On portmanteau-type tests for nonlinear multivariate time series

de Gooijer, J.G.

DOI
10.1016/j.jmva.2023.105157

Publication date
2023

Document Version
Final published version

Published in
Journal of Multivariate Analysis

License
CC BY

Journal of Multivariate Analysis 195 (2023) 105157

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

On portmanteau-type tests for nonlinear multivariate time series

Jan G. De Gooijer

Amsterdam School of Economics, University of Amsterdam, PO Box 15867, 1001 NJ Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 15 October 2022
Received in revised form 8 January 2023
Accepted 8 January 2023
Available online 11 January 2023

AMS 2010 subject classifications:
primary 62M10
secondary 62H15

Keywords:
Mixed multivariate tests
Multiple-lag tests
Multivariate nonlinear time series
Residual autocovariances
Single-lag tests

ABSTRACT

A general framework to devise portmanteau-type test statistics for a general class of multivariate nonlinear time series models with vector martingale difference errors is formulated. Based on this framework a suite of individual and mixed multivariate test statistics is considered. Two applications are developed: single- and multiple-lag test statistics. In each case, the resulting portmanteau test statistic is based on multivariate residuals and multivariate squared residuals. Moreover, single- and multiple-lag mixed multivariate portmanteau-type tests are introduced. These test statistics are designed to detect different forms of inadequacies in the model residuals jointly. All proposed tests take uncertainty due to model estimation properly into account. The asymptotic null distribution of each test statistic follows from the asymptotic distribution of the general portmanteau-type test statistic in a natural way. Some considerations are given to the empirical size and power of six test statistics via a simulation study. All tests have satisfactory size and power properties in finite samples. To demonstrate their practical use, the proposed test statistics are applied to the residuals of a vector bivariate nonlinear threshold model fitted to U.S. interest rates.

1. Introduction

One of the first questions that one can ask in time series model evaluation: "Are the residuals white noise?" A large number of portmanteau-type tests based on the autocorrelation or the autocovariance function are proposed for this purpose; see the monographs by Li [12] and by Akashi et al. [1] and the references therein. These tests often assume that the data generating process (DGP) under study is linear and univariate. An example is the well-known Ljung–Box portmanteau test statistic. For multivariate (vector) linear autoregressive (VAR) models with independent and Gaussian errors various improvements of the Hosking test and the Li-McLeod test have been considered; see, e.g., Mahdi and McLeod [17]. In addition, several works consider testing residuals in VAR and VARMA models with uncorrelated but nonindependent errors; see, e.g., Francq and Raïssi [6] and Maïnassara [18]. Testing for high-dimensional white noise is another important problem which has recently been addressed; see, e.g., De Gooijer and Yuan [3] and Li et al. [14].

With the increasing interest in modeling and forecasting multivariate nonlinear time series there is also an urgent need to develop portmanteau-type test statistics for a wide range of error specifications. Indeed, Ling and Li [16] proposed such a test for multivariate DGPs with conditional nonlinearity in the mean and with multivariate autoregressive conditional heteroskedastic (ARCH) errors (Section 4.3.2). Other multivariate tests, albeit designed for particular nonlinear DGPs, are proposed by Wong and Li [22], Duchesne and Lalancette [5], and Chabot-Hallé and Duchesne [2] among others.

E-mail address: j.g.degooijer@uva.nl.

https://doi.org/10.1016/j.jmva.2023.105157

In this paper, we offer a general framework for constructing portmanteau-type test statistics in a multivariate nonlinear time series context. Based on this framework, we consider both individual and mixed multivariate test statistics. For the proposed individual multivariate portmanteau tests, we distinguish between single- and multiple-lag tests using autocovariances of residuals and autocovariances of squared residuals. However, tests based on other types of residual autocovariances (e.g., bicovariances or autocovariances of absolute residuals) may also be derived from the general testing framework. The proposed mixed multivariate portmanteau tests, combine the individual single- and multiple-lag tests. These latter tests are designed to detect different forms of inadequacies in the multivariate residuals jointly. All proposed tests take uncertainty due to model parameter estimation properly into account, and under some regularity conditions the tests are asymptotically chi-square distributed.

The rest of the paper is organized as follows. Section 2 outlines the multivariate nonlinear model. Section 3 presents the general testing framework, the corresponding multivariate test statistic and its asymptotic null distribution. Section 4 considers various single- and multiple-lag multivariate test statistics, using the general testing framework of Section 3. In Section 5, we propose two mixed multivariate portmanteau-type test statistics. Section 6 reports results of a simulation study. Section 7 presents an illustrative example, using residuals of a vector bivariate nonlinear threshold model fitted to U.S. interest rates. Some concluding remarks are given in Section 8. The Appendix contains a proof of a theoretical result.

Unless otherwise stated all limit results assume that the sample size n goes to ∞. The symbols $\xrightarrow{D}$ and $\xrightarrow{P}$ signify convergence in distribution and convergence in probability, respectively. The symbol $o_p(\mathbf{1})$ denotes a multivariate sequence of random variables converging to zero in probability. The superscript $\top$ denotes matrix or vector transposition. We use boldface to denote (possibly random) vector or matrix variables.

2. Preliminaries

Let $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$ be a stationary and ergodic multivariate stochastic process, where $\mathbf{Y}_t = (Y_{1,t}, \ldots, Y_{m,t})^\top$, $m \geq 1$. Assume that the process is generated by the following multivariate nonlinear time series model, defined by

$$\mathbf{Y}_t = \mathbf{f}(\boldsymbol{\theta}_0 | \mathcal{F}_{t-1}) + \boldsymbol{\varepsilon}_{t,\boldsymbol{\theta}_0}, \tag{1}$$

where $\mathcal{F}_{t-1}$ represents the information set generated by $\{\mathbf{Y}_s, s < t\}$, $\mathbf{f} = (f_1, \ldots, f_m)^\top$ is a known real-valued measurable function with values in $\mathbb{R}^m$, and $\boldsymbol{\theta}_0$ denotes the true, but unknown, value of the parameter vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)^\top$. It is assumed that each $f_i \equiv f_i(\boldsymbol{\theta}|\mathcal{F}_{t-1})$, $i \in \{1, \ldots, m\}$, denote a function of the previous $\mathbf{Y}_t$'s and of $\boldsymbol{\theta}$. In addition, we denote by $f_{t-1}(\mathbf{Y}_t; \boldsymbol{\theta})$ the conditional density function of $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$. The following assumption is about the error process.

Assumption 1. The process $\{\boldsymbol{\varepsilon}_{t,\boldsymbol{\theta}_0} = (\varepsilon_{1,t,\boldsymbol{\theta}_0}, \ldots, \varepsilon_{m,t,\boldsymbol{\theta}_0})^\top, t \in \mathbb{Z}\}$ is supposed to be an m-dimensional vector martingale difference sequence (MDS) satisfying $\mathbb{E}(\boldsymbol{\varepsilon}_{t,\boldsymbol{\theta}_0}|\mathcal{F}_{t-1}) = \mathbf{0}$, $\mathrm{Cov}(\boldsymbol{\varepsilon}_{t,\boldsymbol{\theta}_0}, \boldsymbol{\varepsilon}_{t,\boldsymbol{\theta}_0}|\mathcal{F}_{t-1}) = \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$, $\mathrm{Cov}(\boldsymbol{\varepsilon}_{t,\boldsymbol{\theta}_0}, \boldsymbol{\varepsilon}_{s,\boldsymbol{\theta}_0}|\mathcal{F}_{t-1}) = \mathbf{0}$ $(t \neq s)$, and $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} = (\sigma_{\varepsilon,ij})_{i,j=1}^m$ is a positive definite matrix. The null hypothesis of model adequacy is

$$\mathbb{H}_0 : \mathrm{Cov}(\boldsymbol{\varepsilon}_{t,\boldsymbol{\theta}_0}, \boldsymbol{\varepsilon}_{t-\ell,\boldsymbol{\theta}_0}) = \mathbf{0}, \quad \ell \in \{1, 2, \ldots\}.$$

Remark 1. Unless otherwise stated, Assumption 1 is supposed to hold throughout the paper. The assumption implies that $\boldsymbol{\varepsilon}_{t,\boldsymbol{\theta}_0}$ forms an uncorrelated but not necessarily independent sequence of random vectors with mean zero. If the mean is non-zero, the asymptotic distribution of the test statistics discussed below will remain unchanged. The MDS structure of the error vector $\boldsymbol{\varepsilon}_{t,\boldsymbol{\theta}_0}$ can be directly checked by the MDS-type tests proposed by Wang et al. [21].

Let $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ be a sample time series of length n obtained from the stationary process $\{\mathbf{Y}_t\}$. Conditional on the set of initial values $\{\mathbf{Y}_0, \mathbf{Y}_{-1}, \ldots\}$, the log-likelihood function of the sample takes the form $L_n(\boldsymbol{\theta}_0) = \sum_{t=1}^n \ell_t(\mathbf{Y}_t; \boldsymbol{\theta}_0) = \sum_{t=1}^n \log f_{t-1}(\mathbf{Y}_t; \boldsymbol{\theta}_0)$. Assume that there exists a local maximizer $\widehat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}_0$ such that $\widehat{\boldsymbol{\theta}}_n \xrightarrow{P} \boldsymbol{\theta}_0$. The following Assumption 2 is sufficient for the consistency and asymptotic normality of $\widehat{\boldsymbol{\theta}}_n$.

Assumption 2. Let the following assumptions hold: (a) $\boldsymbol{\Theta} \subset \mathbb{R}^K$ is an open set; (b) Model (1) is correctly specified; (c) The conditional density function $f_{t-1} : \boldsymbol{\Theta} \times \mathbb{R}^m \to \mathbb{R}$ is supposed to have continuous second-order derivatives with respect to $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ almost surely (a.s.); (d) Denote $N_{n,c} = \{\boldsymbol{\theta} \in \boldsymbol{\Theta} : \|\sqrt{n}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\| \leq c\}$ with $\|\cdot\|$ the Euclidean norm, and

$$\mathbf{B}_n(\boldsymbol{\theta}) = -\frac{\partial^2 L_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = -\Big[\sum_{t=1}^n \frac{\partial^2 \ell_t(\mathbf{Y}_t; \boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}\Big]_{i,j=1}^K,$$

there exists a nonrandom positive definite $K \times K$ matrix $\mathcal{I}(\boldsymbol{\theta}_0)$, such that for all $c > 0$,

$$\sup_{\boldsymbol{\theta} \in N_{n,c}} \Big\| \frac{1}{n} \mathbf{B}_n(\boldsymbol{\theta}) - \mathcal{I}(\boldsymbol{\theta}_0) \Big\| \xrightarrow{P} 0;$$

(e) The score (gradient) vector function $\boldsymbol{S}_n(\boldsymbol{\theta}) = \partial L_n(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$ is asymptotically normal, i.e.,

$$\frac{1}{\sqrt{n}} \boldsymbol{S}_n(\boldsymbol{\theta}_0) \xrightarrow{D} \mathcal{I}(\boldsymbol{\theta}_0)^{1/2} \mathbf{Z}, \quad \mathbf{Z} \sim \mathcal{N}_K(\mathbf{0}, \mathbf{I}_K),$$

where $\mathbf{I}_K$ is a $K \times K$ identity matrix.

Assumption 2(a) is standard. It guarantees that the maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_n$ (or an asymptotically equivalent estimator) of $\boldsymbol{\theta}_0$ is an interior point. Assumption 2(c) combined with Assumption 2(a) implies the applicability of the mean-value theorem for the score vector function in any convex set contained in $\boldsymbol{\Theta}$. Assumption 2(d) gives a uniform convergence in probability of the Hessian matrix $\mathbf{B}_n(\boldsymbol{\theta})$ on special compact sets that contain $\boldsymbol{\theta}_0$. Assumption 2(e) is needed to obtain asymptotic normality of $\widehat{\boldsymbol{\theta}}_n$. The correct model Assumption 2(b) is necessary for Proposition 1 and for testing purposes.

Proposition 1 ([19]). *Under Assumption 2, there exists a sequence of local maximizers $\widehat{\boldsymbol{\theta}}_n$ such that $\{\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\}_{n\in\mathbb{N}}$ is bounded in probability, and*

$$\left[\frac{1}{n}\mathbf{B}_n(\boldsymbol{\theta}_0)\right]^{-1}\frac{1}{\sqrt{n}}\mathbf{S}_n(\boldsymbol{\theta}_0) - \sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{P} 0. \tag{2}$$

Furthermore, under Assumption 2 and using Proposition 1 it can be shown that $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{D} \mathcal{N}_K(\mathbf{0}, \mathcal{I}(\boldsymbol{\theta}_0)^{-1})$.

3. General portmanteau-type test statistic

Given $\widehat{\boldsymbol{\theta}}_n$, we denote the $d \times 1$ vector of sample residuals by $\widehat{\boldsymbol{\varepsilon}}_{t,\widehat{\boldsymbol{\theta}}_n} = (\widehat{\varepsilon}_{1,t,\widehat{\boldsymbol{\theta}}_n}, \ldots, \widehat{\varepsilon}_{d,t,\widehat{\boldsymbol{\theta}}_n})^\top = \mathbf{Y}_t - \widehat{\mathbf{f}}_{t-1}$, where $\widehat{\mathbf{f}}_{t-1} \equiv \mathbf{f}(\widehat{\boldsymbol{\theta}}_n|\mathcal{F}_{t-1})$, $(d \in \{1, \ldots, N+1\}; N \ll n, N \in \mathbb{Z}^+)$. We define a vector with unobserved errors and a corresponding vector with residuals by, respectively,

$$\mathbf{U}_{t,\boldsymbol{\theta}_0,L} = [\boldsymbol{\varepsilon}_{t,\boldsymbol{\theta}_0}^\top, \ldots, \boldsymbol{\varepsilon}_{t-L,\boldsymbol{\theta}_0}^\top]^\top \in \mathbb{R}^{m(L+1)}, \quad \mathbf{u}_{t,\widehat{\boldsymbol{\theta}}_n,L} = [\widehat{\boldsymbol{\varepsilon}}_{t,\widehat{\boldsymbol{\theta}}_n}^\top, \ldots, \widehat{\boldsymbol{\varepsilon}}_{t-L,\widehat{\boldsymbol{\theta}}_n}^\top]^\top \in \mathbb{R}^{m(L+1)}.$$

To develop a general framework for testing the adequacy of model (1) based on a transformed vector of residuals, we define a transformation function $g(\cdot)$ in Assumption 3. With different choices of $g(\cdot)$ different types of multivariate portmanteau-type tests can be proposed.

Assumption 3. Let the following assumption hold.

$g : \mathbb{R}^{m(N-1)} \to \mathbb{R}^d$ is a continuously differentiable function such that $\mathbb{E}[g(\mathbf{U}_{t,\boldsymbol{\theta}_0})] = \mathbf{0}$,

where $\mathbf{U}_{t,\boldsymbol{\theta}_0} \equiv \mathbf{U}_{t,\boldsymbol{\theta}_0,N-1}$ is introduced as shorthand notation.

Assumptions 2 and 3 and the following Assumption 4 together are used to establish asymptotic distributions of the proposed test statistics.

