Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/120228

Exponential Weight Portmanteau Tests of Univariate Time Series

*F. B. Dassi** and *M. G. Griffiths*

School of Computing and Mathematics, Faculty of Computing, Engineering and Science, University of South Wales, Cardiff, UK

Received: 22 Aug. 2022, Revised: 22 Sep. 2022, Accepted: 5 Nov. 2022.
Published online: 1 May 2023.

Abstract: This paper presents two new portmanteau tests to evaluate the goodness of fit of ARMA models. The tests are based on exponential weights of the residual autocorrelation function and the residual partial autocorrelation function. A review of previous work on portmanteau tests is given. The performance of the new portmanteau tests is compared with previous portmanteau tests via the use of Monte Carlo experiments with 10,000 replications. The empirical size simulations show that, when an AR(1) process is fitted by an AR(1) model, most portmanteau tests from previous studies do not have significance levels that are stable with respect to lag length. The new residual partial autocorrelation function test is shown to outperform previous tests in terms of its power and its stability with respect to lag length.

Keywords: autoregressive process, moving average process, Monte Carlo experiment, portmanteau test.

1 Introduction

Diagnostic checking is the third stage of the Box and Jenkins methodology. The adequacy of a statistical model is examined, by considering the model's residuals, then the autocorrelation and partial autocorrelation functions are used as diagnostic tools to test the goodness of fit of the model. A portmanteau test is an important method of diagnostic checking, which is used to test the goodness of fit of an ARMA model of a time series, which has been studied by both Box and Pierce [1]and Ljung and Box [2].

A portmanteau test is calculated by summing the residuals of the autocorrelation or partial autocorrelation function of the fitted model. Then the value of the portmanteau test is compared with a critical value. If the value of the portmanteau test is less than the critical value, it means the model is appropriate for the data, and if bigger then the model is considered inappropriate for the data.

Suppose that a time series $\{z_t\}$ is generated by a stationary and invertible $\mathrm{ARMA}(p, q)$ process

$$\phi(B)z_t = \theta(B)e_t$$

where $\{e_t\}$ is a white noise process of mean zero and constant variance σ_e^2, and $\phi(B)$ and $\theta(B)$ are polynomials given by $\phi(B) = 1 - \phi_1 B - \cdots - \phi_P B^p$ and $\theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$, has been fitted by maximum likelihood estimates $(\hat{\phi}, \hat{\theta})$ obtained for the parameters, then it is possible to identify the residuals as

$$\hat{e}_t = \hat{\theta}^{-1}(B)\hat{\phi}(B)z_t, \quad (1)$$

The residuals are computed recursively using Equation (2) in the following form

$$\hat{e}_t = z_t - \sum_{j=1}^{p} \hat{\phi}_j \, z_{t-j} + \sum_{j=1}^{q} \hat{\theta}_j \, \hat{e}_{t-j} \qquad t = 1,2,\cdots,n \quad (2)$$

These residuals $\hat{e}_t$ from the ARMA model will be random if the model is correct, this means that the autocorrelation of the residuals ρ_k will be zero at all lags k. This gives the null hypothesis for all lags k

$H_0: \rho_k = 0$ versus $H_1: \rho_k \neq 0$

When considering the partial autocorrelation ϕ_{kk} of the residual at lags k, the hypothesis test can be given in the equivalent form

$H_0: \phi_{kk} = 0$ versus $H_1: \phi_{kk} \neq 0$

*Corresponding author e-mail: furat.dassi@southwales.ac.uk

All the following tests will use one of the given forms of the hypothesis, depending on whether the statistic relies on ρ_k or ϕ_{kk}.

The rest of the article is organized as follows. In Section 2, provides a review of previous studies of portmanteau tests in univariate time series. Section 3 introduces the new portmanteau tests, which are based on the residual autocorrelation function and the residual partial autocorrelation function combined with exponential weights. In Section 4, the simulation study compares the new portmanteau test statistics with those from previous studies by using the empirical size and the power level methods. In Section 5, we close this article with the conclusion.

2 Previous studies of portmanteau tests

2.1. Box and Pierce, Ljung and Box, and Monti tests

Portmanteau tests have been developed by several researchers, such as, Box and Pierce [1], Ljung and Box [2] and Monti [3]. These tests provide a measure of the accuracy of a fitted model. In the following tests n is the number of observations and m is the maximum lag taken into account.

The sample autocorrelation function can be obtained by

$$\hat{\rho}_k(\hat{e}) = \sum_{t=k+1}^{n} \hat{e}_t \hat{e}_{t-k} \Big/ \sum_{t=1}^{n} \hat{e}_t^2$$

where $\hat{e}_t$ is the residual of the estimated models. Box and Pierce [3] in 1970 showed that if the fitted model is appropriate then the portmanteau test statistic is given by

$$\tilde{Q}_{BP} = n \sum_{k=1}^{m} \hat{\rho}_k^2\,(\hat{e}) \tag{3}$$

Ljung and Box [2] in 1978 proposed a modified form of the portmanteau test statistic given by

$$\tilde{Q}_{LB} = n(n+2) \sum_{k=1}^{m} (n-k)^{-1}\, \hat{\rho}_k^2(\hat{e}) \tag{4}$$

Monti [3] in 1994 suggested the following portmanteau test statistic

$$\tilde{Q}_M = n(n+2) \sum_{k=1}^{m} (n-k)^{-1}\, \hat{\phi}_{kk}^2(\hat{e}) \tag{5}$$

where $\hat{\phi}_{kk}(\hat{e})$ is the residual partial autocorrelation at lag k.

All three of these tests are asymptotically distributed as a chi-squared variable with $(m-p-q)$ degrees of freedom.

2.2. Peña and Rodríguez, and Mahdi and McLeod tests

Peña and Rodríguez [4] in 2002 showed that the portmanteau goodness-of-fit test statistic is based on a general measure of multivariate dependence. Denote the correlation matrix up to order lag m of residual $\hat{e}_t$ from the fitted ARMA(p,q) model by

$$\hat{R}_m(\hat{e}) = \begin{bmatrix} 1 & \hat{\rho}_1(\hat{e}) & \hat{\rho}_2(\hat{e}) & \cdots & \hat{\rho}_k(\hat{e}) \\ \hat{\rho}_1(\hat{e}) & 1 & \hat{\rho}_2(\hat{e}) & \cdots & \hat{\rho}_{k-1}(\hat{e}) \\ \hat{\rho}_2(\hat{e}) & \hat{\rho}_1(\hat{e}) & 1 & \cdots & \hat{\rho}_{k-2}(\hat{e}) \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \hat{\rho}_k(\hat{e}) & \hat{\rho}_{k-1}(\hat{e}) & \hat{\rho}_{k-2}(\hat{e}) & \cdots & 1 \end{bmatrix}$$

Their proposed portmanteau test statistic is based on the determinant of this correlation matrix, a general measure of dependence in multivariate analysis, and is given by

$$\hat{D}_m = n\left(1 - \left|\hat{R}_m(\hat{e})\right|^{1/m}\right) \tag{6}$$

where n is a length of time series. If the model is correctly identified, then $\hat{D}_m$ is asymptotically distributed as a linear combination of chi-squared random variables and is approximately a gamma distribution random variable for large value of

m with parameter α and β. The distribution of $\widehat{D}_m$ can be approximated by the gamma distribution [4].

Peña and Rodríguez [5] in 2006 provided a new portmanteau test statistic, which is the log of the determinant of the mth autocorrelation matrix

$$D_m^* = -\frac{n}{m+1}\log\left|\left|\hat{R}_m\right|\right| \tag{7}$$

where $\hat{R}_m$ is the residual correlation matrix of order m. The gamma distribution is the approximation distribution of D_m^*,

Mahdi and McLeod [6] in 2012 introduced another test statistics based on the residual correlation matrix $\hat{R}_m$, given by

$$\tilde{Q}_{MM} = -\frac{3n}{2m+1}\log\left|\hat{R}_m\right| \tag{8}$$

They show that for large n, this portmanteau test statistic's distribution can be approximated by a chi-squared distribution.

