} \sum_{j=1}^{k_N} \frac{\langle \hat{v}_j, \cdot \rangle}{\hat{\lambda}_j} \langle X_{i-1}, \hat{v}_j \rangle X_i.$$

Let

$$\hat{\varepsilon}_i(t) = X_i(t) - \widehat{\Psi}_N(X_{i-1})(t), \quad i \in \{2, ..., N\}$$

and

$$\hat{\gamma}_{\hat{\varepsilon},h}(u,v) = \frac{1}{N} \sum_{t=1+h}^{N} \hat{\varepsilon}_t(u)\hat{\varepsilon}_{t-h}(v).$$

The weak convergence of $\hat{\gamma}_{\hat{\varepsilon},h}$ can be established under the following assumptions:

Assumption B.1. $\|\Psi\| < 1$

Assumption B.2. $\|\widehat{\Psi}_N - \Psi\| = o_P(N^{-1/4})$.

Assumption B.3. $\{\varepsilon_i,\ -\infty < i < \infty\}$ *are i.i.d., with* $E\varepsilon_i = 0$, $E\|\varepsilon_i\|^4 < \infty$.

Below we let

$$\pi_k(\cdot) = \sum_{i=1}^{k} \langle \cdot, v_i \rangle v_i$$

denote the projection operator on the closed linear span of the first k FPC's, and

$$C^{-1}\pi_k(\cdot) = \sum_{i=1}^{k} \frac{\langle \cdot, v_i \rangle}{\lambda_i} v_i,$$

denote the Moore-Penrose inverse of C.

Assumption B.4. *The array of scalars*

$$\xi_{m,N} = \sum_{\ell=1}^{\infty} \|C^{1/2}\Psi_*^{m+1}C^{-1}\pi_{k_N}\Psi^{h-1}\Gamma_{\varepsilon,0}(v_\ell)\|^2,$$

satisfies

$$\sum_{m=1}^{\infty} \sup_{N\geq 1} \xi_{m,N} < \infty.$$

Assumption B.1 implies that a stationary, causal solution X_i to the FAR(1) equation exists, which takes the form of a linear process

$$X_i(t) = \mu(t) + \sum_{j=0}^{\infty} \Psi^j\left(\varepsilon_{i-j}\right)(t); \tag{B.1}$$

see Bosq (2000) (or Section 8.8 of Kokoszka and Reimherr (2017)). Assumption B.2 coincides with the conclusion of Lemma 3.2 of Zhang (2016), and holds for the estimator Ψ_N under decay rate conditions on the eigenvalues of C in (2.3) and for corresponding rates of increase on the parameter k_N; see Assumption 3.6 of Zhang (2016).

Assumption B.4 is used in order to show that the variables $\hat{\varepsilon}_t(u)\hat{\varepsilon}_{t-h}(v)$ are well approximated with a uniformly L^2-m-approximable triangular array as discussed in Section B.3. This property is not addressed in detail in Zhang (2016). We note that if the operators Ψ_* and C commute, then some simple arithmetic using Parseval's identity and the fact that $\{v_j\}_{j\geq 1}$ are orthonormal gives that

$$\begin{aligned}
\xi_{m,N} &= \sum_{\ell=1}^{\infty} \|\Psi_*^{m+1}C^{1/2}C^{-1}\pi_{k_N}\Psi^{h-1}\Gamma_{\varepsilon,0}(v_\ell)\|^2 \\
&\leq \|\Psi\|^{2m+2} \sum_{\ell=1}^{\infty} \left\| \sum_{j=1}^{k_N} \frac{\langle \Psi^{h-1}\Gamma_{\varepsilon,0}(v_\ell), v_j\rangle}{\lambda_j^{1/2}} v_j \right\|^2 \\
&= \|\Psi\|^{2m+2} \sum_{\ell=1}^{\infty}\sum_{j=1}^{k_N} \frac{\langle v_\ell, \Gamma_{\varepsilon,0}\Psi_*^{h-1}(v_j)\rangle^2}{\lambda_j} \\
&= \|\Psi\|^{2m+2} \sum_{j=1}^{k_N} \frac{\|\Gamma_{\varepsilon,0}\Psi_*^{h-1}(v_j)\|^2}{\lambda_j} \\
&\leq \|\Psi\|^{2m+2} \|C^{-1/2}\Psi^{h-1}\Gamma_{\varepsilon,0}^2\Psi_*^{h-1}C^{-1/2}\|_1,
\end{aligned}$$

where $\|\cdot\|_1$ denotes the trace norm. Hence, in this case, Assumption B.4 holds in light of Assumption B.1 with $\sup_{N\geq 1}\xi_{m,N}$ decreasing to zero geometrically, so long as

$$\|C^{-1/2}\Psi^{h-1}\Gamma_{\varepsilon,0}^2\Psi_*^{h-1}C^{-1/2}\|_1 < \infty. \tag{B.2}$$

Condition (B.2) coincides with the assumption of Proposition 3.1 in equation (29) of Zhang (2016). In the absence of commutativity of Ψ and C, it is unclear what conditions are needed in addition to (B.2) in order for Assumption B.4 to hold. We also note that Assumption B.4 implies that for each $1 \leq i, j \leq H$, the kernels

$$c_{i,j}(u,v,u',v') = \mathrm{Cov}\left(\varepsilon_t(u)\varepsilon_{t-i}(v) - \varepsilon_t(u) f_{t,i}^{(N)}(v), \varepsilon_t(u')\varepsilon_{t-j}(v') - \varepsilon_t(u') f_{t,j}^{(N)}(v')\right) \tag{B.3}$$

where

$$f_{t,i}^{(N)}(v) = \Gamma_{\varepsilon,0}\Psi_*^{i-1}C^{-1}\pi_{k_N}(X_{t-1})(v).$$

The function $f_{t,i}$ characterizes the effect of the estimation of Ψ has on the distribution of the residual autocovariances.

Theorem B.1. *Under Assumptions B.1–B.4, then jointly for $h \in \{1, ..., H\}$,*

$$\sqrt{N}\hat{\gamma}_{\hat{\varepsilon},h}(u,v) := \frac{1}{\sqrt{N}} \sum_{t=1+h}^{N} \hat{\varepsilon}_t(u)\hat{\varepsilon}_{t-h}(v) \xrightarrow{d} G_h(u,v), \tag{B.4}$$

where $G_h(u,v)$, $h \in \{1, ..., H\}$ are mean-zero jointly Gaussian process in $L^2([0,1]^2)$, with cross-covariance kernels $E[G_i(u,v)G_j(u',v')] = c_{i,j}(u,v,u',v')$, where the $c_{i,j}$ are defined in (B.3).

Corollary B.1. *Under the assumptions of Theorem B.1,*

$$\mathrm{KRS}_{N,H}^{(GF)} := N \sum_{h=1}^{H} \|\hat{\gamma}_{\hat{\varepsilon},h}\|^2 \xrightarrow{d} V_H,$$

where V_H is a weighted sum of independent χ_1^2 random variables, with the weights determined by the kernels in (B.3).