Assumption 4. Let the following assumptions hold: (a) For all $c > 0$

$$\sup_{\boldsymbol{\theta}\in N_{n,c}}\left\|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g(\mathbf{U}_{t,\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}^\top} - \mathbf{G}\right\| \xrightarrow{P} 0, \quad \sup_{\boldsymbol{\theta}\in N_{n,c}}\left\|\frac{1}{n}\sum_{t=1}^{n}g(\mathbf{U}_{t,\boldsymbol{\theta}})g(\mathbf{U}_{t,\boldsymbol{\theta}})^\top - \boldsymbol{\Delta}\right\| \xrightarrow{P} 0,$$

and

$$\sup_{\boldsymbol{\theta}\in N_{n,c}}\left\|\frac{1}{n}\sum_{t=1}^{n}g(\mathbf{U}_{t,\boldsymbol{\theta}})\left(\frac{\partial\ell_t(\mathbf{Y}_t;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)^\top - \boldsymbol{\Psi}\right\| \xrightarrow{P} 0,$$

where $\mathbf{G} = \mathbb{E}[\partial g(\mathbf{U}_{t,\boldsymbol{\theta}_0})/\partial\boldsymbol{\theta}_0^\top]$ and $\boldsymbol{\Psi} = \mathbb{E}[g(\mathbf{U}_{t,\boldsymbol{\theta}_0})\left(\partial\ell_t(\mathbf{Y}_t;\boldsymbol{\theta})/\partial\boldsymbol{\theta}\right)^\top]$ are $d \times K$ matrices, and $\boldsymbol{\Delta} = \mathbb{E}[g(\mathbf{U}_{t,\boldsymbol{\theta}_0})g(\mathbf{U}_{t,\boldsymbol{\theta}_0})^\top]$. These matrices exist and are finite. Moreover, the $d \times d$ matrix $\boldsymbol{\Delta}$ is positive definite; (b)

$$\frac{1}{\sqrt{n}}\left[\mathbf{S}_n(\boldsymbol{\theta}_0)^\top, \left(\sum_{t=1}^{n}g(\mathbf{U}_{t,\boldsymbol{\theta}_0})\right)^\top\right]^\top \xrightarrow{D} \boldsymbol{\Sigma}^{1/2}\boldsymbol{\mathcal{Z}}, \quad \boldsymbol{\mathcal{Z}} \sim \mathcal{N}_{d+K}(\mathbf{0}, \mathbf{I}_{d+K}),$$

where

$$\boldsymbol{\Sigma} = \begin{bmatrix}\mathcal{I}(\boldsymbol{\theta}_0) & \boldsymbol{\Psi}^\top\\ \boldsymbol{\Psi} & \boldsymbol{\Delta}\end{bmatrix}$$

is a $(d+K)\times(d+K)$ positive definite matrix.

Assumption 4(a) imposes uniform convergence on special compact sets similar to that in Assumption 2(d). Both assumptions define the matrix $\boldsymbol{\Sigma}$ in Assumption 4(b). The joint weak convergence in Assumption 3(b) can be verified using an appropriate central limit theorem see, e.g., Sweeting [19]. Note, Assumption 2(e) is a special case of Assumption 3(b).

Now we can state a central limit theorem from which the limiting distributions of the proposed general portmanteau test statistic can be obtained.

Theorem 1. *Under Assumptions 2–4,*

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} g(\mathbf{U}_{t,\widehat{\theta}_n}) \xrightarrow{D} \mathcal{N}_d(\mathbf{0}, \boldsymbol{\Omega}),$$

where $\boldsymbol{\Omega}$ is a symmetric and positive definite (commonly abbreviated as SPD*) matrix defined by*

$$\boldsymbol{\Omega} = \left[\mathbf{G}\mathcal{I}(\boldsymbol{\theta}_0)^{-1} \;\; : \mathbf{I}_d\right]\begin{bmatrix}\mathcal{I}(\boldsymbol{\theta}_0) & \boldsymbol{\Psi}^\top \\ \boldsymbol{\Psi} & \boldsymbol{\Delta}\end{bmatrix}\begin{bmatrix}\mathcal{I}(\boldsymbol{\theta}_0)^{-1}\mathbf{G}^\top \\ \mathbf{I}_d\end{bmatrix} = \mathbf{G}\mathcal{I}(\boldsymbol{\theta}_0)^{-1}\mathbf{G}^\top + \boldsymbol{\Psi}\mathcal{I}(\boldsymbol{\theta}_0)^{-1}\mathbf{G}^\top + \mathbf{G}\mathcal{I}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{\Psi}^\top + \boldsymbol{\Delta}. \tag{3}$$

Proof. The proof of Theorem 1 follows as a special case from the proof of the asymptotic distribution of a function of multivariate quantile residuals (Kalliovirta [11]) and, hence, has been omitted. However, the Appendix of this paper contains a similar proof for a special case of the transformation function $g(\cdot)$ introduced in Section 5. □

Let $\widehat{\boldsymbol{\Omega}}_n$ denote an estimator of $\boldsymbol{\Omega}$. Under the additional assumption that $\widehat{\boldsymbol{\Omega}}_n$ is invertible, a consistent estimator of $\boldsymbol{\Omega}$ is given by

$$\widehat{\boldsymbol{\Omega}}_n = \widehat{\mathbf{G}}_n\widehat{\mathcal{I}}_n^{-1}\widehat{\mathbf{G}}_n^\top + \widehat{\boldsymbol{\Psi}}_n\widehat{\mathcal{I}}_n^{-1}\widehat{\mathbf{G}}_n^\top + \widehat{\mathbf{G}}_n\widehat{\mathcal{I}}_n^{-1}\widehat{\boldsymbol{\Psi}}_n^\top + \widehat{\boldsymbol{\Delta}}_n, \tag{4}$$

where

$$\widehat{\mathbf{G}}_n = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial g(\mathbf{u}_{t,\widehat{\theta}_n})}{\partial\boldsymbol{\theta}^\mathrm{T}}, \;\; \widehat{\boldsymbol{\Psi}}_n = \frac{1}{n}\sum_{t=1}^{n} g(\mathbf{u}_{t,\widehat{\theta}_n})\Big(\frac{\partial \ell_t(\mathbf{Y}_t;\widehat{\boldsymbol{\theta}}_n)}{\partial\boldsymbol{\theta}}\Big)^\top, \;\; \widehat{\boldsymbol{\Delta}}_n = \frac{1}{n}\sum_{t=1}^{n} g(\mathbf{u}_{t,\widehat{\theta}_n})g(\mathbf{u}_{t,\widehat{\theta}_n})^\top,$$

and $\widehat{\mathcal{I}}_n$ is a consistent estimator of $\mathcal{I}(\boldsymbol{\theta}_0)$. Consequently, under the null hypothesis of model adequacy, we have

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} g(\mathbf{U}_{t,\widehat{\theta}_n})^\mathrm{T}\widehat{\boldsymbol{\Omega}}_n^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n} g(\mathbf{U}_{t,\widehat{\theta}_n}) \xrightarrow{D} \chi^2(d).$$

This yields the following *general* portmanteau-type test statistic

$$\mathcal{Q} = \frac{1}{n-m+1}\sum_{t=m}^{n} g(\mathbf{u}_{t,\widehat{\theta}_n})^\top\,\widehat{\boldsymbol{\Omega}}_n^{-1}\sum_{t=m}^{n} g(\mathbf{u}_{t,\widehat{\theta}_n}) \overset{\mathbb{H}_0}{\approx} \chi^2(d), \tag{5}$$

where m and d are the dimensions defined in Assumption 3.

Remark 2. If $\{\boldsymbol{\varepsilon}_{t,\theta_0}\}$ is a strong white noise process, i.e., a sequence of independent and identically distributed (i.i.d.) random vectors with mean zero and variance $\boldsymbol{\Sigma}_{\varepsilon}$, then $\boldsymbol{\Delta}_{ij} = \boldsymbol{\Sigma}_{\varepsilon}\otimes\boldsymbol{\Sigma}_{\varepsilon}$, $i,j\in\{1,\ldots,d\}$, where $\otimes$ denotes the Kronecker product. In this case (4) can be simplified, since $\boldsymbol{\Delta} \equiv (\boldsymbol{\Delta}_{ij})_{i,j=1}^{d} = \mathbf{I}_d\otimes\boldsymbol{\Sigma}_{\varepsilon}\otimes\boldsymbol{\Sigma}_{\varepsilon}$.

Remark 3. Note, $\boldsymbol{\Omega}$ is SPD which ensures that $\widehat{\boldsymbol{\Omega}}_n$ is invertible for large sample sizes. However, there are cases where $\boldsymbol{\Omega}$ is not invertible. For this reason several authors have proposed portmanteau test statistics using a weighted sum of independent chi-squared random variables; see, e.g., Francq and Raïssi [6] and Maïnassara [18]. Throughout Sections 4 and 5, it is implicitly assumed that all variants of $\widehat{\boldsymbol{\Omega}}_n$ are invertible.

4. Individual multivariate portmanteau tests

In this section we devise individual multiple-lag and single-lag portmanteau-type tests. The outcomes of single-lag tests may give useful hints of the reasons of a potential model mis-specification. They can be used to complement the information provided by, for instance, the matrix sample auto- and cross-correlation function. Moreover, considering single-lag tests has the advantage to reduce the dimension of the process used to build the tests. Multiple-lag portmanteau tests may not have good power when the value of m is large. Also, it is well known that some widely used multivariate multiple-lag tests are sensitive to mis-specifications in the conditional variance.

In Sections 4.1–4.4, we devise portmanteau-type test statistics based on the lag-ℓ sample autocovariances of residuals $\{\widehat{\varepsilon}_{i,t,\widehat{\theta}_n}\}_{i=1}^{m}$ and the lag-ℓ sample autocovariance of squared residuals $\{\widehat{\varepsilon}^2_{i,t,\widehat{\theta}_n}\}_{i=1}^{m}$. To distinguish between both cases, we introduce the superscripts (1) and (2) in the notations when appropriate.

4.1. Multiple-lag test based on autocovariances of residuals

The null hypothesis for testing multiple-lag residual autocovariances is given by

$$\mathbb{H}_0^{(1)}\colon\; \mathrm{Cov}(\varepsilon_{i,t,\theta_0}, \varepsilon_{j,t-\ell,\theta_0}) = 0, \quad i,j\in\{1,\ldots,m\},\ \text{all } t,\ \text{and } \ell > 0. \tag{6}$$

The proposed test is based on the lag-ℓ autocovariances

$$c^{(1)}_{ij,\widehat{\theta}_n}(\ell)=\frac{1}{n-\ell}\sum_{t=\ell+1}^{n}\widehat{\varepsilon}_{i,t,\widehat{\theta}_n}\widehat{\varepsilon}_{j,t-\ell,\widehat{\theta}_n},\quad i,j\in\{1,\ldots,m\},\ \ell\in\{1,\ldots,K_1\},\ K_1\ll n. \tag{7}$$

For stationary time series, $c^{(1)}_{ij,\widehat{\theta}_n}(\ell)$ is a natural estimator of $\mathrm{Cov}(\varepsilon_{i,t,\theta_0},\varepsilon_{j,t-\ell,\theta_0})$. Given $c^{(1)}_{ij,\widehat{\theta}_n}(\ell)$, we define the corresponding residual autocovariance matrix as $\mathbf{C}^{(1)}_{\widehat{\theta}_n}(\ell)=\big(c^{(1)}_{ij,\widehat{\theta}_n}(\ell)\big)_{i,j=1}^{m}$. Furthermore, we denote the $m^2K_1\times 1$ vector of residual autocovariances by $\mathbf{c}^{(1)}_{\widehat{\theta}_n}=\big(\mathbf{c}^{(1)\top}_{\widehat{\theta}_n}(1),\ldots,\mathbf{c}^{(1)\top}_{\widehat{\theta}_n}(K_1)\big)^{\top}$, where $\mathbf{c}^{(1)}_{\widehat{\theta}_n}(\ell)=\mathrm{vec}(\mathbf{C}^{(1)}_{\widehat{\theta}_n}(\ell))$ with "vec" denoting the column wise vectorization of a matrix.

Let $\mathbf{U}_{t,\theta}\equiv\mathbf{U}_{t,\theta,K_1}$ and $\mathbf{u}_{t,\theta}\equiv\mathbf{u}_{t,\theta,K_1}$. Define the continuously differentiable transformation function $g^{(1)}:\mathbb{R}^{m(K_1+1)}\to\mathbb{R}^{m^2K_1}$ as

$$g^{(1)}(\mathbf{u}_{t,\theta})=\mathrm{vec}[\boldsymbol{\varepsilon}_{t,\theta}\boldsymbol{\varepsilon}^{\top}_{t-1,\theta},\ldots,\boldsymbol{\varepsilon}_{t,\theta}\boldsymbol{\varepsilon}^{\top}_{t-K_1,\theta}]. \tag{8}$$

Then $\mathbb{E}[g^{(1)}(\mathbf{U}_{t,\theta_0})]=\mathbf{0}$. Since the limiting distribution of $g^{(1)}(\mathbf{u}_{t,\theta})$ is multivariate standard normal, it is easy to see that the matrix $\boldsymbol{\Delta}$ in (3) is given by the identity matrix $\mathbf{I}_{m^2K_1}$ with dimension $d=m^2K_1$.

Using the general testing framework of Section 3, the portmanteau-type test statistic for testing residual autocovariances is given by

$$\mathcal{Q}^{(1)}_{K_1}=\frac{1}{n-K_1}\sum_{t=1+K_1}^{n}g^{(1)}(\mathbf{u}_{t,\widehat{\theta}_n})^{\top}(\widehat{\boldsymbol{\Omega}}^{(1)}_n)^{-1}\sum_{t=1+K_1}^{n}g^{(1)}(\mathbf{u}_{t,\widehat{\theta}_n})\xrightarrow{D}\chi^2(m^2K_1), \tag{9}$$

where

$$\widehat{\boldsymbol{\Omega}}^{(1)}_n=\widehat{\mathbf{G}}^{(1)}_n\widehat{\mathcal{I}}^{-1}_n\widehat{\mathbf{G}}^{(1)\top}_n+\widehat{\boldsymbol{\Psi}}^{(1)}_n\widehat{\mathcal{I}}^{-1}_n\widehat{\mathbf{G}}^{(1)\top}_n+\widehat{\mathbf{G}}^{(1)}_n\widehat{\mathcal{I}}^{-1}_n\widehat{\boldsymbol{\Psi}}^{(1)\top}_n+\widehat{\boldsymbol{\Delta}}^{(1)}_n, \tag{10}$$

and where by replacing $g(\mathbf{u}_{t,\widehat{\theta}_n})$ in (4) by $g^{(1)}(\mathbf{u}_{t,\widehat{\theta}_n})$, we have

$$\widehat{\mathbf{G}}^{(1)}_n=\frac{1}{n}\sum_{t=1}^{n}\frac{\partial g^{(1)}(\mathbf{u}_{t,\widehat{\theta}_n})}{\partial\boldsymbol{\theta}^{\top}},\quad\widehat{\boldsymbol{\Psi}}^{(1)}_n=\frac{1}{n}\sum_{t=1}^{n}g^{(1)}(\mathbf{u}_{t,\widehat{\theta}_n})\Big(\frac{\partial\ell_t(\mathbf{Y}_t;\widehat{\boldsymbol{\theta}}_n)}{\partial\boldsymbol{\theta}}\Big)^{\top},\quad\widehat{\boldsymbol{\Delta}}^{(1)}_n=\frac{1}{n}\sum_{t=1}^{n}g^{(1)}(\mathbf{u}_{t,\widehat{\theta}_n})g^{(1)}(\mathbf{u}_{t,\widehat{\theta}_n})^{\top}. \tag{11}$$

These statistics are the estimators of, respectively, $\mathbf{G}^{(1)}=\mathbb{E}[\partial g^{(1)}(\mathbf{U}_{t,\theta_0})/\partial\boldsymbol{\theta}_0^{\top}]$, $\boldsymbol{\Psi}^{(1)}=\mathbb{E}[g^{(1)}(\mathbf{U}_{t,\theta_0})\big(\partial\ell_t(\mathbf{Y}_t;\boldsymbol{\theta})/\partial\boldsymbol{\theta}\big)^{\top}]$, and $\boldsymbol{\Delta}^{(1)}=\mathbb{E}[g^{(1)}(\mathbf{U}_{t,\theta_0})g^{(1)}(\mathbf{U}_{t,\theta_0})^{\top}]$. The matrix $\widehat{\boldsymbol{\Omega}}^{(1)}_n$ is a consistent estimator of the covariance matrix $\boldsymbol{\Omega}$ defined in (3).