2.3. Fisher and Gallagher (2012) and Gallagher and Fisher (2015) tests

Fisher and Gallagher [7] in 2012 introduced a new portmanteau test statistic $\tilde{Q}_{FGLB}$ that is based on the square of the mth-order autocorrelation matrix.

$$\tilde{Q}_{FGLB} = n(n+2)\sum_{k=1}^{m}\frac{(m-k+1)}{m}\frac{\hat{\rho}_k^2}{n-k} \tag{9}$$

where $\hat{\rho}_k^2$ is the autocorrelation at lag k. The statistic is a weighted sum of the squares of the sample autocorrelation coefficients, where the weights consist of a convolution of the Ljung-Box standardizing weights with the sequence

$$\left\{\frac{1}{m},\frac{2}{m},\ldots,\frac{m-1}{m},1\right\}.$$

The $\tilde{Q}_{FGLB}$ is approximately distributed as a gamma distribution [4].

Gallagher and Fisher [8] in 2015 introduced new portmanteau test statistics created by taking general weighted sums of the first $m = m(n)$ squared sample autocorrelations:

$$Q_w = n\sum_{k=1}^{m} w_k\hat{\rho}_k^2 \tag{10}$$

where w_k are the weights, n is the number of observations and m is the maximum lag taken into account. They considered several schemes for the weights w_k, in particular, the kernel-based weights and data adaptive weights:

Gallagher and Fisher [8] proved that the theoretical asymptotic distributions of all weighting schemes of the form given in Equation (10) approach the normal distribution for large values of n.

Kernel-Based Weights

Gallagher and Fisher [8] introduced a new portmanteau test employing a kernel weight. It is based on the square of a kernel function and blended with the Ljung-Box standardizing terms to construct sequence of weights $w_k = \big((n+2)/(n-k)\big)[\mathcal{K}(k/m)]^2$, where $\mathcal{K}(\cdot)$ is the Daniell kernel function, which is defined as.

$$\mathcal{K}(k/m) = \begin{cases} \dfrac{\sin\left(\sqrt{3}\pi(k/m)\right)}{\sqrt{3}\pi(k/m)} & : \ |k/m| < 1 \\ 0 & : \ |k/m| \geq 1 \end{cases} \tag{11}$$

The use of the Daniell kernel function was based on the work of Hong [9, 10] who the first to use it to test a residual series from an unknown distribution. The asymptotic distribution of this test is the normal distribution approximation [8].

Data Adaptive Weights

In this portmanteau test Gallagher and Fisher [8] use the sample partial autocorrelation function $\hat{\rho}$ to define

$$\tilde{Q}_{GFD} = n\sum_{k=1}^{m_0}\frac{n+2}{n-k}\hat{\rho}^2 + n\sum_{k=m_0+1}^{m} w_k\,\hat{\rho}^2, \tag{12}$$

The first m_0 terms obtain the standardizing weight $(n+2)/(n-k)$ from the Ljung-Box statistic, and they choose the remaining weights to be summable

$\left(w_k = -\log\left(1-\left|\hat{\phi}_{kk}\right|\right) \text{ and } \quad m_0 = \min\,(\log(n), M)\right)$,

where M is a finite bound.

Approximation Distribution

Gallagher and Fisher [8] consider the asymptotic behaviour of general weighted portmanteau statistics satisfying Equation (10). The gamma approximation is used for geometrically decaying weights and data adaptive weights, in particular, it has distribution $\Gamma(\alpha,\beta)$ with shape parameter

$$\alpha = \frac{(\sum w_k)^2}{2(\sum w_k^2 - p - q)} \tag{13}$$

and scale parameter

$$\beta = \frac{2(\sum w_k^2 - p - q)}{\sum w_k} \tag{14}$$

as given in [5].

3 New weighted portmanteau tests

3.1. Exponential weighted portmanteau tests

This paper introduces two new portmanteau tests that are based on exponential weights. The first new test is a development of Ljung and Box's test and the second test is a development of Monti's test. These new portmanteau test statistics are defined as

$$\tilde{Q}_{EXLB} = n(n+2)\sum_{k=1}^{m} w_k\frac{\hat{\rho}_k^2}{n-k}. \tag{15}$$

$$\tilde{Q}_{EXM} = n(n+2)\sum_{k=1}^{m} w_k\frac{\hat{\phi}_{kk}^2}{n-k}, \tag{16}$$

where $\hat{\rho}_k^2$ is the sample autocorrelation and $\hat{\phi}_{kk}^2$ is the sample partial autocorrelation at lag m, and w_k is an exponential weight.

3.2 Derivation of the exponential weight w_k

Consider an exponential function of the form

$f(x) = a^x, \quad 0 < a < 1, \quad 0 \le x < 1$

where a is the base and x is the exponent.

Constrain x to the values $(k-1)/m$, that is, terms from $\{0, 1/m, 2/m, \dots, (m-1)/m\}$, where k is the length of lag used in the autocorrelation function and the partial autocorrelation function, and m is the maximum lag.

Also, constrain a to take the value $\frac{1}{m}$.

The exponential function now takes the form

$$f\left(\frac{k-1}{m}\right) = \left(\frac{1}{m}\right)^{\left(\frac{k-1}{m}\right)}$$

this can be rearranged as,

$$= e^{\ln\left(\frac{1}{m}\right)^{\left(\frac{k-1}{m}\right)}}$$

$$= e^{\left(\frac{k-1}{m}\right)\ln\left(\frac{1}{m}\right)}$$

$$= e^{-\left(\frac{k-1}{m}\right)\ln m}$$

Since, $f\left(\frac{k-1}{m}\right)$ is now only a function of the variable k, and m is a constant, it can be rearranged as a function of the lag k, $w(k)$. This can be written as the exponential weight w_k as in Equations (16) and (17).

This exponential weight has similar distribution behaviour to the weights employed by Fisher and Gallagher [7, 8].

3.3 Approximation distribution for the New Portmanteau Tests

The new portmanteau tests have a gamma distribution, which can be shown by application of Theorem 3 from Gallagher and Fisher [8]. The exponential weights and squared exponential weights are summed, and the shape and scale are calculated from Equations (13) and (14). The distribution is similar to those of Peña and Rodríguez, Fisher and Gallagher.

4 Monte Carlo experiment

4.1 Generating Gaussian Random Numbers

Normally distributed N(0,1) pseudo random numbers were generated using the R program. This package employs the Mersenne-Twister generator [11] to generate uniformly distributed pseudo random numbers. These numbers are then transformed into normally distributed numbers N(0,1) by the application of the Box-Muller [12] transformation.

$$y_1 = (-2\log x_1)^{1/2}\cos 2\pi x_1 \tag{17}$$

$$y_2 = (-2\log x_1)^{1/2}\sin 2\pi x_2 \tag{18}$$

where y_1 and y_2 are a Gaussian (or normal) distribution, and x_1, x_2 are independent random variables from the same rectangular density function on the interval (0,1).

4.2 Simulation studies

The aim of the simulation study was to compare the new exponential portmanteau tests against the portmanteau tests used in previous studies, developed by Ljung and Box $\tilde{Q}_{\mathrm{LB}}$, Monti $\tilde{Q}_{\mathrm{M}}$, Mahdi and McLeod $\tilde{Q}_{\mathrm{MM}}$, Fisher and Gallagher $\tilde{Q}_{\mathrm{FGLB}}$, Gallagher and Fisher Kernel-based weights $\tilde{Q}_{\mathrm{GFK}}$ and Data Adaptive Weights $\tilde{Q}_{\mathrm{GFD}}$. The empirical size and the power level of the tests were investigated by conducting simulations studies using the R program.

4.3 Empirical size

A Monte Carlo experiment was conducted with 10,000 replications. The aim was to simulate n = 100 observations under an AR(1) process, $z_t - \phi\, z_{t-1} = e_t$, employing different parameters, $\phi = 0.1$, 0.3, 0.5, 0.7 and 0.9. Next, an AR(1) model was fitted to the generated data producing an estimate $\hat{\phi}$ of the underlying parameter ϕ. The method employed to obtain the fitted model uses the maximum likelihood function, using Approximation 2 from Box, Jenkins and Reinsel [13].

$$(n-2)(n-1)^{-1}\sum_{t=2}^{n} z_t z_{t-1} \Big/ \sum_{t=2}^{n-1} z_t^2 \tag{19}$$

Following this, the autocorrelations of the fitted model were calculated using the residuals $\hat{e}_t = z_t - \hat{\phi}\, z_{t-1}\ (t = 2, \dots, n)$. The test statistics $\tilde{Q}_{LB}$, $\tilde{Q}_{M}$, $\tilde{Q}_{MM}$, $\tilde{Q}_{FGLB}$, $\tilde{Q}_{GFK}$, $\tilde{Q}_{GFD}$, $\tilde{Q}_{EXLB}$ and $\tilde{Q}_{EXM}$ were subsequently calculated. This was repeated for lags of autocorrelations and partial autocorrelations for maximum lags m = 10 and 20.