Theorem B.1 is proven in Section B.4 following the formulation and a proof of a required CLT for triangular arrays of functions in Section B.3. We first explain in Section B.2 how to approximate the distribution of the limit V_H in Corollary B.1

B.2. Approximation of the limit V_H in Corollary B.1

The function $c_{i,j}$ may be estimated by

$$\begin{aligned}
&\hat{c}_{i,j}(u,v,u',v') \qquad\qquad (B.5)\\
&= \frac{1}{N} \sum_{k=1+\max\{i,j\}}^{N} (\hat{\varepsilon}_k(u)\hat{\varepsilon}_{k-i}(v) - \hat{\varepsilon}_k(u)\hat{f}_{k,i}(v))(\hat{\varepsilon}_k(u')\hat{\varepsilon}_{k-j}(v') - \hat{\varepsilon}_k(u')\hat{f}_{k,j}(v'))\\
&\quad - \left(\frac{1}{N}\sum_{k=1+i}^{N} \hat{\varepsilon}_k(u)\hat{\varepsilon}_{k-i}(v) - \hat{\varepsilon}_k(u)\hat{f}_{k,i}(v)\right)
\end{aligned}$$

$$\times \left(\frac{1}{N} \sum_{k=1+j}^{N} \hat{\varepsilon}_k(u') \hat{\varepsilon}_{k-j}(v') - \hat{\varepsilon}_k(u') \hat{f}_{k,j}(v') \right),$$

where

$$\hat{f}_{t,h} = \frac{1}{N} \sum_{j=h+1}^{N} \sum_{k=1}^{k_N} \hat{\lambda}_k^{-1} \left\langle X_{t-1}, \hat{\phi}_k \right\rangle \left\langle X_{j-1}, \hat{\phi}_k \right\rangle \hat{\varepsilon}_{j-h}.$$

The limiting random variable V_H in Corollary B.1 takes the form of a squared norm of a Gaussian process

$$V_H \stackrel{d}{=} \sum_{h=1}^{H} \iint G_h^2(u,v) du dv,$$

and hence may be further expressed as a weighted sum of independent χ_1^2 random variables with weights determined by the covariance kernels $c_{i,j}$ in (B.3). Using then the estimators $\hat{c}_{i,j}$, our goal is to approximate the distribution of V_H using the Welch-Satterthwaite (WS) approximation that approximates V_H by $X \sim \beta \chi_v^2$ where χ_v^2 is a chi–square random variable with v degrees of freedom and $\beta > 0$. To estimate β, v, the WS method equates the first two moments of V_H and X; $E[V_H] = E[X]$ and $Var[V_H] = Var[X]$. From this, we have

$$\hat{\beta} = \frac{\hat{\sigma}_H^2}{2\hat{\mu}_H}, \qquad \hat{v} = \frac{2\hat{\mu}_H^2}{\hat{\sigma}_H^2}.$$

The $\hat{\mu}_H$ is the estimated first moment of V_H, which is obtained from

$$\hat{\mu}_H = \sum_{i=1}^{H} \iint \hat{c}_{i,i}(t,s,t,s) dt ds \approx \frac{1}{J^2} \sum_{t,s} \sum_{i=1}^{H} \hat{c}_{i,i}(t,s,t,s),$$

where J is the number of points on the grid on which the functional data are observed, and $\hat{c}_{i,j}$ are defined in (B.5). The $\hat{\sigma}_H^2$ is the estimated second moment of V_H, calculated from

$$\begin{aligned} \hat{\sigma}_H^2 &= 2 \sum_{1 \le i,j \le H} \iiint \hat{c}_{i,j}(t,s,u,v) dt ds du dv \\ &\approx \frac{2}{M} \sum_{k=1}^{M} \sum_{1 \le i,j \le H} \hat{c}_{i,j}(v_{k1}, v_{k2}, v_{k3}, v_{k4}), \end{aligned}$$

where $\{(v_{k1}, v_{k2}, v_{k3}, v_{k4}),\ 1 \le k \le M\}$ is a set of M randomly selected points on the grid (Monte Carlo integration). A typical choice is $M = 10,000$.

B.3. Central limit theorem for triangular arrays of dependent random variables in $\mathbb{H}$

In this section we formulate a limit result needed to prove Theorem B.1. All functions are assumed to be measurable with respect to suitable σ-algebras. Let $\mathbb{H}$ denote a separable Hilbert space.

Definition B.1. We say that $\{X_{t,n} : t \in \mathbb{Z}, n \in \mathbb{N}\}$ is a *uniformly* $L^\nu - m$*-approximable triangular array* in $\mathbb{H}$ if

a) $EX_{t,n} = 0$,

b) For each n, there exists a measurable space $\mathbb{E}_n$ and a function $g_n : \mathbb{E}_n^\infty \mapsto \mathbb{H}$ such that

$$X_{t,n} = g_n(\varepsilon_{t,n}, \varepsilon_{t-1,n}, ...), \tag{B.6}$$

where the $\varepsilon_{t,n}, t \in \mathbb{Z}$, are independent and identically distributed elements of $\mathbb{E}_n$,

c) if $\theta_m = \sup_{n\geq 1}[E\|X_{0,n} - X_{0,n}^{(m)}\|^\nu)]^{1/\nu}$, then

$$\sum_{m=0}^{\infty} \theta_m < \infty, \tag{B.7}$$

where $X_{0,n}^{(m)} = g_n(\varepsilon_{0,n}, \varepsilon_{-1,n}, ..., \varepsilon_{-m,n}, \varepsilon'_{-(m+1),n}, \varepsilon'_{-(m+2),n}, ...)$, with $\{\varepsilon'_{t,n}\ t \in \mathbb{Z}\}$ an independent and identically distributed sequence in $\mathbb{E}_n$, independent of, but with the same distribution as, $\{\varepsilon_{t,n},\ t \in \mathbb{Z}\}$.

In the definition of stationary m-approximable time series, the analog of representation (B.6) takes the form $X_t = g(\varepsilon_t, \varepsilon_{t-1}, ...)$ and the assumptions in part c) take a correspondingly simpler form, see Hörmann and Kokoszka (2010) or Chapter 16 of Horváth and Kokoszka (2012), among many contributions that have used this concept. In Definition B.1, a different function g_n can be used in every row, indexed by n, of the array $X_{t,n}$, but the "tails" of the functions g_n must "decay" uniformly fast, as specified in condition c).

We aim to establish conditions under which the partial sum

$$S_n = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_{t,n}$$

of variables forming a uniformly $L^\nu - m$-approximable triangular array converge weakly in $\mathbb{H}$ to a Gaussian element. We begin by noting that if $\nu \geq 2$, the covariance operator of S_n is

$$\mathcal{F}_n = \sum_{h=1-n}^{n-1} \left(1 - \frac{|h|}{n}\right) \Gamma_{h,n}, \quad \Gamma_{h,n} = E[X_{0,n} \otimes X_{h,n}].$$

The following theorem refers to trace class operators. The definition and a brief exposition of their chief properties are given in Section 13.5 of Horváth and Kokoszka (2012), but such operators are studied in many textbooks and monographs. In particular, every covariance operator is trace class, and its trace is equal to $\sum_{j=1}^{\infty} \lambda_j$, cf. (2.4). Gaussian elements of $\mathbb{H}$ are concisely explained in Section 11.3 of Kokoszka and Reimherr (2017), as well as in many other textbooks.