Remark 4. Under the hypothesis that model (1) is correctly specified, and using a conditional least squares estimator $\widehat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}_0$, Chabot-Hallé and Duchesne [2, Theorem 1] showed that

$$\sqrt{n}\mathbf{c}^{(1)}_{\widehat{\theta}_n}\xrightarrow{D}\mathcal{N}_{m^2K_1}(\mathbf{0},\boldsymbol{\Omega}^*), \tag{12}$$

where

$$\begin{aligned}\boldsymbol{\Omega}^*&=\big[\mathbf{J}(\mathbf{U}^{-1}\mathbf{R}\mathbf{U}^{-1})\ :\ \mathbf{I}_{m^2K_1}\big]\begin{bmatrix}\mathbf{U}\mathbf{R}^{-1}\mathbf{U} & -\mathbf{U}\mathbf{R}^{-1}\mathbf{J}^{*\top}\\ -\mathbf{J}^*\mathbf{R}^{-1}\mathbf{U} & \boldsymbol{\Delta}_{m^2K_1}\end{bmatrix}\begin{bmatrix}(\mathbf{U}^{-1}\mathbf{R}\mathbf{U}^{-1})\mathbf{J}^{\top}\\ \mathbf{I}_{m^2K_1}\end{bmatrix},\\ &=\mathbf{J}(\mathbf{U}^{-1}\mathbf{R}\mathbf{U}^{-1})\mathbf{J}^{\top}-\mathbf{J}^*\mathbf{R}^{-1}\mathbf{U}(\mathbf{U}^{-1}\mathbf{R}\mathbf{U}^{-1})\mathbf{J}^{\top}-\mathbf{J}(\mathbf{U}^{-1}\mathbf{R}\mathbf{U}^{-1})\mathbf{U}\mathbf{R}^{-1}\mathbf{J}^{*\top}+\boldsymbol{\Delta}_{m^2K_1},\end{aligned} \tag{13}$$

and

$$\begin{aligned}\mathbf{J}&=[\mathbf{J}^{\top}(1),\ \ldots,\mathbf{J}^{\top}(K_1)]^{\top},\quad\mathbf{J}(\ell)=\mathbb{E}\Big[(\mathbf{Y}_t-\mathbf{f}_{t-\ell-1})\otimes\frac{\partial\mathbf{f}_{t-1}}{\partial\boldsymbol{\theta}^{\top}}\Big],\\ \mathbf{J}^*&=[\mathbf{J}^{*\top}(1),\ \ldots,\mathbf{J}^{*\top}(K_1)]^{\top},\ \mathbf{J}^*(\ell)=-\mathbb{E}\Big[(\mathbf{Y}_t-\mathbf{f}_{t-\ell-1})\otimes\boldsymbol{\varepsilon}_{t,\theta_0}\boldsymbol{\varepsilon}^{\top}_{t,\theta_0}\boldsymbol{\Sigma}^{-1}_{\varepsilon}\frac{\partial\mathbf{f}_{t-1}}{\partial\boldsymbol{\theta}^{\top}}\Big],\\ \boldsymbol{\Delta}_{m^2K_1}&=(\boldsymbol{\Delta}_{ij})_{i,j=1}^{K_1},\quad\boldsymbol{\Delta}_{ij}=\mathbb{E}(\boldsymbol{\varepsilon}_{t-i,\theta_0}\boldsymbol{\varepsilon}^{\top}_{t-j,\theta_0}\otimes\boldsymbol{\varepsilon}_{t,\theta_0}\boldsymbol{\varepsilon}^{\top}_{t,\theta_0}),\\ \mathbf{U}&=\mathbb{E}\Big(\frac{\partial\mathbf{f}^{\top}_{t-1}}{\partial\boldsymbol{\theta}}\boldsymbol{\Sigma}^{-1}_{\varepsilon}\frac{\partial\mathbf{f}_{t-1}}{\partial\boldsymbol{\theta}^{\top}}\Big),\quad\mathbf{R}=\mathbb{E}\Big(\frac{\partial\mathbf{f}^{\top}_{t-1}}{\partial\boldsymbol{\theta}}\boldsymbol{\Sigma}^{-1}_{\varepsilon}(\mathbf{Y}_t-\mathbf{f}_{t-1})(\mathbf{Y}_t-\mathbf{f}_{t-1})^{\top}\boldsymbol{\Sigma}^{-1}_{\varepsilon}\frac{\partial\mathbf{f}_{t-1}}{\partial\boldsymbol{\theta}^{\top}}\Big).\end{aligned}$$

When comparing (13) with (3) we see that the $m^2K_1\times K$ matrix $\mathbf{J}$ replaces $\mathbf{G}$, $-\mathbf{J}^*\mathbf{R}^{-1}\mathbf{U}$ replaces $\boldsymbol{\Psi}$, $(\mathbf{U}^{-1}\mathbf{R}\mathbf{U}^{-1})$ replaces $\mathcal{I}(\boldsymbol{\theta}_0)^{-1}$, and $\boldsymbol{\Delta}_{m^2K_1}$ replaces $\boldsymbol{\Delta}$. Furthermore, if $\{\boldsymbol{\varepsilon}_{t,\theta_0}\}$ is a strong white noise process, $\mathbf{J}^*\equiv\mathbf{J}$ and $\mathbf{R}\equiv\mathbf{U}$. In that case the right-hand equation in (13) simplifies to $-\mathbf{J}\mathbf{U}^{-1}\mathbf{J}^{\top}+\boldsymbol{\Delta}_{m^2K_1}$.

Remark 5. A level-adjusted multiple-lag portmanteau test statistic follows by replacing $g^{(1)}(\mathbf{u}_{t,\theta})$ in (9) by $\widetilde{g}^{(1)}(\mathbf{u}_{t,\theta})=\mathrm{vec}[\sqrt{n/(n-1)}\boldsymbol{\varepsilon}_{t,\theta}\boldsymbol{\varepsilon}^{\top}_{t-1,\theta},\ldots,\sqrt{n/(n-K_1)}\boldsymbol{\varepsilon}_{t,\theta}\boldsymbol{\varepsilon}^{\top}_{t-K_1,\theta}]$. Following similar arguments as in Hosking [10], the resulting test statistic should have a small-sample distribution which is more nearly $\chi^2(m^2K_1)$ than that of $\mathcal{Q}^{(1)}_{K_1}$. Following Zhou et al. [25], one may also change $g^{(1)}(\mathbf{u}_{t,\theta})$ to $\widehat{\widehat{g}}^{(1)}(\mathbf{u}_{t,\theta})$:

$$\widehat{\widehat{g}}^{(1)}(\mathbf{u}_{t,\theta}) = \mathrm{vec}[\boldsymbol{\varepsilon}^{\top}_{t,\theta}\boldsymbol{\varepsilon}_{t-1,\theta}, \ldots, \boldsymbol{\varepsilon}^{\top}_{t,\theta}\boldsymbol{\varepsilon}_{t-K_1,\theta}] \in \mathbb{R}^{K_1}.$$

Then the resulting portmanteau test converges to $\chi^2(K_1)$, which has a much smaller degree of freedom than (9).

4.2. Single-lag test based on autocovariances of residuals

Clearly, (5) is a multiple-lag test statistic. It integrates several lags jointly. The test statistic does not provide insight in the possible residual dependence at each individual lag ℓ. A single-lag test statistic can be obtained as a special case of (9). To this end, let

$$\widehat{\mathbf{G}}^{(1)}_{n,\ell} = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}^{\top}}\mathrm{vec}(\widehat{\boldsymbol{\varepsilon}}_{t,\widehat{\theta}_n}\widehat{\boldsymbol{\varepsilon}}^{\top}_{t-\ell,\widehat{\theta}_n}),\quad \widehat{\boldsymbol{\Psi}}^{(1)}_{n,\ell} = \frac{1}{n}\sum_{t=1}^{n}\mathrm{vec}(\widehat{\boldsymbol{\varepsilon}}_{t,\widehat{\theta}_n}\widehat{\boldsymbol{\varepsilon}}^{\top}_{t-\ell,\widehat{\theta}_n})\Big(\frac{\partial\ell_t(\mathbf{Y}_t;\widehat{\boldsymbol{\theta}}_n)}{\partial\boldsymbol{\theta}}\Big)^{\top},$$
$$\widehat{\boldsymbol{\Delta}}^{(1)}_{n,\ell} = \frac{1}{n}\sum_{t=1}^{n}\mathrm{vec}(\widehat{\boldsymbol{\varepsilon}}_{t,\widehat{\theta}_n}\widehat{\boldsymbol{\varepsilon}}^{\top}_{t-\ell,\widehat{\theta}_n})(\mathrm{vec}(\widehat{\boldsymbol{\varepsilon}}_{t,\widehat{\theta}_n}\widehat{\boldsymbol{\varepsilon}}^{\top}_{t-\ell,\widehat{\theta}_n}))^{\top},$$

be consistent estimators of, respectively, $\mathbf{G}^{(1)}_{\ell} = \mathbb{E}[\partial\mathrm{vec}(\boldsymbol{\varepsilon}_{t,\theta_0}\boldsymbol{\varepsilon}^{\top}_{t-\ell,\theta_0})/\partial\boldsymbol{\theta}^{\top}]$, $\boldsymbol{\Psi}^{(1)}_{\ell} = \mathbb{E}[\mathrm{vec}(\boldsymbol{\varepsilon}_{t,\theta_0}\boldsymbol{\varepsilon}^{\top}_{t-\ell,\theta_0})(\partial\ell_t(\mathbf{Y}_t;\boldsymbol{\theta})/\partial\boldsymbol{\theta})^{\top}]$, and $\boldsymbol{\Delta}^{(1)}_{\ell} = \mathbb{E}[\mathrm{vec}(\boldsymbol{\varepsilon}_{t,\theta_0}\boldsymbol{\varepsilon}^{\top}_{t-\ell,\theta_0})(\mathrm{vec}(\boldsymbol{\varepsilon}_{t,\theta_0}\boldsymbol{\varepsilon}^{\top}_{t-\ell,\theta_0}))^{\top}]$, $\ell \in \{1,\ldots,K_1\}$. Given this setup, we propose the following level-adjusted single-lag test statistic

$$\mathcal{S}^{(1)}_{n,\ell} = \frac{n^2}{n-\ell}\mathbf{c}^{(1),\top}_{\widehat{\theta}_n}(\ell)\big(\widehat{\boldsymbol{\Omega}}^{(1)}_{n,\ell}\big)^{-1}\mathbf{c}^{(1)}_{\widehat{\theta}_n}(\ell),\quad \ell \in \{1,\ldots,K_1\}, \tag{14}$$

where

$$\widehat{\boldsymbol{\Omega}}^{(1)}_{n,\ell} = \widehat{\mathbf{G}}^{(1)}_{n,\ell}\widehat{\mathcal{I}}_n^{-1}\widehat{\mathbf{G}}^{(1)\top}_{n,\ell} + \widehat{\boldsymbol{\Psi}}^{(1)}_{n,\ell}\widehat{\mathcal{I}}_n^{-1}\widehat{\mathbf{G}}^{(1)\top}_{n,\ell} + \widehat{\mathbf{G}}^{(1)}_{n,\ell}\widehat{\mathcal{I}}_n^{-1}\widehat{\boldsymbol{\Psi}}^{(1)\top}_{n,\ell} + \widehat{\boldsymbol{\Delta}}^{(1)}_{n,\ell}. \tag{15}$$

Under the null hypothesis $\mathbb{H}^{(1,s)}_0$: $\mathrm{Cov}(\varepsilon_{t,\theta_0}, \varepsilon_{t-\ell,\theta_0}) = 0$, and assuming model (1) is correctly specified, it follows that $\mathcal{S}^{(1)}_{n,\ell} \xrightarrow{D} \chi^2(m^2)$.

4.3. Multiple-lag test based on autocovariances of squared residuals

4.3.1. General case

The null hypothesis for testing multiple-lag squared residual autocovariances is given by

$$\mathbb{H}^{(2)}_0:\ \mathrm{Cov}(\varepsilon^2_{i,t,\theta_0}, \varepsilon^2_{j,t-\ell,\theta_0}) = 0,\quad i,j \in \{1,\ldots,m\},\ \text{all } t,\ \text{and } \ell > 0. \tag{16}$$

Now consider an ARCH error process $\varepsilon_{t,\theta_0} = \eta_t h^{1/2}_{t,\theta_0}$ where $h_{t,\theta_0} = \alpha_0 + \sum_{i=1}^{r}\alpha_i\varepsilon^2_{t-i}$ ($\alpha_0 > 0, \alpha_i \geq 0$, $i \in \{1,\ldots,r\}$, $\theta_0 = (\alpha_0,\alpha_1,\ldots,\alpha_r)$ and $\{\eta_t\}$ is a sequence of i.i.d. random variables with mean zero and variance one. Let $\widehat{\boldsymbol{\theta}}_n$, $\widehat{\varepsilon}^2_{t,\widehat{\theta}_n}$ and $\widehat{h}_{t,\widehat{\theta}_n}$ be the corresponding definitions of θ_0, $\varepsilon^2_{t,\theta_0}$ and h_{t,θ_0}. Then it can be shown that $\bar{\varepsilon}_{\widehat{\theta}_n} = \sum_t \widehat{\varepsilon}^2_{t,\widehat{\theta}_n}/\widehat{h}_{t,\widehat{\theta}_n}$ converges to one in probability. Hence, the proposed test is based on the autocovariance-type statistic

$$\begin{aligned} c^{(2)}_{ij,\widehat{\theta}_n}(\ell) &= \frac{1}{n-\ell}\sum_{t=\ell+1}^{n}(\widehat{\varepsilon}^2_{i,t,\widehat{\theta}_n} - 1)(\widehat{\varepsilon}^2_{j,t-\ell,\widehat{\theta}_n} - 1),\\ &= \frac{1}{n-\ell}\sum_{t=\ell+1}^{n}(\widehat{\eta}_{i,t,\widehat{\theta}_n} - 1)(\widehat{\eta}_{j,t-\ell,\widehat{\theta}_n} - 1),\quad i,j \in \{1,\ldots,m\},\ \ell \in \{1,\ldots,K_2\},\ K_2 \ll n, \end{aligned} \tag{17}$$

where $\widehat{\eta}_{i,t,\widehat{\theta}_n} \equiv \widehat{\varepsilon}^2_{i,t,\widehat{\theta}_n}$, $i = 1,\ldots,m$. We define the corresponding residual autocovariance matrix as $\mathbf{C}^{(2)}_{\widehat{\theta}_n}(\ell) = \big(c^{(2)}_{ij,\widehat{\theta}_n}(\ell)\big)^m_{i,j=1}$. Furthermore, let $\widehat{\boldsymbol{\eta}}_{t,\widehat{\theta}_n} = (\widehat{\eta}_{1,t,\widehat{\theta}_n},\ldots,\widehat{\eta}_{m,t,\widehat{\theta}_n})^{\top}$ and $\mathbf{c}^{(2)}_{\widehat{\theta}_n} = \big(\mathbf{c}^{(2)\mathrm{T}}_{\widehat{\theta}_n}(1),\ldots,\mathbf{c}^{(2)\mathrm{T}}_{\widehat{\theta}_n}(K_2)\big)^{\top}$, where $\mathbf{c}^{(2)}_{\widehat{\theta}_n}(\ell) = \mathrm{vec}(\mathbf{C}^{(2)}_{\widehat{\theta}_n}(\ell))$.