Method of a Monte Carlo experiment to calculate the empirical size of a range of portmanteau tests

Below are the steps of a Monte Carlo experiment in which data are generated by an AR(1) process, $z_t = \phi\, z_{t-1} + e_t$, then fitted under an AR(1) model to find the empirical size of the following portmanteau tests $\tilde{Q}_{LB}$, $\tilde{Q}_{M}$, $\tilde{Q}_{MM}$, $\tilde{Q}_{FGLB}$, $\tilde{Q}_{GFK}$, $\tilde{Q}_{GFD}$, $\tilde{Q}_{EXLB}$ and $\tilde{Q}_{EXM}$.

1. Select the value of the process parameter ϕ and maximum lag m. In this example, $\phi = 0.1$ and m = 10.
2. Generate n = 100 values from a normal distribution (e_t white noise).

3. Use the e_t values to generate observations z_t from an AR(1) process with parameter ϕ.
4. Fit an AR(1) model to the observations by estimating its parameters using the maximum likelihood function.
5. Find the residuals $\hat{e}_t$.
6. Find the residual autocorrelation and partial autocorrelation functions for the model.
7. Calculate the various portmanteau test statistics. For example, $\phi = 0.1$, gives

ϕ	$\tilde{Q}_{LB}$	$\tilde{Q}_{M}$	$\tilde{Q}_{MM}$	$\tilde{Q}_{FGLB}$	$\tilde{Q}_{GFK}$	$\tilde{Q}_{GFD}$	$\tilde{Q}_{EXLB}$	$\tilde{Q}_{EXM}$
0.1	16.6580	17.0012	11.890	8.120	0.917	8.976	6.744	7.058

8. Look up the 5 percentage point of the χ^2_{m-1} distribution and the gamma distribution.

Distributions	χ^2_{m-1}			Gamma				
Tests	$\tilde{Q}_{LB}$	$\tilde{Q}_{M}$	$\tilde{Q}_{MM}$	$\tilde{Q}_{FGLB}$	$\tilde{Q}_{GFK}$	$\tilde{Q}_{GFD}$	$\tilde{Q}_{EXLB}$	$\tilde{Q}_{EXM}$
$m = 10$	16.9		14.1	9.92	1.6	11.55	7.764	

Reject the fitted AR(1) model if the value of the portmanteau test is bigger than the critical value in step 7 (using the appropriate distribution for each portmanteau test).

ϕ	$\tilde{Q}_{LB}$	$\tilde{Q}_{M}$	$\tilde{Q}_{MM}$	$\tilde{Q}_{FGLB}$	$\tilde{Q}_{GFK}$	$\tilde{Q}_{GFD}$	$\tilde{Q}_{EXLB}$	$\tilde{Q}_{EXM}$
0.1	Accept	Reject	Accept	Accept	Accept	Accept	Accept	Accept

9. Repeat 10,000 times for steps 1-8.
10. For each portmanteau test use the number of rejected AR(1) models (out of 10,000) to find the percentage rejected.

Table 1 and 2 give the results of the Monte Carlo experiment and show the proportion of the test statistics $\tilde{Q}_{LB}$, $\tilde{Q}_{M}$, $\tilde{Q}_{MM}$, $\tilde{Q}_{FGLB}$, $\tilde{Q}_{GFK}$, $\tilde{Q}_{GFD}$, $\tilde{Q}_{EXLB}$ and $\tilde{Q}_{EXM}$ values that are above the upper 5 percentage point of the χ^2_{m-1} distribution or gamma distribution. The tables show data fitted under the AR(1) with different parameters, $\phi = 0.1$, 0.3, 0.5, 0.7 and 0.9, with *n* = 100, and autocorrelations and partial autocorrelations lags of *m* = 10 and *m* = 20.

Table 1: Empirical size of the test statistics $\tilde{Q}_{LB}$, $\tilde{Q}_{M}$, $\tilde{Q}_{MM}$, $\tilde{Q}_{FGLB}$, $\tilde{Q}_{GFK}$, $\tilde{Q}_{GFD}$, $\tilde{Q}_{EXLB}$ and $\tilde{Q}_{EXM}$ at 5% significance level for fitted AR(1) models, $n = 100$ and $m = 10$.

ϕ	$\tilde{Q}_{LB}$	$\tilde{Q}_{M}$	$\tilde{Q}_{MM}$	$\tilde{Q}_{FGLB}$	$\tilde{Q}_{GFK}$	$\tilde{Q}_{GFD}$	$\tilde{Q}_{EXLB}$	$\tilde{Q}_{EXM}$
0.1	**0.0531**	0.0548	0.0294	0.0311	0.0122	0.0301	0.0292	0.0287
0.3	**0.0516**	0.0541	0.0347	0.0352	0.0165	0.0336	0.0335	0.0355
0.5	0.0574	**0.0531**	0.0312	0.0369	0.0167	0.0357	0.0369	0.0378
0.7	0.0518	**0.0514**	0.0308	0.0343	0.0185	0.0317	0.0347	0.0349
0.9	0.0605	0.0568	0.0429	0.0474	0.0364	0.0475	0.0535	**0.0504**

Table 1 shows the values of the significance levels when $\alpha = 0.05$, *n* = 100 and *m* = 10. The value of the $\tilde{Q}_{LB}$ test is closer to the 0.05 significance level in two cases, i.e., when $\phi = 0.1$ and 0.3. The value of the $\tilde{Q}_{M}$ test is closer to 0.05 in two cases, i.e., $\phi = 0.5$ and 0.7. The value of the $\tilde{Q}_{EXM}$ test is closer to 0.05 in one case, i.e., when $\phi = 0.9$. Overall, the $\tilde{Q}_{M}$ test is superior to the other tests in most cases.

Table 2: Empirical size of the test statistics $\tilde{Q}_{LB}$, $\tilde{Q}_{M}$, $\tilde{Q}_{MM}$, $\tilde{Q}_{FGLB}$, $\tilde{Q}_{GFK}$, $\tilde{Q}_{GFD}$, $\tilde{Q}_{EXLB}$ and $\tilde{Q}_{EXM}$ at 5% significance level for fitted AR(1) models, $n = 100$ and $m = 20$.

ϕ	$\tilde{Q}_{LB}$	$\tilde{Q}_{M}$	$\tilde{Q}_{MM}$	$\tilde{Q}_{FGLB}$	$\tilde{Q}_{GFK}$	$\tilde{Q}_{GFD}$	$\tilde{Q}_{EXLB}$	$\tilde{Q}_{EXM}$
0.1	0.0616	**0.0557**	0.025	0.0412	0.0223	0.0385	0.0353	0.0316
0.3	0.0616	**0.0545**	0.0255	0.0398	0.0232	0.0397	0.0344	0.0336
0.5	0.0622	**0.0513**	0.0249	0.0438	0.0255	0.044	0.0404	0.0337
0.7	0.0671	0.0555	0.0281	**0.0504**	0.0284	0.0465	0.0467	0.0404

0.9	0.0706	**0.0500**	0.0284	0.0514	0.0318	0.0456	0.0491	0.0438

Table 2 shows the values of the significance levels when $\alpha = 0.05$, $n = 100$ and $m = 20$. The value of the $\tilde{Q}_M$ test is closest to 0.05 in four cases, i.e., when $\phi = 0.1, 0.3, 0.5$ and 0.9. The values of the $\tilde{Q}_{FGLB}$ test is closest to 0.05 in one case, i.e., when $\phi = 0.7$. Overall, the $\tilde{Q}_M$ test is superior than the other tests in most cases.

