Bootstrapping a powerful mixed portmanteau test for time series

Esam Mahdi and Thomas J. Fisher

Department of Statistical Sciences, University of Toronto, Toronto, ON, Canada; Department of Statistics, Miami University, Oxford, OH, USA

ABSTRACT
A new portmanteau test statistic is proposed for detecting nonlinearity in time series data. The new portmanteau statistic is calculated from the log of the determinant of a matrix comprised of the autocorrelations and cross-correlations of the residuals and squared residuals of a fitted time series. The asymptotic distribution of the proposed test statistic is derived as a linear combination of chi-square distributed random variables and can be approximated by a gamma distribution. A bootstrapping approach is shown to be robust when distributional assumptions are relaxed. The efficacy of the statistic is studied against linear and nonlinear dependency structures of some stationary time series models. It is shown that the new test can provide higher power than other tests in many situations. We demonstrate the advantages of the proposed test by investigating linear and nonlinear effects in an economic series and two environmental time series.

1. Introduction
Whether in environmental or economic applications, in the modern practice of time series the detection of nonlinear dynamics, and the modeling thereof, is of fundamental importance. Specifically, after a practitioner accounts for non-stationarity in a time series they will typically model any autocorrelation (or linear dynamics) as this is known to improve standard errors and forecasts. A logical step in this model fitting process is determining the adequacy of the fitted linear model. In many cases, this determination is performed with a so-called Portmanteau statistic; see [4]. Due to the skewness and heavy-tailed data we often incur in modern data analysis, computational methods are often needed to relax the distribution assumptions of many popular statistical approaches including the techniques for time series analysis.

Bootstrapping is a well-known computational method to approximate the variance, and other properties, of a sample statistic. Since the seminal work of Efron [9], many variants of the bootstrap have been proposed and it is known to have many desirable properties [10,13]. In the field of time series it is important to retain any temporal structure in the
bootstrap samples. The block bootstrap [17,23], and stationary bootstrap [42], were developed to work for time series. Other variations exist [18, for example] and a review of the topic can be found in Härdle et al. [19]. For regression applications with a heteroskedasticity, the Wild Bootstrap [52] was developed and its application has been well studied [31,32,35]. This technique has been used in time series when blue heteroskedasticity is present [16,22,54] and in other applications [see 27, for example]. Recently, [26] proposed using the Wild bootstrap on the well-known Ljung-Box portmanteau test [30] in time series and demonstrated it retains adequate type I error rates when heteroskedasticity is present. A Randomly Weighted Bootstrap (RWB) (a variant of the Wild bootstrap proposed in Jin et al. [20]) was proposed for several time series portmanteau statistics in the literature in the presence of heteroskedasticity in Zhu [53]. Recently, the RWB was used in the estimation of the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model in [54].

This article proposes a new portmanteau test that combines several existent results from the literature. Utilizing the method of [53] and [54], we use the RWB technique to approximate the distribution of the statistic under some fairly general scenarios (situations where the underlying stochastic process is heavy-tailed and skewed). The article is organized as follows: Section 2 provides a brief review of standard time series models and some popular portmanteau test statistics that have been used for detecting linear and nonlinear dependency in time series. In Section 3 we propose a new portmanteau test statistic, derive its asymptotic distribution as a linear combination of chi-square random variables and discuss some of its properties. An extension of the RWB procedure is then outlined to approximate the distribution of the proposed statistic. Section 4 reports a Monte Carlo study comparing the empirical findings with the theoretical results and demonstrates that the empirical significance level of the proposed test statistic is accurately estimated by the percentiles of its asymptotic distribution. Simulations also show that the power of the test is often higher than that of other test statistics. Two illustrative applications are given in Section 5 to demonstrate the usefulness of the proposed test for real world datasets. We end the article in Section 6 with some discussions on the advantages and limitations of the new statistic.

2. Time series modeling

The autoregressive-moving average (ARMA) model is arguably the most fundamental of all time series models. An ARMA$(p,q)$ for $n$ observations $z_1, z_2, \ldots, z_n$ of a stationary mean $\mu$ time series can be expressed as

$$\Phi_p(B)(z_t - \mu) = \Theta_q(B)\epsilon_t$$

with

$$\Phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p,$$

$$\Theta_q(B) = 1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q,$$

where $B$ is the backshift operator and the polynomials $\Phi_p(B)$ and $\Theta_q(B)$ are assumed to have all roots outside the unit circle on the complex plain and have no common roots. The noise sequence $\{\epsilon_t\}$ may have further structure or be independent and identically distributed (iid) with mean 0 and constant variance, $\sigma^2 > 0$. 
Let $\beta = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, \mu)$ denote the true parameter values and let $\hat{\beta} = (\hat{\phi}_1, \ldots, \hat{\phi}_p, \hat{\theta}_1, \ldots, \hat{\theta}_q, \hat{\mu})$ denote the $\sqrt{n}$ consistent estimated values, so that the residuals $\hat{\varepsilon}_t$ denote the estimated values of $\varepsilon_t$ for $t = 1, \ldots, n$. If the model in (1) is correctly identified and the noise terms $\{\varepsilon_t\}$ are uncorrelated, then for all non-zero lag time $k$, the residual autocorrelation function, $\text{corr}(\varepsilon_t, \varepsilon_{t+k})$, and the squared residual (or absolute-residual) autocorrelation function, $\text{corr}(\varepsilon_t^2, \varepsilon_{t+k}^2)$ (or $\text{corr}(|\varepsilon_t|, |\varepsilon_{t+k}|)$), should show no specific pattern and the correlation coefficient values should be approximately equal to zero. In addition, the cross-correlation function between the residuals and their squares, $\text{corr}(\varepsilon_t, \varepsilon_{t+k}^2)$, should be approximately uncorrelated with zero values. On the other hand, if the model is not adequately identified, the autocorrelation may take on non-zero values. Further, if there are nonlinear effects in the time series or if the residuals are not independent, these features may appear in the autocorrelation function of the squared (or the absolute) residuals or the cross-correlation of the residuals and their squares. The case for absolute residuals is beyond the scope of this article and we focus our attention on methods using the squared residuals.

Many nonlinear models have been proposed and can be used for analyzing nonlinear time series [see Ch. 10 in 41]. For example, when the model is linear in mean but nonlinear in variance, [11] proposed the Autoregressive Conditional heteroskedasticity, ARCH, that is widely used for analyzing financial time series. This model was generalized by Bollerslev [2], the so-called GARCH process. The innovations $\{\varepsilon_t\}$ in (1) follows a GARCH $(b, a)$ process if

$$\varepsilon_t = \xi_t \sigma_t, \quad \sigma_t^2 = \omega + \sum_{i=1}^{b} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{a} \beta_j \sigma_{t-j}^2,$$

where $\xi_t$ is a sequence of iid random variables with a mean value of 0 and variance of 1, $\omega > 0$, $\alpha_i \geq 0$ for $i = 1, \ldots, a$ and $\alpha_i = 0$ for all $i > a$, $\beta_j \geq 0$ for all $j = 1, \ldots, b$ and $\beta_j = 0$ for all $j > b$, and $\sum_{i=1}^{\text{max}(b,a)} (\alpha_i + \beta_i) < 1$. The summation on $\alpha_i \varepsilon_{t-i}^2$ terms comprises the original ARCH model in Engle [11] with the second summation as the generalization. Many variants of the GARCH process have been proposed to model the dynamic behavior of conditional heteroskedasticity in real time series. Nice reviews of these models are available in Tsay [6,49].

Typically, a practitioner may use a so-called portmanteau test to check the adequacy of a fitted model of form (1) or for the presence of GARCH-type effects in (2). Many authors have proposed such tests, including the commonly employed test statistics by Box and Pierce [4], Ljung and Box [30]. In [36], a portmanteau test for detecting nonlinearity (e.g., presence of ARCH or GARCH-effects) based on the squared residual autocorrelations are proposed. Several authors have improved on the portmanteau statistics by considering different functions of the autocorrelations of a fitted ARMA model or autocorrelations of the squared residuals (see [12,28,39,40,46]). Simulation studies show that these statistics respond well in the detection of ARCH models but tend to lack power compared to other types of nonlinear models that do not have the ARCH effects.

The idea of using empirical generalized correlations (correlation between $\hat{\varepsilon}_t^j$ and $\hat{\varepsilon}_{t+k}^j$, where $k$ is the lag time and $i, j$ are positive integers) to inspect for nonlinear dependency in time series models without considering the effects of the parameter estimation were
introduced in [24,25]. Under the ARMA assumptions in (1), [43] recently proposed portmanteau tests based on the generalized correlations and showed that the test based on the cross-correlation between the residuals and their squares can be more powerful than that of McLeod and Li [36], which is based on the autocorrelation of the squared residuals.

2.1. Portmanteau review

Define $\hat{\rho}_{ij}(k)$ to be the correlation coefficient at lag time $k$ between $\hat{\varepsilon}_i^t$ and $\hat{\varepsilon}_j^{t+k}$, where we focus our attention on $i, j = 1, 2$. At lag time $k$, $\hat{\rho}_{11}(k)$ denotes the autocorrelation coefficient of the residuals, $\hat{\rho}_{22}(k)$ denotes the autocorrelation coefficient of the squared residuals, and $\hat{\rho}_{12}(k)$ (or $\hat{\rho}_{21}(k)$) are the cross-correlation between the residuals and their squares at positive (or negative) lag $k$. Thus, $\hat{\rho}_{ij}(k)$ is given by

$$\hat{\rho}_{ij}(k) = \frac{\hat{\gamma}_{ij}(k)}{\sqrt{\hat{\gamma}_{ii}(0)}} \sqrt{\hat{\gamma}_{jj}(0)},$$

where

$$\hat{\gamma}_{ij}(k) = \frac{1}{n} \sum_{t=1}^{n-k} f_i(\hat{\varepsilon}_t) f_j(\hat{\varepsilon}_{t+k}) \text{ for } k = 0, \pm 1, \pm 2, \ldots \quad (3)$$

Note that $\gamma_{ij}(k) = \gamma_{ji}(-k)$ for $k > 0$ is the autocovariance (cross-covariance) at lag $k$ between the residuals to the power $i$ and the residuals to the power $j$ for $i, j = 1, 2$, for $f_i(x_i) = x_i^i - n^{-1} \sum_{t=1}^{n} x_i^t$, for $i = 1, 2$.