Let $\mathbf{U}_{t,\theta} \equiv \mathbf{U}_{t,\theta,K_2}$ and $\mathbf{u}_{t,\theta} \equiv \mathbf{u}_{t,\theta,K_2}$. Define the continuously differentiable transformation function $g^{(2)}: \mathbb{R}^{m(K_2+1)} \to \mathbb{R}^{m^2K_2}$ as

$$g^{(2)}(\mathbf{u}_{t,\theta}) = \mathrm{vec}[\mathbf{w}_{t,\theta}\mathbf{v}^{\top}_{t-1,\theta},\ldots,\mathbf{w}_{t,\theta}\mathbf{v}^{\top}_{t-K_2,\theta}], \tag{18}$$

where $\mathbf{w}_{t,\theta} = [\varepsilon^2_{1,t,\theta} - 1,\ldots,\varepsilon^2_{m,t,\theta} - 1]^{\top}$ and $\mathbf{v}_{t-\ell,\theta} = [\varepsilon^2_{1,t-\ell,\theta} - 1,\ldots,\varepsilon^2_{m,t-\ell,\theta} - 1]^{\top}$, $\ell \in \{1,\ldots,K_2\}$. Note that $\mathbb{E}[g^{(2)}(\mathbf{U}_{t,\theta_0})] = \mathbf{0}$.

Using the general testing framework of Section 3, we propose the following multiple-lag portmanteau-type test statistic for testing squared residual autocovariances

$$\mathcal{Q}_{K_2}^{(2)} = \frac{1}{n-K_2}\sum_{t=1+K_2}^{n} g^{(2)}(\mathbf{u}_{t,\widehat{\theta}_n})^\top (\widehat{\boldsymbol{\Omega}}_n^{(2)})^{-1} \sum_{t=1+K_2}^{n} g^{(2)}(\mathbf{u}_{t,\widehat{\theta}_n}) \xrightarrow{D} \chi^2(m^2K_2), \tag{19}$$

where $\widehat{\boldsymbol{\Omega}}_n^{(2)}$ has the same structure as $\widehat{\boldsymbol{\Omega}}_n$ in (4) but with $g(\mathbf{u}_{t,\widehat{\theta}_n})$ replaced by $g^{(2)}(\mathbf{u}_{t,\widehat{\theta}_n})$. Thus,

$$\widehat{\mathbf{G}}_n^{(2)} = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial g^{(2)}(\mathbf{u}_{t,\widehat{\theta}_n})}{\partial \boldsymbol{\theta}^\top},\quad \widehat{\boldsymbol{\Psi}}_n^{(2)} = \frac{1}{n}\sum_{t=1}^{n} g^{(2)}(\mathbf{u}_{t,\widehat{\theta}_n})\Big(\frac{\partial \ell_t(\mathbf{Y}_t;\widehat{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}}\Big)^\top,\quad \widehat{\boldsymbol{\Delta}}_n^{(2)} = \frac{1}{n}\sum_{t=1}^{n} g^{(2)}(\mathbf{u}_{t,\widehat{\theta}_n}) g^{(2)}(\mathbf{u}_{t,\widehat{\theta}_n})^\top. \tag{20}$$

These statistics are consistent estimators of, respectively, $\mathbf{G}^{(2)} = \mathbb{E}[\partial g^{(2)}(\mathbf{U}_{t,\theta_0})/\partial \boldsymbol{\theta}_0^\top]$, $\boldsymbol{\Psi}^{(2)} = \mathbb{E}[g^{(2)}(\mathbf{U}_{t,\theta_0})(\partial \ell_t(\mathbf{Y}_t;\boldsymbol{\theta})/\partial\boldsymbol{\theta})^\top]$, and $\boldsymbol{\Delta}^{(2)} = \mathbb{E}[g^{(2)}(\mathbf{U}_{t,\theta_0}) g^{(2)}(\mathbf{U}_{t,\theta_0})^\top]$. Since the limiting distribution of $g^{(2)}(\mathbf{u}_{t,\theta})$ is $\mathcal{N}_{m^2K_2}(\mathbf{0}, \mathbf{I}_{m^2K_2})$, it is easy to show that $\boldsymbol{\Delta} = \mathbb{E}[g^{(2)}(\mathbf{U}_{t,\theta_0}) g^{(2)}(\mathbf{U}_{t,\theta_0})^\top] = \mathbf{I}_{m^2} \otimes (4\mathbf{I}_{K_2} + \mathbf{1}\mathbf{1}^\top)$, where $\mathbf{1}$ is a $K_2 \times 1$ vector with all elements equal to 1. The sth row of the $m^2K_2 \times K$ matrix $\mathbf{G}^{(2)}$ is given by $2\mathbb{E}[(\varepsilon_{i,t-s,\theta_0}^2 - 1)\varepsilon_{j,t,\theta_0}(\partial\varepsilon_{j,t,\theta_0}/\partial\boldsymbol{\theta}^\top) + (\varepsilon_{j,t-s,\theta_0}^2 - 1)\varepsilon_{i,t,\theta_0}(\partial\varepsilon_{i,t-s,\theta_0}/\partial\boldsymbol{\theta}^\top)]$.

Remark 6. As an alternative to (19), ARCH effects can also be tested using cross-correlations of powers of residuals; see, e.g., Section 8.4 of Francq and Zakoïan [7].

4.3.2. Special case

It is interesting to apply the general testing framework to a special case of (1). Specifically, consider a nonlinear multivariate time series model with m-dimensional conditional mean function and m-dimensional ARCH errors, i.e., $\mathbf{Y}_t = \boldsymbol{\mu}_t + \boldsymbol{\varepsilon}_{t,\theta_0}$, $\boldsymbol{\varepsilon}_{t,\theta_0} = \mathbf{V}_t^{1/2}\boldsymbol{\eta}_{t,\theta_0}$, where $\boldsymbol{\mu}_t = \mathbb{E}(\mathbf{Y}_t|\mathcal{F}_{t-1})$, $\mathbf{V}_t = \mathrm{Var}(\mathbf{Y}_t|\mathcal{F}_{t-1})$ and $\mathbf{V}_t^{1/2}$ denotes the (positive) square root of $\mathbf{V}_t$. Assume that $\{\boldsymbol{\eta}_{t,\theta_0} = (\eta_{1,t},\ldots,\eta_{m,t})^\top\}$ is a sequence of i.i.d. random vectors with mean $\mathbf{0}$ and covariance matrix $\mathbf{I}_m$, and $\boldsymbol{\eta}_t$ is independent of $\{\mathbf{Y}_{t-\ell}, \ell \geq 1\}$ for all t. Furthermore, assume that $\boldsymbol{\mu}_t$ and $\mathbf{V}_t$ have continuous second-order derivatives with respect to $\boldsymbol{\theta}_0$ a.s., and that the stationarity, invertibility and identifiability conditions hold for $\boldsymbol{\mu}_t$ and an identifiability condition holds for $\mathbf{V}_t$. Using the notations and definitions introduced in Section 2, let

$$\mathbf{A} \equiv \mathbb{E}\Big[\Big(\frac{1}{\sqrt{n}}\mathbf{S}_n(\boldsymbol{\theta}_0)\Big)\Big(\frac{1}{\sqrt{n}}\mathbf{S}_n(\boldsymbol{\theta}_0)\Big)^\top\Big],\quad \mathbf{B} \equiv \mathbb{E}[\mathbf{B}_n(\boldsymbol{\theta}_0)],\quad \mathbf{X} = [\mathbf{X}(1),\ldots,\mathbf{X}(M)]^\top$$

with $\mathbf{X}(\ell) = \mathbb{E}[(\partial\mathbf{V}_t/\partial\boldsymbol{\theta})\,\mathrm{vec}\{\mathbf{V}_t^{-1}(\boldsymbol{\varepsilon}_{t-\ell,\theta}^\top\mathbf{V}_{t-\ell}^{-1}\boldsymbol{\varepsilon}_{t-\ell,\theta} - 1)\}]$, $\ell \in \{1,\ldots,M\}$. The null hypothesis of interest is given by $\mathbb{H}_0: \mathrm{Cov}(\eta_{i,t,\theta_0}, \eta_{j,t-\ell,\theta_0}) = 0$ $i,j \in \{1,\ldots,m\}$, all t, and $\ell > 0$. Suppose the corresponding test statistic is based on the lag-ℓ residual autocovariance-type statistic $c_{\widehat{\theta}_n}(\ell) = (n-\ell)^{-1}\sum_{t=\ell+1}^{n}(\widehat{\boldsymbol{\eta}}_t^\top\widehat{\boldsymbol{\eta}}_t - 1)(\widehat{\boldsymbol{\eta}}_{t-\ell}^\top\widehat{\boldsymbol{\eta}}_{t-\ell} - 1)$, $\ell \in \{1,\ldots,M\}$, $M \ll n$. Ling and Li [16] showed that the asymptotic joint distribution of $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ and $\sqrt{n}\mathbf{c}_{\widehat{\theta}_n} = \sqrt{n}(c_{\widehat{\theta}_n}(1),\ldots,c_{\widehat{\theta}_n}(M))^\top$, is normal with mean $\mathbf{0}$ and covariance matrix

$$\begin{bmatrix} \mathbf{B}^{-1}\mathbf{A}\mathbf{B}^{-1} & \kappa\mathbf{B}^{-1}\mathbf{X}^\top/2 \\ \kappa\mathbf{X}\mathbf{B}^{-1}/2 & \kappa^2\mathbf{I}_M \end{bmatrix},$$

where $\kappa = \mathbb{E}[\eta_{i,t}^2(\boldsymbol{\eta}_t^\top\boldsymbol{\eta}_t - 1)] = \mathbb{E}[\eta_{i,t}^4 - 1]$, $i \in \{1,\ldots,m\}$. From these results it can be shown that $\sqrt{n}\mathbf{c}_{\widehat{\theta}_n} \xrightarrow{D} \mathcal{N}_M(\mathbf{0}, \kappa^2\boldsymbol{\Omega})$, where $\boldsymbol{\Omega} = \mathbf{I}_M - \mathbf{X}(\kappa\mathbf{B}^{-1} - \mathbf{B}^{-1}\mathbf{A}\mathbf{B}^{-1})\mathbf{X}^\top/\kappa^2$ while the corresponding multiple-lag test statistic is asymptotically χ_M^2 distributed if the model is correct. Remark 7 shows how this test relates to the results in Section 3.

Remark 7. Let $\mathbf{U}_{t,\theta_0} = [\boldsymbol{\eta}_t^\top\boldsymbol{\eta}_t - 1,\ldots,\boldsymbol{\eta}_{t-M}^\top\boldsymbol{\eta}_{t-M} - 1]^\top \in \mathbb{R}^{M+1}$. In addition, we define the continuously differential transformation function $g: \mathbb{R}^{M+1} \to \mathbb{R}^M$ as

$$g(\mathbf{u}_{t,\theta}) = [\mathbf{w}_{t,\theta}^\top\mathbf{v}_{t-1,\theta},\ldots,\mathbf{w}_{t,\theta}^\top\mathbf{v}_{t-M,\theta}], \tag{21}$$

where $\mathbf{w}_{t,\theta} = [\eta_{1,t} - 1,\ldots,\eta_{m,t} - 1]^\top$ and $\mathbf{v}_{t-\ell,\theta} = [\eta_{1,t-\ell} - 1,\ldots,\eta_{m,t-\ell} - 1]^\top$, $\ell \in \{1,\ldots,M\}$. Then, under Assumptions 2–4, the multiple-lag portmanteau test statistic for testing multivariate ARCH residuals is given by

$$\mathcal{Q}_M = \frac{1}{n-M}\sum_{t=1+M}^{n} g(\mathbf{u}_{t,\widehat{\theta}_n})^\top\,\widehat{\boldsymbol{\Omega}}_n^{-1}\sum_{t=1+M}^{n} g(\mathbf{u}_{t,\widehat{\theta}_n}) \xrightarrow{D} \chi^2(M), \tag{22}$$

where

$$\widehat{\boldsymbol{\Omega}}_n = \big[-\widehat{\mathbf{X}}_n\ :\mathbf{I}_M\big]\begin{bmatrix} \widehat{\mathbf{B}}_n^{-1}\widehat{\mathbf{A}}_n\widehat{\mathbf{B}}_n^{-1} & \widehat{\kappa}_n\widehat{\mathbf{B}}_n^{-1}\widehat{\mathbf{X}}_n^\top/2 \\ \widehat{\kappa}_n\widehat{\mathbf{X}}_n\widehat{\mathbf{B}}_n^{-1}/2 & \widehat{\kappa}_n^2\mathbf{I}_M \end{bmatrix}\begin{bmatrix} -\widehat{\mathbf{X}}_n^\top \\ \mathbf{I}_M \end{bmatrix} = \widehat{\mathbf{X}}_n\widehat{\mathbf{B}}_n^{-1}\widehat{\mathbf{A}}_n\widehat{\mathbf{B}}_n^{-1}\widehat{\mathbf{X}}_n^\top - \widehat{\kappa}_n\widehat{\mathbf{X}}_n\widehat{\mathbf{B}}_n^{-1}\widehat{\mathbf{X}}_n^\top + \widehat{\kappa}_n^2\mathbf{I}_M. \tag{23}$$

Here $\widehat{\mathbf{A}}_n$, $\widehat{\mathbf{B}}_n$, and $\widehat{\mathbf{X}}_n$ are consistent estimators of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{X}$, respectively. The constant $\widehat{\kappa}_n = n^{-1}\sum_{t=1}^{n}(\widehat{\boldsymbol{\eta}}_t^\top\widehat{\boldsymbol{\eta}}_t - 1)^2$ is a consistent estimator of κ. When comparing (23) with (4) we see that $\widehat{\kappa}_n^2\mathbf{I}_M$ replaces $\widehat{\boldsymbol{\Delta}}_n$, $\widehat{\mathbf{B}}_n^{-1}\widehat{\mathbf{A}}_n\widehat{\mathbf{B}}_n^{-1}$ replaces $\widehat{\mathcal{I}}_n^{-1}$, $-\widehat{\mathbf{X}}_n$

replaces $\widehat{\mathbf{G}}_n$, and $\widehat{\kappa}_n\widehat{\mathbf{X}}_n\widehat{\mathbf{B}}_n^{-1}/2$ replaces $\widehat{\boldsymbol{\Psi}}_n\widehat{\mathcal{I}}_n^{-1}$. If $\{\boldsymbol{\eta}_t, t \in \mathbb{Z}\}$ follows a multivariate normal distribution, $\widehat{\mathbf{B}}_n^{-1}\widehat{\mathbf{A}}_n\widehat{\mathbf{B}}_n^{-1}$ can be replaced by $\widehat{\mathbf{B}}_n^{-1}$ or $\widehat{\mathbf{A}}_n^{-1}$.

