



Projection-based white noise and goodness-of-fit tests for functional time series

Mihyun Kim¹ · Piotr Kokoszka² · Gregory Rice³

Received: 2 October 2023 / Accepted: 11 July 2024

© The Author(s), under exclusive licence to Springer Nature B.V. 2024

Abstract

We develop two significance tests in the setting of functional time series. The null hypothesis of the first test is that the observed data are sampled from a general weak white noise sequence. The null hypothesis of the second test is that the observed data are sampled from a functional autoregressive model of order one (FAR(1)), which can be used as a goodness-of-fit test. Both tests are based on projections on functional principal components. Such projections are used in a great many functional data analysis (FDA) procedures, so our tests are practically relevant. We derive test statistics for each test that are quadratic forms of lagged autocovariance estimates constructed from principal component projections, and establish the requisite asymptotic theory. A simulation study shows that the tests have complimentary advantages against relevant benchmarks.

Keywords Autoregressive process · Functional principal components · Goodness-of-fit · White noise

1 Introduction

White noise and related goodness-of-fit tests play a vital role in time series analysis. In the context of scalar time series, this is explained in many textbooks, see e.g., Section 1.6 of Brockwell and Davis (2016), with a more detailed treatment in Li (2004). The idea is that if a time series model explains the realization well, the residuals should not contain any additional information, that is, they should form a white noise sequence. Ideally, they should

✉ Mihyun Kim
mihyun.kim@mail.wvu.edu

Piotr Kokoszka
Piotr.Kokoszka@colostate.edu

Gregory Rice
grice@uwaterloo.ca

¹ School of Mathematical and Data Sciences, West Virginia University, Armstrong Hall, Morgantown, WV 26505, USA

² Department of Statistics, Colorado State University, 102 Statistics Building, Fort Collins, CO 80523, USA

³ Department of Statistics and Actuarial Science, University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada

form a sequence of mean zero iid random variables, but standard time series models, like ARIMA, do not require that the model errors be iid. They can form a weak white noise, i.e., a sequence of mean zero uncorrelated random variables with the same variance, as in commonly used ARIMA-GARCH models, see e.g., Section 5.2 of Francq and Zakoian (2010). At a more fundamental level, every purely nondeterministic stationary time series can be represented as moving average with weak white noise errors, i.e., it admits the *Wold decomposition*, see e.g., Section 5.7 of Brockwell and Davis (1991), whereas iid errors form part of model formulation. These issues and their impact on standard goodness-of-fit tests are discussed in some depth in Kokoszka and Politis (2011). It is well-known that formulas involving residuals of many time series models generate extra terms that are not present in analogous formulas based on the unobservable errors, see Horváth et al. (2001) and Koul (2002), among many others. For this reason, goodness-of-fit tests based on residuals require very careful asymptotic analysis.

The objective of this paper is to develop asymptotic theory of weak white noise testing for functional time series (FTS) as well as the theory of goodness-of-fit testing for the FAR(1) model with iid errors. These and similar problems have been extensively studied of late, see Kim et al. (2023) and González-Manteiga et al. (2023) for recent reviews. In particular, Zhang (2016); Bagchi et al. (2018); Characiejus and Rice (2020), and Hlávka et al. (2021) develop spectral domain tests. We note a related recent paper of Bücher et al. (2023) who consider testing for weak white noise in locally stationary functional time series. They do not use projections, but, like we do, use functional covariances, and then bootstrap to approximate the null distribution. The tests we propose do not require bootstrap (we get standard chi-square null limits), but are only applicable under the assumption of stationarity.

Our objective is to extend the test and the requisite theory of Gabrys and Kokoszka (2007) from the setting of testing the assumption of iid functional observations to testing the hypothesis of a weak white noise and the goodness of fit of the FAR(1) model. As explained in Bosq (2000), the FAR(1) model is much more general than its scalar AR(1) counterpart because the functions can capture a lot of additional temporal dependence, and every FAR(p) model can be easily represented as an FAR(1) with a larger functional domain, see also Kokoszka and Reimherr (2013). The test of Gabrys and Kokoszka (2007) is based on commonly used projections on functional principal components (FPCs), the functional scores, rather than the more complex spectral analysis of functional time series. Such projections are used in a great many FDA procedures because they allow to heuristically convert a functional inference problem to a multivariate problem with relatively little effort. However, mathematical analysis and justification of such approaches is not easy, as explained in several chapters of Horváth and Kokoszka (2012) and a large number of papers. It is therefore important to investigate carefully how suitable extensions of the test of Gabrys and Kokoszka (2007) should be formulated and justified. In our context, the impact of estimation on the test statistics needs to be carefully studied for two reasons: (1) the sample projections are based on estimated FPCs rather than on the unobservable population FPCs, (2) the FAR(1) estimators are finite dimensional whereas the errors can be infinite dimensional. Both these aspects make the requisite theory substantially different from the corresponding multivariate theory. The second aspect is particularly challenging. It turns out that to obtain pivotal chi-square limits, an additional centering of the products of the projected residuals is needed, see formulas (4.7) and (4.8). Our approach could be extended to other linear functional models, in particular to various regression models, but specific forms of the test statistics would need to be worked out. The FAR(1) model presents key difficulties due to temporal dependence.

The remainder of the paper is organized as follows. Section 2 presents the basic notation of FDA, and reviews functional principal component analysis and FAR(1) estimation. Sections 3

and 4 are dedicated, respectively, to weak white noise testing and the FAR(1) goodness-of-fit testing. The results of a simulation study comparing the methods put forward in the paper to benchmark methods are presented in Sect. 5. The proofs are collected in Sect. 6.

2 Functional data analysis background and notation

The monograph of Bosq (2000) contains sufficient background, with elaborations in Horváth and Kokoszka (2012) and Hsing and Eubank (2015). The purpose of this section is to define the objects we study, for ease of reference, and to fix the notation.

We assume that functional observations are elements of $L^2 := L^2(\mathcal{T})$, the Hilbert space of real-valued square integrable functions defined on a compact interval $\mathcal{T}$. The inner product in L^2 is defined by $\langle x, y \rangle = \int_{\mathcal{T}} x(t)y(t)dt$, and the norm by $\|x\| = \sqrt{\langle x, x \rangle}$.

Our testing procedure is based on multivariate coefficients obtained by projecting functional observations onto the space spanned by FPCs. Suppose that $X_1, X_2, \dots, X_N$ are identically distributed functions in L^2 with $E\|X_i\|^2 < \infty$, and denote by X a generic random function with the same distribution as each X_i . The FPCs v_j , $j \geq 1$, are the eigenfunctions of the covariance operator of X , defined by

$$\Gamma(x) = E[(X - \mu, x)(X - \mu)], \quad \mu = EX, \quad x \in L^2, \quad (2.1)$$

i.e., $\Gamma(v_j) = \lambda_j v_j$, where the λ_j are the corresponding eigenvalues of Γ . The v_j are arranged in non-increasing order of λ_j , so that typically the first three or four v_j explain most of the variability that the functions have. The FPCs lead to the Karhunen–Lo  ve expansion

$$X_i(t) = \mu(t) + \sum_{j=1}^{\infty} \xi_{ij} v_j(t), \quad (2.2)$$

where the coefficients ξ_{ij} , called scores, are defined by projecting X_i onto the FPCs, i.e., $\xi_{ij} = \langle X_i - \mu, v_j \rangle$, $j \geq 1$. The scores satisfy that $E\xi_{ij} = 0$, $E\xi_{ij}^2 = \lambda_j$, and $E[\xi_{ij}\xi_{ij'}] = 0$ for $j \neq j'$.

Expansion (2.2) 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

$$\hat{\Gamma}(\hat{v}_j)(t) = \hat{\lambda}_j \hat{v}_j(t), \quad 1 \leq j \leq N, \quad (2.3)$$

where $\hat{\Gamma}$ is the sample covariance operator defined by

$$\hat{\Gamma}(x) = \frac{1}{N} \sum_{i=1}^N \langle X_i - \bar{X}_N, x \rangle (X_i - \bar{X}_N), \quad x \in L^2.$$

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 $\hat{\xi}_{ij} = \langle X_i - \bar{X}_N, \hat{v}_j \rangle$ 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 shape of each infinite-dimensional function X_i can be described, to a good approximation, by the finite-dimensional vector of the sample scores, $[\hat{\xi}_{i1}, \hat{\xi}_{i2}, \dots, \hat{\xi}_{ip}]^T$.

The FPCs can be also used to estimate kernel operators in FAR(1) models. A sequence of random functions $\{X_i\}$ in L^2 with mean μ is said to follow an FAR(1) if

$$X_i(t) - \mu(t) = \Psi(X_{i-1} - \mu)(t) + \varepsilon_i(t), \quad i \in \mathbb{Z}, \quad (2.4)$$

where $\Psi(X)(t) = \int \psi(t, s)X(s)ds$ for a kernel function $\psi \in L^2(\mathcal{T}^2)$, and $\{\varepsilon_i\}$ is a mean-zero white noise sequence in L^2 . To ease notation we assume that $\mu(t) = 0$ in the following discussion. When implementing the techniques described below, we start with centering by the sample mean function.

We consider the following estimator for Ψ

$$\widehat{\Psi}_N(\cdot) = \widehat{\Gamma}_1 \tilde{\Gamma}^{-1}(\cdot) = \frac{1}{N} \sum_{j=2}^N \sum_{q=1}^{q_N} \frac{\langle \cdot, \hat{v}_q \rangle}{\hat{\lambda}_q} \langle X_{j-1}, \hat{v}_q \rangle X_j, \quad (2.5)$$

where $\widehat{\Gamma}_1$ is the sample autocovariance operator at lag 1 defined by

$$\widehat{\Gamma}_1(x) = \frac{1}{N} \sum_{j=2}^N \langle X_{j-1}, x \rangle X_j, \quad x \in L^2,$$

and $\tilde{\Gamma}^{-1}$ is the Moore–Penrose inverse of $\widehat{\Gamma}$ defined by

$$\tilde{\Gamma}^{-1}(\cdot) = \sum_{q=1}^{q_N} \frac{\langle \hat{v}_q, \cdot \rangle}{\hat{\lambda}_q} \hat{v}_q. \quad (2.6)$$

For more detailed derivation of $\widehat{\Psi}_N$, see e.g., Chapter 8 of Bosq (2000).

We end this section by introducing some conventions that are often used throughout this paper. We use $\{X_i\}$ to denote the sequence $\{X_i, i \in \mathbb{Z}\}$, and $\int$ to denote $\int_{\mathcal{T}}$. The Frobenius norm of matrices is denoted by $\|\cdot\|_F$, and the Hilbert–Schmidt norm of linear operators acting on L^2 is denoted by $\|\cdot\|_S$. (If a matrix is viewed as an operator, the Frobenius norm is the Hilbert–Schmidt norm.)

3 Weak white noise testing

Suppose that $\{X_i\}$ is a mean-zero second-order stationary FTS taking values in L^2 . We aim to develop a testing procedure to assess the validity of the hypothesis

$$\mathcal{H}_0 : \iint \gamma_h^2(t, s) dt ds = 0, \quad \forall h \neq 0, \quad (3.1)$$

where $\gamma_h(t, s) := E[X_i(t)X_{i+h}(s)]$ is the lag- h autocovariance function of $\{X_i\}$. Note that $\mathcal{H}_0$ contains functional processes that are uncorrelated, but possibly dependent, for example fARCH and fGARCH processes, see Hörmann et al. (2013); Aue et al. (2017); Cerovecki et al. (2019) and Kühnert (2020), as well as functional stochastic volatility models, see Jang et al. (2022) and Kokoszka et al. (2024).

The main idea of our method for detecting serial correlation occurring in *infinite-dimensional* curves is to explore correlation structures in their *finite-dimensional* summary vectors. For efficient dimension reduction, we project $X_1, \dots, X_N$ onto the FPCs that are functional axes capturing most variance of curves. The shape of each curve X_i can then be concisely, but still sufficiently, encoded by its projections. We will comment in the following how in the case of iid X_i our test statistic reduces to the test statistic introduced by Gabrys and Kokoszka (2007); we refer to their test as the GK test.

Recall that the FPCs v_j are the eigenfunctions of Γ in (2.1), and the scores are the projections of each curve X_i onto the v_j , i.e., $\xi_{ij} = \langle X_i, v_j \rangle$, $j \geq 1$. We define the p -dimensional score vectors by

$$\mathbf{Y}_i = \mathbf{Y}_i(p) = [\xi_{i,1}, \dots, \xi_{i,p}]^\top, \quad i = 1, 2, \dots, N. \quad (3.2)$$

If a sequence of functions $\{X_i\}$ is uncorrelated, then for any nonzero-lags, $\{\mathbf{Y}_i\}$ is also uncorrelated since for $h > 0$

$$E[\xi_{i,k}\xi_{i+h,l}] = \iint E[X_i(t)X_{i+h}(s)]v_k(t)v_l(s)dt ds = 0, \quad \forall k, l = 1, \dots, p.$$

The above condition implies zero correlations of the functions X_i only if they live in the subspace spanned by the first p FPCs.

Our goal is to develop a test statistic from the score vectors to detect serial correlation in $\{X_i\}$. It will be a suitably defined quadratic form. For a fixed $0 < H \leq N - 1$, define the row expansion of block matrices $\mathbf{Y}_{i-h}\mathbf{Y}_i^\top$, $1 \leq h \leq H$, by

$$\begin{aligned} \mathbf{S}_i &= \mathbf{S}_i(H, p) \\ &= \text{vec} \left([\mathbf{Y}_{i-1}\mathbf{Y}_i^\top, \mathbf{Y}_{i-2}\mathbf{Y}_i^\top, \dots, \mathbf{Y}_{i-H}\mathbf{Y}_i^\top]^\top \right) \\ &= [\xi_{i-1,1}\xi_{i,1}, \xi_{i-1,1}\xi_{i,2}, \dots, \xi_{i-1,p}\xi_{i,p}, \xi_{i-2,1}\xi_{i,1}, \dots, \\ &\quad \xi_{i-2,p}\xi_{i,p}, \dots, \xi_{i-H,1}\xi_{i,1}, \dots, \xi_{i-H,p}\xi_{i,p}]^\top. \end{aligned} \quad (3.3)$$

Then, under $\mathcal{H}_0$, $\mathbf{S}_i$ is a Hp^2 -dimensional random vector with mean zero and covariance $\Sigma := \Sigma(H, p) = E\mathbf{S}_0\mathbf{S}_0^\top$. The covariance matrix Σ is a $Hp^2 \times Hp^2$ -dimensional fourth cumulant matrix consisting of cross-covariances of $\xi_{-h,k}\xi_{0,l}$ in ascending lexicographic order of h, k, l and $\xi_{-h',k'}\xi_{0,l'}$ in the order of h', k', l' for $h, h' = 1, \dots, H$, and $k, l, k', l' = 1, \dots, p$, i.e.,

$$\Sigma = \begin{bmatrix} E\xi_{-1,1}^2\xi_{0,1}^2 & E\xi_{-1,1}\xi_{0,1}\xi_{-1,1}\xi_{0,2} & \dots & E\xi_{-1,1}\xi_{0,1}\xi_{-H,p}\xi_{0,p} \\ E\xi_{-1,1}\xi_{0,2}\xi_{-1,1}\xi_{0,1} & E\xi_{-1,1}^2\xi_{0,2}^2 & \dots & E\xi_{-1,1}\xi_{0,2}\xi_{-H,p}\xi_{0,p} \\ & & E\xi_{-h,k}\xi_{0,l}\xi_{-h',k'}\xi_{0,l'} & \\ E\xi_{-H,p}\xi_{0,p}\xi_{-1,1}\xi_{0,1} & E\xi_{-H,p}\xi_{0,p}\xi_{-1,1}\xi_{0,2} & \dots & E\xi_{-H,p}^2\xi_{0,p}^2 \end{bmatrix}. \quad (3.4)$$

We note that for iid sequences, Σ takes a diagonal form. Making use of that form in the iid case leads back to the original GK test, but for general stationary, weak white noise series $\{X_i\}$ we do not expect this matrix to be diagonal.