Theorem B.2. *Suppose* $\{X_{t,n} : t \in \mathbb{Z}, n \in \mathbb{N}\}$ *is a uniformly* $L^\nu - m$*-approximable triangular array for some* $\nu \geq 2$. *If, in addition, there exists a positive definite operator* $\mathcal{F}$ *on* $\mathbb{H}$ *satisfying*

i) $\mathcal{F}$ *is trace class,*
ii) For each $v \in \mathbb{H}$, $\langle \mathcal{F}_n(v), v\rangle \to \langle \mathcal{F}(v), v\rangle$, *as* $n \to \infty$,
iii) $\mathrm{tr}(\mathcal{F}_n) \to \mathrm{tr}(\mathcal{F})$, *as* $n \to \infty$.
Then

$$S_n \overset{d(\mathbb{H})}{\to} Z,$$

where Z *is a Gaussian element in* $\mathbb{H}$ *with covariance operator* $\mathcal{F}$, *and* $\overset{d(\mathbb{H})}{\to}$ *denotes weak convergence in* $\mathbb{H}$.

Before proceeding with the proof, we state a proposition that facilitates the application of Theorem B.2 in case of integral operators acting on $\mathbb{H} = L^2([0,1]^p)$, a case directly applicable to the tests we consider.

Proposition B.1. *Suppose* $X_{t,n}$ *satisfy Definition B.1 with* $\mathbb{H} = L^2([0,1]^p)$ *and* $\nu \geq 2$. *Set* $\gamma_{h,n}(\mathbf{t},\mathbf{s}) = Cov(X_{0,n}(\mathbf{t}), X_{h,n}(\mathbf{s}))$.
If there exist autocovariance operators $\Gamma_{h,\infty}$ *with kernels* $\gamma_{h,\infty}$ *satisfying*
1) $\|\gamma_{h,n} - \gamma_{h,\infty}\| \to 0$, *as* $n \to \infty$,
2)

$$\int |\gamma_{h,n}(\mathbf{t},\mathbf{t}) - \gamma_{h,\infty}(\mathbf{t},\mathbf{t})| d\mathbf{t} \to 0, \quad as\ n \to \infty,$$

then conditions i)-iii) of Theorem B.2 hold with

$$\mathcal{F} = \sum_{h\in\mathbb{Z}} \Gamma_{h,\infty}$$

and

$$\mathrm{tr}(\mathcal{F}) = \sum_{h=-\infty}^{\infty} \int \gamma_{h,\infty}(\mathbf{t},\mathbf{t}) d\mathbf{t}.$$

Proof. We first verify condition i). In this proof we let $c_\dagger$ denote unimportant positive numeric constants that may change from line to line. Using the notation in Definition B.1, we note that for each n, $X_{0,n}$ and $X_{h,n}^{(h)}$ are independent. It then follows by the Cauchy-Schwarz inequality that

$$\begin{aligned}
\|\Gamma_{h,n}\|_{\mathcal{S}} &= \left(\iint \gamma_{h,n}^2(\mathbf{t},\mathbf{s}) d\mathbf{t} d\mathbf{s} \right)^{1/2} \\
&= \left(\iint [\mathrm{Cov}(X_{0,n}(\mathbf{t}), X_{h,n}(\mathbf{s}) - X_{h,n}^{(h)}(\mathbf{s}))]^2 d\mathbf{t} d\mathbf{s} \right)^{1/2} \\
&\leq \left(\iint E X_{0,n}^2(\mathbf{t}) E[X_{h,n}(\mathbf{s}) - X_{h,n}^{(h)}(\mathbf{s}))]^2 d\mathbf{t} d\mathbf{s} \right)^{1/2} \\
&= (E\|X_{0,n}^2\|)^{1/2} (E\|X_{h,n} - X_{h,n}^{(h)}\|^2)^{1/2} \\
&\leq c_\dagger \theta_h.
\end{aligned}$$

It follows similarly that $\left|\int \gamma_{h,n}(t,t) dt\right| \leq c_\dagger \theta_h$. To verify ii), observe that condition 1) implies

$\langle \Gamma_{h,n}(v), v\rangle \to \langle \Gamma_{\infty,h}(v), v\rangle$. Therefore, by the dominated convergence theorem,

$$\langle \mathcal{F}_n(v), v\rangle = \sum_{h=1-n}^{n-1} \left(1 - \frac{|h|}{n}\right) \langle \Gamma_{n,h}(v), v\rangle \to \sum_{h\in\mathbb{Z}} \langle \Gamma_{\infty,h}(v), v\rangle = \langle \mathcal{F}(v), v\rangle.$$

The dominated convergence theorem is applicable here since as a result of the Cauchy-Schwarz inequality,

$$|\langle \Gamma_{n,h}(v), v\rangle| \le \|\Gamma_{n,h}(v)\|\|v\| \le \|\Gamma_{n,h}\|_{\mathcal{S}}\|\|v\|^2 \le c_\dagger \theta_h \|v\|^2, \tag{B.8}$$

and the right hand side of (B.8) is summable for each $v \in \mathbb{H}$ as a result of the assumption in (B.7). A similar application of the Cauchy-Schwarz inequality and the dominated convergence theorem also give that

$$\mathrm{tr}(\mathcal{F}_n) = \sum_{h=1-n}^{n-1} \left(1 - \frac{|h|}{n}\right) \int \gamma_{n,h}(\mathbf{t}, \mathbf{t}) d\mathbf{t} \to \mathrm{tr}(\mathcal{F}),$$

which implies *iii)*. □

We next state Theorem 6.1 of Neumann and Paparoditis (2008), which is an important tool in the proof of Theorem B.2. We first formulate a technical assumption they use.

Assumption B.5. *Scalar random variables $X_{k,n}$ form a stationary sequence for each n, $EX_{k,n} = 0$,*
$EX_{k,n}^2 \le C < \infty$ and for $n \ge n_0$ there is a nonincreasing and summable sequence of $\theta_r \ge 0$ such that for all indices

$$1 \le s_1 < s_2 < \cdots < s_u < s_u + r = t_1 \le t_2 \le n,$$

$$|\mathrm{Cov}\left(f(X_{s_1,n}, \ldots, X_{s_u,n}), X_{t_1,n}\right)| \le \theta_r \left\{Ef^2(X_{s_1,n}, \ldots, X_{s_u,n}\right\}^{1/2} \tag{B.9}$$

for all square integrable functions $f : \mathbb{R}^u \to \mathbb{R}$, and for all bounded functions $f : \mathbb{R}^u \to \mathbb{R}$

$$|\mathrm{Cov}\left(f(X_{s_1,n}, \ldots, X_{s_u,n}), X_{t_1,n}X_{t_2,n}\right)| \le \theta_r \|f\|_\infty. \tag{B.10}$$

Assumption B.5 quantifies dependence by θ_r that is indexed by the separation r between s_u and t_1.