4.4. Single-lag test based on autocovariances of squared residuals

Similar to the single-lag test of Section 4.2, a one-lag test statistic based on squared residuals can be obtained from (5). In particular, for a fixed lag ℓ the level-adjusted single-lag test statistic is given by

$$\mathcal{S}_{n,\ell}^{(2)} = \frac{n^2}{n-\ell}\mathbf{c}_{\widehat{\boldsymbol{\theta}}_n}^{(2)\top}(\ell)\left(\widehat{\boldsymbol{\Omega}}_{n,\ell}^{(2)}\right)^{-1}\mathbf{c}_{\widehat{\boldsymbol{\theta}}_n}^{(2)}(\ell), \tag{24}$$

where

$$\widehat{\boldsymbol{\Omega}}_{n,\ell}^{(2)} = \widehat{\mathbf{G}}_{n,\ell}^{(2)}\widehat{\mathcal{I}}_n^{-1}\widehat{\mathbf{G}}_{n,\ell}^{(2)\top} + \widehat{\boldsymbol{\Psi}}_{n,\ell}^{(2)}\widehat{\mathcal{I}}_n^{-1}\widehat{\mathbf{G}}_{n,\ell}^{(2)\mathrm{T}} + \widehat{\mathbf{G}}_{n,\ell}^{(2)}\widehat{\mathcal{I}}_n^{-1}\widehat{\boldsymbol{\Psi}}_{n,\ell}^{(2)\top} + \widehat{\boldsymbol{\Delta}}_{n,\ell}^{(2)}, \tag{25}$$

with

$$\widehat{\mathbf{G}}_{n,\ell}^{(2)} = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}^{\mathrm{T}}}\mathrm{vec}(\widehat{\boldsymbol{\eta}}_{t,\widehat{\boldsymbol{\theta}}_n}\widehat{\boldsymbol{\eta}}_{t-\ell,\widehat{\boldsymbol{\theta}}_n}^{\top}), \quad \widehat{\boldsymbol{\Psi}}_{n,\ell}^{(2)} = \frac{1}{n}\sum_{t=1}^{n}\mathrm{vec}(\widehat{\boldsymbol{\eta}}_{t,\widehat{\boldsymbol{\theta}}_n}\widehat{\boldsymbol{\eta}}_{t-\ell,\widehat{\boldsymbol{\theta}}_n}^{\top})\left(\frac{\partial\ell_t(\mathbf{Y}_t;\widehat{\boldsymbol{\theta}}_n)}{\partial\boldsymbol{\theta}}\right)^{\top},$$

$$\widehat{\boldsymbol{\Delta}}_{n,\ell}^{(2)} = \frac{1}{n}\sum_{t=1}^{n}\mathrm{vec}\big(\widehat{\boldsymbol{\eta}}_{t,\widehat{\boldsymbol{\theta}}_n}\widehat{\boldsymbol{\eta}}_{t-\ell,\widehat{\boldsymbol{\theta}}_n}^{\top}\big)\left(\mathrm{vec}(\widehat{\boldsymbol{\eta}}_{t,\widehat{\boldsymbol{\theta}}_n}\widehat{\boldsymbol{\eta}}_{t-\ell,\widehat{\boldsymbol{\theta}}_n}^{\top})\right)^{\top}.$$

The statistics $\widehat{\mathbf{G}}_{n,\ell}^{(2)}$, $\widehat{\boldsymbol{\Psi}}_{n,\ell}^{(2)}$ and $\widehat{\boldsymbol{\Delta}}_{n,\ell}^{(2)}$ are consistent estimators of, respectively, $\mathbf{G}_\ell^{(2)} = \mathbb{E}[\partial\mathrm{vec}(\boldsymbol{\eta}_{t,\boldsymbol{\theta}_0}\boldsymbol{\eta}_{t-\ell,\boldsymbol{\theta}_0}^{\top})/\partial\boldsymbol{\theta}^{\top}]$, $\boldsymbol{\Psi}_\ell^{(2)} = \mathbb{E}[\mathrm{vec}(\boldsymbol{\eta}_{t,\boldsymbol{\theta}_0}\boldsymbol{\eta}_{t-\ell,\boldsymbol{\theta}_0}^{\top})(\partial\ell_t(\mathbf{Y}_t;\boldsymbol{\theta})/\partial\boldsymbol{\theta})^{\top}]$, and $\boldsymbol{\Delta}_\ell^{(2)} = \mathbb{E}[\mathrm{vec}(\boldsymbol{\eta}_{t,\boldsymbol{\theta}_0}\boldsymbol{\eta}_{t-\ell,\boldsymbol{\theta}_0}^{\top})(\mathrm{vec}(\boldsymbol{\eta}_{t,\boldsymbol{\theta}_0}\boldsymbol{\eta}_{t-\ell,\boldsymbol{\theta}_0}^{\top}))^{\top}]$, $\ell \in \{1,\ldots,K_2\}$. If the null hypothesis $\mathbb{H}_0^{(2,s)}: \mathrm{Cov}(\varepsilon_{t,\boldsymbol{\theta}_0}^2, \varepsilon_{t-\ell,\boldsymbol{\theta}_0}^2) = 0$ is satisfied and model (1) is correctly specified, it follows that $\mathcal{S}_{n,\ell}^{(2)} \xrightarrow{D} \chi^2(m^2)$.

Remark 8. Using the general testing framework of Section 3, the tests constructed in Sections 4.3 and 4.4 can be extended to powers of residuals. However, for the case of absolute residuals the so-called quasi-maximum likelihood approach is recommended for estimating model (1) with i.i.d. (non-Gaussian) errors $\boldsymbol{\varepsilon}_{t,\boldsymbol{\theta}_0}$; see, e.g., Li and Li [15] and Zhu [26].

5. Mixed multivariate portmanteau tests

For *univariate* ARMA-GARCH models, mixed portmanteau test statistics have been proposed by Wong and Ling [23], and Zhu [26]. Also, for a univariate nonlinear conditional mean and conditional variance model, Li [13] utilized different forms of autocorrelations to devise portmanteau test statistics. In this section, we combine the individual multiple- and single-lag test statistics of Sections 4.1 and 4.3, respectively, and construct mixed multivariate portmanteau-type tests based on residual autocovariances and squared residual autocovariances. To distinguish the mixed-type tests from the individual tests of Section 4, we introduce the superscript (1,2) in the notations when appropriate.

5.1. Mixed multiple-lag test

Let $\mathbf{U}_{t,\boldsymbol{\theta}} \equiv \mathbf{U}_{t,\boldsymbol{\theta},K_1+K_2}$ and $\mathbf{u}_{t,\boldsymbol{\theta}} \equiv \mathbf{u}_{t,\boldsymbol{\theta},K_1+K_2}$. For a fixed integer $M \equiv K_1 + K_2$, define the continuously differentiable transformation function $g: \mathbb{R}^{m(M+1)} \to \mathbb{R}^{m^2M}$ as

$$g(\mathbf{u}_{t,\boldsymbol{\theta}}) = [g^{(1)}(\mathbf{u}_{t,\boldsymbol{\theta}})^{\top}, g^{(2)}(\mathbf{u}_{t,\boldsymbol{\theta}})^{\top}]^{\top}, \tag{26}$$

where $g^{(1)}(\mathbf{u}_{t,\boldsymbol{\theta}})$ is given by (8) and $g^{(2)}(\mathbf{u}_{t,\boldsymbol{\theta}})$ by (18). Note that $\mathbb{E}[g(\mathbf{U}_{t,\boldsymbol{\theta}_0})] = \mathbf{0}$. The null hypothesis for testing mixed multiple-lag residual model inadequacies is given by

$$\mathbb{H}_0^{(1,2)}: \ \mathrm{Cov}(\varepsilon_{i,t,\boldsymbol{\theta}_0}, \varepsilon_{j,t-\ell,\boldsymbol{\theta}_0}^2) = 0, \ \ i,j \in \{1,\ldots,m\}, \text{ all } t, \text{ and } \ell > 0. \tag{27}$$

Assume without loss of generality that $\{\boldsymbol{\varepsilon}_{t,\boldsymbol{\theta}_0}\} \overset{i.i.d.}{\sim} \mathcal{N}_M(\mathbf{0}, \mathbf{I}_M)$. Then the joint distribution of $g^{(1)}(\mathbf{u}_{t,\widehat{\boldsymbol{\theta}}_n})$ and $g^{(2)}(\mathbf{u}_{t,\widehat{\boldsymbol{\theta}}_n})$ is given by the following theorem.

Theorem 2. *Under Assumptions 2–4,*

$$\frac{1}{\sqrt{n}}\left[g^{(1)}(\mathbf{u}_{t,\widehat{\boldsymbol{\theta}}_n})^{\top}, g^{(2)}(\mathbf{u}_{t,\widehat{\boldsymbol{\theta}}_n})^{\top}\right]^{\top} \xrightarrow{D} \mathcal{N}_{m^2M}(\mathbf{0}, \mathbf{P}\boldsymbol{\Omega}^{(1,2)}\mathbf{P}^{\top}),$$

where the $m^2M \times (m^2M + K)$ block matrix $\mathbf{P}$ and the $(m^2M + K) \times (m^2M + K)$ block matrix $\boldsymbol{\Omega}^{(1,2)}$ are given by

$$\mathbf{P} = \begin{bmatrix} \mathbf{I}_{m^2K_1} & \mathbf{0}_{m^2K_1\times m^2K_2} & \mathbf{G}^{(1)}\mathcal{I}(\boldsymbol{\theta}_0)^{-1} \\ \mathbf{0}_{m^2K_2\times m^2K_1} & \mathbf{I}_{m^2K_2} & \mathbf{G}^{(2)}\mathcal{I}(\boldsymbol{\theta}_0)^{-1} \end{bmatrix}, \quad \boldsymbol{\Omega}^{(1,2)} = \begin{bmatrix} \mathbf{I}_{m^2K_1} & \mathbf{0}_{m^2K_1\times m^2K_2} & \boldsymbol{\Psi}^{(1)} \\ \mathbf{0}_{m^2K_2\times m^2K_1} & \mathbf{I}_{m^2K_2} & \boldsymbol{\Psi}^{(2)} \\ \boldsymbol{\Psi}^{(1)\mathrm{T}} & \boldsymbol{\Psi}^{(2)\mathrm{T}} & \mathcal{I}(\boldsymbol{\theta}_0) \end{bmatrix}, \tag{28}$$

and where the $m^2K_1 \times K$ *matrices* $\mathbf{G}^{(1)}$ *and* $\boldsymbol{\Psi}^{(1)}$ *are defined in Section 4.1, the* $m^2K_2 \times K$ *matrices* $\mathbf{G}^{(2)}$ *and* $\boldsymbol{\Psi}^{(2)}$ *are defined in Section 4.3.1, and* $\mathcal{I}(\boldsymbol{\theta}_0)$ *is defined in Section 2.*

Proof. See Appendix. □

The ith, $m^2K_i \times m^2K_i$ principal diagonal matrix $\boldsymbol{\Omega}_{i,i}^{(1,2)}$ of $\mathbf{P}\boldsymbol{\Omega}^{(1,2)}\mathbf{P}^\top$ is given by

$$\boldsymbol{\Omega}_{i,i}^{(1,2)} = \mathbf{G}^{(i)}\mathcal{I}(\boldsymbol{\theta}_0)^{-1}\mathbf{G}^{(i)\top} + \boldsymbol{\Psi}^{(i)}\mathcal{I}(\boldsymbol{\theta}_0)^{-1}\mathbf{G}^{(i)\top} + \mathbf{G}^{(i)}\mathcal{I}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{\Psi}^{(i)\top} + \mathbf{I}_{m^2K_i}, \ i \in \{1, 2\}. \tag{29}$$

We see that for $i \in \{1, 2\}$, (29) is a special case of the SPD matrix (3). It is easy to see that the upper off-diagonal matrix of $\mathbf{P}\boldsymbol{\Omega}^{(1,2)}\mathbf{P}^\top$ is given by $\boldsymbol{\Omega}_{1,2}^{(1,2)} = \mathbf{G}^{(1)}\mathcal{I}(\boldsymbol{\theta}_0)^{-1}\mathbf{G}^{(2)\top} + \boldsymbol{\Psi}^{(1)}\mathcal{I}(\boldsymbol{\theta}_0)^{-1}\mathbf{G}^{(2)\top} + \mathbf{G}^{(1)}\mathcal{I}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{\Psi}^{(2)\top}$, and $\boldsymbol{\Omega}_{2,1}^{(1,2)} = (\boldsymbol{\Omega}_{1,2}^{(1,2)})^\top$ is the lower off-diagonal matrix. But $\boldsymbol{\Omega}_{1,2}^{(1,2)}$ and $\boldsymbol{\Omega}_{2,1}^{(1,2)}$ are not positive definite matrices, and hence $\mathbf{P}\boldsymbol{\Omega}^{(1,2)}\mathbf{P}^\top$ is not (semi) positive definite. However, from Higham [9] we know that the nearest symmetric positive semidefinite matrix in the sense of the Frobenius norm to an arbitrary real matrix $\mathbf{A}$ is $(\mathbf{B} + \mathbf{H})/2$ where $\mathbf{H}$ is the symmetric polar factor of $\mathbf{B} = (\mathbf{A} + \mathbf{A}^\top)/2$. This result has been implemented in the MATLAB code nearestSPD written by D'Errico [4]. The code is able to convert $\mathbf{P}\boldsymbol{\Omega}^{(1,2)}\mathbf{P}^\top$ into something that is indeed SPD. We denote the consistent estimator of the resulting SPD matrix by $\widehat{\boldsymbol{\Omega}}^{*,(1,2)}$.