We then define a partial sum of $\mathbf{S}_i$ by

$$\begin{aligned} \mathbf{T}_N &= \mathbf{T}_N(H, p) \\ &= \sqrt{N} [c_{N,1}(1, 1), c_{N,1}(1, 2), \dots, \\ &\quad c_{N,1}(p, p), c_{N,2}(1, 1), \dots, c_{N,2}(p, p) \dots, c_{N,H}(1, 1), \dots, c_{N,H}(p, p)]^\top, \end{aligned} \quad (3.5)$$

where

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

We further define a fourth cumulant matrix by

$$\begin{aligned}\Sigma_N &= \Sigma_N(H, p) = [c_{N,h,h'}(k, l, k', l')]_{h,h'=1,\dots,H,k,l,k',l'=1,\dots,p}, \\ c_{N,h,h'}(k, l, k', l') &= \frac{1}{N} \sum_{i=1+\max\{h,h'\}}^N \xi_{i-h,k} \xi_{i,l} \xi_{i-h',k'} \xi_{i,l'}.\end{aligned}\quad (3.6)$$

Based on $\mathbf{T}_N$ and Σ_N , a Hotelling's T -squared type statistic can be defined by

$$Q_N = Q_N(H, p) = \mathbf{T}_N^\top \Sigma_N^{-1} \mathbf{T}_N.$$

The quadratic form Q_N takes into account the covariance structure of $\mathbf{T}_N$ element-wise by down-weighting elements in $\mathbf{T}_N$ with a large variance and by up-weighting elements in $\mathbf{T}_N$ with a small variance. In this way, we might expect Q_N to perform well in detecting serial correlation even if it lies in projections that have a small variance.

The form Q_N is however not a statistic because it uses projections on the unobservable FPCs v_j . We emphasize that in our functional setting, a test statistic must be computed from 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 defined in (2.3). We thus introduce the following p -dimensional sample score vectors

$$\hat{\mathbf{Y}}_i = \hat{\mathbf{Y}}_i(p) = [\hat{\xi}_{i,1}, \dots, \hat{\xi}_{i,p}]^\top, \quad i = 1, 2, \dots, N.$$

Based on the $\hat{\mathbf{Y}}_i$, we can compute the sample counterpart of $\mathbf{T}_N$ by

$$\begin{aligned}\hat{\mathbf{T}}_N &= \hat{\mathbf{T}}_N(H, p) \\ &= \sqrt{N} [\hat{c}_{N,1}(1, 1), \hat{c}_{N,1}(1, 2), \dots, \\ &\quad \hat{c}_{N,1}(p, p), \hat{c}_{N,2}(1, 1), \dots, \hat{c}_{N,2}(p, p), \dots, \hat{c}_{N,H}(1, 1), \dots, \hat{c}_{N,H}(p, p)]^\top,\end{aligned}\quad (3.7)$$

where $\hat{c}_{N,h}(k, l) = \frac{1}{N} \sum_{i=1+h}^N \hat{\xi}_{i-h,k} \hat{\xi}_{i,l}$. We can also compute the sample counterpart of Σ_N by

$$\begin{aligned}\hat{\Sigma}_N &= \hat{\Sigma}_N(H, p) = [\hat{c}_{N,h,h'}(k, l, k', l')]_{h,h'=1,\dots,H,k,l,k',l'=1,\dots,p}, \\ \hat{c}_{N,h,h'}(k, l, k', l') &= \frac{1}{N} \sum_{i=1+\max\{h,h'\}}^N \hat{\xi}_{i-h,k} \hat{\xi}_{i,l} \hat{\xi}_{i-h',k'} \hat{\xi}_{i,l'}.\end{aligned}\quad (3.8)$$

We now define our test statistic as the quadratic form

$$\hat{Q}_N = \hat{Q}_N(H, p) = \hat{\mathbf{T}}_N^\top \hat{\Sigma}_N^{-1} \hat{\mathbf{T}}_N. \quad (3.9)$$

The statistic $\hat{Q}_N$ quantifies serial correlation in the sequence of the score vectors up to lag H and up to the dimension of projections p . If the X_i are white noise, all elements in $\hat{\mathbf{T}}_N$ should be close to 0. Therefore, a large value of $\hat{Q}_N$ indicates that the observed curves might not be white noise.

Another remark is that the test statistic introduced by Gabrys and Kokoszka (2007), where strong white noise is tested, can be derived as a special case of $\hat{Q}_N$. To see this, observe that under the iid assumption, Σ is a block diagonal matrix consisting of H identical blocks of $\mathbf{V} \otimes \mathbf{V}$ where $\mathbf{V} := \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$, and the λ_j are the eigenvalues of Γ in (2.1). Then, Σ can be estimated using $\hat{\mathbf{V}} := \text{diag}(\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_p)$, and the quadratic form $\hat{Q}_N$ becomes

$$\text{GK}_N(H, p) = \sum_{h=1}^H \sum_{k,l=1}^p \frac{\hat{c}_{N,h}^2(k, l)}{\hat{\lambda}_k \hat{\lambda}_l}. \quad (3.10)$$

Since $\mathcal{H}_0$ covers processes that could exhibit some sort of dependence, our assumptions should involve general, nonlinear, temporal dependence. We use the notion of L^P - m approximability, see Hörmann and Kokoszka (2010). Even though any positive P can be considered, it is generally assumed that $P \geq 1$.

Definition 3.1 A sequence $\{X_i, -\infty < i < \infty\}$ is said to be L^P - m approximable if the following two conditions hold:

(i) Each X_i admits the Bernoulli shift representation

$$X_i = g(\varepsilon_i, \varepsilon_{i-1}, \dots),$$

where $\{\varepsilon_i\}$ are iid elements in a measurable space $\mathbb{S}$, and g is a measurable function such that $g : \mathbb{S}^\infty \rightarrow L^2$.

(ii)

$$\sum_{m=0}^{\infty} \left(E \|X_i - X_i^{(m)}\|^P \right)^{1/P} < \infty, \quad (3.11)$$

where

$$X_i^{(m)} = g(\varepsilon_i, \varepsilon_{i-1}, \dots, \varepsilon_{i-m+1}, \varepsilon'_{i-m}, \varepsilon'_{i-m-1}, \dots) \quad (3.12)$$

with $\{\varepsilon'_i\}$ being an independent copy of $\{\varepsilon_i\}$ defined on the same probability space.

Definition 3.1 indicates that each process admits a causal representation, possibly nonlinear, and the effect of innovations far in the past on the present value decays fast, as specified in (3.11). Most known functional processes, including functional ARCH type models, satisfy Definition 3.1.

We state assumptions under which the limiting distribution of $\widehat{Q}_N$ in (3.9) is established.

Assumption 3.1 The sequence of functions $\{X_i\}$ is L^4 - m -approximable according to Definition 3.1.

Assumption 3.2 The sequence of functions $\{X_i\}$ satisfies the following first, second, and fourth order moment conditions:

(i) if $w \in L^2(\mathcal{T})$, then for all i , $E \int X_i(t) w(t) dt = 0$;

(ii) if $w \in L^2(\mathcal{T}^2)$, then for $i \neq j$, $E \int X_i(t) X_j(s) w(t, s) dt ds = 0$;

(iii) if $w \in L^2(\mathcal{T}^4)$, and if the indices $i, j, k, l \in \mathbb{Z}$ have a unique maximum, then $E \int X_i(t) X_j(s) X_k(u) X_l(v) w(t, s, u, v) dt ds du dv = 0$.

Condition (i) is equivalent to requiring that $EX_i = 0$ in L^2 , i.e., for almost all t , $EX_i(t) = 0$. It is imposed to simplify proofs. It is easy to verify that estimating a constant mean function by the sample mean adds $o_P(1)$ terms and so does not change limiting distributions in our theorems. Condition (ii) states that the function $(t, s) \mapsto E[X_i(t) X_j(s)]$ is the zero element of $L^2(\mathcal{T}^2)$, i.e., for almost all s, t , $E[X_i(t) X_j(s)] = 0$. These two conditions are the precise definition of functional weak white noise. We say that $\{X_i\}$ is a functional martingale difference sequence if $E[X_i | \mathcal{F}_{i-1}] = 0$ a.s. in L^2 , with $\mathcal{F}_{i-1}$ denoting the σ -algebra generated by the variables $\{X_j, j \leq i-1\}$. Condition (iii) appears in Kokoszka et al. (2017), and is meant to restrict the FTS to have the fourth order moment properties of a functional martingale difference sequence or causal GARCH-type process. Indeed it is possible for a function valued white noise sequence to satisfy Conditions (i) and (ii), but not (iii), although we are not aware of any examples that also satisfy Assumption 3.1,

or more generally that are causal Bernoulli shifts. As such we view Condition (iii) as a mild assumption for general functional white noise tests, although it is apparently slightly stronger than the comparable cumulant summability condition of Theorem 2.1 in Zhang (2016).

It is easy to verify that the functional GARCH models of Hörmann et al. (2013) and Aue et al. (2017) satisfy Assumptions 3.1 and 3.2 under standard conditions on the parameters and model innovations. We note that Assumption 3.1 also appears in Zhang (2016), and implies that $E\|X_i\|^4 < \infty$, and so implies that the moment conditions in Assumption 3.2 are well defined.

We now present the asymptotic justification of the test that rejects the null hypothesis (3.1) if $\widehat{Q}_N > \chi_{Hp^2}^2(1 - \alpha)$. To establish the null asymptotic distribution of $\widehat{Q}_N$, we first study the asymptotic behavior of $\widehat{\mathbf{T}}_N$ in the following theorem. We emphasize that our asymptotic results hold for fixed maximum lag, H , and fixed dimension, p .

Theorem 3.1 *Under Assumptions 3.1 and 3.2,*

$$\widehat{\mathbf{T}}_N \xrightarrow{d} \mathbf{G}, \quad \text{as } N \rightarrow \infty,$$

where $\widehat{\mathbf{T}}_N$ is defined in (3.7), and $\mathbf{G}$ is an Hp^2 -dimensional mean zero Gaussian random vector with covariance matrix Σ defined in (3.4).

From Theorem 3.1, we can derive the limiting distribution of $\widehat{Q}_N$.

Theorem 3.2 *Suppose Σ in (3.4) is invertible. Then, under Assumptions 3.1 and 3.2,*

$$\widehat{Q}_N \xrightarrow{d} \chi_{Hp^2}^2, \quad \text{as } N \rightarrow \infty,$$

where $\chi_{Hp^2}^2$ is a chi-squared distribution with Hp^2 degrees of freedom.

A sufficient condition for Σ (or Σ_r introduced in (4.6)) to be invertible is that they are positive-definite. The estimation of the inverses can however be challenging, especially in high-dimensional settings. We apply a banding method, as explained in Sect. 5.

We note that the limiting distribution of $\widehat{Q}_N$ is the same as that of the simpler statistic (3.10) under the assumption of iid white noise. The quadratic form $\widehat{Q}_N$ allows us to test the larger null hypothesis of the weak white noise. It is robust against nonlinear weak dependence that does not produce autocorrelation.

Under the alternative, the functions X_i must have nonzero autocovariances, as formulated in Assumption 3.3.

Assumption 3.3 For some $i \neq j$ and some $k, l = 1, \dots, p$, there exist FPCs $v_k \otimes v_l \in L^2(\mathcal{T}^2)$ such that $E \int X_i(t) X_j(s) v_k(t) v_l(s) dt ds \neq 0$.

Theorem 3.3 *Suppose Σ in (3.4) is invertible. Then, under Assumptions 3.1, 3.2 (i), and 3.3,*
 $\widehat{Q}_N \xrightarrow{P} \infty$, *as* $N \rightarrow \infty$.

4 Goodness-of-fit testing of the FAR(1) model

Recall the definition of the FAR(1) model given in (2.4). To simplify many formulas, we assume without loss of generality that $\mu = 0$. In this section we assume that the errors ε_i are iid with covariance operator Γ_ε . All assumptions will be precisely formulated in the following.

Our objective in this section is to derive a test of the null hypothesis

$$\mathcal{H}_0 : \text{The observations } X_i \text{ follow model (2.4).} \quad (4.1)$$

Our approach is to further develop the test of Sect. 3 so that it can be applied to model residuals. Unlike the unobservable errors ε_i , the residuals $\hat{\varepsilon}_i$ have a complex temporal and cross-sectional (after projections) dependence structure.

Using the estimated autoregressive coefficient $\hat{\Psi}_N$ defined in (2.5), the residuals are

$$\hat{\varepsilon}_i = X_i - \hat{\Psi}_N(X_{i-1}) = X_i - \frac{1}{N} \sum_{j=2}^N \sum_{q=1}^{q_N} \frac{\langle X_{i-1}, \hat{v}_q \rangle}{\hat{\lambda}_q} \langle X_{j-1}, \hat{v}_q \rangle X_j, \quad i = 2, \dots, N, \quad (4.2)$$

For each i , we consider projections of the residual curve onto the first p empirical FPCs:

$$\hat{r}_{i,k} = \langle \hat{\varepsilon}_i, \hat{v}_k \rangle = \hat{\xi}_{i,k} - \frac{1}{N} \sum_{j=2}^N \sum_{q=1}^{q_N} \frac{\hat{\xi}_{i-1,q} \hat{\xi}_{j-1,q}}{\hat{\lambda}_q} \hat{\xi}_{j,k}, \quad k = 1, \dots, p. \quad (4.3)$$

Note that both $\hat{\varepsilon}_i$ and $\hat{v}_k$ are estimators that depend on all $X_1, X_2, \dots, X_N$.

We define p -dimensional vectors of the projected residuals by

$$\hat{\mathbf{R}}_i = \hat{\mathbf{R}}_i(p) = [\hat{r}_{i,1}, \dots, \hat{r}_{i,p}]^\top, \quad i = 2, \dots, N.$$

When constructing $\hat{\mathbf{R}}_i$, two distinct dimension reduction levels are used: q_N and p . The sequence q_N is the number of FPCs used to approximate X_i and compute the residual curves in (4.2). It must increase with N in order for $\hat{\Psi}_N$ to be consistent for Ψ . The residual curves are then projected to the first p empirical FPCs to obtain the p -dimensional vector $\hat{\mathbf{R}}_i$. The asymptotic results in this section assume that q_N increases with N and p is fixed.