Theorem B.3 (Neumann and Paparoditis)**.** *Suppose scalar random variables $X_{k,n}$ satisfy Assumption B.5, and, as $n \to \infty$,*

$$\forall\ \epsilon > 0 \quad \frac{1}{n}\sum_{k=1}^{n} EX_{k,n}^2 I\left\{|n^{-1/2}X_{k,n} > \epsilon\right\} \to 0 \tag{B.11}$$

and

$$\frac{1}{n}\mathrm{Var}\left[X_{1,n} + \cdots + X_{n,n}\right] \to \sigma^2 \in [0, \infty). \tag{B.12}$$

Then,

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n} X_{k,n} \xrightarrow{d} N(0,\sigma^2). \tag{B.13}$$

Proof of Theorem B.2. We let $c_\dagger$ denote an unimportant positive numerical constant that may change between uses. For each $v \in \mathbb{H}$, let $X_{t,n}(v) = \langle X_{t,n}, v\rangle$. We first show that

$$\langle S_n, v\rangle = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_{t,n}(v) \xrightarrow{d} N(0, \langle \mathcal{F}(v), v\rangle), \tag{B.14}$$

by showing that the $X_{t,n}(v)$ satisfy the conditions of Theorem B.3. By (B.6), for each n, the $X_{t,n}(v)$ form a stationary sequence. By conditions a) and c) of Definition B.1, and the assumption $\nu \geq 0$, $EX_{t,n}(v) = 0$ and $EX_{t,n}^2(v) \leq E\|X_{t,n}\|^2\|v\|^2 \leq c_\dagger\|v\|^2 (= C)$.

Let $X_{t,n}^{(m)}(v) = \langle X_{t,n}^{(m)}, v\rangle$. We then have that, for each positive integer u, square integrable $f : \mathbb{R}^u \to \mathbb{R}$, and indices $1 \leq s_1 < \cdots < s_u < s_u + m = t_1 \leq t_2 \leq n$, by the Cauchy-Schwarz inequality,

$$\begin{aligned}
&|\mathrm{Cov}(f(X_{s_1,n}(v), ..., X_{s_u,n}(v)), X_{t_1,n}(v))| \\
&\quad = |\mathrm{Cov}(f(X_{s_1,n}(v), ..., X_{s_u,n}(v)), X_{t_1,n}(v) - X_{t,n}^{(m)}(v))| \\
&\quad \leq \sqrt{Ef^2(X_{s_1,n}(v), ..., X_{s_u,n}(v))}E\langle X_{t_1,n} - X_{t,n}^{(m)}, v\rangle^2 \\
&\quad \leq \sqrt{Ef^2(X_{s_1,n}(v), ..., X_{s_u,n}(v))}\theta_m\|v\|.
\end{aligned}$$

Moreover, for each $f : \mathbb{R}^u \to \mathbb{R}$ such that $\|f\|_\infty = \sup_{x\in\mathbb{R}^u} |f(x)| < \infty$, we have using the triangle inequality that

$$\begin{aligned}
&|\mathrm{Cov}(f(X_{s_1,n}(v), ..., X_{s_u,n}(v)), X_{t_1,n}(v)X_{t_2,n}(v)))| \\
&\quad = |\mathrm{Cov}(f(X_{s_1,n}(v), ..., X_{s_u,n}(v)), X_{t_1,n}(v)X_{t_2,n}(v) - X_{t,n}^{(m)}(v)X_{t_2,n}^{(m)}(v))| \\
&\quad = |E\{[f(X_{s_1,n}(v), ..., X_{s_u,n}(v)) \\
&\qquad - Ef(X_{s_1,n}(v), ..., X_{s_u,n}(v))][X_{t_1,n}(v)(X_{t_2,n}(v) - X_{t_2,n}^{(m)}(v))]\}| \\
&\qquad + |E\{[f(X_{s_1,n}(v), ..., X_{s_u,n}(v)) \\
&\qquad - Ef(X_{s_1,n}(v), ..., X_{s_u,n}(v))][X_{t_2,n}(v)^{(m)}(X_{t_1,n}(v) - X_{t_1,n}^{(m)}(v))]\}| \\
&\quad =: R_{1,n} + R_{2,n}.
\end{aligned}$$

By the Cauchy Schwarz inequality,

$$\begin{aligned}
R_{1,n} &\leq \left(E[f(X_{s_1,n}(v), ..., X_{s_u,n}(v)) - Ef(X_{s_1,n}(v), \ldots, X_{s_u,n}(v))]^2 X_{t_1,n}^2(v)\right)^{1/2} \\
&\qquad \times \left(E(X_{t_2,n}(v) - X_{t_2,n}^{(m)}(v))^2\right)^{1/2} \\
&\leq c_\dagger\|f\|_\infty\theta_m.
\end{aligned}$$

It follows similarly that $R_{2,n} \leq c_\dagger \|f\|_\infty \theta_m$, and hence

$$|\mathrm{Cov}(f(X_{s_1,n}(v), ..., X_{s_u,n}(v)), X_{t_1,n}(v)X_{t_2,n}(v)))| \leq c_\dagger \|f\|_\infty \theta_m.$$

This completes the verification of Assumption B.5. It remains to verify the Lindeberg condition (B.11) and the convergence (B.12). Using dominated convergence theorem, for all $\epsilon > 0$,

$$\frac{1}{n}\sum_{t=1}^{n} EX_{t,n}^2(v) I\{|X_{t,n}(v)|/\sqrt{n} > \epsilon\} = EX_{0,n}^2(v) I\{|X_{0,n}(v)|/\sqrt{n} > \epsilon\} \to 0.$$

By condition *ii)*,

$$\frac{1}{n}\mathrm{Var}\left[\sum_{t=1}^{n} X_{t,n}(v)\right] = \mathrm{Var}[\langle S_n, v\rangle] = \langle \mathcal{F}_n(v), v\rangle \to \langle \mathcal{F}(v), v\rangle.$$

As a result, by Theorem B.3, (B.14) holds.

We now aim to show that the sequence S_n is uniformly tight in $\mathbb{H}$ by applying Lemmas 14-16 of Cerovecki and Hörmann (2017). For any orthonormal basis $\{v_j\}_{j\geq 1}$ of $\mathbb{H}$, using their notation let $p_j^{(n)} = \langle \mathcal{F}_n(v_j), v_j\rangle$, and $p_j^{(0)} = \langle \mathcal{F}(v_j), v_j\rangle$. Then $p_j^{(n)}, p_j^{(0)}$ satisfy the following conditions of Lemma 14 of Cerovecki and Hörmann (2017): a) $p_j^{(n)} \geq 0$, which follows since $\mathcal{F}_n$ is positive definite. b) $\lim_{n\to\infty} p_j^{(n)} = p_j^{(0)}$, which follows from condition *ii)*. (c) $\sum_{j=1}^{\infty} p_j^{(0)} = p < \infty$, which follows with $p = \mathrm{tr}(\mathcal{F})$. (d) $\lim_{n\to\infty} \sum_{j=1}^{\infty} p_j^{(n)} = p$, which follows from condition *iii)*. (e) $\sum_{j=1}^{\infty} p_j^{(n)} < \infty$ for all $n \geq 1$, which follows since by the Cauchy-Schwarz inequality, $|E\langle X_{0,n}, X_{h,n}\rangle| \leq c_\dagger \theta_{|h|}$, and so

$$\sum_{j=1}^{\infty} p_j^{(n)} = \mathrm{tr}(\mathcal{F}_n) = \sum_{h=1-n}^{n-1} \left(1 - \frac{|h|}{n}\right) E\langle X_{0,n}, X_{h,n}\rangle \leq c_\dagger \sum_{h=1}^{\infty} \theta_h < \infty.$$