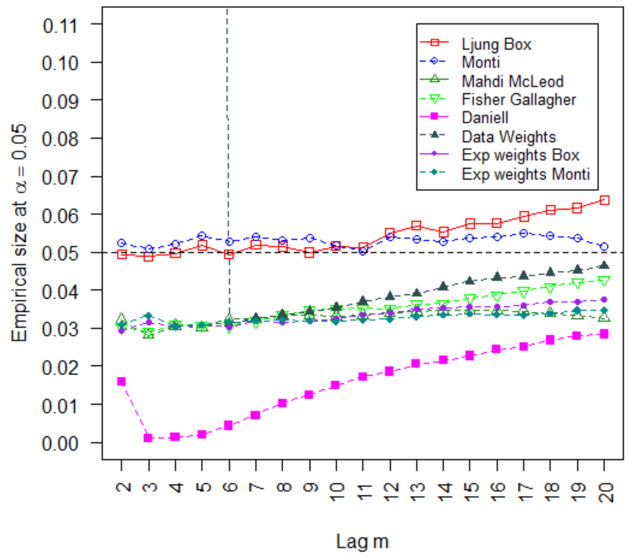

Fig. 1. Empirical size for lags from 2 to 20 for a correctly fitted AR(1) model, with data generated by an AR(1) process with $\phi = 0.5$, at 5% significance level, series of length $n = 150$.

Figure 1 shows the empirical size of lags from 2 to 20 of the $\tilde{Q}_{LB}$, $\tilde{Q}_M$, $\tilde{Q}_{MM}$, $\tilde{Q}_{FGLB}$, $\tilde{Q}_{GFK}$, $\tilde{Q}_{GFD}$, $\tilde{Q}_{EXLB}$ and $\tilde{Q}_{EXM}$ tests based on a 5% significance level, when data are generated by an AR(1) process $\phi = 0.5$ and fitted under an AR(1) model with $n = 150$ and 10,000 replications. The $\tilde{Q}_{LB}$, $\tilde{Q}_{\mathrm{FGLB}}$, and $\tilde{Q}_{GFD}$ tests' empirical sizes increase as the lag increases, while those of the $\tilde{Q}_{MM}$, $\tilde{Q}_{\mathrm{EXLB}}$ and $\tilde{Q}_{EXM}$ tests also increase but more slowly. The $\tilde{Q}_{GFK}$ test rapidly decreases at lag 3, then increases as the lag increases. The $\tilde{Q}_{\mathrm{M}}$ test is unaffected by the lag increase. The $\tilde{Q}_{\mathrm{GFD}}$ test always rejects the correct models at lags 2, 3, 4 and 5 (these points are off the scale in Figure 1), then from lag 6 the $\tilde{Q}_{\mathrm{GFD}}$ increases as the lag increases.

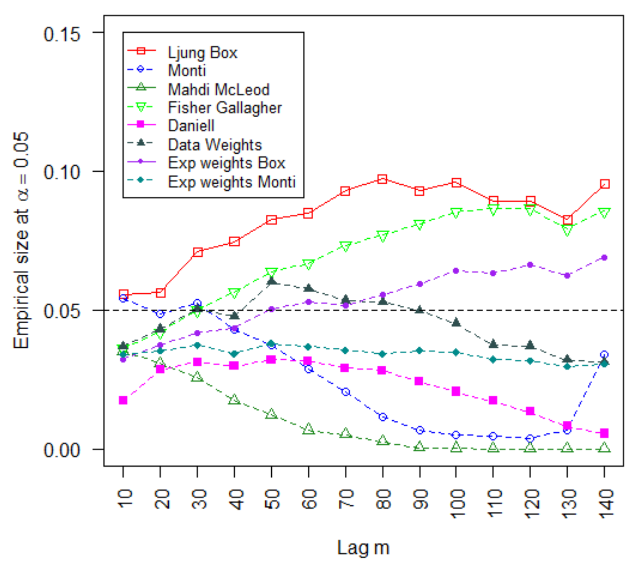

Fig. 2. Empirical size for maximum lags for a properly fitted AR(1) model, with data generated by an AR(1) process with $\phi = 0.5$, at 5% significance level, series of length $n = 150$.

Figure 2 shows the empirical size of large lags based on a 5% significance level, when data are generated by an AR(1) process with $\phi = 0.5$, and fitted by an AR(1) model with n = 100 and 10,000 replications. The $\tilde{Q}_{\mathrm{LB}}$, $\tilde{Q}_{\mathrm{FGLB}}$ and $\tilde{Q}_{\mathrm{EXLB}}$ tests increase as the lag increases. Other tests such as, $\tilde{Q}_{\mathrm{M}}$, $\tilde{Q}_{\mathrm{MM}}$ and $\tilde{Q}_{\mathrm{GFK}}$ decrease when the lag increases and the $\tilde{Q}_{\mathrm{GFD}}$ test initially increases with increasing lag but then decreases for larger lags. The $\tilde{Q}_{\mathrm{EXM}}$ test remains approximately constant as the lag increases.

4.4 Power Studies

In the power studies, the statistics $\tilde{Q}_{\mathrm{LB}}$, $\tilde{Q}_{\mathrm{M}}$, $\tilde{Q}_{\mathrm{MM}}$, $\tilde{Q}_{\mathrm{FGLB}}$, $\tilde{Q}_{\mathrm{GFK}}$, $\tilde{Q}_{\mathrm{GFD}}$, $\tilde{Q}_{\mathrm{EXLB}}$ and $\tilde{Q}_{\mathrm{EXM}}$ were compared using the processes and parameters employed in Monti [11].

Data was generated using a number of alternative ARMA(2,2) processes,

$$z_t = e_t + \phi_1 z_{t-1} + \phi_2 z_{t-2} - \theta_1 e_{t-1} - \theta_2 e_{t-2}$$

and fitted by an AR(1) model and a MA(1) model. Next the residual of the data was obtained, and the ACF and PACF were calculated. For each alternative set of parameters for the ARMA(2,2) process, 10,000 replications of 100 observations were generated. For each test the power was computed with lags m = 10 and 20.

In these experiments the AR(1) model parameter was estimated by using Equation (21). The MA(1) model parameter was estimated by using the maximum likelihood function [13], which is

$$\ln(\theta) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \sum_{t=1}^{n}\frac{e_t^2}{2\sigma^2}. \tag{21}$$

Table 3: Power level of the test statistics $\tilde{Q}_{\mathrm{LB}}$, $\tilde{Q}_{\mathrm{M}}$, $\tilde{Q}_{\mathrm{MM}}$, $\tilde{Q}_{\mathrm{FGLB}}$, $\tilde{Q}_{\mathrm{GFK}}$, $\tilde{Q}_{\mathrm{GFD}}$, $\tilde{Q}_{\mathrm{EXLB}}$ and $\tilde{Q}_{\mathrm{EXM}}$ where data are generated under various alternative ARMA(2,2) processes and fitted by an AR(1) model.

					$m = 10$							
Model	ϕ_1	ϕ_2	θ_1	θ_2	$\tilde{Q}_{LB}$	$\tilde{Q}_{M}$	$\tilde{Q}_{MM}$	$\tilde{Q}_{FGLB}$	$\tilde{Q}_{GFK}$	$\tilde{Q}_{GFD}$	$\tilde{Q}_{EXLB}$	$\tilde{Q}_{EXM}$
1	**---**	**---**	**-0.50**	**---**	0.2634	0.3081	0.3886	0.3314	0.3395	0.3266	0.3702	**0.4326**
2	**---**	**---**	**-0.80**	**---**	0.7444	0.9659	0.9872	0.8995	0.9392	0.9034	0.9336	**0.9901**
3	**---**	**---**	**-0.60**	**0.30**	0.7792	0.9880	0.9935	0.9187	0.9300	0.9275	0.9431	**0.9953**
4	**0.10**	**0.30**	**---**	**---**	0.4283	0.4269	0.5239	0.5290	0.5424	0.5295	0.5612	**0.5619**
5	**1.30**	**-0.35**	**---**	**---**	0.7211	0.7088	0.8467	0.8454	**0.9089**	0.8238	0.8929	0.8972
6	**0.70**	**---**	**-0.40**	**---**	0.5541	0.6179	0.7605	0.6958	0.7821	0.6519	0.7713	**0.8263**
7	**0.70**	**---**	**-0.90**	**---**	0.9872	**1**	**1**	0.9997	**1**	0.9996	**1**	**1**
8	**0.40**	**---**	**-0.60**	**0.30**	0.8414	0.9975	**0.9992**	0.9649	0.9813	0.9669	0.983	**0.9992**
9	**0.70**	**---**	**0.70**	**-0.15**	0.1742	0.1630	0.1822	0.1929	0.1395	**0.2024**	0.1999	0.1928
10	**0.70**	**0.20**	**0.50**	**---**	0.7506	0.7456	0.8150	0.8121	0.7543	0.8066	0.8258	**0.8322**
11	**0.70**	**0.20**	**-0.50**	**---**	0.3915	0.4801	0.6468	0.5482	0.6764	0.5012	0.6489	**0.7268**
12	**0.90**	**-0.40**	**1.20**	**-0.30**	0.7201	0.9735	0.9800	0.8529	0.7698	0.8746	0.8713	**0.9813**
	Average				0.6130	0.6979	0.7603	0.7159	0.7303	0.7095	0.7501	**0.7863**