Under the assumption that the data has been generated from an ARMA process, [4] proposed to the time series literature the nominal portmanteau test in order to check the adequacy of the fitted model. Ljung and Box [30] improved their test by utilizing a multiplicative factor on each squared autocorrelation term. The two respective tests are

$$\tilde{Q}_{11} = n \sum_{k=1}^{m} \hat{\rho}_{11}^2(k) \quad \text{and} \quad Q_{11} = n(n+2) \sum_{k=1}^{m} (n-k)^{-1} \hat{\rho}_{11}^2(k), \quad (4)$$

where $0 < m < n/2$ is the maximum lag considered for significant autocorrelation. Both $\tilde{Q}_{11}$ and $Q_{11}$ share the same asymptotic $\chi^2_{m-p-q}$ distribution but $Q_{11}$ generally has more power.

If the assumptions in (1) are satisfied, [36] proposed a portmanteau test for detecting the presence of the ARCH-effects, based on the autocorrelations of the squared residuals:

$$Q_{22} = n(n+2) \sum_{k=1}^{m} (n-k)^{-1} \hat{\rho}_{22}^2(k). \quad (5)$$

[36] showed that the limiting distribution of $Q_{22}$ can be approximated by a chi-square distribution with $m$ degrees of freedom which different than $Q_{11}$ in (4) since the limiting distribution of $Q_{22}$ does not depend on the order of the fitted ARMA model.

Simulation studies show that the portmanteau statistics based on the squared residuals autocorrelations, such as $Q_{22}$, respond well to ARCH models but tend to lack power in the presence of other types of nonlinear models. One possible reason for the lack of power
could be the fact that these statistics ignore the generalized cross-correlation between the residuals to different powers; i.e., \( \hat{\rho}_{ij}(k) \) in (3) for \( i \neq j \). In this respect, [24,25] proposed the idea of testing for nonlinearity in time series models using the cross-correlation between the residuals and squared residuals. Psaradakis and Vávra [43] develop portmanteau tests to detect nonlinearity from stationary linear models, based on the generalized correlation. Their test statistics are given by

\[
Q_{ij} = n(n + 2) \sum_{k=1}^{m} (n - k)^{-1} \hat{\rho}_{ij}^2(k),
\]

where \( i, j = 1, 2, \ i \neq j \). These tests can be seen as modified versions of the [4,36] tests that utilize the cross-correlation tests by \( \chi_{n}^2 \) and suggested that the tests based on the cross-correlations tend to be more powerful in detecting many types of nonlinearity compared to other statistics based on squared residual autocorrelations.

Many other portmanteau statistics have been developed for time series modeling. Peña and Rodríguez [39] proposed a test to check the adequacy of the fitted ARMA model based on the \( m \)th root of the determinant of a \( m \)th sample residual autocorrelations matrix. They extended this statistic to test for nonlinearity by replacing the sample residual autocorrelations with the squared residual autocorrelations in the \( m \)th autocorrelation matrix. In [40], they considered the log of the determinant of a \( m \)th sample residual autocorrelations matrix. These statistics are shown to be functions of the partial autocorrelation function, similar to [37]. Mahdi and Ian McLeod [34] extend the result to multivariate time series and [12] uses the same matrix to derive a Weighted Ljung-Box and Weighted McLeod-Li test that are asymptotically similar to that of [39,40].

### 3. Proposed test statistic

Motivated by the results in Lawrance and Lewis [24], Lawrance and Lewis [25,39,40,43], we propose a new test for determining the adequacy of a fitted ARMA model. For a stationary time series, consider the block matrix of autocorrelations and cross-correlations of residuals and squared residuals,

\[
\hat{R}(m) = \begin{bmatrix}
\hat{R}_{11}(m) & \hat{R}_{12}(m) \\
\hat{R}_{12}(m) & \hat{R}_{22}(m)
\end{bmatrix}_{2(m+1) \times 2(m+1)}
\]

where

\[
\hat{R}_{ii}(m) = \begin{bmatrix}
1 & \hat{\rho}_{ii}(1) & \ldots & \hat{\rho}_{ii}(m) \\
\hat{\rho}_{ii}(-1) & 1 & \ldots & \hat{\rho}_{ii}(m-1) \\
\vdots & \vdots & \ddots & \vdots \\
\hat{\rho}_{ii}(-m) & \hat{\rho}_{ii}(1-m) & \ldots & 1
\end{bmatrix}, \text{ for } i = 1, 2
\]
and

\[
\hat{R}_{12}(m) = \begin{bmatrix}
0 & \hat{\rho}_{12}(1) & \ldots & \hat{\rho}_{12}(m) \\
\hat{\rho}_{12}(-1) & 0 & \ldots & \hat{\rho}_{12}(m-1) \\
\vdots & \ddots & \ddots & \vdots \\
\hat{\rho}_{12}(-m) & \hat{\rho}_{12}(1-m) & \ldots & 0
\end{bmatrix}.
\]

Note that \(\hat{R}_{21}(m) = \hat{R}_{12}'(m)\) is the matrix of cross-correlations between residuals and their squares, \(\hat{R}_{11}(m)\) is the residual autocorrelation Toeplitz matrix as defined in Peña and Rodriguez [39], and \(\hat{R}_{22}(m)\) is the matrix of the autocorrelation coefficients based on the squared residuals.

Block matrices of this form have many desirable properties; e.g., this matrix is positive definite and \(0 < |\hat{R}(m)| \leq 1\), where \(|·|\) denotes the determinant of the matrix. Variants of \(\hat{R}(m)\) have been used to build other test statistics [see 34, 39, 45, for example] and is the foundation for our proposed portmanteau test:

\[
C_m = -\frac{n}{m} \log|\hat{R}(m)|. \tag{8}
\]

Under the null hypothesis that an ARMA model is adequately modeling the linear effects and no nonlinear effects are present, \(H_0: R(m) = I_{2(m+1)}\), where \(I_{2(m+1)}\) is the identity matrix of dimension \(2(m+1)\). That is, the sample autocorrelations of the residuals/squared residuals and the sample cross-correlations between the residuals and their squares will not significantly differ from zero, and \(C_m\) will take on a value near zero. Under the alternative of an inadequate ARMA model or the presence of nonlinear effects, \(R(m)\) will deviate from an identity, \(|\hat{R}(m)| < 1\), and will approach 0 [see 39, 45], thus \(C_m\) will increase and the fitted linear model will be rejected as inadequate.

### 3.1. Distribution of proposed statistic

We now derive the asymptotic distribution of the proposed statistic \(C_m\), while also discussing some complexities regarding it. Through a sequence of Lemmas and Theorems we demonstrate that under the assumption of normality, \(C_m\) is asymptotically distributed as a linear combination of \(\chi^2\) random variables that can be approximated by a Gamma distribution. We also discuss the distribution in the more general case. In the derivation below \(\mathbf{0}\) represents an appropriately sized vector or matrix of zeroes.

**Lemma 3.1:** The quantity \(-n \log|\hat{R}(m)|\) can asymptotically be decomposed into three components

\[
-n \log|\hat{R}(m)| = -n \log|\hat{R}_{11}(m)| - n \log|\hat{R}_{22}(m)| + n \left(\text{vec } \hat{R}_{12}(m)\right)' \left(\hat{R}_{22}^{-1}(m) \otimes \hat{R}_{11}^{-1}(m)\right) \text{vec } \hat{R}_{12}(m), \tag{9}
\]

**Proof:** The determinant of the block matrix \(\hat{R}(m)\) defined in (7) is

\[
|\hat{R}(m)| = |\hat{R}_{11}(m)| \times |\hat{R}_{22}(m) - \hat{R}_{12}'(m)\hat{R}_{11}^{-1}(m)\hat{R}_{12}(m)|.
\]
Take the natural logarithm of the determinant and note that \( \log(|A|) = \text{tr}(\log(A)) \), where \( \text{tr}(A) \) denotes the trace of matrix \( A \), thus

\[
\log|\hat{R}(m)| = \log|\hat{R}_{11}(m)| + \text{tr} \left( \log \left( \hat{R}_{22}(m) - \hat{R}'_{12}(m)\hat{R}^{-1}_{11}(m)\hat{R}_{12}(m) \right) \right).
\]

Use a Taylor expansion [1] of \( \log(\hat{R}_{22}(m) - \hat{R}'_{12}(m)\hat{R}^{-1}_{11}(m)\hat{R}_{12}(m)) \) and note that

\[
-\text{tr} \left( \hat{R}_{22}(m) \right) < \text{tr} \left( \hat{R}'_{12}(m)\hat{R}^{-1}_{11}(m)\hat{R}_{12}(m) \right) < \text{tr} \left( \hat{R}_{22}(m) \right),
\]

this leads to

\[
\log|\hat{R}(m)| \approx \log|\hat{R}_{11}(m)| + \log|\hat{R}_{22}(m)| - \text{tr} \left( \hat{R}'_{12}(m)\hat{R}^{-1}_{11}(m)\hat{R}_{12}(m) \hat{R}_{22}^{-1}(m) \right). \tag{10}
\]

Recall that \( \text{tr}(Z'BZC) = (\text{vec } Z)'(C' \otimes B)(\text{vec } Z) \), where \( (\text{vec } Z) = [Z_1 : Z_2 : \cdots : Z_m]' \) is a column vector length \( m^2 \) formed by stacking the columns, \( Z_j, j = 1, 2, \ldots, m \), of the \( m \times m \) matrix \( Z \), and \( \otimes \) is the Kronecker product of matrices [38]. Thus (10) becomes

\[
\log|\hat{R}(m)| = \log|\hat{R}_{11}(m)| + \log|\hat{R}_{22}(m)| - \left( \text{vec } \hat{R}_{12}(m) \right)' \left( \hat{R}^{-1}_{22}(m) \otimes \hat{R}^{-1}_{11}(m) \right) \left( \text{vec } \hat{R}_{12}(m) \right). \tag{11}
\]