As a result of these conditions, it may be shown as in Lemma 16 of Cerovecki and Hörmann (2017) that S_n is a tight sequence in $\mathbb{H}$. The result now follows from Prohorov's Theorem, see e.g. p. 46 of Bosq (2000). □

B.4. Proof of Theorem B.1

Let

$$Y_{j,N}^{(h)}(u,v) = \varepsilon_j(u)\varepsilon_{j-h}(v) - \varepsilon_j(u)\Gamma_{\varepsilon,0}\Psi_*^{h-1}C^{-1}\pi_{k_N}(X_{j-1})(v).$$

Using that $\hat{\varepsilon}_j = \varepsilon_j + (\Psi - \widehat{\Psi}_N)(X_{j-1})$, it follows from Assumption B.2 as in the proof of Theorem 3.3 of Zhang (2016) that

$$\iint \left[\frac{1}{\sqrt{N}}\sum_{t=1+h}^{N} \hat{\varepsilon}_t(u)\hat{\varepsilon}_{t-h}(v) - \frac{1}{\sqrt{N}}\sum_{t=1+h}^{N} Y_{t,N}^{(h)}(u,v)\right]^2 dudv = o_P(1).$$

As a result, and since H is fixed, the conclusion of the theorem follows upon showing that

$$\frac{1}{\sqrt{N}} \sum_{t=1+h}^{N} Y_{t,N}^{(h)}(u,v) \xrightarrow{d} G_h(u,v),$$

jointly for $h \in \{1,...,H\}$. We note that $Y_{t,N}^{(h)}$ forms a triangular array of row-wise stationary variables, and the same holds for the vector valued functional process $V_{t,N} = (Y_{t,N}^{(1)},...,Y_{t,N}^{(H)})^\top$, and for any linear combination of projections $P_{t,N} = a_1\langle Y_{t,N}^{(1)}, v_1\rangle_2 + \cdots + a_H\langle Y_{t,N}^{(H)}, v_H\rangle_2$, for any $a_1,...,a_H \in \mathbb{R}$, and $v_1,...,v_H \in L^2([0,1]^2)$, where $\langle\cdot,\cdot\rangle_2$ denotes the standard innerproduct on $L^2([0,1]^2)$. Moreover, if $Y_{t,N}^{(h)}$ is uniformly $L^2 - m$-approximable for each $h \in \{1,...,H\}$, then so are $V_{t,N}$ and $P_{t,N}$. Hence it is enough to show that for any $h \in \{1,..,H\}$, $Y_{t,N}^{(h)}$ is uniformly L^2-m-approximable, and satisfies conditions i)-iii) of Theorem B.2.

Let

$$X_j^{(m)} = \sum_{\ell=0}^{m} \Psi^\ell(\varepsilon_{j-\ell}) + \sum_{\ell=m+1}^{\infty} \Psi^\ell(\varepsilon'_{j-\ell}),$$

where $\{\varepsilon'_\ell,\ -\infty < \ell < \infty\}$ is independent of $\{\varepsilon_\ell,\ -\infty < \ell < \infty\}$, but with the same distribution. Let

$$Y_{t,N}^{(h,m)}(u,v) = \varepsilon_t(u)\varepsilon_{t-h}(v) - \varepsilon_t(u)\Gamma_{\varepsilon,0}\Psi_*^{h-1}C^{-1}\pi_{k_N}(X_{t-1}^{(m)})(v).$$

Then if $m > H$, using that ε_t and $X_{t-1}, X_{t-1}^{(m)}$ are uncorrelated, we obtain that

$$\begin{aligned} &E\|Y_{t,N}^{(h)} - Y_{t,N}^{(h,m)}\|^2 \\ &\quad = E\|\varepsilon_t \otimes \Gamma_{\varepsilon,0}\Psi_*^{h-1}C^{-1}\pi_{k_N}(X_{t-1}) - \varepsilon_t \otimes \Gamma_{\varepsilon,0}\Psi_*^{h-1}C^{-1}\pi_{k_N}(X_{t-1}^{(m)})\|^2 \\ &\quad = E\|\varepsilon_t\|^2 E\|\Gamma_{\varepsilon,0}\Psi_*^{h-1}C^{-1}\pi_{k_N}(X_{t-1} - X_{t-1}^{(m)})\|^2. \end{aligned} \tag{B.15}$$

Since

$$\begin{aligned} X_t - X_t^{(m)} &= \sum_{\ell=m+1}^{\infty} \Psi^\ell(\varepsilon_{t-\ell}) + \sum_{\ell=m+1}^{\infty} \Psi^\ell(\varepsilon'_{t-\ell}) \\ &= \Psi^{m+1}\left(\sum_{\ell=0}^{\infty} \Psi^\ell \varepsilon_{t-(m+1)-\ell}\right) + \Psi^{m+1}\left(\sum_{\ell=0}^{\infty} \Psi^\ell \varepsilon'_{t-(m+1)-\ell}\right) \\ &\stackrel{d}{=} \Psi^{m+1}(X_0) + \Psi^{m+1}(X'_0), \end{aligned}$$

where X'_0 is independent of, but has the same distribution as, X_0. Continuing then on the right hand side of (B.15), we have by expanding $E\|\Gamma_{\varepsilon,0}\Psi_*^{h-1}C^{-1}\pi_{k_N}$ $[\Psi^{m+1}(X_0) + \Psi^{m+1}(X'_0)]\|^2$ that

$$E\|Y_{t,N}^{(h)} - Y_{t,N}^{(h,m)}\|^2 = 2E\|\varepsilon_t\|^2 E\|\Gamma_{\varepsilon,0}\Psi_*^{h-1}C^{-1}\pi_{k_N}\Psi^{m+1}(X_0)\|^2.$$

Using Parseval's identity and the fact that the adjoint of a linear operator R, R^*, satisfies

$\langle Rx, y\rangle = \langle x, R^* y\rangle$, we get that,

$$\begin{aligned}
E\|\Gamma_{\varepsilon,0}\Psi_*^{h-1}C^{-1}\pi_{k_N}\Psi^{m+1}(X_0)\|^2 &= \sum_{\ell=1}^{\infty} E\langle \Gamma_{\varepsilon,0}\Psi_*^{h-1}C^{-1}\pi_{k_N}\Psi^{m+1}(X_9), v_\ell\rangle^2 \\
&= \sum_{\ell=1}^{\infty} E\langle X_0, \underbrace{\Psi_*^{m+1}C^{-1}\pi_{k_N}\Psi^{h-1}\Gamma_{\varepsilon,0}(v_\ell)}_{=:\, A(v_\ell)}\rangle^2 \\
&= \sum_{\ell=1}^{\infty} \iint EX_0(t)X_0(s)A(v_\ell)(t)A(v_\ell)(s)dtds \\
&= \sum_{\ell=1}^{\infty} \langle CA(v_\ell), A(v_\ell)\rangle \\
&= \sum_{\ell=1}^{\infty} \langle C^{1/2}A(v_\ell), C^{1/2}A(v_\ell)\rangle \\
&= \sum_{\ell=1}^{\infty} \|C^{1/2}\Psi_*^{m+1}C^{-1}\pi_{k_N}\Psi^{h-1}\Gamma_{\varepsilon,0}(v_\ell)\|^2 \\
&= \xi_{m,N}.
\end{aligned}$$

Hence $E\|Y_{t,N}^{(h)} - Y_{t,N}^{(h,m)}\|^2 \le c_\dagger \xi_{m,N}$, and so by Assumption B.4, $Y_{t,N}^{(h)}$ is uniformly $L^2 - m$-approximable.