Table 3 displays the power levels based on a 5% significance level when data are generated from an ARMA(2,2) process and an AR(1) model is fitted, with n = 100 and m = 10. Table 3 shows that the $\tilde{Q}_{\mathrm{EXM}}$ test is the most powerful in 10 of the 12 cases, being superior to all other tests in 8 cases. In the case of model 8, the $\tilde{Q}_{\mathrm{EXM}}$ was jointly the most powerful with the $\tilde{Q}_{\mathrm{MM}}$ test. Furthermore, in model 7, the $\tilde{Q}_{\mathrm{EXM}}$ test is equally the most powerful test, alongside tests $\tilde{Q}_{\mathrm{M}}$, $\tilde{Q}_{\mathrm{MM}}$, $\tilde{Q}_{\mathrm{GFK}}$, $\tilde{Q}_{\mathrm{EXLB}}$. However, the $\tilde{Q}_{\mathrm{GFK}}$ and $\tilde{Q}_{\mathrm{GFD}}$ tests are each the best test in one case, that is, for models 5 and 9 respectively. Nevertheless, the $\tilde{Q}_{\mathrm{EXM}}$ test is still the 2nd and 4th best, with power level values close the leading test in both cases. The average value has been taken for each test when data are fitted under an AR(1) model with m = 10, illustrating the superiority of the $\tilde{Q}_{\mathrm{EXM}}$ test compared to the other tests.

Table 4: Power level of the test statistics $\tilde{Q}_{\mathrm{LB}}$, $\tilde{Q}_{\mathrm{M}}$, $\tilde{Q}_{\mathrm{MM}}$, $\tilde{Q}_{\mathrm{FGLB}}$, $\tilde{Q}_{\mathrm{GFK}}$, $\tilde{Q}_{\mathrm{GFD}}$, $\tilde{Q}_{\mathrm{EXLB}}$ and $\tilde{Q}_{\mathrm{EXM}}$ where data are generated under various alternative ARMA(2,2) processes and fitted by a MA (1) model.

					$m = 10$							
Model	ϕ_1	ϕ_2	θ_1	θ_2	$\tilde{Q}_{LB}$	$\tilde{Q}_{M}$	$\tilde{Q}_{MM}$	$\tilde{Q}_{FGLB}$	$\tilde{Q}_{GFK}$	$\tilde{Q}_{GFD}$	$\tilde{Q}_{EXLB}$	$\tilde{Q}_{EXM}$
13	**0.50**	**---**	**---**	**---**	0.2836	0.2689	0.3215	0.3438	0.3552	0.3369	**0.3765**	0.3631
14	**0.80**	**---**	**---**	**---**	0.9823	0.9758	0.9903	0.9926	**0.9942**	0.9921	0.9940	0.9933

15	**1.10**	**-0.35**	---	---	0.9961	0.9957	0.9989	0.9989	**0.9997**	0.9989	0.9993	0.9995
16	---	---	**0.80**	**-0.50**	0.8389	0.9375	0.9734	0.9415	0.9487	0.9481	0.9584	**0.9791**
17	---	---	**-0.60**	**0.30**	0.3868	0.4626	0.5940	0.5209	0.6002	0.4784	0.6001	**0.6727**
18	**0.50**	---	**-0.70**	---	0.8773	0.8606	0.9365	0.9405	**0.9648**	0.9338	0.9613	0.9575
19	**-0.50**	---	**0.70**	---	0.8933	0.8763	0.9452	0.9516	**0.9697**	0.9458	0.9660	0.9633
20	**0.30**	---	**0.80**	**-0.50**	0.6265	0.7602	0.8378	0.7518	0.7323	0.7807	0.7857	**0.8579**
21	**0.80**	---	**-0.50**	**0.30**	0.9786	0.9626	0.9847	0.9897	**0.9931**	0.9886	0.9928	0.9898
22	**1.20**	**-0.50**	**0.90**	---	0.4685	**0.7108**	0.6157	0.4761	0.1232	0.4932	0.4300	0.5735
23	**0.30**	**-0.20**	**-0.70**	---	0.2649	0.2852	0.3491	0.3268	0.3088	0.3237	0.3721	**0.3962**
24	**0.90**	**-0.40**	**1.20**	**-0.30**	0.7888	0.9335	0.9571	0.8958	0.8121	0.9085	0.9076	**0.9615**
	Average				0.6988	0.7525	0.7920	0.7608	0.7335	0.7607	0.7787	**0.8090**

Table 4 shows the power levels based on a 5% significance level when data are generated from an ARMA(2,2) process and a MA(1) model is fitted, with *n* = 100 and *m* = 10. As is apparent in Table 4, the $\tilde{Q}_{\mathrm{EXM}}$ test is again the most powerful test in 5 cases. While the $\tilde{Q}_{\mathrm{GFK}}$ test is the most powerful test in 5 cases, suggesting it is comparable to the $\tilde{Q}_{\mathrm{EXM}}$ test is a close second in most cases. The average power level across all models illustrates that the $\tilde{Q}_{\mathrm{EXM}}$ test is consistently superior under a MA(1) model with *m* = 10.

Table 5: Power level of the test statistics $\tilde{Q}_{\mathrm{LB}}$, $\tilde{Q}_{\mathrm{M}}$, $\tilde{Q}_{\mathrm{MM}}$, $\tilde{Q}_{\mathrm{FGLB}}$, $\tilde{Q}_{\mathrm{GFK}}$, $\tilde{Q}_{\mathrm{GFD}}$, $\tilde{Q}_{\mathrm{EXLB}}$ and $\tilde{Q}_{\mathrm{EXM}}$ where data are generated under various alternative ARMA(2,2) processes and fitted by an AR(1) model.

					$m = 20$							
Model	ϕ_1	ϕ_2	θ_1	θ_2	$\tilde{Q}_{LB}$	$\tilde{Q}_{M}$	$\tilde{Q}_{MM}$	$\tilde{Q}_{FGLB}$	$\tilde{Q}_{GFK}$	$\tilde{Q}_{GFD}$	$\tilde{Q}_{EXLB}$	$\tilde{Q}_{EXM}$
1	---	---	**-0.50**	---	0.2197	0.2038	0.2688	0.2614	0.3021	0.3434	0.3292	**0.3859**
2	---	---	**-0.80**	---	0.5932	0.8651	0.9636	0.7804	0.8928	0.9128	0.8919	**0.9879**
3	---	---	**-0.60**	**0.30**	0.6239	0.9461	0.9838	0.8140	0.9138	0.9363	0.9109	**0.9931**
4	**0.10**	**0.30**	---	---	0.3624	0.2908	0.3951	0.4487	0.5105	**0.5610**	0.5272	0.5175
5	**1.30**	**-0.35**	---	---	0.6289	0.5488	0.7288	0.7604	0.8337	0.8392	0.8534	**0.8537**
6	**0.70**	---	**-0.40**	---	0.4765	0.4828	0.6347	0.6030	0.6810	0.6977	0.7204	**0.7861**
7	**0.70**	---	**-0.90**	---	0.9302	0.9992	**1**	0.9951	0.9998	0.9998	0.9998	**1**
8	**0.40**	---	**-0.60**	**0.30**	0.6874	0.9764	0.9960	0.8748	0.9596	0.9682	0.9577	**0.9988**
9	**0.70**	---	**0.70**	**-0.15**	0.1596	0.1192	0.1248	0.1762	0.1770	**0.2262**	0.1955	0.1806
10	**0.70**	**0.20**	**0.50**	---	0.6378	0.5977	0.7210	0.7500	0.7931	**0.8247**	0.8038	0.8107
11	**0.70**	**0.20**	**-0.50**	---	0.3114	0.2987	0.4392	0.4045	0.4870	0.5102	0.5344	**0.6194**
12	**0.90**	**-0.40**	**1.20**	**-0.30**	0.5661	0.8960	0.9604	0.7313	0.8379	0.8851	0.8320	**0.9817**
	Average				0.5164	0.6021	0.6847	0.6333	0.6990	0.7254	0.7130	**0.7596**

Table 5 shows power levels based on a 5% significance level when data are generated from an ARMA(2,2) process and an AR(1) model is fitted, with *n* = 100 and *m* = 20. Table 5 shows that once again the $\tilde{Q}_{\mathrm{EXM}}$ test is more powerful than other portmanteau tests in 9 of the 12 cases. The average power level across the 12 models shows that the $\tilde{Q}_{\mathrm{EXM}}$ test is consistently effective, when data are fitted under an AR(1) model with *m* = 20.