We define the sample autocovariance column matrix calculated from the $\hat{\mathbf{R}}_i$ by

$$\begin{aligned} \hat{\mathbf{T}}_{r,N} &= \hat{\mathbf{T}}_{r,N}(H, p) \\ &= \sqrt{N} [\hat{c}_{N,1}^r(1, 1), \hat{c}_{N,1}^r(1, 2), \dots, \hat{c}_{N,1}^r(p, p), \dots \dots, \\ &\quad \hat{c}_{N,H}^r(1, 1), \dots, \hat{c}_{N,H}^r(p, p)]^\top, \end{aligned} \quad (4.4)$$

where

$$\hat{c}_{N,h}^r(k, l) = \frac{1}{N} \sum_{i=1+h}^N \hat{r}_{i,k} \hat{r}_{i-h,l}.$$

The model residuals $\hat{\varepsilon}_i$ (and the projected residuals $\hat{r}_i$) typically have some common dependence on estimators of the model parameters, so white noise tests applied to the residuals must be adjusted to account for this. To discuss how our quadratic-form-based testing procedure is adjusted, we first define the projection of ε_i onto the population FPCs v_k by $r_{i,k} = \langle \varepsilon_i, v_k \rangle$ for $k = 1, \dots, p$. Also, let

$$f_{i,l}(h) = \langle \Gamma_\varepsilon \Psi_*^{h-1} \Gamma^{-1}(X_{i-1}), v_l \rangle, \quad l = 1, \dots, p, \quad (4.5)$$

where Ψ_* is the adjoint operator of Ψ . The variable $f_{i,l}(h)$ quantifies the effect of the estimation of Ψ on the distribution of the projected residuals. From $r_{i,k}$, $f_{i,l}(h)$, we define the fourth order cumulant matrix by

$$\Sigma_r = \Sigma_r(H, p)$$

$$= \left[E \left(r_{0,k} r_{-h,l} - r_{0,k} f_{0,l}(h) \right) \left(r_{0,k'} r_{-h',l'} - r_{0,k'} f_{0,l'}(h') \right) \right]_{1 \leq h, h' \leq H, 1 \leq k, l, k', l' \leq p}, \quad (4.6)$$

which is an $Hp^2 \times Hp^2$ matrix consisting of cross-covariances of $r_{0,k} r_{-h,l} - r_{0,k} f_{0,l}(h)$ and $r_{0,k'} r_{-h',l'} - r_{0,k'} f_{0,l'}(h')$, in ascending lexicographic order of h, k, l and h', k', l' . To estimate Σ_r , we need to first estimate $f_{i,l}(h)$ in (4.5). Note that since $X_{j-1} = \sum_{i=0}^{\infty} \Psi^i(\varepsilon_{j-1-i})$ and the ε_j are uncorrelated, we have

$$\Gamma_\varepsilon \Psi_*^{h-1} \Gamma^{-1}(X_{i-1}) = E \left\langle \Psi^{h-1}(\varepsilon_{j-h}), \Gamma^{-1}(X_{i-1}) \right\rangle \varepsilon_{j-h} = E \left\langle X_{j-1}, \Gamma^{-1}(X_{i-1}) \right\rangle \varepsilon_{j-h}.$$

Therefore, $f_{i,l}(h)$ may be estimated by

$$\begin{aligned} \hat{f}_{i,l}^{(N)}(h) &= \left\langle \frac{1}{N} \sum_{j=h+1}^N \left\langle X_{j-1}, \tilde{\Gamma}^{-1}(X_{i-1}) \right\rangle \hat{\varepsilon}_{j-h}, v_l \right\rangle \\ &= \frac{1}{N} \sum_{j=h+1}^N \sum_{q=1}^{q_N} \hat{\lambda}_q^{-1} \langle X_{i-1}, \hat{v}_q \rangle \langle X_{j-1}, \hat{v}_q \rangle \hat{r}_{j-h,l}, \end{aligned} \quad (4.7)$$

where $\tilde{\Gamma}^{-1}$ is defined in (2.6). The matrix Σ_r may then be estimated by

$$\begin{aligned} \hat{\Sigma}_{r,N} &= \left[\hat{\mathcal{C}}_{N,h,h'}^r(k, l, k', l') \right]_{h,h'=1,\dots,H, k,l,k',l'=1,\dots,p}, \\ \hat{\mathcal{C}}_{N,h,h'}^r(k, l, k', l') &= \frac{1}{N} \sum_{i=1+\max\{h,h'\}}^N \left(\hat{r}_{i,k} \hat{r}_{i-h,l} - \hat{r}_{i,k} \hat{f}_{i,l}^{(N)}(h) \right) \left(\hat{r}_{i,k'} \hat{r}_{i-h',l'} - \hat{r}_{i,k'} \hat{f}_{i,l'}^{(N)}(h') \right). \end{aligned} \quad (4.8)$$

We now define the quadratic-form-based test statistic for testing model residuals by

$$\hat{Q}_N^{(GF)} = \hat{Q}_N^{(GF)}(H, p) = \hat{\mathbf{T}}_{r,N}^\top \hat{\Sigma}_{r,N}^{-1} \hat{\mathbf{T}}_{r,N}. \quad (4.9)$$

We consider the following assumptions to establish asymptotic properties of $\hat{Q}_N^{(GF)}$.

Assumption 4.1 $\|\Psi\|_S < 1$.

Assumption 4.2 $\|\hat{\Psi}_N - \Psi\|_S = o_P(N^{-1/4})$.

Assumption 4.3 $\{\varepsilon_i\}$ are iid with $E\varepsilon_i = 0$, $E\|\varepsilon_i\|^4 < \infty$.

Assumption 4.4 Ψ and Γ are commutative, and

$$\left\| \Gamma^{-1/2} \Psi^{h-1} \Gamma_\varepsilon^2 \Psi_*^{h-1} \Gamma^{-1/2} \right\|_1 = \sum_{q=1}^{\infty} \frac{\left\| \Gamma_\varepsilon \Psi_*^{h-1}(v_q) \right\|^2}{\lambda_q} < \infty,$$

where $\|\cdot\|_1$ is the trace norm.

Assumption 4.1 implies that there exists a stationary and causal solution to the FAR(1) model, which is $X_i(t) = \sum_{j=1}^{\infty} \Psi^j(\varepsilon_{i-j})(t)$. It could be replaced with a more general condition stated on page 7 of Bosq (2000) that there is $j_0 \geq 1$ such that $\|\Psi^{j_0}\|_{\mathcal{L}} < 1$ at the expense of adding an additional constant in our proofs. It is known by Lemma 2.3 of Zhang (2016) that $\hat{\Psi}_N$ converges to Ψ at the rate of $o_P(N^{-1/4})$ under Assumption 4.3 and some rate conditions on the eigenvalues λ_j and the increasing q_N . To avoid listing these technical

assumptions, we simply formulate Assumption 4.2. Assumption 4.4 guarantees that a central limit theorem for uniformly L^2 - m -approximable triangular arrays can be applied, see Kim et al. (2023), which is a key technique to establish the asymptotic distribution of $\widehat{Q}_N^{(GF)}$.

We now state our main result in the following theorem. Proofs of the results in this section are presented in Sect. 6.2.

Theorem 4.1 *Consider the FAR(1) model defined in (2.4) with the ε_i satisfying Assumption 4.3. Under Assumptions 4.1, 4.2, and 4.4,*

$$\widehat{\mathbf{T}}_{r,N} \xrightarrow{d} \mathbf{G}_r, \quad N \rightarrow \infty,$$

where $\widehat{\mathbf{T}}_{r,N}$ is defined in (4.4), and $\mathbf{G}_r$ is a Hp^2 -dimensional mean zero Gaussian random vector with covariance matrix Σ_r defined in (4.6).

The theorem states that the sample autocovariance column matrix computed from the projected residuals converges in distribution to a Gaussian random vector. The fourth cumulant structure (4.6) of the limiting random vector has a more complex form than when it is simply based on a stationary white noise sequence. As discussed earlier, this complexity is attributable to the adjustment made to account for the effect of estimation of the autoregressive operator Ψ .

From Theorem 4.1, we can establish the asymptotic null distribution of $\widehat{Q}_N^{(GF)}$. By normalizing $\widehat{\mathbf{T}}_{r,N}$ using $\widehat{\Sigma}_r$, $\widehat{Q}_N^{(GF)}$ converges in distribution to $\chi_{Hp^2}^2$.

Theorem 4.2 *Suppose Σ_r in (4.6) is invertible. Then, under the assumptions of Theorem 4.1, $\widehat{Q}_N^{(GF)} \xrightarrow{d} \chi_{Hp^2}^2$, as $N \rightarrow \infty$.*

Remark 4.1 Establishing the consistency of tests based on $\widehat{Q}_N^{(GF)}$, namely that

$$\widehat{Q}_N^{(GF)} \xrightarrow{P} \infty, \quad \text{as } N \rightarrow \infty, \quad (4.10)$$

when $X_1, \dots, X_N$ is drawn from a serially correlated FTS not following an FAR(1) model, is a difficult problem. Analogous results have not yet been obtained for the norm-based statistics considered in Kim et al. (2023) and Zhang (2016), and showing (4.10) appears somewhat more challenging. We leave it as an open problem for further research.

5 A small numerical study

The goal of this section is to evaluate finite sample performance of the proposed tests and to compare them to the tests introduced by Gabrys and Kokoszka (2007); Kokoszka et al. (2017), and Zhang (2016). Extensive comparisons to other tests are presented in Kim et al. (2023). They show that the test of Kokoszka et al. (2017) is one of the best, and the spectral domain test of Zhang (2016) is the other excellent choice. Since the test of Gabrys and Kokoszka (2007) applies only to iid null hypothesis, we expect to see an advantage of our tests over that early test. In Sect. 5.1, we assess the weak white noise tests, and in Sect. 5.2 we conduct the goodness-of-fit testing for FAR(1) models.

The application of our tests requires the selection of tuning parameters p , H and q_N . Based on the findings reported in this section, as a rule of thumb, we recommend using

$p = 3, 4, 5$ and $H \approx 5$, without claiming that these values are optimal. We give below a specific rule for selecting q_N .

Another decision to make is how to transform raw data into functional objects using a basis system. We report the results for the B -spline basis and note that practically identical results are obtained for the Fourier basis. Intuitively, this is because these expansions do not substantially impact temporal dependence.

We recall the norm-based test introduced by Kokoszka et al. (2017). Its test statistic is defined as the sum of the L^2 norm of sample autocovariance functions $\hat{\gamma}_{N,h}$ up to lag H :

$$\text{KRS}_{N,H} = N \sum_{h=1}^H \|\hat{\gamma}_{N,h}\|^2, \quad \hat{\gamma}_{N,h}(t, s) = \frac{1}{N} \sum_{i=1+h}^n X_{i-h}(t) X_i(s).$$

The spectral domain test of Zhang (2016) is based on the following statistics:

$$Z_N = \frac{N}{8\pi^2} \sum_{h=1}^{N-1} h^{-2} \iint \{\hat{\gamma}_{N,h}(t, s) + \hat{\gamma}_{N,h}(s, t)\}^2 dt ds.$$

While our quadratic form involves a dimension reduction, the tests based on $\text{KRS}_{N,H}$ and Z_N utilize full-dimensional information on serial correlation in functional observations. Their inference procedures are also developed under weak white noise assumption and can be applied for goodness-of-fit testing for FAR(1) models. We will show that our quadratic-based test has comparable size to the tests based on $\text{KRS}_{N,H}$ and Z_N , and can be more powerful in detecting autocorrelation.

In order to compute $\hat{Q}_N$ in (3.9) and $\hat{Q}_N^{(GF)}$ in (4.9), the fourth order cumulant matrices Σ and Σ_r need to be estimated. Their reliable estimation is however often challenging especially in high dimensional settings where either the total number of lags H or the dimension of projections p is large compared to sample size N . To improve estimation of the covariance matrices, we apply the convex banding estimator introduced by Bien et al. (2016).

The banding estimator developed in Bien et al. (2016) is the solution to a convex optimization problem. To apply it to our setting, let $g_l = \{(j, k) \in [Hp^2]^2 : |j - k| \geq Hp^2 - l\}$ be a subset of matrix indices of $[Hp^2]^2 = \{1, \dots, Hp^2\} \times \{1, \dots, Hp^2\}$. Then, g_l indicates $l(l+1)$ indices in the two right triangles farthest from the diagonal of Σ (or Σ_r). Bien et al. (2016) consider the following optimization problem

$$\hat{\mathbf{S}} = \arg \min_{\mathbf{S}} \left\{ \frac{1}{2} \|\mathbf{S} - \Sigma\|_F^2 + \lambda \sum_{l=1}^{Hp^2-1} f(\mathbf{S}, g_l) \right\},$$

where $\lambda > 0$ and $f(\mathbf{S}, g_l)$ is some function measuring the magnitude of entries in the two right triangles corresponding to g_l in $\mathbf{S}$. It is shown that the solution $\hat{\mathbf{S}}$ is sparsely banded, positive definite with high probability and asymptotically close to Σ in terms of the Frobenius norm. An algorithm for computing $\hat{\mathbf{S}}$ with a data-driven bandwidth has been developed in their R package `hierband`. Letting $\mathfrak{d} = Hp^2$, we apply the method when the number of parameters to be estimated in the covariance matrices, $\mathfrak{d}(\mathfrak{d}+1)/2$, is larger than $\sqrt{N}$.

The work presented in this section was performed using R, R Development Core Team (2008). We used R package `wnntests` to implement the tests from Gabrys and Kokoszka (2007); Kokoszka et al. (2017), and Zhang (2016).

5.1 The weak white noise test

In this section we simulate several functional processes and assess the empirical size and power of the test based on $\widehat{Q}_N$ introduced in Sect. 3. We consider the following data generating processes (DGPs).

- *Standard Brownian motion (BM)*: the X_i are the iid standard Brownian motions on $[0, 1]$. We generate their trajectories as rescaled cumulative sums of independent normal variables.
- *Functional GARCH process (fGARCH)*: 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 $\alpha, \beta : L^2 \rightarrow L^2$ are linear operators satisfying

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

The ε_i are iid Ornstein–Uhlenbeck processes given by $\varepsilon_i(t) = e^{-t/2}B_i(e^t)$, where the B_i are iid standard BMs. The particular settings for σ_i and ε_i are from Cerovecki et al. (2019).

- *Fourier-vector autoregressive models of order 1 process (F-VAR(1))*: The X_i satisfy

$$X_i(t) = \sum_{j=1}^9 x_{i,j} \phi_j(t),$$

where $\{\phi_j\}$ is a set of Fourier basis elements in L^2 , having the form of $\phi_0(t) = 1$, $\phi_{2k-1}(t) = \sqrt{2} \cos(2\pi kt)$, and $\phi_{2k}(t) = \sqrt{2} \sin(2\pi kt)$, for $k = 1, 2, \dots$. The vector valued process $\mathbf{X}_i := [x_{i,1}, \dots, x_{i,9}]^\top$ satisfy vector autoregressive models of order 1

$$\mathbf{X}_i = \mathbf{A}\mathbf{X}_{i-1} + \mathbf{E}_i,$$

where $\mathbf{E}_i := [e_{i,1}, \dots, e_{i,9}]^\top$ and for fixed j , $\{e_{i,j}\}$ follows a scalar GARCH process such that $e_{i,j} = \sigma_{i,j}\varepsilon_{i,j}$, $\sigma_{i,j}^2 = 0.1 + 0.8\sigma_{i-1,j}^2 + 0.1e_{i-1,j}^2$, and the $\varepsilon_{i,j}$ follow the standard normal distribution. The matrix $\mathbf{A}$ is a diagonal matrix with zeros on all diagonals except the first d diagonals taking a value of 0.5, e.g., when $d = 3$, $\mathbf{A} = \text{diag}(0.5, 0.5, 0.5, 0, 0, 0, 0, 0, 0)$. For d , we consider $d \in \{1, 3, 5, 7, 9\}$.

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. We simulated 1000 replications of each process with sample sizes $N \in \{100, 250\}$.

We compute the tests statistics $\widehat{Q}_N(H, p)$ and $\text{GK}_N(H, p)$ with $p \in \{3, 4, 5\}$, i.e., we use between 3 and 5 FPCs, typical choices encountered in applications. We set $H = 5$ for the tests based on $\widehat{Q}_N(H, p)$, $\text{GK}_N(H, p)$, and $\text{KRS}_{N,H}$. The test based on Z_N requires bootstrap procedures with block size b and the number of resamples M to be approximate the null distribution of Z_N . We use $b = 5$ and $M = 100$.

In order to investigate empirical sizes, we consider the **BM** and **fGARCH** DGPs. Table 1 shows that our test has sizes comparable with the $\text{KRS}_{N,H}$ and Z_N tests, which are among the most accurate tests. As expected, the $\text{GK}_N(H, p)$ test performs poorly when the white noise is not iid (the **fGARCH** DGP). The quadratic form we derived effectively addresses this deficiency.

Table 1 Empirical sizes (in percent) based on 1000 replications. The tests $\hat{Q}_N(H, p)$, $GK_N(H, p)$, $KRS_{N,H}$, and Z_N are applied to evaluate the empirical size of data generated from **BM** and **fGARCH**; $N \in \{100, 250\}$

DGP	BM				fGARCH			
	100		250		100		250	
N	5%	1%	5%	1%	5%	1%	5%	1%
Nominal level								
$\hat{Q}_N(5, 3)$	3.9	0.7	5.0	0.8	3.6	0.7	4.8	0.8
$\hat{Q}_N(5, 4)$	4.3	0.4	4.6	1.3	3.9	0.6	4.6	0.6
$\hat{Q}_N(5, 5)$	3.9	0.1	4.4	0.8	4.2	0.6	3.8	0.8
$GK_N(5, 3)$	2.9	0.7	4.5	0.8	28.9	16.1	45.9	30.9
$GK_N(5, 4)$	3.2	0.6	4.4	0.6	35.0	21.9	53.9	38.1
$GK_N(5, 5)$	2.2	0.2	4.3	0.4	41.4	26.3	61.4	44.8
$KRS_{N,5}$	4.9	1.1	4.0	1.0	3.2	0.6	4.8	0.7
Z_N	6.3	1.0	7.0	2.9	4.5	1.7	4.8	1.1

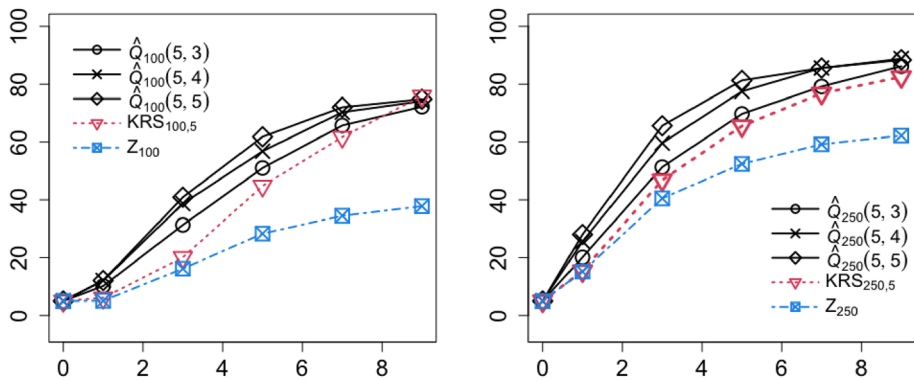


Fig. 1 Empirical power (in percent) for increasing values of d (proxy of correlation to noise ratio) (x-axis) for **F-VAR(1)** to which tests $\hat{Q}_N(H, p)$ and $KRS_{N,H}$ are applied with $H = 5$, and Z_N is applied; $N = 100$ (left), $N = 250$ (right); significance level = 5%

To assess empirical power, we apply the tests based on $\hat{Q}_N$ and $KRS_{N,H}$, Z_N to data generated from **F-VAR(1)**. The rejection rates are displayed in Fig. 1. We see that the new test has a power advantage over the $KRS_{N,H}$ and Z_N tests.