Multiply (11) by \(-n\), the ‘ into three components plus a constant term

\[
-n \log|\hat{R}(m)| = -n \log|\hat{R}_{11}(m)| - n \log|\hat{R}_{22}(m)| + n \left( \text{vec } \hat{R}_{12}(m) \right)' \left( \hat{R}^{-1}_{22}(m) \otimes \hat{R}^{-1}_{11}(m) \right) \left( \text{vec } \hat{R}_{12}(m) \right). \tag{12}
\]

Lemma 3.2: The third component in (9) can be expressed as

\[
n \left( \text{vec } \hat{R}_{12}(m) \right)' \left( \hat{R}^{-1}_{22}(m) \otimes \hat{R}^{-1}_{11}(m) \right) \left( \text{vec } \hat{R}_{12}(m) \right) = n \text{tr} \left( \hat{R}'_{12}(m)\hat{R}_{12}(m) \right).
\]

Proof: By \( \text{vec } (ABC) = (C' \otimes A')(\text{vec } B) \) and the fact \((\text{vec } Z)'(A' \otimes B)(\text{vec } Z) = \text{tr}(AZ'\ BZ)\), we note that \( \hat{R}_{12}(m) = D_{11}\hat{C}_{12}(m)D_{22} \), where \( D_{11} \) and \( D_{22} \) are, respectively, diagonal matrices with \( i \)th diagonal elements \((\hat{\gamma}'_{11}(0))^{-1/2} \) and \((\hat{\gamma}'_{22}(0))^{-1/2} \), and \( \hat{C}_{12}(m) \) is the cross-covariance matrix of the residuals and their squares with order \( m \). Now write the third component in (9) as

\[
n \left( \text{vec } \hat{R}_{12}(m) \right)' \left( \hat{R}^{-1}_{22}(m) \otimes \hat{R}^{-1}_{11}(m) \right) \left( \text{vec } \hat{R}_{12}(m) \right) = n \left( \text{vec } \hat{C}_{12}(m) \right)' \left( D_{22} \otimes D_{11} \right) \left( \hat{R}^{-1}_{22}(m) \otimes \hat{R}^{-1}_{11}(m) \right) \left( D_{22} \otimes D_{11} \right) \text{vec } \hat{C}_{12}(m)
\]
\[
\begin{align*}
&= n \left( \text{vec} \hat{C}_{12}(m) \right)' \left[ D_{22} \hat{R}_{22}^{-1}(m) D_{22} \otimes D_{11} \hat{R}_{11}^{-1}(m) D_{11} \right] \text{vec} \hat{C}_{12}(m) \\
&= n \left( \text{vec} \hat{C}_{12}(m) \right)' \left( \hat{C}_{12}^{-1}(m) \otimes \hat{C}_{11}^{-1}(m) \right) \text{vec} \hat{C}_{12}(m) \\
&= n \text{tr} \left( \hat{C}_{22}^{-1}(m) \hat{C}_{12}(m) \hat{C}_{11}^{-1}(m) \hat{C}_{12}(m) \right) \\
&= n \text{tr} \left( \hat{R}_{12}(m) \hat{R}_{12}(m) \right).
\end{align*}
\]

Substituting the results of Lemma 3.2 into (9) of Lemma 3.1 and multiply the results by the normalizing term $1/m$, the proposed a statistic is asymptotically given by

\[
C_m = -\frac{n}{m} \log|\hat{R}(m)| = -\frac{n}{m} \log|\hat{R}_{11}(m)| - \frac{n}{m} \log|\hat{R}_{22}(m)| + \frac{n}{m} \text{tr} \left( \hat{R}_{12}(m) \hat{R}_{12}(m) \right).
\] (12)

The proposed statistic has an interesting interpretation as it can be seen as an omnibus version of three existent tests. The first component is asymptotically equivalent to the statistic proposed in [40], $\hat{D}_{11}$, which can be used to test for linear autocorrelation in the residuals. The second is also asymptotically equivalent to the test in Peña and Rodríguez [40], $\hat{D}_{22}$, which can be used to test for the nonlinearity, or heteroskedasticity, models (uncorrelated but not independent). The third is asymptotically equivalent to a weighted variant [in the vein of 12] of the tests proposed in Psaradakis and Vávra [43], $Q_{12}$ and $Q_{21}$, which can be used to detect whether the cross correlations between the residuals and their squared values deviate from zero.

**Lemma 3.3:** The first and the second components in (12) are given by

\[
-\frac{n}{m} \log|\hat{R}_{ii}(m)| = -n \sum_{k=1}^{m} \frac{m + 1 - k}{m} \log(1 - \hat{\pi}_{i,k}^2), \quad i = 1, 2
\] (13)

where $\hat{\pi}_{1,k}$ is the $k$th partial autocorrelation of residuals and $\hat{\pi}_{2,k}$ is the $k$th partial autocorrelation of the squared residuals.

**Proof:** The proof follows those presented in Peña and Rodriguez [39,46].

**Lemma 3.4:** The third component in (12) can be written as the sum of squares of cross-correlation between residuals and their squares.

\[
\frac{n}{m} \text{tr} \left( \hat{R}_{12}(m) \hat{R}_{12}(m) \right) = \frac{n}{m} \sum_{k=-m, \ k\neq 0}^{m} (m + 1 - |k|) \hat{\rho}_{12}^2(k).
\] (14)
Proof:

\[
\frac{n}{m} \text{tr} \left( \hat{R}'_{12}(m) \hat{R}_{12}(m) \right) \\
= \frac{n}{m} \left( \text{vec} \hat{R}_{12}(m) \right)' \text{vec} \hat{R}_{12}(m) \\
= \frac{n}{m} \left[ (m+1)(0) + m[\hat{\rho}_{12}^2(-1) + \hat{\rho}_{12}^2(1)] \\
+ (m-1)[\hat{\rho}_{12}^2(-2) + \hat{\rho}_{12}^2(2)] + \cdots + [\hat{\rho}_{12}^2(-m) + \hat{\rho}_{12}^2(m)] \right] \\
= \frac{n}{m} \sum_{k=-m, \ k \neq 0}^m (m+1-|k|) \hat{\rho}_{12}^2(k). \quad \blacksquare
\]

From the previous Lemmas, we rewrite the proposed test statistic as a linear combination of tests based on the partial autocorrelations of the residuals, partial autocorrelations of squared residuals, cross-correlation between the residuals and their squares at positive and negative lags as follows

\[
-\frac{n}{m} \log |\hat{R}(m)| = -\frac{n}{m} \sum_{k=1}^m (m+1-k) \log(1 - \hat{\rho}_{1,k}^2) \\
-\frac{n}{m} \sum_{k=1}^m (m+1-k) \log(1 - \hat{\rho}_{2,k}^2) \\
+ \frac{n}{m} \sum_{k=-m, \ k \neq 0}^m (m+1-|k|) \hat{\rho}_{12}^2(k). \quad (15)
\]

The marginal distribution of the first two components in (15) can be found in Peña and Rodriguez [39]. Based on an application of the delta-method [see 37,39], the distribution of the partial autocorrelations is asymptotically equivalent to the autocorrelation and the first two components are asymptotically equivalent to the results in Fisher and Gallagher [12]. For the third component, one can apply the methodology of [45] on the theoretical results of Psaradakis and Vávra [43] to find a similar marginal distribution. The asymptotic distribution of \( C_m \) depends on the joint distribution of all terms \( \hat{\rho}_{ij}(k) \) for \( i, j = 1, 2 \) and \( k = 1, \ldots, m \).

To derive the distribution of \( C_m \) first note the equation in (15) is asymptotically equivalent to a quadratic form \( n\hat{r}'_m W\hat{r}_m \) where \( \hat{r}_m \) is a 4\( m \) vector comprised of autocorrelation and cross-correlation terms of residuals and their squares, and \( W \) is an appropriate diagonal matrix with elements corresponding to the weights associated with the components in (15), given by \( (m, m-1, \ldots, 1, m, m-1, \ldots, 1, m, m-1, \ldots, 1, m, m-1, \ldots, 1)/m \).

The distribution of statistics of the form \( n\hat{r}'_m W\hat{r}_m \) is well understood when \( \hat{r}_m \) is asymptotically normal, see [3,47,48]. Thus, determining the distribution of \( C_m \) is equivalent to determining the joint distribution of a vector mixed with autocorrelations of residuals, autocorrelation of squared residuals and cross correlation of residuals and their squares. This problem is essentially an extension of Wong and Ling [51] and was recently studied.
in Mahdi [33]. In a general setting, knowledge of the joint third and fourth cumulants is necessary [see 5] but in the case of iid Normal innovations, \( \sqrt{n}\hat{r}_m \) will be asymptotically normally distributed as follows.

**Theorem 3.5:** If the assumptions in (1) hold and the underlying stochastic process is normally distributed, then for any fixed integer \( m < n \) the asymptotic distribution of

\[
\sqrt{n}\hat{r}_m = \sqrt{n}\left(\hat{\rho}_{11}(1), \ldots, \hat{\rho}_{11}(m), \hat{\rho}_{22}(1), \ldots, \hat{\rho}_{22}(m), \hat{\rho}_{12}(1), \ldots, \hat{\rho}_{12}(m), \hat{\rho}_{12}(-1), \ldots, \hat{\rho}_{12}(-m)\right)'
\]

is \( \mathcal{N}(0, \Sigma) \) where \( \Sigma \) is an \((4m) \times (4m)\) covariance matrix of the form

\[
\Sigma = \begin{bmatrix}
I_m - Q & 0 & 0 & 0 \\
0 & I_m & 0 & 0 \\
0 & 0 & I_m & 0 \\
0 & 0 & 0 & I_m
\end{bmatrix}
\]

where \( 0 \) is \( m \times m \) zero matrix, \( Q = X_mV^{-1}X_m' \) is an idempotent matrix with rank \( p + q \), \( V \) is the information matrix for the parameters \( \beta \) and \( X_m \) is an \( m \times (p + q) \) matrix, with coefficients \( \phi'_i \) and \( \theta'_i \) defined by \( 1/\phi(B) = \sum_{i=0}^{\infty} \phi'_i B^i \) and \( 1/\theta(B) = \sum_{i=0}^{\infty} \theta'_i B^i \) as defined in [5, pp. 296–304].