We now aim to show that conditions i)-iii) of Theorem B.2 hold for the sequence $Y_{t,N}^{(h)}$. By expanding the products and using that ε_i is a white noise sequence, we obtain that for all $k \neq 0$,

$EY_{0,N}^{(h)}(u,v)Y_{k,N}^{(h)}(u',v') = 0$, and

$$\begin{aligned}
&EY_{0,N}^{(h)}(u,v)Y_{0,N}^{(h)}(u',v') \\
&\quad = C(u,u')C(v,v') - C(u,u')E\varepsilon_{-h}(v)f_{0,h}^{(N)}(v') \\
&\qquad - C(u,u')E\varepsilon_{-h}(v')f_{0,h}^{(N)}(v) + C(u,u')Ef_{0,h}^{(N)}(v)f_{0,h}^{(N)}(v').
\end{aligned}$$

For $N > M$, we have that using again Parseval's identity

$$\begin{aligned}
&\int\int \left[E\varepsilon_{-h}(v)f_{0,h}^{(N)}(v') - E\varepsilon_{-h}(v)f_{0,h}^{(M)}(v')\right]^2 dvdv' \\
&\quad \le E\|\varepsilon_{-h}\|^2 E\|\Gamma_{\varepsilon,0}\Psi_*^{h-1}C^{-1}[\pi_{k_N}(X_{-1}) - \pi_{k_M}(X_{-1})]\|^2 \\
&\quad = E\|\varepsilon_{-h}\|^2 E\left\|\sum_{\ell=k_M+1}^{k_N} \frac{1}{\lambda_\ell}\langle X_{t-1}, v_\ell\rangle \Gamma_{\varepsilon,0}\Psi_*^{h-1}(v_\ell)\right\|^2 \\
&\quad = E\|\varepsilon_{-h}\|^2 E\sum_{j=1}^{\infty}\left\langle \sum_{\ell=k_M+1}^{k_N} \frac{1}{\lambda_\ell}\langle X_{t-1}, v_\ell\rangle \Gamma_{\varepsilon,0}\Psi_*^{h-1}(v_\ell), v_j\right\rangle^2
\end{aligned}$$

$$
\begin{aligned}
&= E\|\varepsilon_{-h}\|^2 E\sum_{j=1}^{\infty}\sum_{\ell,\ell'=k_M+1}^{k_N}\frac{1}{\lambda_\ell\lambda'_\ell}\underbrace{E\langle X_{t-1},v_\ell\rangle\langle X_{t-1},v_{\ell'}\rangle}_{=\lambda_\ell\mathbb{1}\{\ell=\ell'\}}\\
&\quad\times\langle v_\ell,\Psi^{h-1}\Gamma_{\varepsilon,0}(v_j)\rangle\langle v_{\ell'},\Psi^{h-1}\Gamma_{\varepsilon,0}(v_j)\rangle\\
&= \sum_{\ell=k_M+1}^{k_N}\frac{1}{\lambda_\ell}\sum_{j=1}^{\infty}\langle\Gamma_{\varepsilon,0}\Psi_*^{h-1}(v_\ell),v_j\rangle^2\\
&= \sum_{\ell=k_M+1}^{k_N}\frac{\|\Gamma_{\varepsilon,0}\Psi_*^{h-1}(v_\ell)\|^2}{\lambda_\ell}\to 0
\end{aligned}
$$

by Assumption B.4 as $\min\{N,M\}\to\infty$, since k_N is non-decreasing and tends to infinity. Therefore $E\varepsilon_{-h}(v)f_{0,h}^{(N)}(v')$ is Cauchy in $L^2([0,1]\times[0,1])$, and hence has a limit. It may be shown similarly that $Ef_{0,h}^{(N)}(v)f_{0,h}^{(N)}(v')$ is Cauchy, and hence there exists a covariance kernel $c_{h,h}$ so that

$EY_{0,N}^{(h)}(u,v)Y_{0,N}^{(h)}(u',v') \xrightarrow{L^2} c_{h,h}(u,v,u',v')$. By the Cauchy-Schwarz and Jensen's inequality,

$$
\begin{aligned}
&\int\left|E\varepsilon_{-h}(v)f_{0,h}^{(N)}(v)-E\varepsilon_{-h}(v)f_{0,h}^{(M)}(v)\right|dv\\
&\quad\le(E\|\varepsilon_{-h}\|^2)^{1/2}(E\|f_{0,h}^{(N)}-f_{0,h}^{(M)}\|^2)^{1/2}\to 0
\end{aligned}
$$

as $\min\{N,M\}\to\infty$. It hence also follows similarly that

$$
\iint|EY_{0,N}^{(h)}(u,v)Y_{0,N}^{(h)}(u,v)-c_{h,h}(u,v,u,v)|dudv\to 0,\quad N\to\infty.
$$

Therefore the conditions of Proposition B.1 are satisfied with $\gamma_{0,\infty}=c_{h,h}$ and $\gamma_{h,\infty}=0$, which in turn imply that condition i)-iii) of Theorem B.2 hold. Applying this result gives

$$
\frac{1}{\sqrt{N}}\sum_{t=1+h}^{N}Y_{t,N}^{(h)}(u,v)\xrightarrow{d}G_h(u,v),
$$

where G_h is Gaussian with covariance $c_{h,h}$. The joint result follows upon noting that for each $h,r\in\{1,...,H\}$, $EY_{0,N}^{(h)}(u,v)Y_{k,N}^{(r)}(u',v')=0$, if $k\neq 0$, and

$$
EY_{0,N}^{(h)}(u,v)Y_{0,N}^{(r)}(u',v')\to c_{h,r}(u,v,u',v').
$$

Appendix C: Additional tables

TABLE C.1
Empirical sizes (in percent) based on 1000 replications. The tests $\mathrm{KRS}_{N,H}^{(GF)}$, $Z_N^{(GF)}(b)$ *are applied to evaluate the goodness-of-fit of an FAR(1) model with the Gaussian kernel.*

	DGP	FAR(1)–BM				FAR(1)–fGARCH			
S	N	100		250		100		250	
	Nominal level	5%	1%	5%	1%	5%	1%	5%	1%
0.5	$\mathrm{KRS}_{N,1}$	4.9	1.2	5.2	0.9	4.6	1.4	5.3	1.7
	$\mathrm{KRS}_{N,5}$	5.3	1.1	4.2	1.4	4.0	0.6	3.9	1.1
	$\mathrm{KRS}_{N,20}$	7.1	1.7	5.6	1.3	5.5	1.8	5.2	1.2
	$Z_N(10)$	4.3	1.0	4.6	0.8	5.0	0.7	4.9	1.5
0.8	$\mathrm{KRS}_{N,1}$	5.3	1.7	5.2	0.9	4.9	1.3	5.3	1.1
	$\mathrm{KRS}_{N,5}$	5.2	1.4	4.2	1.4	3.8	0.6	4.1	1.0
	$\mathrm{KRS}_{N,20}$	6.8	1.7	5.6	1.3	4.6	1.3	5.5	1.2
	$Z_N(10)$	3.9	0.8	4.4	0.7	5.1	1.4	4.7	1.2

TABLE C.2
Empirical sizes (in percent) based on 1000 replications for goodness-of-fit test of the FLM model. The tests $\mathrm{KRS}_{N,H}^{(GF)}$, $Z_N^{(GF)}(b)$ *are applied to evaluate the goodness-of-fit of an FLM model with the Gaussian kernel* (5.2).