To summarize, our simulation results show that the test based on $\hat{Q}_N$ is robust to nonlinear dependence that does not violate the weak white noise null hypothesis, and so is a significant advance over the test of Gabrys and Kokoszka (2007). Its empirical size is comparable to the test of Kokoszka et al. (2017), Zhang (2016) over which it has a power advantage.

5.2 Diagnostics for the FAR(1) model

We now evaluate the goodness-of-fit test based on the statistic $\hat{Q}_N^{(GF)}$ defined in (4.9). We first calculate the residual curves $\hat{\varepsilon}_i$ according to (4.2), and then project them onto the first p FPCs and compute $\hat{Q}_N^{(GF)}$. If $\hat{Q}_N^{(GF)}$ is not significant, the FAR(1) model is declared to be adequate. As benchmarks, we use the tests denoted here as $KRS_{N,H}^{(GF)}$, $Z_N^{(GF)}$. These are the same tests as those used in Sect. 5.1, but applied to the residuals $\hat{\varepsilon}_i$ rather than the observations X_i and adjusted to account for the effect of estimation of the autoregressive operator Ψ . More

Table 2 Empirical sizes (in percent) based on 1000 replications. The tests $\widehat{Q}_N^{(GF)}(H, p)$, $\text{KRS}_{N,H}^{(GF)}$, and $Z_N^{(GF)}$ are applied to evaluate the goodness-of-fit of data generated from **FAR(1)-BM** with the level of serial dependence $S \in \{0.3, 0.7\}$; $N \in \{100, 250\}$

N	$S = 0.3$				$S = 0.7$			
	100	250	100	250	100	250	100	250
Nominal level	5%	1%	5%	1%	5%	1%	5%	1%
$\widehat{Q}_N^{(GF)}(5, 3)$	8.6	3.2	6.1	1.9	1.7	1.1	3.4	2.4
$\widehat{Q}_N^{(GF)}(5, 4)$	8.6	3.4	6.2	2.5	1.1	0.6	1.0	0.8
$\widehat{Q}_N^{(GF)}(5, 5)$	9.2	3.0	5.9	2.4	1.0	0.8	1.0	0.5
$\text{KRS}_{N,5}^{(GF)}$	5.3	0.8	5.1	1.8	6.4	1.6	6.1	1.8
$Z_N^{(GF)}$	5.1	1.8	5.8	2.1	5.2	1.2	6.0	1.8

details including their theoretical justification are available in Appendix B of Supplementary Material in Kim et al. (2023) on $\text{KRS}_{N,H}^{(GF)}$, and in Section 3.2 of Zhang (2016) on $Z_N^{(GF)}$.

We assess the empirical size and power by considering the following DGPs:

- **Functional autoregressive models of order 1 with BM errors (FAR(1)-BM):** recall (2.4). The X_i satisfy

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

where the ε_i follow the **BM** in Sect. 5.1. 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 $S \in \{0.3, 0.7\}$.

- **Fourier-vector autoregressive models of order 2 process (F-VAR(2)):** The X_i satisfy

$$X_i(t) = \sum_{j=1}^9 x_{i,j} \phi_j(t),$$

where $\{\phi_j\}$ is a set of Fourier basis elements in L^2 , having the form of $\phi_0(t) = 1$, $\phi_{2k-1}(t) = \sqrt{2} \cos(2\pi kt)$, and $\phi_{2k}(t) = \sqrt{2} \sin(2\pi kt)$, for $k = 1, \dots$. The vector valued process $\mathbf{X}_i := [x_{i,1}, \dots, x_{i,9}]^\top$ satisfy vector autoregressive models of order 1

$$\mathbf{X}_i = \mathbf{A}_1 \mathbf{X}_{i-1} + \mathbf{A}_2 \mathbf{X}_{i-2} + \mathbf{E}_i,$$

where $\mathbf{E}_i := [e_{i,1}, \dots, e_{i,9}]^\top$ and for fixed j , $\{e_{i,j}\}$ follows a scalar GARCH process such that $e_{i,j} = \sigma_{i,j} \varepsilon_{i,j}$, $\sigma_{i,j}^2 = 0.1 + 0.8\sigma_{i-1,j}^2 + 0.1e_{i-1,j}^2$, and the $\varepsilon_{i,j}$ follow the standard normal distribution. The matrix $\mathbf{A}_1 = 0.5\mathbf{I}$ where $\mathbf{I}$ is an 9×9 identity matrix, and $\mathbf{A}_2$ is a diagonal matrix with zeros on all diagonals except the first d diagonals taking a value of 0.5. For d , we consider $d \in \{1, 3, 5, 7, 9\}$.

We note that **FAR(1)-BM** is a FAR(1) model as in (2.4), whereas **F-VAR(2)** is not. To assess the empirical size, we consider the FAR(1) model. Table 2 reports the empirical rejection rates from the tests based on $\widehat{Q}_N^{(GF)}$, $\text{KRS}_{N,H}^{(GF)}$, and $Z_N^{(GF)}$ at nominal levels of 5% and 1%. To evaluate the empirical power, we apply the tests to data generated from **F-VAR(2)**. The power curves as a function of d are presented in Fig. 2. The truncation level q_N , used in computing the kernel estimator $\widehat{\Psi}_N$, was chosen to be $q_N = \max\{[N^{1/3}], q_N^*\}$ where $[x]$ is the integer closest to x , and q_N^* is the smallest q_N such that $\sum_{j=1}^{q_N} \hat{\lambda}_j / \sum_{j=1}^N \hat{\lambda}_j > 0.90$. The conclusions from Table 2 and Fig. 2 are summarized as follows.

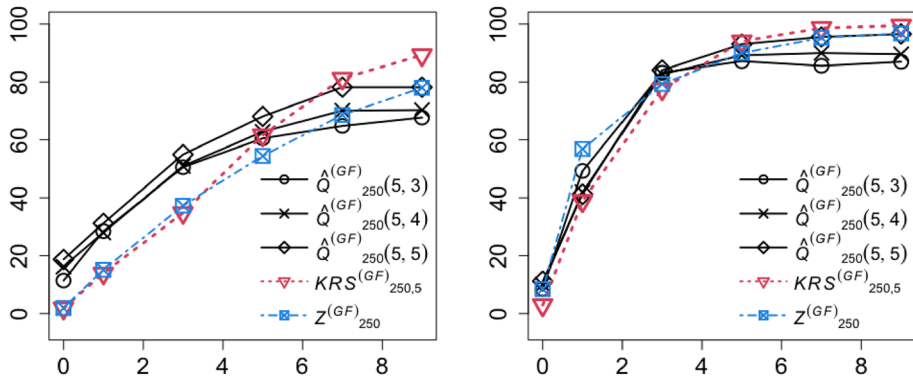


Fig. 2 Empirical power (in percent) for increasing values of d (proxy of correlation to noise ratio) (x-axis) for **F-VAR(2)** to which tests $\hat{Q}_N^{(GF)}(H, p)$ and $KRS_{N,H}^{(GF)}$ are applied with $H = 5$, and $Z_N^{(GF)}$ is applied; $N = 100$ (left), $N = 250$ (right); significance level = 5%

1. The test based on $\hat{Q}_N^{(GF)}$ is oversized for small N and S , but has correct size for larger N , and tends to be conservative for larger S ; see Table 2.
2. In terms of power, we again observe similar performance as in Sect. 5.1 in that $\hat{Q}_N^{(GF)}$ outperforms when the signal to noise ratio is small, e.g., when $d = 1, 3$ in Fig. 2, whereas $KRS_{N,H}^{(GF)}$ and $Z_N^{(GF)}$ perform well when the signal to noise ratio is large, e.g., $d = 7, 9$.

A broad conclusion from our limited numerical experiments is that the test of Kim et al. (2023) might be more reliable for goodness-of fit testing. This paper focuses on the derivation and large sample justification of the test, and a more extensive numerical investigation may be needed.

6 Proofs of the theorems of Sections 3 and 4

6.1 Proofs of the asymptotic results in Sections 3

6.1.1 Proof of Theorem 3.1

The proof of Theorem 3.1 is split into two steps. First we show in Proposition 6.1 that the sample autocovariance column matrix $\hat{\mathbf{T}}_N$ defined in (3.7) is asymptotically close to its population counterpart $\mathbf{T}_N$ defined in (3.5). Then, we investigate the asymptotic distribution of $\mathbf{T}_N$ in Proposition 6.2.

One of the key ideas in the proof of Proposition 6.1 is that the difference between the sample FPCs $\hat{v}_j$ and the population FPCs v_j is asymptotically negligible. It is known that this result holds not only for iid sequences but also when the X_i are L^4 - m -approximable, see Theorem 7 of Hörmann and Kokoszka (2012). We state the result in the following lemma.

Lemma 6.1 *Under Assumption 3.1,*

$$E \|\hat{v}_j - v_j\|^2 = O(N^{-1}), \quad 1 \leq j \leq p.$$

The following lemma is needed to prove Proposition 6.1, which mainly uses the moment conditions described in Assumption 3.2 for its proof.

Lemma 6.2 Under Assumptions 3.1, 3.2,

$$E \|\widehat{\Gamma}_h\|_S^2 = O(N^{-1}),$$

where $\widehat{\Gamma}_h(\cdot) := \frac{1}{N} \sum_{i=1+h}^N \langle X_i, \cdot \rangle X_{i-h}$ is the sample autocovariance operator.

Proof For any set of orthonormal basis $\{e_l\}$, we get that

$$\begin{aligned} \|\widehat{\Gamma}_h\|_S^2 &= \sum_{l=1}^{\infty} \|\widehat{\Gamma}_h(e_l)\|^2 \\ &= \sum_{l=1}^{\infty} \frac{1}{N^2} \left(\sum_{i=1+h}^N \langle X_i, e_l \rangle^2 \|X_{i-h}\|^2 + \sum_{i \neq j} \langle X_i, e_l \rangle \langle X_j, e_l \rangle \langle X_{i-h}, X_{j-h} \rangle \right). \end{aligned}$$

By Assumption 3.2 (iii) we have that $E \sum_{i \neq j} \langle X_i, e_l \rangle \langle X_j, e_l \rangle \langle X_{i-h}, X_{j-h} \rangle = 0$. It then follows from Parseval's identity, Cauchy–Schwarz inequality and Assumption 3.1 that

$$E \|\widehat{\Gamma}_h\|_S^2 = \frac{1}{N^2} \sum_{i=1+h}^N E \|X_i\|^2 \|X_{i-h}\|^2 \leq \frac{N-h}{N^2} E \|X_0\|^4 = O(N^{-1}).$$

□

The following proposition shows that the difference between the autocovariance column matrix $\widehat{\mathbf{T}}_N$ and its population counterpart $\mathbf{T}_N$ is asymptotically negligible.

Proposition 6.1 Under Assumptions 3.1 and 3.2,

$$\widehat{\mathbf{T}}_N - \mathbf{T}_N = O_P(N^{-1/2}),$$

where $\widehat{\mathbf{T}}_N$ is defined in (3.7), and $\mathbf{T}_N$ is defined in (3.5).

Proof It is sufficient to show that for any $h = 1, \dots, H$ and any $k, l = 1, \dots, p$,

$$\sqrt{N} \{\widehat{c}_{N,h}(k, l) - c_{N,h}(k, l)\} \xrightarrow{P} 0. \quad (6.1)$$

To show (6.1), consider the decomposition

$$\begin{aligned} &\sqrt{N} \{\widehat{c}_{N,h}(k, l) - c_{N,h}(k, l)\} \\ &= \frac{1}{\sqrt{N}} \sum_{i=1+h}^N (\widehat{\xi}_{i-h,k} \widehat{\xi}_{i,l} - \xi_{i-h,k} \xi_{i,l}) \\ &= \frac{1}{\sqrt{N}} \sum_{i=1+h}^N (\widehat{\xi}_{i-h,k} - \xi_{i-h,k}) \xi_{i,l} + \frac{1}{\sqrt{N}} \sum_{i=1+h}^N \xi_{i-h,k} (\widehat{\xi}_{i,l} - \xi_{i,l}) \\ &\quad + \frac{1}{\sqrt{N}} \sum_{i=1+h}^N (\widehat{\xi}_{i-h,k} - \xi_{i-h,k}) (\widehat{\xi}_{i,l} - \xi_{i,l}) \\ &=: P_1(N) + P_2(N) + P_3(N) \end{aligned}$$

It follows from Cauchy–Schwarz inequality that

$$|P_1(N)| = \left| \left\langle \frac{1}{\sqrt{N}} \sum_{i=1+h}^N \langle X_i, v_l \rangle X_{i-h}, \widehat{v}_k - v_k \right\rangle \right| \leq \|\widehat{\Gamma}_h(v_l)\| \|\sqrt{N}(\widehat{v}_k - v_k)\|.$$

where $\widehat{\Gamma}_h(\cdot) = \frac{1}{N} \sum_{i=1+h}^N \langle X_i, \cdot \rangle X_{i-h}$. Then, by Lemmas 6.1, 6.2, we get that $P_1(N) = O_P(N^{-1/2})$. Analogously, we can show that $P_2(N) = O_P(N^{-1/2})$. To show $P_3(N) \xrightarrow{P} 0$, observe that

$$|P_3(N)| = \left| \left\langle \widehat{\Gamma}_h(\hat{v}_k - v_k), \sqrt{N}(\hat{v}_l - v_l) \right\rangle \right| \leq \frac{1}{\sqrt{N}} \|\widehat{\Gamma}_h\|_{\mathcal{S}} \left\| \sqrt{N}(\hat{v}_k - v_k) \right\| \left\| \sqrt{N}(\hat{v}_l - v_l) \right\|.$$

Then, by Lemmas 6.1, 6.2, $P_3(N) = O_P(N^{-1})$. $\square$

The following two lemmas are used to prove Proposition 6.2. The first lemma states that if random functions are L^4 - m -approximable, then their score vectors preserve the weak dependence structure.

Lemma 6.3 *Under Assumption 3.1, the sequence of vectors $\{\mathbf{Y}_i, i \in \mathbb{Z}\}$ defined in (3.2) is L^4 - m -approximable.*

Proof Since $\mathbf{Y}_i = [\langle X_i, v_1 \rangle, \langle X_i, v_2 \rangle, \dots, \langle X_i, v_p \rangle]^\top$, it follows from Assumption 3.1 that $\mathbf{Y}_i$ has a Bernoulli shift representation such that $\mathbf{Y}_i = f(\varepsilon_i, \varepsilon_{i-1}, \dots)$ where $f : \mathbb{S}^\infty \rightarrow \mathbb{R}^p$. Also, let $\mathbf{Y}_i^{(m)} = [\langle X_i^{(m)}, v_1 \rangle, \langle X_i^{(m)}, v_2 \rangle, \dots, \langle X_i^{(m)}, v_p \rangle]^\top$, where $\{X_i^{(m)}\}$ is defined as in (3.12), then $\{\mathbf{Y}_i^{(m)}\}$ can also be expressed as a Bernoulli shift representation such that $\mathbf{Y}_i^{(m)} = f(\varepsilon_i, \varepsilon_{i-1}, \dots, \varepsilon_{i-m+1}, \varepsilon'_{i-m}, \varepsilon'_{i-m-1}, \dots)$ with the independent copy $\{\varepsilon'_i\}$.