**Proof:** The proof may be established by straightforward calculation following similar arguments to that in Wong and Ling [51] where the results of [4,36] are combined, but now include the results of Psaradakis and Vávra [43].

**Theorem 3.6:** If the assumptions in Theorem 3.5 hold, then the asymptotic distribution of the proposed statistic is

\[
C_m = -\frac{n}{m} \log|\hat{R}(m)| \xrightarrow{D} \sum_{i=1}^{4m} \lambda_i \chi_{1,i}^2
\]

where \( \chi_{1,i}^2 \) are independent \( \chi^2 \) random variables and \( \lambda_i \) are the eigenvalues of \( \Sigma W \), with \( \Sigma \) defined in (16) and \( W \) is a diagonal matrix with elements \((m,m-1, \ldots,1,m,m-1, \ldots,1,m,m-1, \ldots,1)/m\).

**Proof:** This results follows from the asymptotic normality in Theorem 3.5 along with the results in Box [3,39].

**Corollary 3.7:** Under the assumptions of Theorem 3.5, the asymptotic distribution of \( C_m \) can be approximated by a gamma distribution, \( \Gamma(\alpha, \beta) \), where

\[
\alpha = \frac{3m[2(m+1) - (p+q)]^2}{4(m+1)(2m+1) - 6m(p+q)},
\]
and
\[ \beta = \frac{4(m+1)(2m+1) - 6(m+p+q)}{3m[2(m+1) - (p+q)]}, \]
where the distribution has a mean of \( \alpha \beta = 2(m+1) - (p+q) \) and a variance of \( \alpha \beta^2 = (4(m+1)(2m+1) - 6(m+p+q))/3m \).

**Proof:** Note that upper percentiles of the cumulative distribution of the form \( \sum \lambda_i \chi_{1,i}^2 \) can be approximated as \( a \chi_2^2 \), where the parameters \( a \) and \( c \) can be selected so that the mean and variance equal to those of exact distribution of \( C_m \) [see 3,47,48]. Through cumulant matching arguments similar to [12,39], for large \( m \) one can show that \( a \chi_2^2 \) is equivalent to a gamma distribution with shape and scale parameters
\[ \alpha = K_1^2/K_2, \quad \text{and} \quad \beta = K_2/K_1, \]
where
\[ K_1 = \sum \lambda_i = \text{tr}(\Sigma W) = 2(m+1) - (p+q), \]
and
\[ K_2 = 2 \sum \lambda_i^2 = 2\text{tr}(\Sigma W)^2 = \frac{4(m+1)(2m+1)}{3m} - 2(p+q). \]
From here, the result follows.

### 3.2. Bootstrapping algorithm

The asymptotic results presented in Section 3.1 rely on iid normality of the innovations. In the modern practice of time series, violation of this assumption is common (consider the countless empirical examples where the series exhibits skewness and large tails). In these scenarios, the distribution of \( C_m \) is more complex. Specifically, \( r_m \) will asymptotically be normally distributed but a calculation of the covariance matrix requires derivation of the joint third and fourth cumulants of terms in \( r_m \) [see Ch. 7.2 in 5]. To alleviate this complexity we adapt the RWB method proposed in Zhu [53], and recently used in [54], for use with our statistic. Specifically,

1. Estimate the model from (1) using least squares. From the residuals, compute \( \hat{\rho}_{ij}(k) \) for \( i,j = 1,2 \) and \( k = 1,\ldots,m \) and store in a vector \( \hat{r}_m \).
2. Generate a sequence of iid random variables \( w^* = \{w_1^*, w_2^*, \ldots, w_m^*\} \) independent of the data from a common distribution such that \( P(w_i^* \geq 0) = 1 \) with mean and variance both equal to 1 (we use the standard Exponential distribution), and estimate the model (1) using weighted least squares with weights \( w^* \) and compute the residuals, \( e_i^* \).
3. Calculate \( \delta = W \times \{\sqrt{n}(\hat{r}_m - \hat{r}_m)\} \), where \( \hat{r}_m \) is a length 4\( m \) vector with terms
\[ \hat{\rho}_{ij}(k)^* = \frac{\sum_{t=k+1}^{n} w_t^* ((e_t^*)^i - E[(e_t^*)^i]) (e_{t-k}^*)^j - E[(e_t^*)^j])}{\sqrt{\sum_{t=1}^{n} ((e_t^*)^i - E[(e_t^*)^i])^2} \sqrt{\sum_{t=1}^{n} ((e_t^*)^j - E[(e_t^*)^j])^2}} \]
with \( E[(e_t^*)^1] = 0 \) and \( E[(e_t^*)^2] = 1 \), and the matrix \( W \) is defined above.
(4) Repeat steps 2 and 3 a large number of times, $B$ (typically $B \geq 500$), to obtain $\{\delta(1), \ldots, \delta(B)\}$, and compute its covariance matrix and its associated eigenvalues $\hat{\lambda}_i^*$, for $i = 1, \ldots, 4m$.

(5) Generate $N$ iid random samples $\{z_{ij}^{(1)}, \ldots, z_{ij}^{(4m)}\}_{j=1}^N$, where $N$ is a large number (say $N = 1000$), from a multivariate normal distribution with covariance $I_{4m}$ and compute the sequence $\{K^{(j)}\}_{j=1}^N$ by

$$K^{(j)} = \sum_{i=1}^{4m} \hat{\lambda}_i^* (z_i^{(j)})^2$$

(6) The sequence $\{K^{(j)}\}_{j=1}^N$ constitutes a bootstrapped sampling distribution for our proposed statistic $C_m$. Using the sample quantiles of $\{K^{(j)}\}_{j=1}^N$ we can determine critical values or we can approximate a $p$-value for $C_m$ by calculating $(\#(K^{(j)} > C_m) + 1)/(N + 1)$.

The above algorithm is a logical extension of that proposed in Zhu [53], and a special case of that in Zhu et al. [54]. The key to the algorithm is steps 3 and 4 where the covariance matrix of the vector $r_m$ is approximated.

Unlike other bootstrapping methods the RWB approach does not require the practitioner to select a block length or similar parameters. [53] shows that the algorithm does not appear to be sensitive to the distribution of the weights (we found the standard exponential works fairly well). Two parameters must be specified in the algorithm and are largely dependent on the computational resources available. In our simulations we use $B = 2,000$ and $N = 10,000$.

Remark 3.1: The above algorithm can be modified for the other statistics discussed in this article, including $Q_{11}$, $Q_{22}$, $Q_{12}$, and $Q_{21}$ by only working with specific auto/cross-correlation terms in step 3.

4. Computational study

We conduct a simulation study to investigate the appropriateness of the asymptotic distribution of the proposed test for different sample sizes and to compare its performance to the methods from the literature. We also study the effects of skewness and excess kurtosis on the proposed method and demonstrate the bootstrapping algorithm in Section 3.2 provides satisfactory results in approximating the distribution. Portmanteau statistics are known to be sensitive to the maximum lag, $m$, considered [see 14, for a discussion]. For brevity, we limit our study to maximum lags $m = 5$ and $m = 10$.

We focus our attention on testing for the adequacy of a fitted ARMA models. That is, our simulations consider the case of an underfit ARMA model as well as the detection of nonlinear effects (e.g., GARCH-type structures, or others) in the residual series. We compare the proposed statistic, $C_m$, to that of $Q_{11}$ [30], $Q_{22}$ [36], and the two tests $Q_{12}$ and $Q_{21}$ in Psaradakis and Vávra [43]. The primary goals of our simulations are: to show that an omnibus, or mixed, test comprised of autocorrelations of residuals, their squares, and cross-correlation of the residuals and their squares, can gain in detection power of nonlinear models; and to show that the asymmetric structure of $C_m$ (where lag 1 terms appear $m$
Table 1. Empirical sizes at nominal rate of 5% of \( C_m, Q^{**}, Q_{11}, Q_{22}, Q_{12}, \) and \( Q_{21} \) under a Gaussian ar(1) model with \( \phi = 0.8 \) and Gaussian ar(2) with \( \phi_1 = 0.8, \phi_2 = -0.3 \) at different sample sizes and maximum lags \( m \).

| \( m \) | \( n = 250 \) | \( n = 500 \) | \( n = 1000 \) | \( n = 250 \) | \( n = 500 \) | \( n = 1000 \) |
|------|--------|--------|--------|--------|--------|--------|
| 5    | 4.8    | 4.6    | 6.6    | 5.5    | 6.1    | 5.8    |
|      | 5.1    | 5.5    | 5.4    | 5.3    | 5.0    | 4.5    |
|      | 4.2    | 4.7    | 6.1    | 4.0    | 4.5    | 4.4    |
|      | 4.1    | 3.8    | 5.8    | 4.0    | 5.8    | 4.1    |
|      | 5.3    | 4.7    | 4.7    | 4.1    | 4.7    | 4.6    |
| 10   | 5.6    | 5.8    | 5.6    | 5.7    | 5.5    | 5.5    |
|      | 5.3    | 5.7    | 5.3    | 5.5    | 5.0    | 6.2    |
|      | 4.3    | 5.5    | 4.7    | 4.8    | 6.1    | 5.0    |
|      | 5.3    | 4.8    | 4.1    | 3.8    | 4.7    | 4.9    |

Note that the statistic \( Q^{**} \) is a combination of the \( Q_{11}, Q_{22}, Q_{12}, \) and \( Q_{21} \), and following Theorem 3.5 will be approximately \( \chi^2 \) distributed with \( 4m - (p + q) \) degrees of freedom under the null hypothesis of (1) and normal innovations. The bootstrapping algorithm in Section 3.2 can be utilized in the cases of non-normality where the \( W \) matrix in step 3 is replaced by a diagonal matrix with the Ljung-Box correction terms; i.e., \( (n + 2)/(n - k) \) for \( k = 1, \ldots, m \).