Let $\widehat{\mathbf{P}}$ be a consistent estimator of $\mathbf{P}$. Then, from Theorem 2, we have the mixed multiple-lag multivariate portmanteau-type test statistic

$$\begin{aligned}\mathcal{Q}_M^{(1,2)} = &\Big[\frac{1}{n-K_1}\sum_{t=1+K_1}^{n} g^{(1)}(\mathbf{u}_{t,\widehat{\boldsymbol{\theta}}_n})^\top, \frac{1}{n-K_2}\sum_{t=1+K_2}^{n} g^{(2)}(\mathbf{u}_{t,\widehat{\boldsymbol{\theta}}_n})^\top\Big]^\top (\widehat{\mathbf{P}}\widehat{\boldsymbol{\Omega}}^{*,(1,2)}\widehat{\mathbf{P}}^\top)^{-1} \\ &\times \Big[\frac{1}{n-K_1}\sum_{t=1+K_1}^{n} g^{(1)}(\mathbf{u}_{t,\widehat{\boldsymbol{\theta}}_n})^\top, \frac{1}{n-K_2}\sum_{t=1+K_2}^{n} g^{(2)}(\mathbf{u}_{t,\widehat{\boldsymbol{\theta}}_n})^\top\Big] \xrightarrow{D} \chi^2(m^2(K_1+K_2)).\end{aligned} \tag{30}$$

Remark 9. Consider the case $K_2 = 0$, i.e., only residual autocovariances are used for testing the adequacy of model (1). The matrix $\mathbf{P}$ in (28) becomes $\mathbf{P} = [\mathbf{I}_{m^2K_1}\ \mathbf{G}^{(1)}]$, and $\boldsymbol{\Omega}^{(1,2)}$ is given by

$$\boldsymbol{\Omega}^{(1,2)} = \begin{bmatrix} \mathbf{I}_{m^2K_1} & \boldsymbol{\Psi}^{(1)} \\ \boldsymbol{\Psi}^\top & \mathcal{I}(\boldsymbol{\theta}_0) \end{bmatrix}.$$

So, $\mathbf{P}\boldsymbol{\Omega}^{(1,2)}\mathbf{P}^\top$ reduces to expression (29) for $i = 1$. In this case, the sample analogue of $\mathbf{P}\boldsymbol{\Omega}^{(1,2)}\mathbf{P}^\top$ is given by the matrix $\widehat{\boldsymbol{\Omega}}_n^{(1)}$ in (10). Similarly, in the case $K_1 = 0$ the sample analogue of $\mathbf{P}\boldsymbol{\Omega}^{(1,2)}\mathbf{P}^\top$ is given by $\widehat{\boldsymbol{\Omega}}_n^{(2)}$. So, $\mathcal{Q}_M^{(1,2)}$ nests the test statistics $\mathcal{Q}_M^{(1)}$ and $\mathcal{Q}_M^{(2)}$ when $M = K_1 = K_2$.

Remark 10. Consider the case $m = 1$. Assume that $K_1 = K_2 \equiv k$. The matrices $\mathbf{P}$ and $\boldsymbol{\Omega}^{(1,2)}$ in (28) are given by

$$\mathbf{P} = \begin{bmatrix} \mathbf{I}_k & \mathbf{0}_{k\times k} & \mathbf{G}^{(1)}\mathcal{I}(\boldsymbol{\theta}_0)^{-1} \\ \mathbf{0}_{k\times k} & \mathbf{I}_k & \mathbf{G}^{(2)}\mathcal{I}(\boldsymbol{\theta}_0)^{-1} \end{bmatrix}, \quad \boldsymbol{\Omega}^{(1,2)} = \begin{bmatrix} \mathbf{I}_k & \mathbf{0}_{k\times k} & \boldsymbol{\Psi}^{(1)} \\ \mathbf{0}_{k\times k} & \mathbf{I}_k & \boldsymbol{\Psi}^{(2)} \\ \boldsymbol{\Psi}^\top & \boldsymbol{\Psi}^{(2)\top} & \mathcal{I}(\boldsymbol{\theta}_0) \end{bmatrix}.$$

Here $\mathbf{G}^{(i)}$ and $\boldsymbol{\Psi}^{(i)}$ are two $k \times K$ matrices, $i \in \{1, 2\}$. These matrices are special cases of similar matrices defined in Sections 4.1 and 4.3.1. For the construction of the mixed multiple-lag *univariate* test statistic, we define the continuously differentiable transformation function $g^{*,(1,2)}\colon \mathbb{R}^{2k+1} \to \mathbb{R}^{2k}$ as $g^{*,(1,2)}(\mathbf{u}_{t,\boldsymbol{\theta}}) = [g^{*,(1)}(\mathbf{u}_{t,\boldsymbol{\theta}})^\top, g^{*,(2)}(\mathbf{u}_{t,\boldsymbol{\theta}})^\top]^\top$, where $g^{*,(1)}(\mathbf{u}_{t,\boldsymbol{\theta}}) = [\varepsilon_{t,\boldsymbol{\theta}}\varepsilon_{t-1,\boldsymbol{\theta}}, \ldots, \varepsilon_{t,\boldsymbol{\theta}}\varepsilon_{t-k,\boldsymbol{\theta}}]^\top$ and $g^{*,(2)}(\mathbf{u}_{t,\boldsymbol{\theta}}) = [(\varepsilon_{t,\boldsymbol{\theta}}^2 - 1)(\varepsilon_{t-1,\boldsymbol{\theta}}^2 - 1), \ldots, (\varepsilon_{t,\boldsymbol{\theta}}^2 - 1)(\varepsilon_{t-k,\boldsymbol{\theta}}^2 - 1)]^\top$. Assume that the null hypothesis $\mathbb{H}_0^{*,(1,2)}\colon \mathrm{Cov}(\varepsilon_{t,\boldsymbol{\theta}_0}, \varepsilon_{t-\ell,\boldsymbol{\theta}_0}^2) = 0$ is satisfied for all t and $\ell > 0$. In addition, assume that the univariate nonlinear model is correctly specified. Then, similar to (30), the mixed multiple-lag univariate portmanteau-type test statistic is given by

$$\begin{aligned}\mathcal{Q}_k^{*,(1,2)} = &\Big[\frac{1}{n-k}\sum_{t=1+k}^{n} g^{*,(1)}(\mathbf{u}_{t,\widehat{\boldsymbol{\theta}}_n})^\top, \frac{1}{n-k}\sum_{t=1+k}^{n} g^{*,(2)}(\mathbf{u}_{t,\widehat{\boldsymbol{\theta}}_n})^\top\Big]^\top (\widehat{\mathbf{P}}\widehat{\boldsymbol{\Omega}}^{(1,2)}\widehat{\mathbf{P}}^\top)^{-1} \\ &\times \Big[\frac{1}{n-k}\sum_{t=1+k}^{n} g^{*,(1)}(\mathbf{u}_{t,\widehat{\boldsymbol{\theta}}_n})^\top, \frac{1}{n-k}\sum_{t=1+k}^{n} g^{*,(2)}(\mathbf{u}_{t,\widehat{\boldsymbol{\theta}}_n})^\top\Big] \xrightarrow{D} \chi^2(2k).\end{aligned} \tag{31}$$

This test statistic can be viewed as a generalized version of a test proposed by Li [13, Sect. 3.3] used for diagnostic checking of univariate nonlinear DGPs with conditional mean function and ARCH errors.

Table 1
Empirical sizes (in percentages) of single-lag portmanteau-type test statistics at the nominal significance levels $\alpha = 0.05$ and $\alpha = 0.10$ using DGP-0 with Gaussian errors, ARCH errors and weak WN errors.

n	ℓ	Gaussian errors			ARCH	Weak WN	Gaussian errors			ARCH	Weak WN
		$\mathcal{S}_{n,\ell}^{(1)}$	$\mathcal{S}_{n,\ell}^{(2)}$	$\mathcal{S}_{n,\ell}^{(1,2)}$	$\mathcal{S}_{n,\ell}^{(1)}$	$\mathcal{S}_{n,\ell}^{(1)}$	$\mathcal{S}_{n,\ell}^{(1)}$	$\mathcal{S}_{n,\ell}^{(2)}$	$\mathcal{S}_{n,\ell}^{(1,2)}$	$\mathcal{S}_{n,\ell}^{(1)}$	$\mathcal{S}_{n,\ell}^{(1)}$
500	1	6.6	5.9	7.7	6.5	7.1	12.7	11.6	13.8	12.4	12.9
	2	5.7.	5.6	6.1	5.3	4.8	10.8	10.6	11.4	10.9	10.1
	3	5.3	5.7	5.8	5.3	4.8	10.6	10.9	10.9	10.5	10.1
	4	5.2	5.6	5.4	4.8	4.9	10.3	10.8	10.6	10.0	9.7
	5	5.2	5.9	5.4	4.5	4.8	10.4	11.2	11.1	9.9	9.9
1000	1	5.8	6.3	7.1	5.2	6.2	10.9	11.6	12.8	11.1	11.6
	2	4.9	5.7	5.7	4.9	5.2	10.0	11.5	11.1	10.2	10.4
	3	5.1	5.9	5.5	5.5	5.2	10.1	10.7	10.9	10.9	10.6
	4	5.0	6.3	6.2	5.2	5.1	10.0	11.5	11.6	10.4	9.9
	5	5.4	6.0	5.7	5.2	4.8	10.6	11.2	11.0	10.3	10.1

5.2. Mixed single-lag test

Similar to the single-lag test statistics $S_{n,\ell}^{(1)}$ and $S_{n,\ell}^{(2)}$, a mixed single-lag test statistic may be derived. In particular, the mixed single-lag test statistic $S_{n,\ell}^{(1,2)}$ is given by

$$\mathcal{S}_{n,\ell}^{(1,2)} = \frac{1}{n}\big[\mathbf{c}_{\widehat{\theta}_n}^{(1)\mathrm{T}}(\ell),\ \mathbf{c}_{\widehat{\theta}_n}^{(2)\mathrm{T}}(\ell)\big]^{\top}\Big(\mathbf{P}_\ell \boldsymbol{\Omega}_\ell^{(1,2)}\widehat{\mathbf{P}}_\ell^{\top}\Big)^{-1}\big[\mathbf{c}_{\widehat{\theta}_n}^{(1)\mathrm{T}}(\ell),\ \mathbf{c}_{\widehat{\theta}_n}^{(2)\mathrm{T}}(\ell)\big], \tag{32}$$

where

$$\mathbf{P}_\ell = \begin{bmatrix} \mathbf{I}_{m^2} & \mathbf{0}_{m^2\times m^2} & \mathbf{G}_\ell^{(1)}\mathcal{I}(\boldsymbol{\theta}_0)^{-1} \\ \mathbf{0}_{m^2\times m^2} & \mathbf{I}_{m^2} & \mathbf{G}_\ell^{(2)}\mathcal{I}(\boldsymbol{\theta}_0)^{-1} \end{bmatrix}, \quad \boldsymbol{\Omega}_\ell^{(1,2)} = \begin{bmatrix} \mathbf{I}_{m^2} & \mathbf{0}_{m^2\times m^2} & \boldsymbol{\Psi}_\ell^{(1)} \\ \mathbf{0}_{m^2\times m^2} & \mathbf{I}_{m^2} & \boldsymbol{\Psi}_\ell^{(2)} \\ \boldsymbol{\Psi}_\ell^{(1)\mathrm{T}} & \boldsymbol{\Psi}_\ell^{(2)\mathrm{T}} & \mathcal{I}(\boldsymbol{\theta}_0) \end{bmatrix}.$$

On replacing $\mathbf{G}_\ell^{(i)}$, $\boldsymbol{\Psi}_\ell^{(i)}$ and $\mathcal{I}(\boldsymbol{\theta_0})$ by consistent estimators and assuming model (1) is correctly specified, it follows that $S_{n,\ell}^{(1,2)} \xrightarrow{D} \chi^2(2m^2)$ under $\mathbb{H}_0^{(s,1,2)}: \mathrm{Cov}[(\varepsilon_{t,\boldsymbol{\theta}_0}, \eta_{t,\boldsymbol{\theta}_0})(\varepsilon_{t-\ell,\boldsymbol{\theta}_0}, \eta_{t-\ell,\boldsymbol{\theta}_0})^{\top}] = 0$, $\ell \in \{1,\ldots,M\}$.

6. Monte Carlo simulations

In this section, we examine the size and power of the set of multiple-lag test statistics $\{\mathcal{Q}_M^{(1)}, \mathcal{Q}_M^{(2)}, \mathcal{Q}_M^{(1,2)}\}$ and the set of single-lag test statistics $\{\mathcal{S}_{n,\ell}^{(1)}, \mathcal{S}_{n,\ell}^{(2)}, \mathcal{S}_{n,\ell}^{(1,2)}\}$ with $K_1 = K_2 \equiv M \in \{2, 3, 6, 10\}$, and $\ell \in \{1,\ldots,5\}$, and for various sample sizes n. For the empirical sizes reported in Tables 1 and 2 the number of replications is 3000 for each model and sample size combination. The parameters are estimated by conditional least squares. To study the size we employ the following DGP.

DGP-0: Two-regime first-order self-exciting threshold autoregressive model (VSETAR(2;1,1))

$$\mathbf{Y}_t = \{\boldsymbol{\Phi}_1^{(1)}\mathbf{Y}_{t-1} + \boldsymbol{\varepsilon}_t\}\mathbb{I}(Y_{1,t-1} < 0) + \{\boldsymbol{\Phi}_1^{(2)}\mathbf{Y}_{t-1} + \boldsymbol{\varepsilon}_t\}\mathbb{I}(Y_{1,t-1} \geq 0), \tag{33}$$

where $\mathbb{I}(\cdot)$ is the indicator function, and where

$$\boldsymbol{\Phi}_1^{(1)} = \begin{bmatrix} 0.6 & 0 \\ 0.3 & 0.6 \end{bmatrix},$$

$\boldsymbol{\Phi}_1^{(2)} = -\boldsymbol{\Phi}_1^{(1)}$. The random variables $\mathbf{Y}_{t-1}$ and $\boldsymbol{\varepsilon}_t$ are independent for all t. The innovations satisfy the following three specifications:

(i) $\{\boldsymbol{\varepsilon}_t\} \overset{i.i.d.}{\sim} \mathcal{N}_2(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})$ with

$$\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix}.$$

Model (33) is similar in structure as a VSETAR model specified by Tsay [20], albeit using different parameter specifications.

(ii) ARCH errors, i.e., $\boldsymbol{\varepsilon}_t = \mathbf{V}_t^{1/2}\boldsymbol{\eta}_t$, where

$$\mathbf{V}_t = \begin{bmatrix} 0.7 + 0.3\varepsilon_{1,t}^2 & 0 \\ 0 & 0.5 + 0.5\varepsilon_{2,t}^2 \end{bmatrix},$$

(iii) weak white noise (WN) errors, i.e., $\boldsymbol{\varepsilon}_t = (\eta_{1,t}\eta_{1,t-1}, \eta_{2,t}\eta_{2,t-1})^{\top}$, where $\boldsymbol{\eta}_t = (\eta_{1,t}, \eta_{2,t})^{\top} \overset{i.i.d.}{\sim} \mathcal{N}_2(\mathbf{0}, \mathbf{I}_2)$.