Table 6: Power level of the test statistics $\tilde{Q}_{\mathrm{LB}}$, $\tilde{Q}_{\mathrm{M}}$, $\tilde{Q}_{\mathrm{MM}}$, $\tilde{Q}_{\mathrm{FGLB}}$, $\tilde{Q}_{\mathrm{GFK}}$, $\tilde{Q}_{\mathrm{GFD}}$, $\tilde{Q}_{\mathrm{EXLB}}$ and $\tilde{Q}_{\mathrm{EXM}}$ where data are generated under various alternative ARMA(2,2) processes and fitted by a MA(1) model.

					$m = 20$							
Model	ϕ_1	ϕ_2	θ_1	θ_2	$\tilde{Q}_{LB}$	$\tilde{Q}_{M}$	$\tilde{Q}_{MM}$	$\tilde{Q}_{FGLB}$	$\tilde{Q}_{GFK}$	$\tilde{Q}_{GFD}$	$\tilde{Q}_{EXLB}$	$\tilde{Q}_{EXM}$
13	**0.50**	---	---	---	0.2468	0.1816	0.2192	0.2882	0.3194	**0.3593**	0.3465	0.3197
14	**0.80**	---	---	---	0.9632	0.9310	0.9745	0.9841	0.9911	**0.9925**	0.9912	0.9891
15	**1.10**	**-0.35**	---	---	0.9879	0.9840	0.9981	0.9979	0.9994	0.9993	**0.9996**	0.9994
16	---	---	**0.80**	**-0.50**	0.7042	0.8146	0.9260	0.8611	0.9313	0.9522	0.9313	**0.9678**
17	---	---	**-0.60**	**0.30**	0.3168	0.3237	0.4425	0.4110	0.4883	0.5105	0.5340	**0.6110**
18	**0.50**	---	**-0.70**	---	0.7930	0.7223	0.8643	0.8933	0.9355	0.9405	**0.9425**	0.9363
19	**-0.50**	---	**0.70**	---	0.8187	0.7520	0.8831	0.9113	0.9463	0.9497	**0.9512**	0.9451
20	**0.30**	---	**0.80**	**-0.50**	0.4959	0.5731	0.7214	0.6417	0.7273	0.7895	0.7385	**0.8259**
21	**0.80**	---	**-0.50**	**0.30**	0.9624	0.9124	0.9685	0.9826	0.9899	0.9906	**0.9911**	0.9864
22	**1.20**	**-0.50**	**0.90**	---	0.3866	0.5803	0.6031	0.4463	0.4256	0.5129	0.4570	**0.6371**

23	**0.30**	**-0.20**	**-0.70**	**---**	0.2260	0.1939	0.2366	0.2707	0.2964	**0.3516**	0.3307	0.3466
24	**0.90**	**-0.40**	**1.20**	**-0.30**	0.6292	0.8283	0.9189	0.8007	0.8795	0.9141	0.8805	**0.9582**
	Average				0.6276	0.6498	0.7297	0.7074	0.7442	0.7719	0.7578	**0.7936**

Table 6 shows the power levels based on a 5% significance level, when data are generated from an ARMA(2,2) process and a MA(1) model is fitted, with n = 100 and m = 20. It is evident that the $\tilde{Q}_{\text{EXM}}$ test is more powerful than other portmanteau tests in 5 cases. The average power value shows that the $\tilde{Q}_{\text{EXM}}$ test is consistently the most powerful test compared to the others. This is supported by the average of $\tilde{Q}_{\text{EXM}}$ being appreciably larger than for the other tests, when data are fitted under a MA(1) model with m = 20.

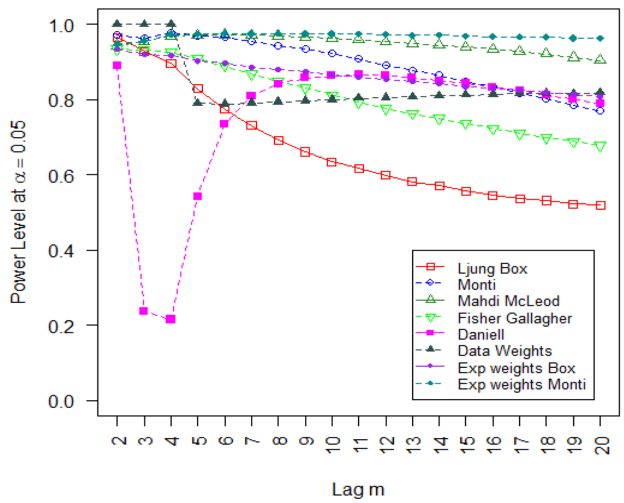

Fig. 3. Power level for lags from 2 to 20 for a fitted AR(1) model, data generated by a MA(1) process with $\theta = -0.8$, at 5% significance level, series of length n = 85.

Figure 3 shows the power level for lags from 2 to 20 based on a 5% significance level, when data are generated by a MA(1) process with $\theta = -0.8$ and fitted under an AR(1) model with n = 85 (following the simulation of Gallagher and Fisher [5] and 10,000 replications. The power of the $\tilde{Q}_{\text{LB}}$, $\tilde{Q}_{\text{M}}$ and $\tilde{Q}_{\text{FGLB}}$ tests decreases as the lag increases. The $\tilde{Q}_{\text{MM}}$ and $\tilde{Q}_{\text{EXLB}}$ tests slowly decrease as the lag increases. The $\tilde{Q}_{\text{GFK}}$ test rapidly decreases at lags 3 and 4, then rapidly increases as the lag increases. The $\tilde{Q}_{\text{GFD}}$ test rapidly decreases at lag 5, then slowly increases as the lag increases. From Figure 3 it is apparent that this test rejects all models, even correct ones, for the lags examined. The $\tilde{Q}_{\text{EXM}}$ test has a high-power level and remains constant as the lag increases, unlike most of the other tests.

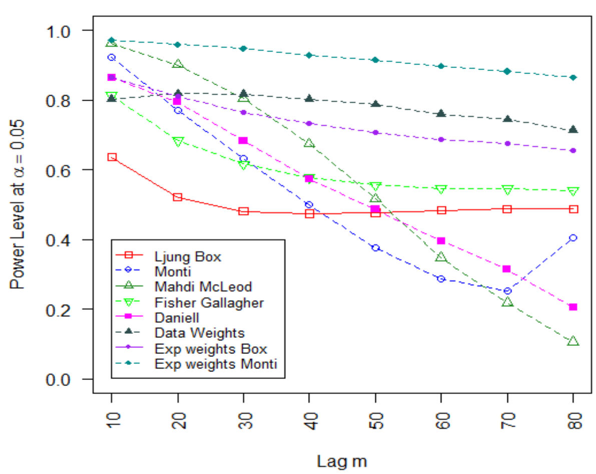

Fig. 4. Power level for lags from 10 to 80 for a fitted AR(1) model, data generated by a MA(1) process with $\theta = -0.8$, at 5% significance level, series of length n = 85.

Figure 4 shows the power level of large lags based on a 5% significance level, when data are generated by a MA(1) process with $\theta = -0.8$ and fitted under an AR(1) model with *n* = 85 and 10,000 replications. The power of the $\tilde{Q}_{\mathrm{GFD}}$, $\tilde{Q}_{\mathrm{EXLB}}$ and $\tilde{Q}_{\mathrm{EXM}}$ tests slowly decreases as the lag increases. The $\tilde{Q}_{\mathrm{LB}}$ and $\tilde{Q}_{\mathrm{FGLB}}$ tests have similar behaviour to each other, initially slowly decreasing and then remaining constant as the lag increases further. The power of the $\tilde{Q}_{\mathrm{M}}$, $\tilde{Q}_{\mathrm{MM}}$ and $\tilde{Q}_{\mathrm{GFK}}$ tests decreases as the lag increases. In all cases, the $\tilde{Q}_{\mathrm{EXM}}$ is the most powerful test.