It follows from Cauchy–Schwarz inequality that

$$\left\| \mathbf{Y}_i - \mathbf{Y}_i^{(m)} \right\|_F^2 = \sum_{j=1}^p \left\langle X_i - X_i^{(m)}, v_j \right\rangle^2 \leq p \left\| X_i - X_i^{(m)} \right\|^2.$$

Therefore by Assumption 3.1 we get that

$$\sum_{m=0}^{\infty} \left(E \left\| \mathbf{Y}_i - \mathbf{Y}_i^{(m)} \right\|_F^4 \right)^{1/4} \leq p^{1/2} \sum_{m=0}^{\infty} \left(E \left\| X_i - X_i^{(m)} \right\|^4 \right)^{1/4} < \infty.$$

$\square$

Using Lemma 6.3 we can further obtain the following result.

Lemma 6.4 *Under Assumption 3.1, the sequence of vectors $\{\mathbf{S}_i, i \in \mathbb{Z}\}$ defined in (3.3) is L^2 - m -approximable.*

Proof Recall that $\mathbf{S}_i$ is the row expansion of the block matrix

$$\left[\mathbf{Y}_{i-1} \mathbf{Y}_i^\top, \mathbf{Y}_{i-2} \mathbf{Y}_i^\top, \dots, \mathbf{Y}_{i-H} \mathbf{Y}_i^\top \right]^\top.$$

Therefore, by Lemma 6.3, $\mathbf{S}_i$ admits a Bernoulli shift representation such that $\mathbf{S}_i = g(\varepsilon_i, \varepsilon_{i-1}, \dots)$ where $g : \mathbb{S}^\infty \rightarrow \mathbb{R}^{Hp^2}$. Also, let $\mathbf{S}_i^{(m)}$ be the row expansion of the block matrix $[\mathbf{Y}_{i-1}^{(m-1)} \mathbf{Y}_i^{(m)\top}, \mathbf{Y}_{i-2}^{(m-2)} \mathbf{Y}_i^{(m)\top}, \dots, \mathbf{Y}_{i-H}^{(m-H)} \mathbf{Y}_i^{(m)\top}]^\top$ for $m > H$, then $\mathbf{S}_i^{(m)}$ can be expressed as $\mathbf{S}_i^{(m)} = g(\varepsilon_i, \varepsilon_{i-1}, \dots, \varepsilon_{i-m+1}, \varepsilon'_{i-m}, \varepsilon'_{i-m-1}, \dots)$ with the independent copy $\{\varepsilon'_i\}$.

It then follows from Minkowski inequality and Cauchy–Schwarz inequality that

$$\left(E \left\| \mathbf{S}_i - \mathbf{S}_i^{(m)} \right\|_F^2 \right)^{1/2}$$

$$\begin{aligned}
 &\leq \sum_{h=1}^H \left(E \| \mathbf{Y}_{i-h} \mathbf{Y}_i^\top - \mathbf{Y}_{i-h}^{(m-h)} \mathbf{Y}_i^{(m)\top} \|_F^2 \right)^{1/2} \\
 &\leq \sum_{h=1}^H \left\{ \left(E \| \mathbf{Y}_{i-h} (\mathbf{Y}_i - \mathbf{Y}_i^{(m)})^\top \|_F^2 \right)^{1/2} + \left(E \| (\mathbf{Y}_{i-h} - \mathbf{Y}_{i-h}^{(m-h)}) \mathbf{Y}_i^{(m)\top} \|_F^2 \right)^{1/2} \right\} \\
 &\leq \sum_{h=1}^H \left\{ \left(E \| \mathbf{Y}_{i-h} \|_F^4 \right)^{1/4} \left(E \| \mathbf{Y}_i - \mathbf{Y}_i^{(m)} \|_F^4 \right)^{1/4} + \left(E \| \mathbf{Y}_{i-h} - \mathbf{Y}_{i-h}^{(m-h)} \|_F^4 \right)^{1/4} \left(E \| \mathbf{Y}_i^{(m)} \|_F^4 \right)^{1/4} \right\}.
 \end{aligned}$$

Thus, by Lemma 6.3 we obtain that $\sum_{m=0}^\infty (E \| \mathbf{S}_i - \mathbf{S}_i^{(m)} \|_F^2)^{1/2} < \infty$. $\square$

Proposition 6.2 Under Assumptions 3.1 and 3.2,

$$\mathbf{T}_N \xrightarrow{d} \mathbf{G}, \quad N \rightarrow \infty,$$

where $\mathbf{T}_N$ is defined in (3.5), and $\mathbf{G}$ is a Hp^2 -dimensional mean zero Gaussian random vector with covariance matrix Σ defined in (3.4).

Proof By the Cramér–Wold theorem it is sufficient to show that for any $\mathbf{t} \in \mathbb{R}^{Hp^2}$

$$\mathbf{t}^\top \mathbf{T}_N \xrightarrow{d} \mathbf{t}^\top \mathbf{G}, \quad N \rightarrow \infty. \quad (6.2)$$

Take any $\mathbf{t} \in \mathbb{R}^{Hp^2}$ and let $\mathbf{t} = [t_{111}, t_{112}, \dots, t_{1pp}, t_{211}, \dots, t_{2pp}, \dots, t_{hkl}, \dots, t_{Hpp}]^\top$. Define

$$Z_i(\mathbf{t}) = \mathbf{t}^\top \mathbf{S}_i = \sum_{h=1}^H \sum_{k,l=1}^p t_{hkl} \xi_{i-h,k} \xi_{i,l},$$

where $\mathbf{S}_i$ is defined in (3.3).

By Lemma 6.4, $\{Z_i(\mathbf{t}), i \in \mathbb{Z}\}$ is L^2 - m -approximable since $\{Z_i(\mathbf{t})\}$ is a sequence of projections of L^2 - m -approximable sequence $\{\mathbf{S}_i\}$. It follows from Assumption 3.2 (ii) that $E[Z_i(\mathbf{t})] = \langle \mathbf{t}, E\mathbf{S}_i \rangle = 0$ since $E\xi_{i-h,k}\xi_{i,l} = 0$ for any $h \neq 0$ and any $k, l = 1, \dots, p$. Also, by Assumption 3.2 (iii) we get the covariance structure of $Z_i(\mathbf{t})$ such that

$$EZ_0(\mathbf{t})Z_j(\mathbf{t}) = E[\mathbf{t}^\top \mathbf{S}_0 \mathbf{S}_j^\top \mathbf{t}] = \mathbf{t}^\top E[\mathbf{S}_0 \mathbf{S}_j^\top] \mathbf{t} = \begin{cases} \mathbf{t}^\top \Sigma \mathbf{t}, & j = 0, \\ 0, & j \neq 0. \end{cases}$$

Therefore we have $\sum_{j=-\infty}^\infty EZ_0(\mathbf{t})Z_j(\mathbf{t}) = \mathbf{t}^\top \Sigma \mathbf{t}$, and by Theorem 3 of Wu (2005) we can conclude that

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N Z_i(\mathbf{t}) \xrightarrow{d} \mathbf{t}^\top \mathbf{G}.$$

Now observe that

$$\begin{aligned}
 \mathbf{t}^\top \mathbf{T}_N &= \frac{1}{\sqrt{N}} \sum_{h=1}^H \sum_{i=1+h}^N \sum_{k,l=1}^p t_{hkl} \xi_{i-h,k} \xi_{i,l} \\
 &= \frac{1}{\sqrt{N}} \sum_{i=1}^N Z_i(\mathbf{t}) - \frac{1}{\sqrt{N}} \sum_{h=1}^H \sum_{i=1}^h \sum_{k,l=1}^p t_{hkl} \xi_{i-h,k} \xi_{i,l}
 \end{aligned}$$

$$=: \frac{1}{\sqrt{N}} \sum_{i=1}^N Z_i(\mathbf{t}) - P_4(N).$$

It follows from Cauchy–Schwarz inequality that $E|\xi_{i-h,k}\xi_{i,l}| \leq (E\xi_{i-h,k}^2)^{1/2}(E\xi_{i,l}^2)^{1/2} = \lambda_k^{1/2}\lambda_l^{1/2} < \infty$. Therefore by Markov’s inequality we have that $P_4(N) = O_P(N^{-1/2})$, and (6.2) is then proven by Slutsky’s theorem. $\square$

PROOF OF THEOREM 3.1: Using Slutsky’s theorem, the proof follows from Propositions 6.1, 6.2.

6.1.2 Proof of Theorem 3.2

The key for the proof is to show that $\widehat{\Sigma}_N$ in (3.8) is asymptotically close to Σ in (3.4), which will be shown in Proposition 6.3. The following lemma is a preliminary result, which will be used to prove Propositions 6.3 and 6.6.

Lemma 6.5 *Suppose that $\{U_i\}$, $\{\widehat{U}_i\}$, $\{V_i\}$, $\{\widehat{V}_i\}$, $\{W_i\}$, $\{\widehat{W}_i\}$, $\{Z_i\}$, and $\{\widehat{Z}_i\}$ are sequences of random variables and satisfy the followings*

$$\begin{aligned} |U_i| &\leq C_i^u B_i^u; \quad |V_i| \leq C_i^v B_i^v; \quad |W_i| \leq C_i^w B_i^w; \quad |Z_i| \leq C_i^z B_i^z; \\ |\widehat{U}_i - U_i| &\leq D_i^u B_i^u; \quad |\widehat{V}_i - V_i| \leq D_i^v B_i^v; \quad |\widehat{W}_i - W_i| \leq D_i^w B_i^w; \quad |\widehat{Z}_i - Z_i| \leq D_i^z B_i^z, \end{aligned}$$

where $\{C_i^u\}$, $\{C_i^v\}$, $\{C_i^w\}$, $\{C_i^z\}$, $\{D_i^u\}$, $\{D_i^v\}$, $\{D_i^w\}$, and $\{D_i^z\}$ are sequences of non-negative random variables depending on N , $\{B_i^u\}$, $\{B_i^v\}$, $\{B_i^w\}$, $\{B_i^z\}$ are sequences of random variables not depending on N , and as $N \rightarrow \infty$

$$C_i^u(N) = O_P(1); \quad C_i^v(N) = O_P(1); \quad C_i^w(N) = O_P(1); \quad C_i^z(N) = O_P(1); \quad (6.3)$$

$$D_i^u(N) = o_P(1); \quad D_i^v(N) = o_P(1); \quad D_i^w(N) = o_P(1); \quad D_i^z(N) = o_P(1); \quad (6.4)$$

$$\sup_i E(B_i^u)^4 < \infty; \quad \sup_i E(B_i^v)^4 < \infty; \quad \sup_i E(B_i^w)^4 < \infty; \quad \sup_i E(B_i^z)^4 < \infty. \quad (6.5)$$

Then it holds that

$$\frac{1}{N} \sum_{i=1}^N (\widehat{U}_i \widehat{V}_i \widehat{W}_i \widehat{Z}_i - U_i V_i W_i Z_i) \xrightarrow{P} 0, \quad N \rightarrow \infty.$$

Proof Consider the following decomposition

$$\begin{aligned} &\frac{1}{N} \sum_{i=1}^N (\widehat{U}_i \widehat{V}_i \widehat{W}_i \widehat{Z}_i - U_i V_i W_i Z_i) \\ &= \frac{1}{N} \sum_{i=1}^N (\widehat{U}_i - U_i)(\widehat{V}_i - V_i)(\widehat{W}_i - W_i)(\widehat{Z}_i - Z_i) \\ &\quad + \frac{1}{N} \sum_{i=1}^N \{(\widehat{U}_i - U_i)V_i + U_i(\widehat{V}_i - V_i) + (\widehat{U}_i - U_i)(\widehat{V}_i - V_i)\} W_i Z_i \\ &\quad + \frac{1}{N} \sum_{i=1}^N U_i V_i \{(\widehat{W}_i - W_i)Z_i + W_i(\widehat{Z}_i - Z_i) + (\widehat{W}_i - W_i)(\widehat{Z}_i - Z_i)\} \end{aligned}$$

$$- \frac{1}{N} \sum_{i=1}^N \{(\widehat{U}_i - U_i)V_i + U_i(\widehat{V}_i - V_i)\} \{(\widehat{W}_i - W_i)Z_i + W_i(\widehat{Z}_i - Z_i)\}.$$

We will show that the following holds

$$\frac{1}{N} \sum_{i=1}^N (\widehat{U}_i - U_i)(\widehat{V}_i - V_i)(\widehat{W}_i - W_i)(\widehat{Z}_i - Z_i) = o_p(1); \quad (6.6)$$

$$\frac{1}{N} \sum_{i=1}^N (\widehat{U}_i - U_i)(\widehat{V}_i - V_i)W_iZ_i = o_p(1); \quad (6.7)$$

$$\frac{1}{N} \sum_{i=1}^N (\widehat{U}_i - U_i)V_iW_iZ_i = o_p(1). \quad (6.8)$$

Analogous arguments prove that the other partial sums in the decomposition converge to zero in probability as well.

Observe that

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N (\widehat{U}_i - U_i)(\widehat{V}_i - V_i)(\widehat{W}_i - W_i)(\widehat{Z}_i - Z_i) &\leq D_i^u D_i^v D_i^w D_i^z \times \left(\frac{1}{N} \sum_{i=1}^N B_i^u B_i^v B_i^w B_i^z \right); \\ \frac{1}{N} \sum_{i=1}^N (\widehat{U}_i - U_i)(\widehat{V}_i - V_i)W_iZ_i &\leq D_i^u D_i^v C_i^w C_i^z \times \left(\frac{1}{N} \sum_{i=1}^N B_i^u B_i^v B_i^w B_i^z \right); \\ \frac{1}{N} \sum_{i=1}^N (\widehat{U}_i - U_i)V_iW_iZ_i &\leq D_i^u C_i^v C_i^w C_i^z \times \left(\frac{1}{N} \sum_{i=1}^N B_i^u B_i^v B_i^w B_i^z \right). \end{aligned}$$

By applying Cauchy–Schwarz inequality multiple times and by (6.5), we have that

$$\sup_i E B_i^u B_i^v B_i^w B_i^z \leq \sup_i \{E(B_i^u)^4\}^{1/4} \{E(B_i^v)^4\}^{1/4} \{E(B_i^w)^4\}^{1/4} \{E(B_i^z)^4\}^{1/4} < \infty.$$

It then follows from Markov inequality that $\frac{1}{N} \sum_{i=1}^N B_i^u B_i^v B_i^w B_i^z = O_P(1)$. Therefore, convergences (6.6), (6.7), and (6.8) can be proven by the conditions (6.3), (6.4), and (6.5). $\square$

Proposition 6.3 *Under Assumption 3.1,*

$$\widehat{\Sigma}_N - \Sigma = o_P(1).$$

Proof It is sufficient to show that for any $h, h' = 1, \dots, H$ and any $k, l, k', l' = 1, \dots, p$,

$$\frac{1}{N} \sum_{i=1+\max\{h,h'\}}^N \widehat{\xi}_{i-h,k} \widehat{\xi}_{i,l} \widehat{\xi}_{i-h',k'} \widehat{\xi}_{i,l'} - E \xi_{-h,k} \xi_{0,l} \xi_{-h',k'} \xi_{0,l'} \xrightarrow{P} 0. \quad (6.9)$$

To show (6.9), consider the following decomposition

$$\begin{aligned} &\frac{1}{N} \sum_{i=1+\max\{h,h'\}}^N \widehat{\xi}_{i-h,k} \widehat{\xi}_{i,l} \widehat{\xi}_{i-h',k'} \widehat{\xi}_{i,l'} - E \xi_{-h,k} \xi_{0,l} \xi_{-h',k'} \xi_{0,l'} \\ &= \frac{1}{N} \sum_{i=1+\max\{h,h'\}}^N (\widehat{\xi}_{i-h,k} \widehat{\xi}_{i,l} \widehat{\xi}_{i-h',k'} \widehat{\xi}_{i,l'} - \xi_{i-h,k} \xi_{i,l} \xi_{i-h',k'} \xi_{i,l'}) \end{aligned}$$

$$\begin{aligned}
 & + \frac{1}{N} \sum_{i=1+\max\{h,h'\}}^N \xi_{i-h,k} \xi_{i,l} \xi_{i-h',k'} \xi_{i,l'} - E \xi_{-h,k} \xi_{0,l} \xi_{-h',k'} \xi_{0,l'} \\
 & =: P_5(N) + P_6(N)
 \end{aligned}$$

For $P_5(N)$, observe that for any $i = 1, \dots, N$, and any $k = 1, \dots, p$, $|\hat{\xi}_{i,k} - \xi_{i,k}| \leq \|\hat{v}_k - v_k\| \|X_i\|$, $|\xi_{i,k}| \leq \|X_i\|$. By Lemma 6.1 and Assumption 3.1, we have that $\|\hat{v}_k - v_k\| = O_P(N^{-1/2})$ and $E\|X_i\|^4 < \infty$. Thus, by Lemma 6.5, $P_5(N) = o_P(1)$.