Table 2
Empirical sizes (in percentages) of multiple-lag portmanteau-type test statistics at the nominal significance levels $\alpha = 0.05$ and $\alpha = 0.10$ using DGP-0 with Gaussian errors, ARCH errors and weak WN errors.

n	M	Gaussian errors			ARCH	Weak WN	Gaussian errors			ARCH	Weak WN
		$\mathcal{Q}_M^{(1)}$	$\mathcal{Q}_M^{(2)}$	$\mathcal{Q}_M^{(1,2)}$	$\mathcal{Q}_M^{(1)}$	$\mathcal{Q}_M^{(1)}$	$\mathcal{Q}_M^{(1)}$	$\mathcal{Q}_M^{(2)}$	$\mathcal{Q}_M^{(1,2)}$	$\mathcal{Q}_M^{(1)}$	$\mathcal{Q}_M^{(1)}$
500	2	5.8	5.2	2.8	7.4	6.9	11.0	10.4	5.2	7.7	12.7
	3	5.6	4.7	4.0	7.3	6.3	11.3	10.3	6.8	7.4	12.4
	6	10.2	3.5	11.7	8.6	6.2	16.1	8.3	16.3	7.3	12.4
	10	12.1	2.5	25.3	7.1	5.9	17.7	6.7	31.1	6.9	10.9
1000	2	5.0	6.0	2.7	5.8	5.6	19.9	11.9	5.0	11.3	11.1
	3	5.4	5.8	5.2	5.7	5.6	10.4	11.0	6.6	11.7	10.7
	6	7.4	5.0	7.1	5.9	5.8	12.7	10.7	11.8	11.1	10.6
	10	9.6	4.2	14.6	5.3	5.3	15.4	9.0	21.7	10.9	10.5

Table 3
Power (in percentages) of the multivariate single-lag test statistics $\mathcal{S}_{n,\ell}^{(i)}$, $(i \in \{1,2\})$ and $\mathcal{S}_{n,\ell}^{(1,2)}$ and the multivariate multiple-lag test statistics $\mathcal{Q}_M^{(i)}$, $(i \in \{1,2\})$ and $\mathcal{Q}_M^{(1,2)}$ at a 5% nominal significance level.

n	ℓ	Single-lag			M	Multiple-lag		
		$\mathcal{S}_{n,\ell}^{(1)}$	$\mathcal{S}_{n,\ell}^{(2)}$	$\mathcal{S}_{n,\ell}^{(1,2)}$		$\mathcal{Q}_M^{(1)}$	$\mathcal{Q}_M^{(2)}$	$\mathcal{Q}_M^{(1,2)}$
250	1	6.5	2.5	43.3	2	100	100	93.7
	2	97.1	96.3	99.8	3	100	100	95.3
	3	99.4	99.2	99.8	6	100	100	99.8
	4	99.9	99.9	100	10	98.2	98.2	99.9
	5	100	99.9	100				
500	1	74.8	66.8	98.5	2	100	100	97.9
	2	100	100	100	3	100	100	96.3
	3	100	100	100	6	100	100	99.5
	4	100	100	100	10	100	100	100
	5	100	100	100				

Table 1 shows the empirical sizes of the single-lag portmanteau test statistics $\mathcal{S}_{n,\ell}^{(\cdot)}$ at the nominal significance levels $\alpha = 0.5$ and $\alpha = 0.10$ using DGP-0 with error processes (i)–(iii). Note that for all values of n and ℓ, there is a close agreement between the empirical and nominal sizes of all test statistics, with some improvements as the sample size increases from $n = 500$ to $n = 1000$. The overall pattern of empirical sizes remains satisfactory for $\ell \in \{6, \ldots, 10\}$ (not shown here). Table 2 displays similar results for the multiple-lag portmanteau test statistics $\mathcal{Q}_M^{(\cdot)}$ using again DGP-0 with error processes (i)–(iii). We see that for $M \in \{2, 3, 6\}$, the test statistics $\mathcal{Q}_M^{(i)}$ $(i \in \{1, 2\})$, defined in (9) and (19), are well approximated by a $\chi^2(m^2 M)$ distribution for all error distributions and both values of n. However, this is not the case for $\mathcal{Q}_M^{(1,2)}$ with severe overrejections of the null hypothesis in the case of ARCH errors and weak WN errors. Nonetheless, we like to point out that in the particular case of VARMA processes with i.i.d. innovations, the chi-square distribution is an approximation as the number of autocorrelations is taken large together with the sample size. It is clear that this phenomenon also holds for model (1).

To study the power of each portmanteau-type tests statistics, we choose DGP-0 as the null hypothesis and the following DGP as the alternative model.

DGP:

$$\begin{aligned}\mathbf{Y}_t &= \{\boldsymbol{\Phi}_1^{(1)}\mathbf{Y}_{t-1} + \boldsymbol{\Phi}_2^{(1)}\mathbf{Y}_{t-2} + \boldsymbol{\varepsilon}_t\}\mathbb{I}(Y_{1,t-1} < 0) + \{\boldsymbol{\Phi}_1^{(2)}\mathbf{Y}_{t-1} + \boldsymbol{\Phi}_2^{(2)}\mathbf{Y}_{t-2} + \boldsymbol{\varepsilon}_t\}\mathbb{I}(Y_{1,t-1} \geq 0),\\ \boldsymbol{\varepsilon}_t &= \boldsymbol{\Pi}\boldsymbol{\varepsilon}_{t-1} + \mathbf{u}_t,\end{aligned} \tag{34}$$

where $\boldsymbol{\Phi}_1^{(i)}$ is defined in DGP-0, $i = 1, 2$,

$$\boldsymbol{\Phi}_2^{(1)} = \begin{bmatrix} 0 & -0.2 \\ 0.5 & -0.3 \end{bmatrix}, \quad \boldsymbol{\Phi}_2^{(2)} = -\boldsymbol{\Phi}_2^{(1)}, \quad \boldsymbol{\Pi} = \begin{bmatrix} 0.10 & 0.06 \\ 0.01 & 0.90 \end{bmatrix},$$

and where $\{\boldsymbol{u}_t\} \overset{i.i.d.}{\sim} \mathcal{N}_2(\mathbf{0}, \mathbf{I}_2)$.

To make the rejection rates comparable across the test statistics, the estimated rejection rates are size-adjusted. Namely, based on 1000 replications, we empirically find the 95% percentiles of each test statistic and use these values as the corrected critical value for the power comparison. Table 3 shows size-adjusted power results for the single- and multiple-lag test statistics. Here, the results are based on 1000 replications of sample sizes $n = 250$ and $n = 500$. It is interesting to see that all tests have good power for all lags ℓ and all values M considered.

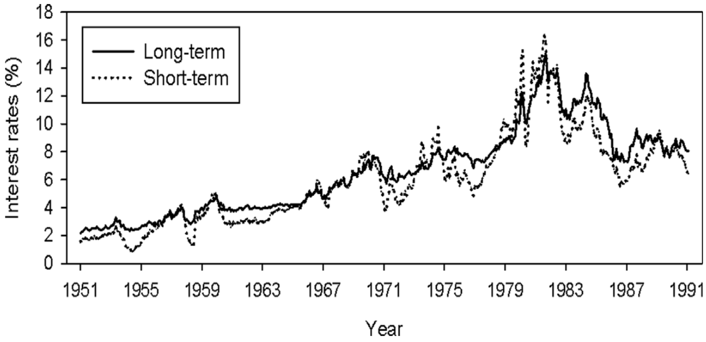

Fig. 1. U.S. interest rates 1951-01–1991-02.

7. Empirical application

To illustrate the usefulness of the proposed multivariate portmanteau test statistics in an empirical setting, we consider a set of residuals obtained from fitting a two-regime bivariate threshold vector error model (TVECM) to the U.S. long-term interest rate (R_t) and the U.S. short-term interest rate (r_t). The dataset can be downloaded from the Huston McCulloch website (https://www.asc.ohio-state.edu/mcculloch.2/ts/mcckwon/mccull.htm). Also it is included in the R-tsDyn package under the name zeroyld. Fig. 1 displays a plot of the data. Following Hansen and Seo [8], the period of interest is 1952-01 to 1991-02 ($n = 482$). We see a strong similarity between the patterns of both interest rate series indicating that R_t and r_t are cointegrated. Indeed, after a preliminary analysis Hansen and Seo [8] find evidence that a TVECM(2;1,1) fits the data better than a linear VECM. Here, we assume that the TVECM is correctly specified and analyze its residuals by the individual and mixed multivariate portmanteau tests discussed in the previous sections. So, to some extent our analysis is of historical interest.

First, we report estimation results for the TVECM(2;1,1). Using the R-tsDyn package, the fitted model is given by

$$\begin{bmatrix} \Delta R_t \\ \Delta r_t \end{bmatrix} = \left\{ \begin{bmatrix} 0.56^{**} \\ 1.94 \end{bmatrix} + \begin{bmatrix} 0.23^{*} \\ 1.15^{***} \end{bmatrix} w_{t-1} + \begin{bmatrix} 0.21 & -0.13 \\ 0.76^{***} & 0.02 \end{bmatrix} \begin{bmatrix} \Delta R_{t-1} \\ \Delta r_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{1,t} \end{bmatrix} \right\} \mathbb{I}(w_{t-1} \leq -1.05)$$
$$+ \left\{ \begin{bmatrix} 0.01 \\ -0.00 \end{bmatrix} + \begin{bmatrix} -0.00 \\ 0.05 \end{bmatrix} w_{t-1} + \begin{bmatrix} -0.03 & 0.08 \\ 0.14 & 0.17^{*} \end{bmatrix} \begin{bmatrix} \Delta R_{t-1} \\ \Delta r_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{2,t} \\ \varepsilon_{2,t} \end{bmatrix} \right\} \mathbb{I}(w_{t-1} > -1.05), \tag{35}$$

where $w_t = R_t - 1.046 r_{t-1}$ and $\Delta Y_t = Y_t - Y_{t-1}$, i.e., the first difference of a time series $\{Y_t\}$. Here, $***$ denotes a p-value in the range $[0, 0.001]$, $**$ denotes a p-value in the range $(0.001, 0.01]$, and $*$ is a p-value in the range $(0.01, 0.05]$.

The estimation results are in close agreement with those reported by Hansen and Seo [8, pp. 310–311]. Based on 8.8% of the observations, we see that the first regime occurs when $R_t \leq 1.05 r_t - 1.046$, i.e., when $\{R_t\}$ is more than 1.046 percentage points below r_t. The second regime (with 91.2% of the observations) is when $R_t > 1.05 r_t - 1.046$.

Next, we perform the proposed tests of Sections 4 and 5. The results are given in Table 4. Clearly, the $S_{n,\ell}^{(i)}$ ($i \in \{1, 2\}$; $\ell \in \{1, \ldots, 5\}$) tests indicate that (35) is adequate in modeling the interest rates. But this observation is not supported by the mixed single-lag test statistics with all p-values equal 0.000 at all lags ℓ. The values of $S_M^{(i)}$ ($i \in \{1, 2\}$; $M \in \{2, 3\}$) indicate that the fitted TVECM(2;1,1) might be suitable to describe the interest rates. But at $M = 6$ these statistics indicate model inadequacies. Similarly, the mixed test statistic $Q_M^{(1,2)}$ rejects model (35). However, as noted in Section 6, there are cases where this latter test statistic performs poorly. Hence, the results for $Q_M^{(1,2)}$ should be interpreted carefully. Finally, we would like to stress that the empirical application is intended as an illustration of the proposed portmanteau test statistics, not as an in-depth analysis of the residuals series based on the fitted TVECM(2;1,1).

8. Concluding remarks

In this paper, we proposed a general framework for testing multivariate white noise in multivariate nonlinear time series models with vector martingale errors. One advantage of the testing framework is that uncertainties due to model parameter estimation are naturally taken into account. Another advantage is that a large number of portmanteau-type tests follow as a special case of the proposed general test statistic. Specifically, we considered individual and mixed multivariate portmanteau-type test statistics. In each case, we distinguished between single- and multiple-lag tests using autocovariances of residuals and autocovariances of squared residuals, resulting in four individual tests and two mixed tests. A summary of these tests is given in Table 5. The results of the empirical application indicated that both individual and mixed tests are indispensable in detecting mis-specifications of the fitted VECM.

No attempt has been made to extend the individual tests to other forms of autocovariances, like bicovariances or, more general, cross-correlations of residuals, and their weighted variants. But by defining different forms of the transition

Table 4
P-values of the single-lag and multiple-lag portmanteau-type test statistics based on residuals obtained from the fitted TVECM(2;1,1) as given by (35).

Single-lag				Multiple-lag			
ℓ	$\mathcal{S}_{n,\ell}^{(1)}$	$\mathcal{S}_{n,\ell}^{(2)}$	$\mathcal{S}_{n,\ell}^{(1,2)}$	M	$\mathcal{Q}_{M}^{(1)}$	$\mathcal{Q}_{M}^{(2)}$	$\mathcal{Q}_{M}^{(1,2)}$
1	0.580	0.579	0.000	2	0.314	0.312	0.000
2	0.957	0.956	0.000	3	0.391	0.388	0.000
3	0.848	0.847	0.000	6	0.025	0.023	0.000
4	0.239	0.235	0.000	10	0.000	0.000	0.000
5	0.063	0.061	0.000				

Table 5
Summary of the individual multiple-lag test statistics $\mathcal{Q}_{K_i}^{(i)}$, the individual single-lag test statistics $\mathcal{S}_{n,\ell}^{(i)}$, ($i \in \{1,2\}$), and the mixed multiple- and single-lag test statistics $\mathcal{Q}_{M}^{(1,2)}$ and $\mathcal{S}_{n,\ell}^{(1,2)}$. Equation numbers are given in parentheses.

Null hypothesis ($\mathbb{H}_0$)	Individual	Distribution	Null hypothesis ($\mathbb{H}_0$)	Mixed	Distribution
$\mathrm{Cov}(\varepsilon_{i,t,\theta_0}, \varepsilon_{j,t-\ell,\theta_0}) = 0$	$\mathcal{Q}_{K_1}^{(1)}$ (9)	$\chi^2(mK_1)$	$\mathrm{Cov}(\varepsilon_{i,t,\theta_0}, \varepsilon^2_{j,t-\ell,\theta_0}) = 0$	$\mathcal{Q}_{M}^{(1,2)}$ (30)	$\chi^2(m^2M)$
$\mathrm{Cov}(\varepsilon_{t,\theta_0}, \varepsilon_{t-\ell,\theta_0}) = 0$	$\mathcal{S}_{n,\ell}^{(1)}$ (14)	$\chi^2(m^2)$			$(M = K_1 + K_2)$
			$\mathrm{Cov}[(\varepsilon_{t,\theta_0}, \eta_{t,\theta_0})(\varepsilon_{t-\ell,\theta_0}, \eta_{t-\ell,\theta_0})^{\mathrm{T}}] = 0$	$\mathcal{S}_{n,\ell}^{(1,2)}$ (32)	$\chi^2(2m^2)$
$\mathrm{Cov}(\varepsilon^2_{i,t,\theta_0}, \varepsilon^2_{j,t-\ell,\theta_0}) = 0$	$\mathcal{Q}_{K_2}^{(2)}$ (19)	$\chi^2(m^2K_2)$			
$\mathrm{Cov}(\varepsilon^2_{i,t,\theta_0}, \varepsilon^2_{j,t-\ell,\theta_0}) = 0$	$\mathcal{S}_{n,\ell}^{(2)}$ (24)	$\chi^2(m^2)$			

function $g(\cdot)$, special cases of the general portmanteau test statistic can be easily obtained. One may also explore the link between the well-known Lagrange multiplier portmanteau-type tests and each test statistic discussed in the paper. Additionally, a reviewer has suggested to study diagnostic tests under the weaker assumption of no correlated errors, i.e., that are not necessarily of an MDS-type. Another issue that will possibly be interesting for future research is to consider structural changes in the unconditional variance of model (1) as in Xu and Philips [24]. These issues could broaden the scope of the present study. Nevertheless, we hope that the proposed general testing framework can serve as a first step to derive other (omnibus) portmanteau-type tests for detecting multivariate white noise errors.