The next study is similar to Gallagher and Fisher [8], where data are generated under an ARMA(2,2) process and are fitted by an ARMA(1,1) model, see Tables 7 and 8. Following on from previously published research in this area, *m* was set at 10 and 20, and *n* was set at 100, and in each case, the critical value was determined from the corresponding asymptotic distribution.

Table 7: Power level of the test statistics $\tilde{Q}_{\mathrm{LB}}$, $\tilde{Q}_{\mathrm{M}}$, $\tilde{Q}_{\mathrm{MM}}$, $\tilde{Q}_{\mathrm{FGLB}}$, $\tilde{Q}_{\mathrm{GFK}}$, $\tilde{Q}_{\mathrm{GFD}}$, $\tilde{Q}_{\mathrm{EXLB}}$ and $\tilde{Q}_{\mathrm{EXM}}$ where data are generated under various alternative ARMA(2,2) processes, and fitted by an ARMA(1,1) model $m = 10$.

					$m = 10$							
Model	ϕ_1	ϕ_2	θ_1	θ_2	$\tilde{Q}_{LB}$	$\tilde{Q}_{M}$	$\tilde{Q}_{MM}$	$\tilde{Q}_{FGLB}$	$\tilde{Q}_{GFK}$	$\tilde{Q}_{GFD}$	$\tilde{Q}_{EXLB}$	$\tilde{Q}_{EXM}$
4	**0.10**	**0.30**	**---**	**---**	0.2988	0.2681	0.3252	0.3497	0.2951	0.3310	**0.4343**	0.4148
5	**1.30**	**-0.35**	**---**	**---**	0.1183	0.1292	0.123	0.1105	0.0659	0.1001	0.1552	**0.1651**
9	**0.70**	**---**	**0.70**	**-0.15**	0.1180	0.1121	0.1032	0.1107	0.0534	0.1068	**0.1483**	0.1446
10	**0.70**	**0.20**	**0.50**	**---**	0.1663	0.1538	0.1737	0.1863	0.1528	0.1715	**0.2490**	0.2407
12	**0.90**	**-0.40**	**1.20**	**-0.30**	0.3856	0.3902	0.4289	0.4338	0.2873	0.4325	**0.5132**	0.5032
15	**1.10**	**-0.35**	---	---	0.1428	**0.1460**	0.1193	0.1168	0.0201	0.097	0.1383	0.1459
16	**---**	**---**	**0.80**	**-0.50**	0.3664	0.4477	0.4691	0.3859	0.1234	0.4323	0.4412	**0.5085**
17	**---**	**---**	**-0.60**	**0.30**	0.1194	0.1179	0.1188	0.1151	0.0598	0.1033	0.1565	**0.1679**
20	**0.30**	**---**	**0.80**	**-0.50**	0.3986	0.4626	0.4941	0.4332	0.1756	0.4683	0.4867	**0.5383**
21	**0.80**	**---**	**-0.50**	**0.30**	0.1195	0.1329	0.1241	0.1062	0.0364	0.0955	0.1421	**0.1619**
22	**1.20**	**-0.50**	**0.90**	**---**	0.4549	**0.7459**	0.6585	0.4605	0.0500	0.4703	0.4895	0.6768
23	**0.30**	**-0.20**	**-0.70**	**---**	0.1836	0.1829	0.1896	0.1948	0.0894	0.2044	0.2391	**0.2412**
	Average				0.2339	0.2741	0.2773	0.2503	0.1174	0.2511	0.2995	**0.3257**

Table 7 shows the power levels based on a 5% significance level, when data are generated from an ARMA(2,2) process and an ARMA(1,1) model is fitted, with *n* = 100 and *m* = 10. Table 7 demonstrates that the $\tilde{Q}_{\mathrm{EXM}}$ test is more powerful than other portmanteau tests in 7 cases, but the $\tilde{Q}_{\mathrm{EXLB}}$ test is better in 4 cases, and the $\tilde{Q}_{\mathrm{EXLB}}$ and $\tilde{Q}_{\mathrm{M}}$ tests are best in 1 case each. The average value has been taken for each test, which illustrates that the $\tilde{Q}_{\mathrm{EXM}}$ test is better than other tests. This means that the new $\tilde{Q}_{\mathrm{EXM}}$ test is more powerful than other tests when data are fitted under an ARMA(1,1) model with $\mathrm{m} = 10$.

Table 8: Power level of the test statistics $\tilde{Q}_{\mathrm{LB}}$, $\tilde{Q}_{\mathrm{M}}$, $\tilde{Q}_{\mathrm{MM}}$, $\tilde{Q}_{\mathrm{FGLB}}$, $\tilde{Q}_{\mathrm{GFK}}$, $\tilde{Q}_{\mathrm{GFD}}$, $\tilde{Q}_{\mathrm{EXLB}}$ and $\tilde{Q}_{\mathrm{EXM}}$ when data are generated under various alternative ARMA(2,2) processes, and fitted by an ARMA(1,1) model $m = 20$.

					$m = 20$							
Model	ϕ_1	ϕ_2	θ_1	θ_2	$\tilde{Q}_{LB}$	$\tilde{Q}_{M}$	$\tilde{Q}_{MM}$	$\tilde{Q}_{FGLB}$	$\tilde{Q}_{GFK}$	$\tilde{Q}_{GFD}$	$\tilde{Q}_{EXLB}$	$\tilde{Q}_{EXM}$
4	**0.10**	**0.30**	**---**	**---**	0.2440	0.1774	0.2106	0.2724	0.2720	**0.3615**	0.3525	0.3247
5	**1.30**	**-0.35**	**---**	**---**	0.1105	0.0885	0.0704	0.0841	0.0595	0.1131	0.1110	**0.1139**
9	**0.70**	**---**	**0.70**	**-0.15**	0.1194	0.0899	0.0706	0.1026	0.0742	**0.1294**	0.1247	0.1078
10	**0.70**	**0.20**	**0.50**	**---**	0.1449	0.1053	0.1052	0.1409	0.1310	**0.1954**	0.1936	0.1795
12	**0.90**	**-0.40**	**1.20**	**-0.30**	0.2996	0.3140	0.3320	0.3461	0.3472	**0.4625**	0.4367	0.4288
15	**1.10**	**-0.35**	---	---	0.1298	0.1141	0.0883	0.1119	0.0711	0.1185	0.1233	**0.1211**
16	**---**	**---**	**0.80**	**-0.50**	0.2940	0.3063	0.3569	0.3201	0.2812	**0.4711**	0.3888	0.4525
17	**---**	**---**	**-0.60**	**0.30**	0.1451	0.1062	0.0999	0.1168	0.0840	0.1375	0.1448	**0.1472**
20	**0.30**	**---**	**0.80**	**-0.50**	0.3143	0.3168	0.3699	0.3422	0.3120	**0.4997**	0.4138	0.4635
21	**0.80**	**---**	**-0.50**	**0.30**	0.1125	0.0999	0.0795	0.0907	0.0600	0.1115	0.1119	**0.1217**
22	**1.20**	**-0.50**	**0.90**	**---**	0.3466	0.6033	0.6279	0.3828	0.3069	0.4911	0.4428	**0.6838**
23	**0.30**	**-0.20**	**-0.70**	**---**	0.1641	0.1307	0.1295	0.1624	0.1384	**0.2365**	0.2005	0.1938
	Average				0.2021	0.2044	0.2117	0.2061	0.1781	0.2773	0.2537	**0.2782**

Table 8 shows the power levels based on a 5% significance level, when data are generated from an ARMA(2,2) process and an ARMA(1,1) model is fitted, with *n* = 100 and *m* = 20. Table 8 shows that the $\tilde{Q}_{\mathrm{EXM}}$ test is more powerful than other portmanteau tests in 5 cases, but the $\tilde{Q}_{\mathrm{GFD}}$ test is better in 7 cases. The average value calculated for each test illustrates that the $\tilde{Q}_{\text{EXM}}$ test is better than other tests. Generally, the new $\tilde{Q}_{\text{EXM}}$ test is more powerful than other tests in most cases when data fitted under an ARMA(1,1) model with m = 10 or 20.