For $P_6(N)$, observe that we have a Bernoulli shift representation such that $\xi_{i-h,k} \xi_{i,l} \xi_{i-h',k'} \xi_{i,l'} = g(\varepsilon_i, \varepsilon_{i-1}, \dots)$, which is stationary and ergodic. Then, it follows from the mean ergodic theorem that $P_6(N) = o_P(1)$. $\square$

The following lemma follows from equation (5.8.6) of Horn and Johnson (2012) and is needed to prove Theorem 3.2.

Lemma 6.6 Let $\kappa(\Sigma) = \|\Sigma^{-1}\|_F \|\Sigma\|_F$, then

$$\frac{\|\Sigma^{-1} - \hat{\Sigma}^{-1}\|_F}{\|\Sigma^{-1}\|_F} \leq \frac{\kappa(\Sigma)}{1 - \kappa(\Sigma) \|\Sigma - \hat{\Sigma}\|_F / \|\Sigma\|_F} \frac{\|\Sigma - \hat{\Sigma}\|_F}{\|\Sigma\|_F}.$$

PROOF OF THEOREM 3.2: By Theorem 3.1, we have that $\hat{\mathbf{T}}_N \xrightarrow{d} \mathbf{G}$, where $\mathbf{G}$ is a mean zero Gaussian random vector with covariance Σ . Consider the map $f: \mathbb{R}^{Hp^2} \rightarrow \mathbb{R}$ defined by $f(\mathbf{t}) = \mathbf{t}^\top \Sigma^{-1} \mathbf{t}$ for $\mathbf{t} \in \mathbb{R}^{Hp^2}$. Applying the continuous mapping theorem, we get that $\hat{\mathbf{T}}_N^\top \Sigma^{-1} \hat{\mathbf{T}}_N = f(\hat{\mathbf{T}}_N) \xrightarrow{d} f(\mathbf{G}) = \chi_{Hp^2}^2$.

Observe that

$$\hat{Q}_N = \hat{\mathbf{T}}_N^\top \hat{\Sigma}_N^{-1} \hat{\mathbf{T}}_N = \hat{\mathbf{T}}_N^\top \Sigma^{-1} \hat{\mathbf{T}}_N + \hat{\mathbf{T}}_N^\top (\hat{\Sigma}_N^{-1} - \Sigma^{-1}) \hat{\mathbf{T}}_N.$$

Since Σ is positive definite, we have that $\hat{\Sigma}_N^{-1} - \Sigma^{-1} = o_P(1)$ by Proposition 6.3 and Lemma 6.6. It thus follows from Slutsky's theorem that $\hat{Q}_N \xrightarrow{d} \chi_{Hp^2}^2$.

6.1.3 Proof of Theorem 3.3

Observe that $\hat{Q}_N = \hat{\mathbf{T}}_N^\top \Sigma^{-1} \hat{\mathbf{T}}_N + \hat{\mathbf{T}}_N^\top (\hat{\Sigma}_N^{-1} - \Sigma^{-1}) \hat{\mathbf{T}}_N$. By Proposition 6.3 and Lemma 6.6, we have that $\hat{\mathbf{T}}_N^\top (\hat{\Sigma}_N^{-1} - \Sigma^{-1}) \hat{\mathbf{T}}_N \xrightarrow{P} 0$. Since Σ^{-1} is positive-definite, we aim to show that $\hat{c}_{N,h}(k, l)$ converges to a nonzero constant when there exists some $h \geq 1$ and some $k, l = 1, \dots, p$, such that $E \xi_{0,k} \xi_{h,l} \neq 0$.

For this, it is sufficient to show that

$$c_{N,h}^2(k, l) \xrightarrow{P} q > 0, \quad \sqrt{N}(\hat{c}_{N,h}(k, l) - c_{N,h}(k, l)) = O_P(1),$$

since $\hat{c}_{N,h}^2(k, l) = c_{N,h}^2(k, l) + 2c_{N,h}(k, l)(\hat{c}_{N,h}(k, l) - c_{N,h}(k, l)) + (\hat{c}_{N,h}(k, l) - c_{N,h}(k, l))^2$. To show $c_{N,h}^2(k, l) \xrightarrow{P} q > 0$, observe that

$$c_{N,h}(k, l)^2 = \left[\frac{1}{N} \sum_{i=1+h}^N \{ \xi_{i-h,k} \xi_{i,l} - E \xi_{i-h,k} \xi_{i,l} + E \xi_{i-h,k} \xi_{i,l} \} \right]^2$$

$$\begin{aligned}
 &= \left[\frac{1}{N} \sum_{i=1+h}^N \{ \xi_{i-h,k} \xi_{i,l} - E \xi_{i-h,k} \xi_{i,l} \} \right]^2 \\
 &\quad + 2 \frac{N-h}{N} E \xi_{i-h,k} \xi_{i,l} \frac{1}{N} \sum_{i=1+h}^N \{ \xi_{i-h,k} \xi_{i,l} - E \xi_{i-h,k} \xi_{i,l} \} \\
 &\quad + \frac{(N-h)^2}{N^2} \{ E \xi_{0,k} \xi_{h,l} \}^2.
 \end{aligned}$$

By Lemma 6.4, $\{\xi_{i-h,k} \xi_{i,l}\}$ is L^2 - m -approximable, and then by Theorem 3 of Wu (2005) $\frac{1}{\sqrt{N}} \sum_{i=1+h}^N \{ \xi_{i-h,k} \xi_{i,l} - E \xi_{i-h,k} \xi_{i,l} \}$ converges to some Gaussian random variable. From this, we obtain $c_{N,h}(k, l)^2 = O_P(N^{-1}) + O_P(N^{-1/2}) + O(1)$. Now, to show $\sqrt{N}(\hat{c}_{N,h}(k, l) - c_{N,h}(k, l)) = O_P(1)$, observe that $\sqrt{N}(\hat{c}_{N,h}(k, l) - c_{N,h}(k, l)) = P_1(N) + P_2(N) + P_3(N)$ where $P_1(N)$, $P_2(N)$, and $P_3(N)$ are defined in the proof of Proposition 6.1. Then by applying Cauchy–Schwarz inequality multiple times and using $E \|X_0\|^4 < \infty$, it can be readily shown that $P_1(N) = O_P(1)$, $P_2(N) = O_P(1)$, and $P_3(N) = O_P(N^{-1/2})$, respectively.

6.2 Proofs of the asymptotic results in Section 4

6.2.1 Proof of Theorem 4.1

Recall that the projection of ε_i onto the FPCs v_k , $r_{i,k} = \langle \varepsilon_i, v_k \rangle$, for $k = 1, \dots, p$. For p -dimensional vectors of the projected errors $[r_{i,1}, \dots, r_{i,p}]^\top$, $i = 2, \dots, N$, define the autocovariance function by

$$c_{N,h}^r(k, l) = \frac{1}{N} \sum_{i=1+h}^N \{ r_{i,k} r_{i-h,l} - r_{i,k} f_{i,l}^{(N)}(h) \}, \quad f_{i,l}^{(N)}(h) = \left\langle \Gamma_\varepsilon \Psi_*^{h-1} \Gamma^{-1} \pi_{q_N}(X_{i-1}), v_l \right\rangle, \quad (6.10)$$

where $\Gamma^{-1} \pi_{q_N}(\cdot) = \sum_{q=1}^{q_N} \frac{\langle \cdot, v_q \rangle}{\lambda_q} v_q$ is the Moore–Penrose inverse of Γ in (2.1) with $\pi_{q_N}(\cdot) = \sum_{q=1}^{q_N} \langle \cdot, v_q \rangle v_q$ being the projection operator on the closed linear span of the first q_N FPCs. Consider the population autocovariance column matrix

$$\mathbf{T}_{r,N} = \sqrt{N} \begin{bmatrix} c_{N,1}^r(1, 1), & c_{N,1}^r(1, 2), & \dots, & c_{N,1}^r(p, p), & \dots \dots, & c_{N,h}^r(1, 1), & \dots, & c_{N,h}^r(p, p) \end{bmatrix}^\top. \quad (6.11)$$

For the proof of Theorem 4.1 we first show that the sample autocovariance column matrix $\hat{\mathbf{T}}_{r,N}$ defined in (4.4) is asymptotically close to its population counterpart $\mathbf{T}_{r,N}$ in (6.11), which will be presented in Proposition 6.4. We then investigate the asymptotic distribution of $\mathbf{T}_{r,N}$ in Proposition 6.5.

The following lemma is used to prove Proposition 6.4.

Lemma 6.7 *Under Assumptions 4.1, 4.2 and 4.3,*

$$\|\hat{\Gamma}_{\varepsilon,h}\|_{\mathcal{S}} = o_P(N^{-1/4}),$$

where $\hat{\Gamma}_{\varepsilon,h}(\cdot) = \frac{1}{N} \sum_{i=1+h}^N \langle \hat{\varepsilon}_{i-h}, \cdot \rangle \hat{\varepsilon}_i$.

Proof Using that $\hat{\varepsilon}_i = \varepsilon_i + \mathcal{D}_N(X_{i-1})$ where $\mathcal{D}_N = \Psi - \hat{\Psi}_N$, we have that

$$\begin{aligned} \|\hat{\Gamma}_{\varepsilon,h}\|_S &\leq \left\| \frac{1}{N} \sum_{i=1+h}^N \langle \varepsilon_{i-h}, \cdot \rangle \varepsilon_i \right\|_S + \left\| \frac{1}{N} \sum_{i=1+h}^N \langle \mathcal{D}_N(X_{i-h-1}), \cdot \rangle \varepsilon_i \right\|_S \\ &\quad + \left\| \frac{1}{N} \sum_{i=1+h}^N \langle \varepsilon_{i-h}, \cdot \rangle \mathcal{D}_N(X_{i-1}) \right\|_S + \left\| \frac{1}{N} \sum_{i=1+h}^N \langle \mathcal{D}_N(X_{i-h-1}), \cdot \rangle \mathcal{D}_N(X_{i-1}) \right\|_S \\ &=: P_7(N) + P_8(N) + P_9(N) + P_{10}(N). \end{aligned}$$

Take any set of orthonormal basis $\{e_l\}$ in L^2 . For P_7 , we have that

$$\begin{aligned} \{P_7(N)\}^2 &= \sum_{l=1}^{\infty} \left\| \frac{1}{N} \sum_{i=1+h}^N \langle \varepsilon_{i-h}, e_l \rangle \varepsilon_i \right\|^2 \\ &= \frac{1}{N^2} \sum_{l=1}^{\infty} \left(\sum_{i=1+h}^N \langle \varepsilon_{i-h}, e_l \rangle^2 \|\varepsilon_i\|^2 + \sum_{i \neq j} \langle \varepsilon_{i-h}, e_l \rangle \langle \varepsilon_{j-h}, e_l \rangle \langle \varepsilon_i, \varepsilon_j \rangle \right). \end{aligned}$$

By Assumption 4.3 we have that $E \sum_{i \neq j} \langle \varepsilon_{i-h}, e_l \rangle \langle \varepsilon_{j-h}, e_l \rangle \langle \varepsilon_i, \varepsilon_j \rangle = 0$. It then follows from Parseval's identity and Cauchy-Schwarz inequality that $E\{P_7(N)\}^2 = O(N^{-1})$.

For P_8 , we have that

$$\begin{aligned} \{P_8(N)\}^2 &= \frac{1}{N^2} \sum_{l=1}^{\infty} \left\| \sum_{i=1+h}^N \langle \mathcal{D}_N(X_{i-h-1}), e_l \rangle \varepsilon_i \right\|^2 \\ &= \frac{1}{N^2} \sum_{l=1}^{\infty} \left(\sum_{i=1+h}^N \langle \mathcal{D}_N(X_{i-h-1}), e_l \rangle^2 \|\varepsilon_i\|^2 \right. \\ &\quad \left. + \sum_{i \neq j} \langle \mathcal{D}_N(X_{i-h-1}), e_l \rangle \langle \mathcal{D}_N(X_{j-h-1}), e_l \rangle \langle \varepsilon_i, \varepsilon_j \rangle \right) \\ &= \frac{1}{N^2} \left(\sum_{i=1+h}^N \|\mathcal{D}_N(X_{i-h-1})\|^2 \|\varepsilon_i\|^2 \right. \\ &\quad \left. + \sum_{i \neq j} \langle \mathcal{D}_N(X_{i-h-1}), \mathcal{D}_N(X_{j-h-1}) \rangle \langle \varepsilon_i, \varepsilon_j \rangle \right) \\ &\leq \frac{1}{N^2} \sum_{i=1+h}^N \|\mathcal{D}_N\|_S^2 \|X_{i-h-1}\|^2 \|\varepsilon_i\|^2 \\ &\quad + \frac{1}{N^2} \sum_{i \neq j} \|\mathcal{D}_N\|_S^2 \|X_{i-h-1}\| \|X_{j-h-1}\| \|\varepsilon_i\| \|\varepsilon_j\|. \end{aligned}$$

It then follows from Cauchy-Schwarz inequality and Assumption 4.2 that $\{P_8(N)\}^2 = o_P(N^{-1/2})$. Analogously, we can get $\{P_9(N)\}^2 = o_P(N^{-1/2})$, and $\{P_{10}(N)\}^2 = o_P(N^{-1})$. $\square$

The following proposition proves that the difference between $\hat{\mathbf{T}}_{r,N}$ and $\mathbf{T}_{r,N}$ is asymptotically negligible.