All numerical studies were conducted using the R software [44] in a parallel framework with the rugarch package [15] for data generation. This allows us to generate data with different nonlinear structures and under some fairly general distribution assumptions. Source code is available in the supplementary material.

### 4.1. Studies on empirical size

First we evaluate the empirical type I error rates of the proposed statistic, \( C_m \), along with the others we will consider here, \( Q_{11}, Q_{22}, Q_{12}, Q_{21} \) and \( Q^{**} \), by calculating the rejection rate of the tests out of 1,000 replications under the null hypothesis. In Table 1 we report the empirical size when the correct model is fit to a series of different sample sizes, \( n = 250, 500 \) and 1000, generated by Gaussian AR(1) with parameter \( \phi = 0.8 \) and AR(2) with \( \phi_1 = 0.8 \) and \( \phi_2 = -0.3 \) processes. In Table 1, generally we see all test report type I error rates within the acceptable range (3.7% to 6.3% based on Wald constructed 95% acceptance regions). Only in a few cases do we see rejection rates outside the acceptable range.

The presented theoretical findings are based on the assumption of an underlying Gaussian process as the distribution of \( C_m \) (and \( Q^{**} \)) in a more general setting is more complicated. Table 2 displays the empirical size at the nominal rate of 5% when utilizing the asymptotic distribution under the same AR(1) process where the innovation is not normally distributed. On the left side we report the type I error rates when the innovations are generated from the Skewed Normal distribution such that the skewness is approximately 0.56 (this corresponds to skewness parameter 1.5 in the \texttt{rdist} function in the rugarch
Table 2. Empirical sizes at nominal rate of 5% of $C_m$, $Q_{**}$, $Q_{11}$, $Q_{22}$, $Q_{12}$, and $Q_{21}$ under an $ar(1)$ process with $\phi = 0.8$ when the innovations are generated from a Skewed Normal or Students’ $t$ distribution, at different sample sizes and maximum lags $m$.

| $m$ | $n$ | Skewed Normal Innovations | | Students’ $t$ Innovations | |
|-----|-----|---------------------------|--------|--------------------------|--------|
|     |     | $C_m$ | $Q_{**}$ | $Q_{11}$ | $Q_{22}$ | $Q_{12}$ | $Q_{21}$ | $C_m$ | $Q_{**}$ | $Q_{11}$ | $Q_{22}$ | $Q_{12}$ | $Q_{21}$ |
| 5   | 250 | 6.9  | 6.5      | 5.6      | 4.7      | 2.8      | 3.9      | 6.6  | 6.2      | 5.4      | 5.9      | 4.6      | 4.9      |
| 5   | 500 | 6.9  | 6.6      | 4.7      | 3.8      | 5.0      | 5.6      | 6.7  | 6.1      | 4.4      | 5.7      | 5.6      | 4.8      |
| 5   | 1000 | 6.0  | 5.7      | 6.1      | 4.9      | 3.8      | 4.1      | 5.7  | 5.0      | 3.9      | 5.3      | 5.5      | 3.9      |
| 10  | 250 | 7.6  | 7.0      | 4.1      | 6.1      | 4.1      | 4.8      | 6.8  | 5.7      | 5.2      | 5.8      | 5.1      | 4.5      |
| 10  | 500 | 8.0  | 6.6      | 5.3      | 4.7      | 5.2      | 4.9      | 6.0  | 5.4      | 4.1      | 4.9      | 5.0      | 3.8      |
| 10  | 1000 | 7.1  | 5.9      | 4.3      | 5.2      | 4.6      | 4.3      | 6.2  | 6.4      | 4.9      | 5.4      | 4.2      | 5.4      |

Table 3. Empirical sizes at nominal rate of 5% of $C_m$, $Q_{**}$, $Q_{11}$, $Q_{22}$, $Q_{12}$, and $Q_{21}$ under an $ar(1)$ model with $\phi = 0.8$ and at different sample sizes and maximum lags $m$ when the innovations are generated from a Skewed Students’ $t$ distribution. Results with the asymptotic distribution and the randomly weighted bootstrap algorithm are presented.

| $m$ | $n$ | Based on asymptotic distribution | | Based on RWB algorithm | |
|-----|-----|---------------------------------|--------|------------------------|--------|
|     |     | $C_m$ | $Q_{**}$ | $Q_{11}$ | $Q_{22}$ | $Q_{12}$ | $Q_{21}$ | $C_m$ | $Q_{**}$ | $Q_{11}$ | $Q_{22}$ | $Q_{12}$ | $Q_{21}$ |
| 5   | 250 | 7.2  | 6.7      | 6.0      | 4.6      | 5.2      | 3.8      | 2.7  | 1.8      | 5.5      | 2.4      | 3.8      | 3.0      |
| 5   | 500 | 7.8  | 7.1      | 4.7      | 5.3      | 3.6      | 5.0      | 2.7  | 1.7      | 4.5      | 3.7      | 3.6      | 1.5      |
| 5   | 1000 | 7.4  | 8.0      | 5.0      | 5.7      | 4.0      | 4.5      | 2.9  | 2.6      | 4.8      | 3.1      | 4.1      | 2.9      |
| 10  | 250 | 8.9  | 9.3      | 6.1      | 5.0      | 4.1      | 5.2      | 1.9  | 1.2      | 4.9      | 1.7      | 3.1      | 1.9      |
| 10  | 500 | 8.0  | 7.0      | 7.0      | 6.0      | 4.1      | 4.0      | 1.6  | 1.3      | 6.4      | 1.4      | 3.1      | 1.4      |
| 10  | 1000 | 8.5  | 8.3      | 4.7      | 5.7      | 5.3      | 3.9      | 1.4  | 1.9      | 4.2      | 2.4      | 4.2      | 1.6      |

package). On the right side the innovations are from the Students’ $t$ distribution where the excess kurtosis is 1 (shape parameter 10 in the rdist function).

Table 2 shows the proposed statistic $C_m$ and the combination statistic $Q_{**}$, which use a combination of autocorrelations of residuals, squared residuals and the cross-correlation of the residuals and their squares, begins to report inflated type I error rates in the presence of skewness. This phenomen appears more problematic at the larger maximum lag of $m = 10$. In the case of heavy tails, we only see a moderate increase in type I error rates.

In Table 3 we consider the robustness of the statistics when the RWB algorithm is utilized. Data is generated from the same $ar(1)$ process above with innovations from the Skewed Students’ $t$ distribution such that the skewness is approximately 0.85 and the excess kurtosis is 1.73, thus the underlying innovations come from a distribution with both heavy-tails and skewness. The table reports the empirical rejection rates at a nominal rate of 5% when the asymptotic distribution is utilized and when the RWB algorithm in Section 3.2 is used. For comparison, we also include the RWB-based type I error rates for the statistics $Q_{11}$, $Q_{22}$, $Q_{12}$ and $Q_{21}$.

Table 3 shows that the proposed statistic and $Q_{**}$, both which are asymptotically equivalent to linear combinations of $Q_{11}$, $Q_{22}$, $Q_{12}$ and $Q_{21}$, report inflated type I error rates when utilizing the asymptotic distribution. This is due to the joint third and fourth cumulants. As before, we also note the increased type I error rate appears larger for the larger maximum lag of $m = 10$. However, when utilizing the RWB algorithm to approximate the distribution, the type I errors do not exceed the nominal level for the proposed statistic. Overall we see most statistics report conservative type I error rates, which is generally preferred in practice compared to inflated type I errors.
Table 4. Empirical power at nominal rate of 1% of $C_m, Q_{ss}, Q_{11}, Q_{22}, Q_{12}, \text{and } Q_{21}$ under two alternatives, a Gaussian arma(1,1) model with $\phi_1 = 0.8$ and $\theta_1 = 0.3$ and a Gaussian ar(1)+arch(1) with $\phi_1 = 0.8, \omega = 1$ and $\alpha_1 = 0.4$ where the process is underfit with an ar(1) at different sample sizes and maximum lags $m$.

| m   | n  | $C_m$ | $Q_{ss}$ | $Q_{11}$ | $Q_{22}$ | $Q_{12}$ | $Q_{21}$ |
|-----|----|-------|----------|----------|----------|----------|----------|
| 5   | 250| 63.0  | 34.0     | 73.1     | 2.4      | 2.1      | 1.5      |
|     | 500| 97.1  | 83.5     | 98.6     | 4.7      | 1.1      | 1.3      |
|     | 1000| 100.0 | 99.9     | 100.0    | 7.1      | 1.2      | 1.3      |
| 10  | 250| 51.4  | 24.2     | 57.8     | 2.1      | 1.8      | 1.8      |
|     | 500| 93.9  | 67.2     | 96.2     | 2.9      | 1.5      | 1.2      |
|     | 1000| 100.0 | 99.4     | 100.0    | 5.4      | 1.5      | 1.4      |

4.2. Power studies

We now consider the empirical power of the proposed method and compare it to some of the statistics in the literature. The first two simulation scenarios are structured to demonstrate the proposed statistic $C_m$ provides comparable power to one of the statistics that should be more powerful. Specifically, we generate data from a Gaussian arma(1,1) model with $\phi_1 = 0.8$ and $\theta_1 = 0.3$ but only an ar(1) model is fit; thus we’ve intentionally underfit the autocorrelation in the series. We would expect temporal correlation in the residuals, thus the Ljung-Box test $Q_{11}$ should detect the underfit. In the second scenario, the data follows a Gaussian ar(1)+arch(1) process with $\phi_1 = 0.8, \alpha_1 = 0.4$ and $\omega = 1$ but only an ar(1) is fit. Here, we would expect substantial temporal correlation in the squares of the residuals and for the McLeod-Li test $Q_{22}$ to be quite powerful. We highlight the most powerful statistic in **boldface**.