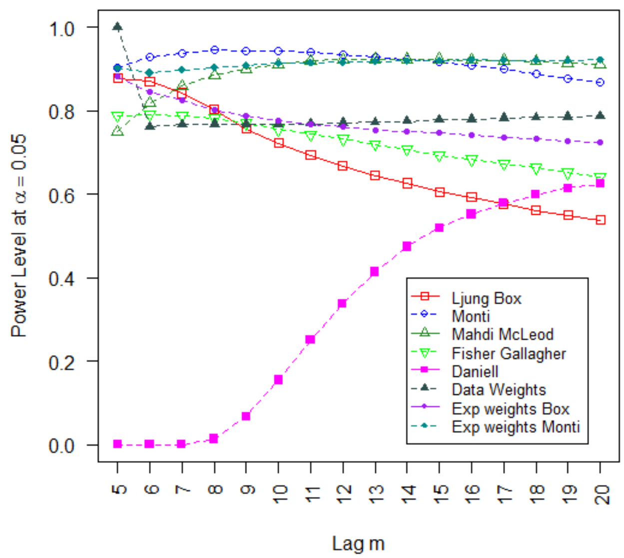

Fig. 5. Power level for lags from 5 to 20 for a fitted ARMA(1,1) model, data generated by an ARMA(2,1) process with ϕ_1=1.2, ϕ_2= -0.5 and θ = -0.9, at 5% significance level, series of length $n = 150$.

Figure 5 shows the power level for lags from 5 to 20 based on a 5% significance level, when data are generated by an ARMA(2,1) process with ϕ_1= 1.2, ϕ_2=-0.5 and $\theta = -0.9$, and fitted under an ARMA(1,1) model with n = 150 and 10,000 replications. The power of the $\tilde{Q}_{\text{LB}}$, $\tilde{Q}_{\text{EXLB}}$ and $\tilde{Q}_{\text{FGLB}}$ tests decreases as the lag increases. The $\tilde{Q}_{\text{M}}$ test increases up to lag 10, then it slowly decreases as the lag increases. The $\tilde{Q}_{\text{MM}}$ test increases as the lag increases. The $\tilde{Q}_{\text{GFK}}$ test is stable at lags 5, 6 and 7, then rapidly increases as the lag increases. The $\tilde{Q}_{\text{GFD}}$ test rapidly decreases at lag 6, then slowly increases as the lag increases further. The power level of the $\tilde{Q}_{\text{EXM}}$ test is approximately constant as the lag increases. In general, the $\tilde{Q}_{\text{EXM}}$ is the most effective test as it is powerful and less influenced by increasing lag size compared to other tests.

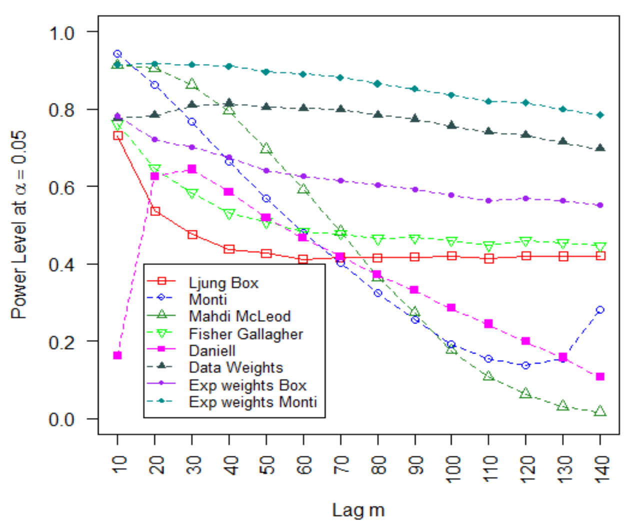

Fig. 6. Power level for maximum lags for a fitted ARMA(1,1) model, data generated by an ARMA(2,1) process with ϕ_1=1.2, ϕ_2= -0.5 and θ = -0.9, at 5% significance level, series of length $n = 150$.

Figure 6 shows the power level for large lags based on a 5% significance level, when data are generated by an ARMA(2,1) process with ϕ_1= 1.2, ϕ_2=-0.5 and $\theta = -0.9$, and fitted under an ARMA(1,1) model with n = 150 and 10,000 replications. The results of Figure 6 are similar to the results of Figure 4, except in the case of the $\tilde{Q}_{\text{GFK}}$ test. The value of the $\tilde{Q}_{\text{GFK}}$ test is less than 0.2 at lag 10, rapidly increasing at lag 20 and then decreasing as lag increases further.

5 Conclusion

The empirical size simulations, see Tables 1 and 2, show that the Monti $\tilde{Q}_{MM}$ test is superior than all other tests when data are generated from an AR(1) process and the parameters are fitted under an AR(1) model. The empirical size simulations, see Figures 1 and 2, shows that portmanteau tests from previous studies and the new weighted autocorrelation test, $\tilde{Q}_{EXLB}$, do not have significance levels that are stable with respect to lag length, particularly for longer lags. However, the proposed exponential weighted partial autocorrelation test, $\tilde{Q}_{\text{EXM}}$, is not affected by lag length. This stability with respect to lag length of the $\tilde{Q}_{\text{EXM}}$, test provides higher reliability compared with other tests.

When considering the power of a test when data are generated from ARMA(2,2) process and fitted under an AR(1) , a MA(1) and a ARMA(1,1) model the new $\tilde{Q}_{EXM}$ test is the most powerful of all the tests examined, see Tables 3, 4, 5, 6, 7 and 8. This is strongly supported by the analysis which shows that it's mean power level, across a range of processes and models, is always the highest. In addition, analysis of the stability of the power levels, see Figures 3, 4, 5 and 6, shows that the $\tilde{Q}_{EXM}$ test is again the most stable with respect to lag lengths. This shows that it has an important feature of being able to reject incorrect models independent of lag length, a feature not found in other portmanteau tests.

Conflict of interest: The authors declare that there is no conflict regarding the publication of this paper.

Acknowledgement

The authors would like to acknowledge the European Regional Development Fund (ERDF) and the Welsh Government for partially funding this study. We would also like to thank all members of the Centre of Excellence in Mobile and Emerging Technologies (CEMET), the University of South Wales, for their contribution in various capacities.

References

[1] G. P. Box and D. A. Pierce, Distribution of residual autocorrelations in autoregressive integrated moving average time series models, *Journal of the American Statistical Association* 65(332): 1509-1526 (1970).

[2] G. M. Ljung and G. E. P. Box. On a measure of lack of fit in time series models, *Biometrika* 65(2): 297-303 (1978).

[3] C. Monti. A proposal for residual autocorrelation test in linear models. *Biometrika* 81(4): 776-780 (1994).

[3] D. Peña and J. Rodríguez. A powerful portmanteau test of lack of fit from time series. *Journal of the American Statistical Association* 97(458): 601-610 (2002).

[4] D. Peña and J. Rodríguez. The log of the autocorrelation matrix for testing goodness of fit in time series, *Journal of statistical planning and inference* 136 (8): 2706-2718 (2006).

[5] E. Mahdi and I. A. McLeod. Improved multivariate portmanteau test, *Journal of Time Series Analysis* 33(2): 211-222 (2012).

[6] T. J. Fisher and C. M. Gallagher, New Weighted Portmanteau Statistics for time series goodness of fit testing, *Journal of the American Statistical Association* 107(498): 777-787 (2012).

[7] C. M. Gallagher and T. J. Fisher, On weighted of portmanteau tests for time-series goodness-of-fit, *Journal of Time Series Analysis* 36: 67-83 (2015).

[8] Y. Hong, Consistent testing for serial correlation of unknown form, *Econometrica* 64 (4): 837–864 (1996a).

[9] Y. Hong, Testing for independence between two covariance stationary time series, *Biometrika* 83(3): 615–625 (1996b).

[10] M. Matsumoto and T. Nishimura. Mersenne Twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation* 8, pp. 3-30 (1998).

[11] G. E. P. Box and M. E. Muller, A note on the generation of random normal deviates, *The Annals of Mathematics Statistics* 29: 610-611(1958).

[12] G. E. P. Box, G. M. Jenkins and G. C. Reinsel, *Time series analysis: Forecasting and control,* 4th ed. Hoboken, NJ: Wiley, 2008.