Proposition 6.4 Under Assumptions 4.1, 4.2 and 4.3,

$$\widehat{\mathbf{T}}_{r,N} - \mathbf{T}_{r,N} = o_P(1)$$

Proof It is sufficient to show that for any $h = 1, \dots, H$ and any $k, l = 1, \dots, p$,

$$\sqrt{N} \{ \tilde{c}_{N,h}^r(k, l) - c_{N,h}^r(k, l) \} \xrightarrow{P} 0.$$

Let $\tilde{c}_{N,h}^r(k, l) = \frac{1}{N} \sum_{i=1+h}^N \tilde{r}_{i,k} \tilde{r}_{i-h,l}$ where $\tilde{r}_{i,k} = \langle \hat{\varepsilon}_i, v_k \rangle$. Then we have that

$$\begin{aligned} \sqrt{N} \{ \tilde{c}_{N,h}^r(k, l) - c_{N,h}^r(k, l) \} &= \sqrt{N} \{ \hat{c}_{N,h}^r(k, l) - \tilde{c}_{N,h}^r(k, l) \} + \sqrt{N} \{ \tilde{c}_{N,h}^r(k, l) - c_{N,h}^r(k, l) \} \\ &=: P_{11}(N) + P_{12}(N). \end{aligned}$$

We aim to show that $P_{11}(N) \xrightarrow{P} 0$, and $P_{12}(N) \xrightarrow{P} 0$. For $P_{11}(N)$, we obtain the following decomposition

$$\begin{aligned} &\sqrt{N} \{ \hat{c}_{N,h}^r(k, l) - \tilde{c}_{N,h}^r(k, l) \} \\ &= \frac{1}{\sqrt{N}} \sum_{i=1+h}^N (\hat{r}_{i,k} \hat{r}_{i-h,l} - \tilde{r}_{i,k} \tilde{r}_{i-h,l}) \\ &= \frac{1}{\sqrt{N}} \sum_{i=1+h}^N (\hat{r}_{i,k} - \tilde{r}_{i,k}) \tilde{r}_{i-h,l} + \frac{1}{\sqrt{N}} \sum_{i=1+h}^N \tilde{r}_{i,k} (\hat{r}_{i-h,l} - \tilde{r}_{i-h,l}) \\ &\quad + \frac{1}{\sqrt{N}} \sum_{i=1+h}^N (\hat{r}_{i,k} - \tilde{r}_{i,k}) (\hat{r}_{i-h,l} - \tilde{r}_{i-h,l}) \\ &=: P_{11,1}(N) + P_{11,2}(N) + P_{11,3}(N). \end{aligned}$$

Observe that

$$|P_{11,1}(N)| = \left| \left\langle \frac{1}{\sqrt{N}} \sum_{i=1+h}^N \langle \hat{\varepsilon}_{i-h}, v_l \rangle \hat{\varepsilon}_i, \hat{v}_k - v_k \right\rangle \right| \leq \|\widehat{\Gamma}_{\varepsilon,h}(v_l)\| \|\sqrt{N}(\hat{v}_k - v_k)\|,$$

where $\widehat{\Gamma}_{\varepsilon,h}(\cdot) = \frac{1}{N} \sum_{i=1+h}^N \langle \hat{\varepsilon}_{i-h}, \cdot \rangle \hat{\varepsilon}_i$. Then by Lemmas 6.1, 6.7, we have that $P_{11,1}(N) = o_P(N^{-1/4})$. Analogous arguments show that $P_{11,2}(N) = o_P(N^{-1/4})$. To show $P_{11,3}(N) \xrightarrow{P} 0$, we observe that

$$\begin{aligned} |P_{11,3}(N)| &= \left| \left\langle \frac{1}{\sqrt{N}} \sum_{i=1+h}^N \langle \hat{\varepsilon}_{i-h}, \hat{v}_l - v_l \rangle \hat{\varepsilon}_i, \hat{v}_k - v_k \right\rangle \right| \\ &\leq \frac{1}{\sqrt{N}} \|\widehat{\Gamma}_{\varepsilon,h}\|_{\mathcal{S}} \|\sqrt{N}(\hat{v}_l - v_l)\| \|\sqrt{N}(\hat{v}_k - v_k)\|. \end{aligned}$$

Then, by Lemmas 6.1, 6.7, $P_{11,3}(N) = o_P(N^{-3/4})$.

For $P_{12}(N)$, let

$$Y_{i,N}^{(h)}(u, v) = \varepsilon_i(u) \varepsilon_{i-h}(v) - \varepsilon_i(u) \Gamma_{\varepsilon} \Psi_*^{h-1} \Gamma^{-1} \pi_{q_N}(X_{i-1})(v). \quad (6.12)$$

Then as in the proof of Theorem 3.3 of Zhang (2016) it can be shown that $\|\frac{1}{\sqrt{N}} \sum_{i=1+h}^N \{\hat{\varepsilon}_i \otimes \hat{\varepsilon}_{i-h} - Y_{i,N}^{(h)}\}\|^2 = o_P(1)$. Take $v_k \otimes v_l \in L^2(\mathcal{T}^2)$, then we have that

$$\begin{aligned} |P_{12}(N)| &= \left| \left\langle \frac{1}{\sqrt{N}} \sum_{i=1+h}^N \{\hat{\varepsilon}_i \otimes \hat{\varepsilon}_{i-h} - Y_{i,N}^{(h)}\}, v_k \otimes v_l \right\rangle \right| \\ &\leq \left\| \frac{1}{\sqrt{N}} \sum_{i=1+h}^N \{\hat{\varepsilon}_i \otimes \hat{\varepsilon}_{i-h} - Y_{i,N}^{(h)}\} \right\| \|v_k \otimes v_l\|. \end{aligned}$$

Therefore, $P_{12}(N) = o_P(1)$. $\square$

We establish the asymptotic distribution of $\mathbf{T}_{r,N}$ in the following proposition.

Proposition 6.5 *Consider the FAR(1) model defined in (2.4) with the ε_i satisfying Assumption 4.3. Under Assumptions 4.1, 4.2, and 4.4, we have*

$$\mathbf{T}_{r,N} \xrightarrow{d} \mathbf{G}_r, \quad N \rightarrow \infty,$$

where $\mathbf{T}_{r,N}$ is defined in (6.11), and $\mathbf{G}_r$ is a Hp^2 -dimensional mean zero Gaussian random vector with covariance matrix Σ_r defined in (4.6).

Proof For any arbitrary vector $\mathbf{t} = [t_{111}, t_{112}, \dots, t_{1pp}, t_{211}, \dots, t_{2pp}, \dots, t_{hkl}, \dots, t_{Hpp}]^\top \in \mathbb{R}^{Hp^2}$, let

$$Z_{r,i}(\mathbf{t}) = \sum_{h=1}^H \sum_{k,l=1}^p t_{hkl} \left(r_{i,k} r_{i-h,l} - r_{i,k} f_{i,l}^{(N)}(h) \right). \quad (6.13)$$

First, we will show that the partial sum of $Z_{r,i}(\mathbf{t})$ converges to $\mathbf{t}^\top \mathbf{G}_r$. To this end, recall $Y_{i,N}^{(h)}$ defined in (6.12). Then, $Z_{r,i}(\mathbf{t}) = \sum_{h=1}^H \sum_{k,l=1}^p t_{hkl} \langle Y_{i,N}^{(h)}, v_k \otimes v_l \rangle$. It is proven in Section B.4 of Supplementary Material in Kim et al. (2023) that $\frac{1}{\sqrt{N}} \sum_{i=1}^N Y_{i,N}^{(h)}$ jointly converges for $h \in \{1, \dots, H\}$ to a mean zero Gaussian process, say G_h , in $L^2(\mathcal{T}^2)$ with covariance

$$\begin{aligned} E[G_h(u, v) G_{h'}(u', v')] \\ = E \left[(\varepsilon_i(u) \varepsilon_{i-h}(v) - \varepsilon_i(u) f_i^{(h)}(v)) (\varepsilon_i(u') \varepsilon_{i-h'}(v') - \varepsilon_i(u') f_i^{(h')}(v')) \right], \end{aligned}$$

where $f_i^{(h)} := \Gamma_\varepsilon \Psi_*^{h-1} \Gamma^{-1}(X_{i-1})$ defined in $L^2(\mathcal{T})$. Then, since $Z_{r,i}(\mathbf{t})$ is a linear combination of projections of $Y_{i,N}^{(h)}$ onto $\{v_k \otimes v_l\}_{k,l=1,\dots,p}$, we have that

$$\begin{aligned} \frac{1}{\sqrt{N}} \sum_{i=1}^N Z_{r,i}(\mathbf{t}) &= \frac{1}{\sqrt{N}} \sum_{i=1}^N \sum_{h=1}^H \sum_{k,l=1}^p t_{hkl} \langle Y_{i,N}^{(h)}, v_k \otimes v_l \rangle \\ &\xrightarrow{d} \sum_{h=1}^H \sum_{k,l=1}^p t_{hkl} \langle G_h, v_k \otimes v_l \rangle \stackrel{d}{=} \mathbf{t}^\top \mathbf{G}_r. \end{aligned}$$

Now observe that

$$\mathbf{t}^\top \mathbf{T}_{r,N} = \frac{1}{\sqrt{N}} \sum_{h=1}^H \sum_{i=1+h}^N \sum_{k,l=1}^p t_{hkl} (r_{i,k} r_{i-h,l} - r_{i,k} f_{i,l}^{(N)}(h))$$

$$\begin{aligned}
 &= \frac{1}{\sqrt{N}} \sum_{i=1}^N Z_{r,i}(\mathbf{t}) - \frac{1}{\sqrt{N}} \sum_{h=1}^H \sum_{i=1}^h \sum_{k,l=1}^p t_{hkl}(r_{i,k}r_{i-h,l} - r_{i,k}f_{i,l}^{(N)}(h)) \\
 &=: \frac{1}{\sqrt{N}} \sum_{i=1}^N Z_{r,i}(\mathbf{t}) - P_{13}(N).
 \end{aligned}$$

We now aim to show that $P_{13}(N) \xrightarrow{P} 0$, which completes the proof. It follows from Cauchy–Schwarz inequality, Parseval’s equality, and Assumption 4.4 that

$$\begin{aligned}
 &E \left| r_{i,k}r_{i-h,l} - r_{i,k}f_{i,l}^{(N)}(h) \right| \\
 &\leq (Er_{i,k}^2)^{1/2}(Er_{i-h,l}^2)^{1/2} + (Er_{i,k}^2)^{1/2} \left(Ef_{i,l}^{(N)}(h)^2 \right)^{1/2} \\
 &\leq E\|\varepsilon_0\|^2 + (E\|\varepsilon_0\|^2)^{1/2} \left(E\|\Gamma_\varepsilon \Psi_*^{h-1} \Gamma^{-1} \pi_{q_N}(X_{i-1})\|^2 \right)^{1/2} \\
 &= E\|\varepsilon_0\|^2 + (E\|\varepsilon_0\|^2)^{1/2} \left(E \sum_{l=1}^\infty \left\langle \Gamma^{-1} \pi_{q_N}(X_{i-1}), \Psi^{h-1} \Gamma_\varepsilon(v_l) \right\rangle^2 \right)^{1/2} \\
 &= E\|\varepsilon_0\|^2 + (E\|\varepsilon_0\|^2)^{1/2} \left(\sum_{q=1}^{q_N} \sum_{l=1}^\infty \frac{\langle v_q, \Psi^{h-1} \Gamma_\varepsilon(v_l) \rangle^2}{\lambda_q} \right)^{1/2} \\
 &= E\|\varepsilon_0\|^2 + (E\|\varepsilon_0\|^2)^{1/2} \left(\sum_{q=1}^{q_N} \frac{\|\Gamma_\varepsilon \Psi_*^{h-1}(v_q)\|^2}{\lambda_q} \right)^{1/2} < \infty.
 \end{aligned}$$

Then we have that $P_{13}(N) = O_P(N^{-1/2})$. $\square$

PROOF OF THEOREM 4.1: Using Slutsky’s theorem, the proof follows from Propositions 6.4, 6.5.

6.2.2 Proof of Theorem 4.2

Recall that

$$r_{i,k} = \langle \varepsilon_i, v_k \rangle; \quad \hat{r}_{i,k} = \langle \hat{\varepsilon}_i, \hat{v}_k \rangle; \quad \tilde{r}_{i,k} = \langle \hat{\varepsilon}_i, v_k \rangle$$

and

$$\begin{aligned}
 f_{i,l}^{(N)}(h) &= \left\langle \Gamma_\varepsilon \Psi_*^{h-1} \Gamma^{-1} \pi_{q_N}(X_{i-1}), v_l \right\rangle \text{ in (6.10);} \\
 \hat{f}_{i,l}^{(N)}(h) &= \frac{1}{N} \sum_{j=h+1}^N \langle X_{j-1}, \tilde{\Gamma}^{-1}(X_{i-1}) \rangle \hat{r}_{j-h,l} \text{ in (4.7),}
 \end{aligned}$$

where $\Gamma^{-1} \pi_{q_N}(\cdot) = \sum_{q=1}^{q_N} \lambda_q^{-1} \langle \cdot, v_q \rangle v_q$ defined in (6.10), and $\tilde{\Gamma}^{-1}(\cdot) = \sum_{q=1}^{q_N} \hat{\lambda}_q^{-1} \langle \cdot, \hat{v}_q \rangle \hat{v}_q$ defined in (2.6). Define

$$\tilde{f}_{i,l}^{(N)}(h) = \frac{1}{N} \sum_{j=h+1}^N \langle X_{j-1}, \tilde{\Gamma}^{-1}(X_{i-1}) \rangle \tilde{r}_{j-h,l}.$$

The following lemmas are preliminary results, which will be used to prove Proposition 6.3.

Lemma 6.8 Under Assumptions 4.1, 4.2 and 4.3, for any $i = 1, \dots, N$ and any $k = 1, \dots, p$, the following holds

$$\begin{aligned} |r_{i,k}| &\leq \|\varepsilon_i\|; & |\tilde{r}_{i,k}| &\leq o_p(1) \times \|X_{i-1}\| + \|\varepsilon_i\|; \\ |\tilde{r}_{i,k} - r_{i,k}| &\leq o_p(1) \times \|X_{i-1}\|; & |\hat{r}_{i,k} - \tilde{r}_{i,k}| &\leq o_p(1) \times \|X_{i-1}\| + o_p(1) \times \|\varepsilon_i\|. \end{aligned}$$

Proof Since $\hat{\varepsilon}_i = \varepsilon_i + \mathcal{D}_N(X_{i-1})$, we have that

$$\begin{aligned} |\tilde{r}_{i,k}| &= |\langle \hat{\varepsilon}_i, v_k \rangle| \leq \|\varepsilon_i + \mathcal{D}_N(X_{i-1})\| \leq \|\mathcal{D}_N\|_{\mathcal{S}} \|X_{i-1}\| + \|\varepsilon_i\|; \\ |\hat{r}_{i,k} - \tilde{r}_{i,k}| &= |\langle \hat{\varepsilon}_i, \hat{v}_k - v_k \rangle| \leq \|\mathcal{D}_N\|_{\mathcal{S}} \|\hat{v}_k - v_k\| \|X_{i-1}\| + \|\hat{v}_k - v_k\| \|\varepsilon_i\|. \end{aligned}$$

Similarly, we have that $|r_{i,k}| \leq \|\varepsilon_i\|$, and $|\tilde{r}_{i,k} - r_{i,k}| = |\langle \hat{\varepsilon}_i - \varepsilon_i, v_k \rangle| = |\langle \mathcal{D}_N(X_{i-1}), v_k \rangle| \leq \|\mathcal{D}_N\|_{\mathcal{S}} \|X_{i-1}\|$. Then, Assumption 4.2 and Lemma 6.1 complete the proof. $\square$

Lemma 6.9 Under Assumptions 4.1, 4.2, 4.3 and 4.4, the following holds for any $i = 1, \dots, N$, $l = 1, \dots, p$, and $h = 1, \dots, H$,

$$\begin{aligned} |\tilde{f}_{i,l}^{(N)}(h)| &\leq O_P(1) \times \|X_{i-1}\|; & |\hat{f}_{i,l}^{(N)}(h) - \tilde{f}_{i,l}^{(N)}(h)| &\leq o_P(1) \times \|X_{i-1}\|; \\ |f_{i,l}^{(N)}(h)| &\leq O(1) \times \|X_{i-1}\|; & |\tilde{f}_{i,l}^{(N)}(h) - f_{i,l}^{(N)}(h)| &\leq o_P(1) \times \|X_{i-1}\|; \\ |f_{i,l}(h)| &\leq O(1) \times \|X_{i-1}\|; & |f_{i,l}^{(N)}(h) - f_{i,l}(h)| &\leq o(1) \times \|X_{i-1}\|. \end{aligned}$$

Proof Before we proceed, we first show that

$$\left\| \frac{1}{N} \sum_{i=h+1}^N \langle X_{i-1}, \tilde{\Gamma}^{-1}(\cdot) \rangle X_{i-1-h} \right\|_{\mathcal{S}} = O_P(1). \quad (6.14)$$

It follows from Parseval's identity that for any set of orthonormal basis $\{e_l\}$

$$\begin{aligned} &\left\| \frac{1}{N} \sum_{i=h+1}^N \langle X_{i-1}, \tilde{\Gamma}^{-1}(\cdot) \rangle X_{i-1-h} \right\|_{\mathcal{S}}^2 \\ &= \frac{1}{N^2} \sum_{l=1}^{\infty} \left\| \sum_{i=1+h}^N \langle X_{i-1}, \tilde{\Gamma}^{-1}(e_l) \rangle X_{i-1-h} \right\|^2 \\ &= \frac{1}{N^2} \sum_{l=1}^{\infty} \left(\sum_{i=1+h}^N \langle \tilde{\Gamma}^{-1}(X_{i-1}), e_l \rangle^2 \|X_{i-1-h}\|^2 \right. \\ &\quad \left. + \sum_{i \neq j} \langle \tilde{\Gamma}^{-1}(X_{i-1}), e_l \rangle \langle \tilde{\Gamma}^{-1}(X_{j-1}), e_l \rangle \langle X_{i-1-h}, X_{j-1-h} \rangle \right) \\ &= \frac{1}{N^2} \left(\sum_{i=1+h}^N \|\tilde{\Gamma}^{-1}(X_{i-1})\|^2 \|X_{i-1-h}\|^2 \right. \\ &\quad \left. + \sum_{i \neq j} \langle \tilde{\Gamma}^{-1}(X_{i-1}), \tilde{\Gamma}^{-1}(X_{j-1}) \rangle \langle X_{i-1-h}, X_{j-1-h} \rangle \right) \\ &\leq \frac{1}{N^2} \sum_{i=1+h}^N \|\tilde{\Gamma}^{-1}\|_{\mathcal{S}}^2 \|X_{i-1}\|^2 \|X_{i-1-h}\|^2 \end{aligned}$$

$$+ \frac{1}{N^2} \sum_{i \neq j} \|\tilde{\Gamma}^{-1}\|_{\mathcal{S}}^2 \|X_{i-1-h}\| \|X_{j-1-h}\| \|X_{i-1}\| \|X_{j-1}\|.$$