Table 4 reports the empirical power at the 1% significance level. For the underfit arma(1,1) process we see that the Ljung-Box statistic is most powerful with the proposed statistic $C_m$ providing comparable power. Given in this scenario all temporal structure should be present in the residuals, and since the proposed $C_m$ is comprised of terms involving the residuals and their squares, it is not overly surprising the traditional Ljung-Box test is more powerful. In the second scenario we see either the proposed $C_m$ or McLeod-Li $Q_{22}$ is most powerful. We also note that the cross-correlations of the residuals and their squares (the components of $Q_{12}$ and $Q_{21}$) provide some detection (although not particularly strong) of the underlying arch process. This partially explains why the proposed method is most powerful in a few scenarios – it combines elements of $Q_{22}$ with $Q_{12}$ and $Q_{21}$. Lastly, we note that the weighted statistic $C_m$ is more powerful than its non-weighted counterpart $Q_{ss}$ in all scenarios and that all statistics demonstrate a reduction in power as the lag increases. This phenomenon is well-known in the literature [see 14, for example] but note that the decrease in power appears less for the proposed statistic, $C_m$.

The next study considers detecting some nonlinear processes studied in the literature. Data is generated from the following five processes:

- **M1**, $z_t = \varepsilon_t - 0.3\varepsilon_{t-1} + 0.2\varepsilon_{t-2} + 0.4\varepsilon_t\varepsilon_{t-2} - 0.25\varepsilon_{t-2}^2$,
- **M2**, $z_t = 0.4z_{t-1} - 0.3z_{t-2} + 0.5z_{t-1}\varepsilon_{t-1} + \varepsilon_t$,
- **M3**, $z_t = 0.4z_{t-1} - 0.3z_{t-2} + 0.5z_{t-1}\varepsilon_{t-1} + 0.8\varepsilon_{t-1} + \varepsilon_t$.
Table 5. Empirical power at nominal rate of 5% of $C_m, Q_{**}, Q_{11}, Q_{22}, Q_{12}$, and $Q_{21}$ under various non-linear alternatives where the process is fit with an $ar(p)$ with $p$ selected based on AIC, at different sample sizes and maximum lags $m$.

| Model | Lag $m = 5$ | Lag $m = 10$ |
|-------|-------------|-------------|
|       | $C_m$   | $Q_{**}$ | $Q_{11}$ | $Q_{22}$ | $Q_{12}$ | $Q_{21}$ | $C_m$ | $Q_{**}$ | $Q_{11}$ | $Q_{22}$ | $Q_{12}$ | $Q_{21}$ |
| M1    | 76.2    | 64.1    | 0.3     | 29.3    | 47.1    | 84.6    | 72.5   | 48.5    | 0.2     | 22.5    | 37.7    | 73.1    |
| M2    | 84.6    | 76.9    | 0.6     | 74.7    | 36.1    | 74.8    | 85.0   | 72.6    | 1.1     | 73.1    | 32.0    | 70.1    |
| M3    | 89.4    | 84.8    | 0.9     | 79.4    | 13.3    | 81.8    | 88.5   | 80.2    | 0.7     | 76.8    | 11.9    | 88.4    |
| M4    | 79.5    | 57.6    | 0.0     | 57.2    | 55.0    | 30.1    | 79.1   | 51.5    | 0.1     | 55.0    | 49.9    | 24.9    |
| M5    | 30.0    | 20.2    | 0.6     | 20.8    | 11.1    | 22.7    | 27.6   | 11.9    | 0.8     | 16.6    | 8.2     | 15.2    |

where $\varepsilon_t$’s are a sequence of independent and identically distributed innovations. The first three models are analyzed by [21] (see also [39,40]), whereas the other models are studied in Psaradakis and Vávra [43]. The nonlinear process is then fit with an $ar(p)$ where $p$ is selected using the Akaike Information Criterion (AIC) and statistics are calculated based on the estimated residuals. To study the robustness of the statistics we generate the innovations from a Skewed Student’s $t$-distribution and utilize the RWB algorithm for all the statistics.

Table 5 reports the empirical power of the statistics using the RWB algorithm at the nominal level of 5%. The results demonstrate the proposed statistics is comparable, or more powerful, than those studied from the literature. In particular, we draw attention to the results for models M2, M4 and M5. There, each of $Q_{22}$, $Q_{12}$ and $Q_{21}$ provides substantial power but $C_m$, which uses information from all three of those statistics, provides a notable increase in power. We also see the expected behavior of an increase in power as $n$ increases and a general decrease in power as the lag $m$ increases.

Our last simulation considers the potential increase in power offered by our statistic, and the potential gains of using an omnibus statistic. Consider a modification of M4 above:

$$z_t = 0.5 - (0.4 - 0.4\varepsilon_{t-1})z_{t-1} + \varepsilon_t,$$

where $\delta$ can be considered a perturbation parameter that controls the amount of nonlinearity. When $\delta = 0$ we have a simple $ar(1)$ process, as $\delta$ increases the process becomes increasingly nonlinear.

Data is generated using model (19) for $n = 500$ with Skewed Student’s $t$-distributed innovations and values of $\delta$ ranging from 0 to 0.5. An $ar(p)$ is fit to the data, where $p$ is selected from AIC (generally we may expect AIC to select $p = 1$ to model the autoregressive part of the process). Figure 1 displays the empirical power of the statistics $C_m, Q_{**},$
Figure 1. Powers of $C_m$ (solid black line), $Q_{**}$ (dotted dashed gray line), $Q_{22}$ (solid gray line), $Q_{12}$ (dashed gray line) and $Q_{21}$ (dotted gray line) at lag $m = 5$ in detecting the nonlinear structure in (19) for a 5% test when data of size 500 are generated and an $ar(p)$ is fit to the data with $p$ selected from AIC.

$Q_{22}$ (second most powerful statistic in Table 5), $Q_{12}$ and $Q_{21}$. The distribution is approximated through the RWB algorithm and the empirical power is calculated based on 1,000 realizations at each $\delta$ value.

Figure 1 shows a very interesting power plot. At smaller $\delta$ values, we see that the statistic $Q_{12}$ is most powerful followed closely by the proposed statistics $C_m$. Yet, as $\delta$ increases the power of $Q_{12}$ starts to decrease while the statistics $Q_{22}$ and $Q_{21}$ gain power. The proposed statistic, $C_m$, which is asymptotically a convolution of the other statistics, effectively combines all the information and is most powerful starting around $\delta = 0.225$. The power of $Q_{**}$ follows the pattern of $C_m$ but is generally less powerful, perhaps suggesting most of the remaining correlation is at lower lags [see 14].

To understand the behavior of the statistics in Figure 1, we conduct some follow up simulations to study the nature of the autocorrelation and cross-correlation of the residuals and their squares from the above study. Three series of length $n = 5,000$ were simulated from model (19) (we chose a large $n$ to ensure some level of consistency in the estimation of the residual autocorrelation and cross-correlation functions), one each for $\delta = 0.15, 0.30$ and 0.45. Each of the three nonlinear series was fit with an $ar(p)$ where $p$ is chosen from AIC (in our simulations $p = 2, 1, 2$ were selected, respectively) and Figure 2 displays the sample autocorrelation of the residuals and their squares (on the diagonal), and the cross correlation of the residuals and their squares (on the off diagonal).

Figure 2 shows that for $\delta = 0.15$ the strongest residual correlations occur at $\hat{\epsilon}_{12}(1)$ and $\hat{\epsilon}_{12}(2)$ which correspond to components of the $Q_{12}$ statistic. There is also a meaningful correlation in the $\hat{\epsilon}_{22}(1)$ term (corresponding to the moderate power in the $Q_{22}$ near $\delta = 0.15$). At $\delta = 0.30$, there is meaningful correlation in $\hat{\epsilon}_{12}(k), \hat{\epsilon}_{21}(k)$, and $\hat{\epsilon}_{22}(k)$ terms up to $k = 3$, which helps demonstrate why all the statistics demonstrate some detection power, and that $C_m$ is most powerful. At $\delta = 0.45$ little correlation remains in the components of $Q_{12}$, which explains its decrease in power, while there is substantial correlation in the components of $Q_{22}$ (and $Q_{21}$ to a lesser extent). This corresponds closely to the results of Table 5 which shows $C_m$ as most powerful, followed by $Q_{22}$ and then $Q_{21}$ slightly more powerful than $Q_{12}$.
This example provides an interesting case study on the usefulness of an omnibus statistic, such as $C_m$. Without oracle type knowledge, a practitioner would be unable to rely on a single statistic such as $Q_{12}$, $Q_{21}$ or $Q_{22}$, to detect nonlinearity in the residual series. The proposed omnibus statistic $C_m$ can encapsulate all the relevant information, all while weighting the components in a way known to increase power as seen compared to the studied $Q_{**}$ statistic [see 14].

In conclusion, the simulations demonstrate the proposed statistic $C_m$ can attain good power (at least comparable to other methods, if not better) compared to many of the proposed statistics in the literature. The simulations demonstrate that by using all the information contained in the autocorrelations of the residuals, autocorrelation of the squared residuals and cross correlation of the residuals and their squares, one can attain more power in detecting nonlinear effects than any statistic based on just one measure, all while retaining adequate type I error rates and providing comparable power in detecting underfit linear effects.

5. Illustrative applications

We demonstrate the usefulness of the proposed test for detecting nonlinear processes in an economic series and some environmental data recently studied in the literature.

5.1. Crude oil prices

Consider a short study on the daily West Texas Intermediate (WTI) Crude Oil Prices [50], in U.S. dollars per barrel, from September 01, 2019 through July 20, 2022 obtained using the
tidyquant package [8]. This time frame encompasses all market days beginning roughly six months before the onset of the lock downs due to the SARS-CoV-2 pandemic and the economic turbulence that has occurred since, and results in a length \( n = 723 \) series. The changes in daily crude prices (VWTI) are seen in Figure 3 along with the normal QQ-plot of the daily returns showing the data is heavy-tailed.