	DGP	FLM–BM				FLM–fGARCH			
S	N	100		250		100		250	
	Nominal level	5%	1%	5%	1%	5%	1%	5%	1%
0.5	$\mathrm{KRS}_{N,1}$	5.8	1.3	4.0	1.3	4.6	0.5	5.3	1.3
	$\mathrm{KRS}_{N,5}$	4.3	0.8	5.1	1.1	3.0	0.5	3.5	0.5
	$\mathrm{KRS}_{N,20}$	6.5	2.1	6.9	1.8	5.0	1.5	4.4	0.8
	$Z_N(10)$	7.2	1.8	5.1	1.2	5.4	1.2	4.6	0.5
0.8	$\mathrm{KRS}_{N,1}$	5.8	1.3	5.1	1.1	5.0	0.6	4.5	1.0
	$\mathrm{KRS}_{N,5}$	4.3	0.8	6.3	1.0	3.9	0.6	3.8	0.5
	$\mathrm{KRS}_{N,20}$	6.5	2.1	8.0	1.8	4.5	1.3	5.3	1.0
	$Z_N(10)$	6.8	1.9	5.5	1.0	4.7	1.0	4.0	0.7

References

Aue, A., Horváth, L. and Pellat, D. (2017). Functional generalized autoregressive conditional heteroskedasticity. *Journal of Time Series Analysis*, **38,** 3–21. MR3601312

Aue, A. and van Delft, A. (2020). Testing for stationarity of functional time series in the frequency domain. *The Annals of Statistics*, **48,** 2505–2547. MR4152111

Bagchi, P., Characiejus, V. and Dette, H. (2018). A simple test for white noise in functional time series. *Journal of Time Series Analysis*, **39,** 54–74. MR3743093

Bartlett, M. S. (1978). *An introduction to stochastic processes: with special reference to methods and applications*. Cambridge University Press, Cambridge. MR0475536

Bosq, D. (2000). *Linear Processes in Function Spaces*. Springer. MR1783138

Box, G. E. P. and Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive moving average time series models. *Journal of the American Statistical Association*, **65,** 1509–1526. MR0273762

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. Springer, New York. MR1093459

Bücher, A., Dette, H. and Heinrichs, F. (2023). A portmanteau-type test for detecting serial correlation in locally stationary functional time series. *Statistical Inference for Stochastic Processes*, 1–24.

Cerovecki, C., Characiejus, V. and Hörmann, S. (2021). The maximum of the periodogram of Hilbert space valued time series. *Journal of the American Statistical Association*, **000,** 000–000; under review.

Cerovecki, C., Francq, C., Hörmann, S. and Zakoïan, J. (2019). Functional GARCH models: the quasi-likelihood approach and its applications. *Journal of Econometrics*, **209,** 353–375. MR3944755

Cerovecki, C. and Hörmann, S (2017). On the CLT for discrete Fourier transforms of functional time series. *Journal of Multivariate Analysis*, **154,** 282–295. MR3588570

Chang, J., Chen, C., Qiao, X. and Yao, Q. (2023). An autocovariance-based learning framework for high-dimensional functional time series. *Journal of Econometrics*.

Characiejus, V. and Rice, G. (2020). A general white noise test based on kernel lag-window estimates of the spectral density operator. *Econometrics and Statistics*, **13,** 175–196. MR4058332

Chen, W. W. and Deo, R. S. (2004). Power transformations to induce normality and their applications. *Journal of the Royal Statistical Society (B).*, **66,** 117–130. MR2035762

Chiou, J-M., Chen, Y-T. and Hsing, T. (2019). Identifying multiple changes for a functional data sequence with application to freeway traffic segmentation. *The Annals of Applied Statistics*, **13,** 1430–1463. MR4019145

Chitturi, R. V. (1976). Distribution of multivariate white noise autocorrelation. *Journal of the American Statistical Association*, **71,** 223–226. MR0405756

Cuesta-Albertos, J. A., Garcí a-Portugués, E., Febrero-Bande, M. and González-Manteiga, W. (2019). Goodness-of-fit tests for the functional linear model based on randomly projected empirical processes. *The Annals of Statistics*, **47,** number 1. MR3909938

Durbin, J. and Brown, R. (1967). Tests of serial independence based on the cumulated periodogram. *Bulletin of the International Statistical Institute*, **42,** 1039–1048.

Durlauf, S. N. (1991). Spectral based testing of the martingale hypothesis. *Journal of Econometrics*, **50,** 355–376. MR1147117

Francq, C. and Zakoian, J-M. (2010). *GARCH models*. Wiley. MR3186556

Gabrys, R., Horváth, L. and Kokoszka, P. (2010). Tests for error correlation in the functional linear model. *Journal of the American Statistical Association*, **105,** 1113–1125. MR2752607

Gabrys, R. and Kokoszka, P. (2007). Portmanteau test of independence for functional observations. *Journal of the American Statistical Association*, **102,**

1338–1348. MR2412554

García-Portugués, E., Álvarez-Liébana, J., Álvarez-Pérez, G. and González-Manteiga, W. (2021). A goodness-of-fit test for the functional linear model with functional response. *Scandinavian Journal of Statistics*, **48,** 502–528. MR4275155

González-Manteiga, W., Ruiz-Medina, M. D., López-Pérez, A. M. and Álvarez Liébana, J. (2023). Testing the goodness of fit of a hilbertian autoregressive model.