Since we have that $\|\tilde{\Gamma}^{-1}\|_{\mathcal{S}} = \|\Gamma^{-1}\pi_{q_N}\|_{\mathcal{S}} + o_P(1)$ which follows from the proof of Lemma 6.8 of Zhang (2016) and $E\|X_0\|^4 < \infty$, we prove (6.14). We also note that as shown in the proof of Lemma 6.8 of Zhang (2016)

$$\left\| \frac{1}{N} \sum_{j=h+1}^N \langle X_{j-1}, \tilde{\Gamma}^{-1}(\cdot) \varepsilon_{j-h} \rangle \right\|_{\mathcal{S}} = \|\Gamma_{\varepsilon} \Psi_*^{h-1} \Gamma^{-1} \pi_{q_N}\|_{\mathcal{S}} + o_P(1). \quad (6.15)$$

Now it follows from Assumption 4.4 that $|f_{i,l}^{(N)}(h)| \leq \|\Gamma_{\varepsilon} \Psi_*^{h-1} \Gamma^{-1} \pi_{q_N}\|_{\mathcal{S}} \|X_{i-1}\| = O(1)\|X_{i-1}\|$ and $|f_{i,l}(h)| \leq \|\Gamma_{\varepsilon} \Psi_*^{h-1} \Gamma^{-1}\|_{\mathcal{S}} \|X_{i-1}\| = O(1)\|X_{i-1}\|$. Also, by Assumption 4.4, (6.14), and (6.15) we have that

$$\begin{aligned} |\tilde{f}_{i,l}^{(N)}(h)| &\leq \left\| \frac{1}{N} \sum_{j=h+1}^N \langle X_{j-1}, \tilde{\Gamma}^{-1}(\cdot) \varepsilon_{j-h} \rangle \right\|_{\mathcal{S}} \\ &\quad + \|\mathcal{D}_N\|_{\mathcal{S}} \left\| \frac{1}{N} \sum_{j=h+1}^N \langle X_{j-1}, \tilde{\Gamma}^{-1}(\cdot) X_{j-1-h} \rangle \right\|_{\mathcal{S}} \Big\} \|X_{i-1}\| \\ &= O_P(1)\|X_{i-1}\|. \end{aligned}$$

It follows from Assumption 4.4 that

$$|f_{i,l}^{(N)}(h) - f_{i,l}(h)| \leq \sum_{q=q_N}^{\infty} \frac{\|\Gamma_{\varepsilon} \Psi_*^{h-1}(v_q)\|}{\lambda_q} \|X_{i-1}\| = o(1)\|X_{i-1}\|.$$

Now define $\tilde{f}_{i,l}^{(N)}(h) = \frac{1}{N} \sum_{j=h+1}^N \langle X_{j-1}, \tilde{\Gamma}^{-1}(X_{i-1}) \rangle \langle \varepsilon_{j-h}, v_l \rangle$, and observe that by (6.15)

$$\begin{aligned} |\tilde{f}_{i,l}^{(N)}(h) - f_{i,l}^{(N)}(h)| &\leq |\tilde{f}_{i,l}^{(N)}(h) - \tilde{f}_{i,l}^{(N)}(h)| + |\tilde{f}_{i,l}^{(N)}(h) - f_{i,l}^{(N)}(h)| \\ &\leq \left(\|\mathcal{D}_N\|_{\mathcal{S}} \left\| \frac{1}{N} \sum_{j=h+1}^N \langle X_{j-1}, \tilde{\Gamma}^{-1}(\cdot) X_{j-1-h} \rangle \right\|_{\mathcal{S}} + o_P(1) \right) \|X_{i-1}\|, \end{aligned}$$

and

$$\begin{aligned} |\hat{f}_{i,l}^{(N)}(h) - \tilde{f}_{i,l}^{(N)}(h)| &\leq \left\| \frac{1}{N} \sum_{j=h+1}^N \langle X_{j-1}, \tilde{\Gamma}^{-1}(X_{i-1}) \rangle \hat{\varepsilon}_{j-h} \right\| \|\hat{v}_l - v_l\| \\ &\leq \left\{ \|\Gamma_{\varepsilon} \Psi_*^{h-1} \Gamma^{-1} \pi_{q_N}\|_{\mathcal{S}} + o_P(1) + \|\mathcal{D}_N\|_{\mathcal{S}} \left\| \frac{1}{N} \sum_{j=h+1}^N \langle X_{j-1}, \tilde{\Gamma}^{-1}(\cdot) X_{j-1-h} \rangle \right\|_{\mathcal{S}} \right\} \\ &\quad \|X_{i-1}\| \|\hat{v}_l - v_l\|. \end{aligned}$$

Then, Assumption 4.2, Lemma 6.1, and (6.14) complete the proof. $\square$

Proposition 6.6 Under the assumptions in Theorem 4.2,

$$\hat{\Sigma}_{r,N} - \Sigma_r = o_P(1).$$

Proof It is sufficient to show that for any $h, h' = 1, \dots, H$ and any $k, l, k', l' = 1, \dots, p$,

$$\begin{aligned} & \frac{1}{N} \sum_{i=1+\max\{h,h'\}}^N (\hat{r}_{i,k} \hat{r}_{i-h,l} - \hat{r}_{i,k} \hat{f}_{i,l}^{(N)}(h)) (\hat{r}_{i,k'} \hat{r}_{i-h',l'} - \hat{r}_{i,k'} \hat{f}_{i,l'}^{(N)}(h')) \\ & - E(r_{i,k} r_{i-h,l} - r_{i,k} f_{i,l}(h)) (r_{i,k'} r_{i-h',l'} - r_{i,k'} f_{i,l'}(h')) \xrightarrow{P} 0. \end{aligned} \quad (6.16)$$

To show (6.16), consider the following decomposition

$$\begin{aligned} & \frac{1}{N} \sum_{i=1+\max\{h,h'\}}^N \left\{ (\hat{r}_{i,k} \hat{r}_{i-h,l} - \hat{r}_{i,k} \hat{f}_{i,l}^{(N)}(h)) (\hat{r}_{i,k'} \hat{r}_{i-h',l'} - \hat{r}_{i,k'} \hat{f}_{i,l'}^{(N)}(h')) \right. \\ & \quad \left. - (\tilde{r}_{i,k} \tilde{r}_{i-h,l} - \tilde{r}_{i,k} \tilde{f}_{i,l}^{(N)}(h)) (\tilde{r}_{i,k'} \tilde{r}_{i-h',l'} - \tilde{r}_{i,k'} \tilde{f}_{i,l'}^{(N)}(h')) \right\} \\ & + \frac{1}{N} \sum_{i=1+\max\{h,h'\}}^N \left\{ (\tilde{r}_{i,k} \tilde{r}_{i-h,l} - \tilde{r}_{i,k} \tilde{f}_{i,l}^{(N)}(h)) (\tilde{r}_{i,k'} \tilde{r}_{i-h',l'} - \tilde{r}_{i,k'} \tilde{f}_{i,l'}^{(N)}(h')) \right. \\ & \quad \left. - (r_{i,k} r_{i-h,l} - r_{i,k} f_{i,l}(h)) (r_{i,k'} r_{i-h',l'} - r_{i,k'} f_{i,l'}(h')) \right\} \\ & + \frac{1}{N} \sum_{i=1+\max\{h,h'\}}^N \left\{ (r_{i,k} r_{i-h,l} - r_{i,k} f_{i,l}(h)) (r_{i,k'} r_{i-h',l'} - r_{i,k'} f_{i,l'}(h')) \right. \\ & \quad \left. - (r_{i,k} r_{i-h,l} - r_{i,k} f_{i,l}(h)) (r_{i,k'} r_{i-h',l'} - r_{i,k'} f_{i,l'}(h')) \right\} \\ & + \frac{1}{N} \sum_{i=1+\max\{h,h'\}}^N \left\{ (r_{i,k} r_{i-h,l} - r_{i,k} f_{i,l}(h)) (r_{i,k'} r_{i-h',l'} - r_{i,k'} f_{i,l'}(h')) \right. \\ & \quad \left. - E(r_{i,k} r_{i-h,l} - r_{i,k} f_{i,l}(h)) (r_{i,k'} r_{i-h',l'} - r_{i,k'} f_{i,l'}(h')) \right\} \\ & =: P_{14}(N) + P_{15}(N) + P_{16}(N) + P_{17}(N) \end{aligned}$$

For $P_{14}(N)$, observe that

$$\begin{aligned} P_{14}(N) &= \frac{1}{N} \sum_{i=1+\max\{h,h'\}}^N (\hat{r}_{i,k} \hat{r}_{i-h,l} \hat{r}_{i,k'} \hat{r}_{i-h',l'} - \tilde{r}_{i,k} \tilde{r}_{i-h,l} \tilde{r}_{i,k'} \tilde{r}_{i-h',l'}) \\ & - \frac{1}{N} \sum_{i=1+\max\{h,h'\}}^N (\hat{r}_{i,k} \hat{r}_{i-h,l} \hat{r}_{i,k'} \hat{f}_{i,l'}^{(N)}(h') - \tilde{r}_{i,k} \tilde{r}_{i-h,l} \tilde{r}_{i,k'} \tilde{f}_{i,l'}^{(N)}(h')) \\ & - \frac{1}{N} \sum_{i=1+\max\{h,h'\}}^N (\hat{r}_{i,k} \hat{f}_{i,l}^{(N)}(h) \hat{r}_{i,k'} \hat{r}_{i-h',l'} - \tilde{r}_{i,k} \tilde{f}_{i,l}^{(N)}(h) \tilde{r}_{i,k'} \tilde{r}_{i-h',l'}) \\ & + \frac{1}{N} \sum_{i=1+\max\{h,h'\}}^N (\hat{r}_{i,k} \hat{f}_{i,l}^{(N)}(h) \hat{r}_{i,k'} \hat{f}_{i,l'}^{(N)}(h') - \tilde{r}_{i,k} \tilde{f}_{i,l}^{(N)}(h) \tilde{r}_{i,k'} \tilde{f}_{i,l'}^{(N)}(h')). \end{aligned}$$

Then each of the four partial sums in $P_{14}(N)$ converges to 0 in probability by Lemmas 6.5, 6.8, and 6.9. Analogous arguments can be used to prove that $P_{15}(N) \xrightarrow{P} 0$ and $P_{16}(N) \xrightarrow{P} 0$. For $P_{17}(N)$, observe that we have a Bernoulli shift representation such that $(r_{i,k} r_{i-h,l} - r_{i,k} f_{i,l}(h)) (r_{i,k'} r_{i-h',l'} - r_{i,k'} f_{i,l'}(h')) = g(\varepsilon_i, \varepsilon_{i-1}, \dots)$, which is stationary and ergodic. Then, it follows from the mean ergodic theorem that $P_{17}(N) = o_P(1)$. $\square$

PROOF OF THEOREM 4.2: Using Theorem 4.1 and Proposition 6.6, the proof is analogous to the Proof of Theorem 3.2.

Acknowledgements Piotr Kokoszka was partially supported by the United States National Science Foundation grant DMS–2123761. Gregory Rice was partially supported by the Natural Science and Engineering Research Council of Canada, NSERC-RGPIN 50503-10477.

Declarations

Conflict of interest The authors declare that they have no financial or non-financial interests that are directly or indirectly related to the work submitted for publication.

References

- Aue A, Horváth L, Pellat D (2017) Functional generalized autoregressive conditional heteroskedasticity. *J Time Series Anal* 38:3–21
- Bagchi P, Characiejus V, Dette H (2018) A simple test for white noise in functional time series. *J Time Series Anal* 39:54–74
- Bien J, Bunea F, Xiao L (2016) Convex banding of the covariance matrix. *J Am Stat Assoc* 111:834–845
- Bosq D (2000) Linear processes in function spaces. Springer, New York
- Brockwell PJ, Davis RA (1991) Time series: theory and methods. Springer, New York
- Brockwell PJ, Davis RA (2016) Introduction to time series and forecasting, 3rd edn. Springer, New York
- Bücher A, Dette H, Heinrichs F (2023) A portmanteau-type test for detecting serial correlation in locally stationary functional time series. *Stat Inference Stoch Process* 26:255–278
- Cerovecki C, Francq C, Hörmann S, Zakoian J (2019) Functional GARCH models: the quasi-likelihood approach and its applications. *J Econom* 209:353–375
- Characiejus V, Rice G (2020) A general white noise test based on kernel lag-window estimates of the spectral density operator. *Econom Stat* 13:175–196
- Francq C, Zakoian JM (2010) GARCH models. Wiley, New York
- Gabrys R, Kokoszka P (2007) Portmanteau test of independence for functional observations. *J Am Stat Assoc* 102:1338–1348
- González-Manteiga W, Crujeiras R, García-Portugués E (2023) A review of goodness-of-fit tests for models involving functional data. *Trends in mathematical, information and data sciences: a tribute to Leandro Pardo*. Springer, New York, pp 349–358
- Hlávka Z, Hušková M, Meintanis S (2021) Testing serial independence with functional data. *Test* 30:603–629
- Hörmann S, Kokoszka P (2010) Weakly dependent functional data. *Ann Stat* 38:1845–1884
- Hörmann S, Kokoszka P (2012) Functional time series. In: Rao CR, Rao TS (eds) Time series, handbook of statistics, vol 30. Amsterdam, The Netherlands
- Hörmann S, Horváth L, Reeder R (2013) A functional version of the ARCH model. *Econom Theory* 29:267–288
- Horn RA, Johnson CR (2012) Matrix Analysis. Cambridge University Press
- Horváth L, Kokoszka P (2012) Inference for functional data with applications. Springer, New York
- Horváth L, Kokoszka P, Teyssiere G (2001) Empirical process of squared residuals of an ARCH sequence. *Ann Stat* 29:445–469
- Hsing T, Eubank R (2015) Theoretical foundations of functional data analysis, with an introduction to linear operators. Wiley, New York
- Jang P, Jauch M, Matteson D (2022) Functional stochastic volatility in financial option surfaces. *Data Sci Sci* 1:6–19
- Kim M, Kokoszka P, Rice G (2023) White noise testing for functional time series. *Stat Surv* 17:119–168
- Kokoszka P, Politis D (2011) Nonlinearity of ARCH and stochastic volatility models and Bartlett’s formula. *Probab Math Stat* 31:47–59
- Kokoszka P, Reimherr M (2013) Determining the order of the functional autoregressive model. *J Time Series Anal* 34:116–121
- Kokoszka P, Rice G, Shang H (2017) Inference for the autocovariance of a functional time series under conditional heteroscedasticity. *J Multivar Anal* 162:32–50
- Kokoszka P, Mohammadi N, Wang H, Wang S (2024) Functional diffusion driven stochastic volatility model. *Bernoulli* Forthcoming
- Koul HL (2002) Weighted empirical processes in dynamic nonlinear models. Springer, New York

- Kühnert S (2020) Functional ARCH and GARCH models: a Yule-Walker approach. *Electron J Stat* 14:4321–4360
- Li WK (2004) Diagnostic checks in time series. Chapman and Hall, London
- R Development Core Team (2008) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>, ISBN 3-900051-07-0
- Wu W (2005) Nonlinear system theory: another look at dependence. *Proc Natl Acad Sci U S A* 102:14150–14154
- Zhang X (2016) White noise testing and model diagnostic checking for functional time series. *J Econom* 194:76–95

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.