Figure 3 shows changes in daily crude prices (VWTI) are reasonably stationary but that the distribution of changes exhibit heavy tails. The series exhibit a decaying autocorrelation function (not shown) and AIC suggest an \( \text{ar}(3) \) will model the linear dependency in the data. The fitted \( \text{ar}(3) \) model has parameters

\[
\phi_1 = -0.3144, \quad \phi_2 = -0.1443, \quad \phi_3 = -0.0733, \quad \sigma^2 = 10.44
\]

and the autocorrelation functions of the resulting residual series and squared residuals series can be seen in Figure 4 (note the off-diagonal terms are the cross correlations of the residuals and their squares). There we see no meaningful temporal correlation in the residuals (indicating an \( \text{ar}(3) \) adequately models the linear dependence structure). There appears to be a meaningful correlation in the \( \hat{\rho}_{12}(k) \) terms and measurable correlation in \( \hat{\rho}_{21}(1) \) and \( \hat{\rho}_{22}(1) \). Given the non-normality of the data we apply the RWB algorithm to compute the various test statistics at lags \( m = 5, m = 10 \) and \( m = 20 \) with the results provided in Table 6 (we use \( B = 10,000 \) and \( N = 10,000 \) in this application). Not surprisingly given Figure 4, we see the Ljung-Box test \( Q_{11} \) confirms we have adequately modeled the linear process. The tests of Pasaradakis-Vávra [43], \( Q_{12} \) and \( Q_{21} \), both reject at the smaller lags but provide differing results at the higher lag of \( m = 20 \), and the McLeod-Li [36] test, \( Q_{22} \) also detects nonlinearity. Both omnibus test, \( C_m \) and \( Q_{**} \), reject the null hypothesis that a linear model is adequate.

We note that the \( p \)-value of \( Q_{**} \) more than doubles when the lag increases from \( m = 5 \) to \( m = 20 \) while that of the \( C_m \) is relatively constant. In general, a practitioner must choose the lag \( m \) and many of the portmanteau test of the Ljung-Box form (including \( Q_{**} \)) are known to be sensitive to the lag [see 14]. To study these effects we compute the \( p \)-value of \( C_m \) and \( Q_{**} \) using the RWB algorithm at lags \( m = 4, \ldots, 40 \) and display them in Figure 5. There,
Figure 4. Correlograms of the residuals and their squares from the ar(3) model fit to the daily returns for the WTI Cude Oil Price series.

Figure 5. p-values of $C_m$ (solid black line) and $Q^{**}$ (dotted dashed gray line) at lags $m = 4, \ldots, 40$ demonstrating the relative stability of the asymmetrically weighted statistic $C_m$ compared to $Q^{**}$ when an ar(3) model is fit to the daily returns for the WTI Cude Oil Price series.

we see that the reported $p$-values of the asymmetrically weighted $C_m$ are relatively stable across all lags studied while the $Q^{**}$ has a noticeable increasing behavior demonstrating it is more sensitive to the chosen lag.

5.2. Air quality measurements

The Nitric Oxide measures (micrograms per cubic meter) at Marylebone Road and North Kensington air quality stations in London were collected from the Department for
Table 6. The statistics and associated $p$-values, based on the RWB algorithm, of the portmanteau tests when an $ar(3)$ is fit to the daily returns for the WTI series.

| Lag $m$ | Stat. RWB | Stat. RWB | Stat. RWB |
|---------|-----------|-----------|-----------|
|         | Stat.     | $p$-value | Stat.     | $p$-value | Stat.     | $p$-value |
| $m = 5$ | 170.02    | 0.008     | 176.26    | 0.009     | 196.19    | 0.010     |
| $Q_{11}$| 174.22    | 0.009     | 184.41    | 0.012     | 226.27    | 0.019     |
| $Q_{22}$| 1.56      | 0.617     | 3.80      | 0.886     | 25.74     | 0.309     |
| $Q_{22}$| 44.64     | 0.002     | 44.70     | 0.002     | 44.92     | 0.002     |
| $Q_{12}$| 106.26    | 0.022     | 110.54    | 0.023     | 116.14    | 0.022     |
| $Q_{21}$| 21.77     | 0.005     | 25.36     | 0.029     | 39.48     | 0.119     |

Figure 6. Standardized logarithm (base 10) daily mean concentrations of Nitric Oxide at the Marylebone Road and North Kensington air quality stations in London, UK from 2015 to 2019.

Environment, Food and Rural Affairs in the United Kingdom. These two series are a subset of a multivariate time series explored in Cirkovic and Fisher [7] and are available in their accompanying R package autocovarianceTesting.

Following the procedure in Cirkovic and Fisher [7] the two $n = 1461$ length marginal series are transformed by a logarithm and then standardized by both month and by weekday/weekend means and standard deviations to achieve stationary. The resulting series can be seen in Figure 6. The automatic lag selecting test procedure from [7] suggests the two series have equivalent autocovariance structures ($H_0$ : two series share a common autocovariance structure, $p$-value of 0.504). Thus, we may expect the two series to have similar linear dynamics and have a similar ARMA fit. In fact, using AIC to select the order, an AR(3) is suggested for both series as an appropriate model. The fitted AR model parameters are quite similar:

Marylebone Road : $\phi_1 = 0.4542$, $\phi_2 = 0.0237$, $\phi_3 = 0.0435$, $\sigma^2 = 0.7591$

North Kensington : $\phi_1 = 0.4252$, $\phi_2 = 0.0032$, $\phi_3 = 0.0503$, $\sigma^2 = 0.7953$

As a follow-up, consider testing the adequacy of the fitted AR(3) models for the two series by applying the proposed statistic and those studied from the literature. The two
Figure 7. Normal QQ-Plots of the standardized logarithm (base 10) daily mean concentrations of Nitric Oxide at the Marylebone Road and North Kensington Air Quality stations.

Table 7. The statistics and associated $p$-values, based on both the asymptotic distribution and RWB algorithm, of the portmanteau tests when an $ar(3)$ is fit to the two Nitric Oxide air quality datasets.

| Lag $m = 5$ | Marylebone Road | North Kensington |
|-------------|-----------------|-----------------|
| $C_m$       | $10^8$          | $10^5$          | $10^5$          |
| $Q_{m^*}$   | $10^{-28}$      | $10^{-5}$       | $10^{-5}$       |
| $Q_{11}$    | $10^{-19}$      | $10^{-5}$       | $10^{-5}$       |
| $Q_{22}$    | $10^{-11}$      | $10^{-4}$       | $10^{-4}$       |
| $Q_{12}$    | $10^{-12}$      | $10^{-5}$       | $10^{-5}$       |
| $Q_{21}$    | $10^{-11}$      | $10^{-4}$       | $10^{-4}$       |

| Lag $m = 10$ | Marylebone Road | North Kensington |
|-------------|-----------------|-----------------|
| $C_m$       | $10^{51}$       | $10^{-9}$       | $10^{-10}$      |
| $Q_{m^*}$   | $10^{-7}$       | $0.002$         | $0.002$         |
| $Q_{11}$    | $0.05$          | $0.988$         | $0.785$         |
| $Q_{22}$    | $0.001$         | $0.025$         | $0.026$         |
| $Q_{12}$    | $0.010$         | $0.022$         | $0.029$         |
| $Q_{21}$    | $10^{-4}$       | $0.002$         | $0.043$         |

series exhibit some minor deviations from normality (see Figure 7) so we utilize both the RWB algorithm to calculate the associated $p$-values of the test statistics as well as reporting the $p$-values based on the asymptotic distribution. Table 7 reports the test statistic values and associated $p$-values for the six studied test statistics at two maximum lags.

We see that the two $AR(3)$ models adequately model the linear relationship based on the Ljung-Box statistic, $Q_{11}$, but there is overwhelming evidence for the presence of nonlinear effects. In particular, the proposed statistic $C_m$ provides unquestionable evidence for the inadequacy of the $AR(3)$ model as does the statistic of McLeod-Li, $Q_{22}$ and Pasaradakis-Vávra, $Q_{21}$. To gain further insight, consider the correlogram plots of the two residual series and squared residuals in Figure 8. The upper-left correlogram is a plot of the autocorrelation of the two residual series. In agreement with the Ljung-Box statistic, there is no meaningful correlation present. However, we see a decaying autocorrelation feature in the
two bottom panels (corresponding to components of the $Q_{21}$ and $Q_{22}$ statistics). Contextually, Figure 8 appears to show that although the linear process of Marylebone Road and North Kensington Nitric Oxide series may be equivalent, the nonlinear processes differ (consider the bottom left panel of Figure 8).

5.3. Discussion on applications

We remind the reader the proposed statistic, $C_m$, is asymptotically equivalent to a convolution of $Q_{11}$, $Q_{22}$, $Q_{12}$ and $Q_{21}$, and that the statistic $Q_{**}$ is essentially the summation of the four. In the WTI series, the statistics $Q_{12}$ and $Q_{22}$ consistently suggest nonlinearity, however with the air quality series, $Q_{21}$ and $Q_{22}$ suggest nonlinearity in the residuals while $Q_{12}$ provides contradictory results. In both studies, the omnibus tests $C_m$ and $Q_{**}$ provide evidence for the presence of a nonlinear temporal structure. Since foretelling which statistic is preferred for a given dataset would require oracle type abilities, these examples demonstrate the utility of using a omnibus statistic.

Further, the construction of $C_m$ follows that of Peña and Rodríguez [39] and has similarities to that of Fisher and Gallagher [12]. Those test essentially weight the correlation at lower lags with more emphasis than those at higher lags, while the test $Q_{**}$ considers all lags equally. Likewise, the [43] tests, $Q_{12}$ and $Q_{21}$, and [36] test, $Q_{22}$, also equally weigh each lag. With the WTI data, Figure 5 demonstrates the consistency of $C_m$ regardless of lag compared to $Q_{**}$. For the air quality data analysis in Figure 8, it appears the strongest evidence of nonlinearity is at lags 1 and 2, and we see that that $C_m$ offers more evidence than either $Q_{21}$, $Q_{22}$ and $Q_{**}$, particularly at the larger lag (Table 7).
6. Discussion

The proposed test statistic has several interesting properties. It can be seen as a combination of four weighted tests. The first test is based on the partial autocorrelation of the residuals that can be used to test for linearity in time series models. The second is based on the partial autocorrelation of the squared residuals that can be used to test for nonlinearity. The third and the fourth tests are based on the cross-correlation between the residuals and their squares at negative and positive lags, respectively. Each term in the test is scaled by \((m - i + 1)/m\), which allows the lower-order autocorrelations and cross-correlations to receive more emphasis than the larger lag terms.