Górecki, T., , Hörmann, S., Horváth, L. and Kokoszka, P. (2018). Testing normality of functional time series. *Journal of Time Series Analysis*, **39,** 471–487. MR3819053

Górecki, T., Horváth, L. and Kokoszka, P. (2020). Tests of normality of functional data. *International Statistical Review*, **88,** 677–697. MR4180673

Grenander, U. and Rosenblatt, M. (1953). Statistical spectral analysis of time series arising from stationary stochastic processes. *The Annals of Mathematical Statistics*, **24,** 537–558. MR0058901

Grenander, U. and Rosenblatt, M. (2008). *Statistical Analysis of Stationary Time Series*, volume 320. American Mathematical Society. MR0084975

Guo, S. and Qiao, X. (2023). On consistency and sparsity for high-dimensional functional time series with application to autoregressions. *Bernoulli*, **29,** 451–472. MR4497254

Hlávka, Z., Hušková, M. and Meintanis, S. (2021). Testing serial independence with functional data. *Test*, **30,** 603–629. MR4297270

Hong, Y. (1996). Consistent testing for serial correlation of unknown form. *Econometrica: Journal of the Econometric Society*, **64,** 837–864. MR1399220

Hörmann, S., , Kokoszka, P. and Nisol, G. (2018). Testing for periodicity of functional time series. *The Annals of Statistics*, **46,** 2960–2984. MR3851761

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

Hörmann, S., Kidziński, L. and Hallin, M. (2015). Dynamic functional principal components. *Journal of the Royal Statistical Society(B).*, **77,** 319–348. MR3310529

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

Hörmann, S., Kuenzer, T. and Kokoszka, P. (2021). Testing normality of spatially indexed functional data. *Canadian Journal of Statistics*, **50,** 304–326. MR4389182

Horváth, L., Hušková, M. and Rice, G. (2013). Test of independence for functional data. *Journal of Multivariate Analysis*, **17,** 100–119. MR3053537

Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. Springer. MR2920735

Horváth, L., Kokoszka, P. and Rice, G. (2014). Testing stationarity of functional time series. *Journal of Econometrics*, **179,** 66–82. MR3153649

Horváth, L., Kokoszka, P., VanderDoes, J. and Wang, S. (2022). Inference in functional factor models with applications to yield curves. *Journal of Time Series Analysis*, **43,** 872–894. MR4524853

Horváth, L. and Rice, G. (2014). Extensions of some classical methods in change point analysis. *Test*, **23,** 219–255. MR3210268

Hosking, J. R. M. (1980). The multivariate portmanteau statistic. *Journal of the American Statistical Association*, **75,** 602–608. MR0590689

Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Wiley. MR3379106

Jarque, C. M. and Bera, A. K. (1980). Efficient tests for normality, homoskedasticity and serial independence of regression residuals. *Economic Letters*, **6,** 255–259. MR0615323

Jarque, C. M. and Bera, A. K. (1987). A test of normality of observations and regression residual. *International Statistical Review*, **55,** 163–172. MR0963337

Kargin, V. and Onatski, A. (2008). Curve forecasting by functional autoregression. *Journal of Multivariate Analysis*, **99,** 2508–2526. MR2463404

Kim, M. and Petoukhov, D. (2022). *wwntests: Hypothesis tests for functional time series*. R package version 1.0.2.

Kokoszka, P., Maslova, I., Sojka, J. and Zhu, L. (2008). Testing for lack of dependence in the functional linear model. *Canadian Journal of Statistics*, **36,** 207–222. MR2431682

Kokoszka, P. and Reimherr, M. (2017). *Introduction to Functional Data Analysis*. CRC Press. MR3793167

Kokoszka, P., Rice, G. and Shang, H L (2017). Inference for the autocovariance of a functional time series under conditional heteroscedasticity. *Journal of Multivariate Analysis*, **162,** 32–50. MR3719333

Kuenzer, T., Hörmann, S. and Kokoszka, P. (2021). Principal component analysis of spatially indexed functions. *Journal of the American Statistical Association*, **116,** 1444–1456. MR4309284

Lee, C.E., Zhang, X. and Shao, X. (2020). Testing conditional mean independence for functional data. *Biometrika*, **107,** 331–346. MR4108934

Li, W. K. (2004). *Diagnostic Checks in Time Series*. Chapman and Hall.

Li, W. K. and McLeod, A. I. (1981). Distribution of the residual autocorrelations in multivariate ARMA time series models. *Journal of the Royal Statistical Society, Series* B, **43,** 231–239. MR0626770

Ljung, G. and Box, G. (1978). On a measure of lack of fit in time series models. *Biometrika*, **66,** 67–72.

Lobato, I. and Velasco, C. (2004). A simple test of normality for time series. *Econometric Theory*, **20,** 671–689. MR2083156

McElroy, T. and Roy, A. (2022). A review of seasonal adjustment diagnostics. *International Statistical Review*, **90,** 259–284. MR4481435

McLeod, A. I. (1978). On the distribution of residual autocorrelations in box-jenkins models. *Journal of the Royal Statistical Society. Series B (Methodological).*, **40,** 296–302. MR0522212

Mestre, G., Portela, J., Rice, G., Roque, A. Muñoz San and Alonso, E. (2021). Functional time series model identification and diagnosis by means of auto- and partial autocorrelation analysis. *Computational Statistics & Data Analysis*, **155,** 107108. MR4163024

Neumann, M. H. and Paparoditis, E. (2008). Goodness-of-fit tests for Marko-

vian time series models: Central limit theory and bootstrap approximations. *Bernoulli*, **14,** 14–46. MR2401652

Panaretos, V. M. and Tavakoli, S. (2013). Cramér–Karhunen–Loève representation and harmonic principal component analysis of functional time series. *Stochastic Processes and their Applications*, **123,** 2779–2807. MR3054545

Panaretos, V. M. and Tavakoli, S. (2013). Fourier analysis of stationary time series in function space. *The Annals of Statistics*, **41,** 568–603. MR3099114

Patilea, V. and Sánchez-Sellero, C. (2020). Testing for lack-of-fit in functional regression models against general alternatives. *Journal of Statistical Planning and Inference*, **209,** 229–251. MR4096265

R Development Core Team. (2008). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer. MR2168993

Rice, G., Wirjanto, T. and Zhao, Y. (2020). Tests for conditional heteroscedasticity of functional data. *Journal of Time Series Analysis*, **41,** 733–758. MR4176358

Shao, X. (2011). Testing for white noise under unknown dependence and its applications to diagnostic checking for time series models. *Econometric Theory*, **27,** 312–343. MR2782041

Shumway, R. H. and Stoffer, D. S. (2017). *Time Series Analysis and Its Applications with R Examples*. Springer. MR3642322

van Delft, A. (2020). A note on quadratic forms of stationary functional time series under mild conditions. *Stochastic Processes and their Applications*, **130,** 4206–4251. MR4102264

van Delft, A. and Eichler, M. (2018). Locally stationary functional time series. *Electronic Journal of Statistics*, **12,** 107–170. MR3746979

Xin, H. and Shang, H.L. (2023). Nonlinear autocorrelation function of functional time series. *Nonlinear Dynamics*, **111,** 2537–2554. MR4468556

Yeh, C. K., Rice, G. and Dubin, J. A. (2023). Functional spherical autocorrelation: A robust estimate of the autocorrelation of a functional time series. *Electronic Journal of Statistics*, **17,** 650–687. MR4545121

Yuan, G., Shang, H.L. and Yanrong, Y. (2019). High-dimensional functional time series forecasting: An application to age-specific mortality rates. *Journal of Multivariate Analysis*, **170,** 232–243. MR3913038

Zamani, A., Hashemi, M. and Haghbin, H. (2019). Improved functional portmanteau tests. *Journal of Statistical Computation and Simulation*, **89,** 1423–1436. MR3929230

Zhang, X. (2016). White noise testing and model diagnostic checking for functional time series. *Journal of Econometrics*, **194,** 76–95. MR3523520