Acknowledgments

We thank the Editor-in-Chief, the Associate Editor and two reviewers for their valuable comments and suggestions.

Appendix

Proof of Theorem 2. Since $\lim_{n\to\infty} \mathbb{P}(\widehat{\boldsymbol{\theta}}_n \neq \infty) = 1$ by Proposition 1, it is assumed that $\widehat{\boldsymbol{\theta}}_n \neq \infty$. Also by Proposition 1, for every $\epsilon > 0$ there exist c_0 and n_0 such that $\mathbb{P}(\widehat{\boldsymbol{\theta}}_n \in N_{n,c_0}) > 1-\epsilon$ for all $n > n_0$. By Assumption 2(a), $n^{-1}\sum_{t=1}^{n} \partial g^{(i)}(\mathbf{u}_{t,\widetilde{\theta}_n})/\partial\boldsymbol{\theta}^{\mathrm{T}} \xrightarrow{P} \mathbf{G}^{(i)}$, and for all $\widetilde{\boldsymbol{\theta}}_n \in N_{n,c_0}$ and $c > 0$, so that especially $n^{-1}\sum_{t=1}^{n} \partial g^{(i)}(\mathbf{u}_{t,\widehat{\theta}_n})/\partial\boldsymbol{\theta}^{\mathrm{T}} \xrightarrow{P} \mathbf{G}^{(i)}$, $i \in \{1,2\}$.

Let $\boldsymbol{\xi}_n = \big[(1/\sqrt{n})\sum_{t=1}^{n} g^{(1)}(\mathbf{u}_{t,\widehat{\theta}_n})^{\mathrm{T}}, (1/\sqrt{n})\sum_{t=1}^{n} g^{(2)}(\mathbf{u}_{t,\widehat{\theta}_n})^{\mathrm{T}}\big]^{\mathrm{T}}$. The mean-value theorem implies that

$$\frac{1}{\sqrt{n}}\boldsymbol{\xi}_n = \frac{1}{\sqrt{n}}\Big[\sum_{t=1}^{n} g^{(1)}(\mathbf{u}_{t,\theta_0})^{\mathrm{T}}, \sum_{t=1}^{n} g^{(2)}(\mathbf{u}_{t,\theta_0})^{\mathrm{T}}\Big]^{\mathrm{T}} + \frac{1}{\sqrt{n}}\Big[\sum_{t=1}^{n}\frac{\partial g^{(1)}(\mathbf{u}_{t,\widetilde{\theta}})}{\partial\boldsymbol{\theta}^{\mathrm{T}}}, \sum_{t=1}^{n}\frac{\partial g^{(2)}(\mathbf{u}_{t,\widetilde{\theta}})}{\partial\boldsymbol{\theta}^{\mathrm{T}}}\Big]^{\mathrm{T}}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0), \tag{A.1}$$

where

$$\frac{\partial g^{(i)}(\mathbf{u}_{t,\widetilde{\theta}})}{\partial\boldsymbol{\theta}^{\mathrm{T}}} = \Big[\frac{\partial g^{(i)}(\mathbf{u}_{t,\widetilde{\theta}_{(1)}})}{\partial\boldsymbol{\theta}}, \ldots, \frac{\partial g^{(i)}(\mathbf{u}_{t,\widetilde{\theta}_{(m)}})}{\partial\boldsymbol{\theta}}\Big]^{\mathrm{T}},$$

with $g^{(i)}(\mathbf{u}_{t,\widetilde{\theta}_{(j)}}) = [g^{(i)}(\mathbf{u}_{t,\widetilde{\theta}_{(j)}})^{\mathrm{T}}, \ldots, g^{(i)}(\mathbf{u}_{t-K_i+1,\widetilde{\theta}_{(j)}})^{\mathrm{T}}]^{\mathrm{T}}$, $\widetilde{\boldsymbol{\theta}} = (\widetilde{\boldsymbol{\theta}}_{(1)}, \ldots, \widetilde{\boldsymbol{\theta}}_{(m)})$ and $\|\widetilde{\boldsymbol{\theta}}_{(j)} - \boldsymbol{\theta}_0\| < \|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|$ for each $i \in \{1,2\}$ and $j \in \{1,\ldots,m\}$. By Assumption 3(a) and Proposition 1, (A.1) can be written as

$$\frac{1}{\sqrt{n}}\boldsymbol{\xi}_n = \mathbf{P}\frac{1}{\sqrt{n}}\Big[\sum_{t=1}^{n} g^{(1)}(\mathbf{u}_{t,\widehat{\theta}_n})^{\mathrm{T}}, \sum_{t=1}^{n} g^{(2)}(\mathbf{u}_{t,\widehat{\theta}_n})^{\mathrm{T}}, \frac{1}{\sqrt{n}}\Big(\sum_{t=1}^{n}\frac{\partial \ell_t(\mathbf{Y}_t;\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\Big)^{\mathrm{T}}\Big]^{\mathrm{T}} + o_p(\mathbf{1}),$$

where $\mathbf{P}$ is an $(m^2(K_1+K_2)\times(m^2(K_1+K_2)+K))$ block matrix given by

$$\mathbf{P} = \begin{bmatrix} \mathbf{I}_{m^2K_1} & \mathbf{0}_{m^2K_1\times m^2K_2} & \mathbf{G}^{(1)}\mathcal{I}(\boldsymbol{\theta}_0)^{-1} \\ \mathbf{0}_{m^2K_2\times m^2K_1} & \mathbf{I}_{m^2K_2} & \mathbf{G}^{(2)}\mathcal{I}(\boldsymbol{\theta}_0)^{-1} \end{bmatrix}.$$

Denote

$$\mathbf{Z}_n = \Big[\frac{1}{\sqrt{n}}\Big(\sum_{t=1}^{n} g^{(1)}(\mathbf{u}_{t,\theta_0})\Big)^{\mathrm{T}}, \frac{1}{\sqrt{n}}\Big(\sum_{t=1}^{n} g^{(2)}(\mathbf{u}_{t,\theta_0})\Big)^{\mathrm{T}}, \frac{1}{\sqrt{n}}\sum_{t=1}^{n} \partial\ell_t(\mathbf{Y}_t;\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}^{\mathrm{T}}\Big]^{\mathrm{T}},$$

an $(m^2(K_1+K_2)+K)\times\ 1$ vector. By Assumption 3(a), we have $n^{-1}\sum_{t=1}^{n}\partial\ell_t(\mathbf{Y}_t;\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}^{\mathrm{T}}\}\cdot o_p(\mathbf{1}) = o_p(\mathbf{1})$. Then

$$\frac{1}{\sqrt{n}}\Big[\Big(\sum_{t=1}^{n} g^{(1)}(\mathbf{u}_{t,\theta_0})\Big)^{\mathrm{T}}, \Big(\sum_{t=1}^{n} g^{(2)}(\mathbf{u}_{t,\theta_0})\Big)^{\mathrm{T}}\Big]^{\mathrm{T}} = \mathbf{P}\mathbf{Z}_n + o_p(\mathbf{1}).$$

Using the concept of stacked matrices, we write $\mathbf{Z}_n = (1/\sqrt{n})\sum_{t=1}^{n}\boldsymbol{\nu}_t + o_p(\mathbf{1})$, where

$$\boldsymbol{\nu}_t = \Big[\mathrm{vec}(\boldsymbol{\varepsilon}_{t,\theta}\boldsymbol{\varepsilon}_{t-1,\theta}^{\mathrm{T}}),\ \ldots,\mathrm{vec}(\boldsymbol{\varepsilon}_{t,\theta}\boldsymbol{\varepsilon}_{t-K_1,\theta}^{\mathrm{T}}), \mathrm{vec}(\mathbf{w}_{t,\theta}\mathbf{v}_{t-1,\theta}^{\mathrm{T}}),\ldots,\mathrm{vec}(\mathbf{w}_{t,\theta}\mathbf{v}_{t-K_2,\theta}^{\mathrm{T}}), (1/\sqrt{n})\partial\ell_t(\mathbf{Y}_t;\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}^{\mathrm{T}}\Big]^{\mathrm{T}}.$$

It is straightforward to check that $\mathbb{E}(\boldsymbol{\nu}_t|\mathcal{F}_{t-1}) = \mathbf{0}$, and furthermore $\{\boldsymbol{\nu}_t\}$ is a martingale difference relative to $\mathcal{F}_1, \mathcal{F}_2, \ldots$. By the martingale central limit theorem, we have

$$\mathbf{Z}_n \xrightarrow{D} \mathcal{N}_{m^2(K_1+K_2)+K}(\mathbf{0}, \boldsymbol{\Omega}^{(1,2)}),$$

where

$$\boldsymbol{\Omega}^{(1,2)} = \mathbb{E}(\boldsymbol{\nu}_t\boldsymbol{\nu}_t^{\mathrm{T}}) = \begin{bmatrix} \mathbf{I}_{m^2K_1} & \mathbf{0}_{m^2K_1\times m^2K_2} & \boldsymbol{\Psi}^{(1)} \\ \mathbf{0}_{m^2K_2\times m^2K_1} & \mathbf{I}_{m^2K_2} & \boldsymbol{\Psi}^{(2)} \\ \boldsymbol{\Psi}^{(1)\mathrm{T}} & \boldsymbol{\Psi}^{(2)\mathrm{T}} & \mathcal{I}(\boldsymbol{\theta}_0)\end{bmatrix}.$$

Therefore, with $M = K_1 + K_2$,

$$\frac{1}{\sqrt{n}}\boldsymbol{\xi}_n \xrightarrow{D} \mathcal{N}_{m^2M}(\mathbf{0}, \mathbf{P}\boldsymbol{\Omega}^{(1,2)}\mathbf{P}^{\top}).\quad\square$$

References

[1] F. Akashi, M. Taniguchi, A.C. Monti, T. Amano, Diagnostic Methods in Time Series, Springer-Verlag, Singapore, 2021.
[2] D. Chabot-Hallé, P. Duchesne, Diagnostic checking of multivariate nonlinear time series models with martingale difference errors, Statist. Probab. Lett. 78 (2008) 997–1005.
[3] J.G. De Gooijer, A. Yuan, Non parametric portmanteau tests for detecting non linearities in high dimensions, Commun. Statist. Theory Methods 45 (2016) 385–399.
[4] J. D'Errico, Nearest SPD, 2022, Available at https://www.mathworks.com/matlabcentral/fileexchange/42885-nearestspd, MATLAB Central File Exchange.
[5] P. Duchesne, S. Lalancette, On testing for multivariate ARCH effects in vector time series models, Canad. J. Statist. 31 (2003) 275–292.
[6] C. Francq, H. Raïssi, Multivariate portmanteau test for autoregressive models with uncorrelated but nonindependent errors, J. Time Ser. Anal. 28 (2007) 454–470.
[7] C. Francq, J.-M. Zakoïan, GARCH Models: Structure, Statistical Inference and Financial Applications, second ed., Wiley, NewYork, 2019.
[8] B.E. Hansen, B. Seo, Testing for two-regime threshold cointegration in vector error-correction models, J. Econometrics 110 (2002) 293–318.
[9] N.J. Higham, Computing a nearest symmetric positive semidefinite matrix, Linear Algebra Appl. 103 (1988) 103–118.
[10] J.R.M. Hosking, The multivariate portmanteau statistic, J. Am. Statist. Assoc. 75 (1980) 602–607.
[11] L. Kalliovirta, Quantile Residuals for Multivariate Models Discussion Paper No. 247, Helsinki Center of Economic Research, University of Helsinki, Finland, 2008, Available at https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.876.3866&rep=rep1&type=pdf.
[12] W.K. Li, Diagnostic Checks in Time Series, Chapman & Hall/CRC, Boca Raton, Florida, 2004.
[13] Y. Li, On Mixed Portmanteau Statistics for the Diagnostic Checking of Time Series Models using Gaussian Quasi-Maximum Likelihood Approach (MSc. thesis), University of Hong Kong, Pokfulam, Hong Kong, 2012, Available at https://hub.hku.hk/handle/10722/173876.
[14] Z. Li, C. Lam, J. Yao, Q. Yao, On testing for high-dimensional white noise, Ann. Statist. 47 (2019) 3382–3412.
[15] G. Li, W.K. Li, Diagnostic checking for time series models with conditional heteroskedasticity estimated by the least absolute deviation approach, Biometrika 92 (2005) 691–701.
[16] S. Ling, W.K. Li, Diagnostic checking of nonlinear multivariate time series with multivariate ARCH errors, J. Time Ser. Anal. 18 (1997) 447–464.
[17] E. Mahdi, A.I. McLeod, Improved multivariate portmanteau test, J. Time Ser. Anal. 33 (2012) 211–222.
[18] Y.B. Maïnassara, Multivariate portmanteau test for structural VARMA models with uncorrelated but non-independent error terms, J. Statist. Plann. Inference 141 (2011) 2961–2975.
[19] T.J. Sweeting, Uniform asymptotic normality of the maximum likelihood estimator, Ann. Statist. 8 (1980) 1375–1381.
[20] R.S. Tsay, Testing and modeling multivariate threshold models, J. Amer. Statist. Assoc. 93 (1989) 1188–1202.
[21] G. Wang, K. Zhu, X. Shao, Testing for the martingale difference hypothesis in multivariate time series models, J. Bus. Econom. Statist. 40 (2022) 980–994.
[22] H. Wong, W.K. Li, Detecting and diagnostic checking multivariate conditional heteroscedastic time series models, Ann. Inst. Statist. Math. 54 (2002) 45–59.
[23] H. Wong, S. Ling, Mixed portmanteau tests for time series models, J. Time Ser. Anal. 26 (2005) 569–579.
[24] K.-L. Xu, P.C.B. Philips, Adaptive estimation of autoregressive models with time-varying variances, J. Econometrics 142 (2008) 265–280.
[25] J. Zhou, F. Jiang, K. Zhu, W.K. Li, Time series models for realized covariance matrices based on the matrix-f distribution, Statist. Sinica 32 (2022) 755–786.
[26] K. Zhu, A mixed portmanteau test for ARMA-GARCH models by the quasi-maximum exponential likelihood estimation approach, J. Time Ser. Anal. 34 (2013) 230–237.