In contrast to some other portmanteau tests, the proposed test responds well to nonlinear models that do not have ARCH-type structures. In particular, the proposed test responds very well to time series where the residuals and their squares are cross correlated. Simulation results show the power of the proposed test is comparable, if not more powerful than, other nonlinear tests studied by [36,43].

Several possible extensions to this article can be pursued. One is to approximate the distribution, and \(p\)-value of the proposed statistic, based on a different bootstrapping, or Monte Carlo, method than the RWB method used here. For instance, Monte Carlo methods are suggested by Lin and McLeod [29] and Mahdi and Ian McLeod [34] to compute \(p\)-values of a portmanteau test statistic based on the determinant autocorrelation matrix, but these methods are computationally expensive as they require repeated calculated of the determinant of a \(2(m + 1) \times 2(m + 1)\) matrix required on the order of \(O(4(m + 1)^2)\) operations.

Another extension to this article could be done by deriving a new test based on extending the block matrix given by (7) to the other generalized-correlation terms. Lastly, generalizing the result for the use of multivariate time series, similar to [34,45] seems like a fairly straightforward calculation but may require very large samples to be tenable.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Esam Mahdi  http://orcid.org/0000-0002-3447-0639

Thomas J. Fisher  http://orcid.org/0000-0001-5885-7646

References

[1] R. Bhatia Matrix Analysis: Graduate Texts in Mathematics, 169. Springer, New York, 1997.
[2] T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, J. Econom. 31 (1986), pp. 307–327. Available at https://www.sciencedirect.com/science/article/pii/0304407686900631.
[3] G.E.P. Box, Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification, Ann. Math. Stat. 25 (1954), pp. 290–302.
[4] G.E.P. Box and D.A. Pierce, Distribution of residual autocorrelations in autoregressive-integrated moving average time series models, J. Am. Stat. Assoc. 65 (1970), pp. 1509–1526. Available at http://www.jstor.org/stable/2284333.
[5] P.J. Brockwell and R.A. Davis, Time Series: Theory and Methods, 2nd ed. Springer-Verlag, New York, 1991.
[6] R. Carmona, Statistical Analysis of Financial Data in R, 2nd ed. Springer, New York, 2014.
[7] D. Cirkovic and T.J. Fisher, On testing for the equality of autocovariance in time series, Environmentics 32 (2021), p. e2680.
[8] M. Dancho and D. Vaughan, tidyquant: Tidy Quantitative Financial Analysis (2022). Available at https://CRAN.R-project.org/package=tidyquant, R package version 1.0.4.
[9] B. Efron, Bootstrap methods: another look at the jackknife, Ann. Stat. 7 (1979), pp. 1–26. DOI:10.1214/aos/1176344552
[10] B. Efron and R. Tibshirani, An Introduction to the Bootstrap, Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Taylor & Francis, 1994. Available at https://books.google.com/books?id=gLlpIUxRntoC.
[11] R.F. Engle, Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation, Econometrica 50 (1982), pp. 987–1007.
[12] T.J. Fisher and C.M. Gallagher, New weighted portmanteau statistics for time series goodness of fit testing, J. Am. Stat. Assoc. 107 (2012), pp. 777–787. DOI:10.1080/01621459.2012.688465
[13] D.A. Freedman, Bootstrapping regression models, Ann. Stat. 9 (1981), pp. 1218–1228. DOI:10.1214/aos/1176345638
[14] C.M. Gallagher and T.J. Fisher, On weighted portmanteau tests for time series goodness-of-fit, J. Time Ser. Anal 36 (2015), pp. 67–83.
[15] A. Ghalanos, rugarch: Univariate GARCH Models. R Package Version 1.4-4 (2020).
[16] P. Hall, J.L. Horowitz, and B.Y. Jing, On blocking rules for the bootstrap with dependent data, Biometrika 82 (1995), pp. 561–574. DOI:10.1093/biomet/82.3.561
[17] W. Härdle, J. Horowitz, and J.P. Kreiss, Bootstrap methods for time series, Int. Stat. Rev. / Revue Internationale De Statistique 71 (2003), pp. 435–459. Available at http://www.jstor.org/stable/1403897.
[18] J.H. Kim, Wild bootstrapping variance ratio tests, Econ. Lett. 92 (2006), pp. 38–43. Available at https://www.sciencedirect.com/science/article/pii/S0165176506000140.
[19] A.J. Lawrence, Higher-order residual analysis for nonlinear time series with autoregressive correlation structures, Int. Stat. Rev. / Revue Internationale De Statistique 55 (1987), pp. 21–35. Available at https://www.sciencedirect.com/science/article/pii/S0165176506000876.
[20] T. Lee, Block wild bootstrap-based cusum tests robust to high persistence and misspecification, Comput. Stat. Data Anal. 150 (2020). Article Id 106996, Available at https://www.sciencedirect.com/science/article/pii/S0167947320300876.
[21] W.K. Li and T.K. Mak, On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity, J. Time Ser. Anal. 15 (1994), pp. 627–636. DOI:10.1111/j.1467-9892.1994.tb00217.x
[22] J.W. Lin and A. McLeod, Improved peña-rodríguez portmanteau test, Comput. Stat. Data Anal. 51 (2006), pp. 1731–1738. Available at https://www.sciencedirect.com/science/article/pii/S0167947306002003.
[30] G.M. Ljung and G.E.P. Box, *On a measure of lack of fit in time series models*, Biometrika 65 (1978), pp. 297–303. Available at http://www.jstor.org/stable/2335207.

[31] J.G. MacKinnon, *Bootstrap methods in econometrics*, Econ. Rec. 82 (2006), pp. S2–S18. DOI:10.1111/j.1467-9892.2006.00328.x

[32] J.G. MacKinnon, M. Ørregaard Nielsen, and M.D. Webb, *Wild bootstrap and asymptotic inference with multiway clustering*, J. Bus. Econ. Stat. 39 (2021), pp. 505–519.

[33] E. Mahdi, *New goodness-of-fit tests for time series models* (2021)

[34] E. Mahdi and A. Ian McLeod, *Improved multivariate portmanteau test*, J. Time Ser. Anal. 33 (2012), pp. 211–222. DOI:10.1111/j.1467-9892.2011.00752.x

[35] E. Mammen, *Bootstrap and wild bootstrap for high dimensional linear models*, Ann. Stat. 21 (1993), pp. 255–285. DOI:10.1214/aos/1176349025

[36] A.I. McLeod and W.K. Li, *Distribution of the residual autocorrelation in multivariate arma time series models*, J. Time Ser. Anal. 4 (1983), pp. 269–273.

[37] A.C. Monti, *A proposal for residual autocorrelation test in linear models*, Biometrika 81 (1994), pp. 776–780.

[38] H. Neudecker, *Some theorems on matrix differentiation with special reference to kronecker matrix products*, J. Am. Stat. Assoc. 64 (1969), pp. 953–962.

[39] D. Peña and J. Rodríguez, *A powerful portmanteau test of lack of fit for time series*, J. Am. Stat. Assoc. 97 (2002), pp. 601–610. Available at http://www.jstor.org/stable/3085675.

[40] D. Peña and J. Rodríguez, *The log of the determinant of the autocorrelation matrix for testing goodness of fit in time series*, J. Stat. Plan. Inference. 136 (2006), pp. 2706–2718. Available at https://www.sciencedirect.com/science/article/pii/S0378375804004574.

[41] D. Peña, G.C. Tiao, and R.S. Tsay, *A Course in Time Series Analysis*, John Wiley, New York, 2001.

[42] D.N. Politis and J.P. Romano, *The stationary bootstrap*, J. Am. Stat. Assoc. 89 (1994), pp. 1303–1313. Available at http://www.jstor.org/stable/2290993.

[43] Z. Psaradakis and M. Vávra, *Portmanteau tests for linearity of stationary time series*, Econom. Rev. 38 (2019), pp. 248–262. DOI:10.1080/07474938.2016.1261015

[44] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2020. Available at https://www.R-project.org/, ISBN 3-900051-07-0.

[45] M.W. Robbins and T.J. Fisher, *Cross-correlation matrices for tests of independence and causality between two multivariate time series*, J. Bus. Econ. Stat. 33 (2015), pp. 459–473. DOI:10.1080/07350015.2014.962699

[46] J. Rodríguez and E. Ruiz, *A powerful test for conditional heteroscedasticity for financial time series with highly persistent volatilities*, Stat. Sin. 15 (2005), pp. 505–525.

[47] F.E. Satterthwaite, *Synthesis of variance*, Psychometrika 6 (1941), pp. 309–316.

[48] F.E. Satterthwaite, *An approximate distribution of estimates of variance components*, Biometrics Bulletin 2 (1946), pp. 110–114.

[49] R.S. Tsay, *Analysis of Financial Time Series*, 2nd ed. Wiley, New York, 2005.

[50] U.S. Energy Information Administration, *Crude Oil Prices: West Texas Intermediate (WTI) – Cushing, Oklahoma*. Available at https://fred.stlouisfed.org/series/DCOILWTICO, Accessed from the Federal Reserve Bank of St. Louis (FRED) on 27 July 2022.

[51] H. Wong and S. Ling, *Mixed portmanteau tests for time-series models*, J. Time Ser. Anal. 26 (2005), pp. 569–579.

[52] C.F.J. Wu, *Jackknife, bootstrap and other resampling methods in regression analysis*, Ann. Stat. 14 (1986), pp. 1261–1295. DOI:10.1214/aos/1176350142

[53] K. Zhu, *Bootstrapping the portmanteau tests in weak auto–regressive moving average models*, J. R. Stat. Soc. Ser. B Stat. Methodol. 78 (2016), pp. 463–485. Available at https://app.dimensions.ai/details/publication/pub.1043057179 and https://mpra.ub.uni-muenchen.de/61930/1/MPRA_paper_61930.pdf.

[54] Q. Zhu, R. Zeng, and G. Li, *Bootstrap inference for garch models by the least absolute deviation estimation*, J. Time Ser. Anal. 41 (2020), pp. 21–40.