

Comparative study of portmanteau tests for the residuals autocorrelation in ARMA models

Samir K. Safi¹, Alaa A. Al-Reqeb²

¹Dept. of Economics and Statistics, Faculty of Commerce, the Islamic University of Gaza, Gaza, Palestine

²Dept. of Statistics, Statistician Research, Gaza, Palestine

Email address:

samirsafi@gmail.com (S. K. Safi), alaa.kytc@gmail.com (A. A. Al-Reqeb)

To cite this article:

Samir K. Safi, Alaa A. Al-Reqeb. Comparative Study of Portmanteau Tests for the Residuals Autocorrelation in ARMA Models. *Science Journal of Applied Mathematics and Statistics*. Vol. 2, No. 1, 2014, pp. 1-13. doi: 10.11648/j.sjams.20140201.11

Abstract: The portmanteau statistic for testing the adequacy of an autoregressive moving average (*ARMA*) model is based on the first m autocorrelations of the residuals from the fitted model. We consider some of portmanteau tests for univariate linear time series such as Box and Pierce [2], Ljung and Box [9], Monti [12], Peña and Rodríguez [13 and 14], Generalized Variance Test (Gvtest) by Mahdi and McLeod [11] and Fisher [4]. We conduct an extensive computer simulation time series data, to make comparison among these tests. We consider different model parameters for small, moderate and large samples to examine the effect of lag m on the power of the selected tests, and determine the most powerful test for *ARMA* models. The similar portmanteau tests models was evaluated for the real data set on electricity consumption in Khan Younis, Palestine (April 2009 - May 2013). We found that, portmanteau tests have the highest values of power for large sample data ($N = 500$) comparing to small and moderate samples ($N = 50$ and 200). We found that the portmanteau tests are sensitive to the chosen for m value. Indeed there are loss of the power values for lags m ranging from $m = 5$ to 20 , where Box-Pierce, Ljung-Box and Monti tests have more power loss than the other selected tests. The power loss reaches its minimum values for large sample data comparing to small and moderate samples. In addition, the results of the simulation study and real data analysis showed that the most powerful tests varies between Gvtest and Fisher tests.

Keywords: ARMA Model, Portmanteau Test, Residual Analysis, Autocorrelation, Model Diagnostic, Simulation

1. Introduction and Literature Review

1.1. Introduction

Time series model diagnostic checking is the most important stage of time series model building. In examining the adequacy of a statistical model, an analysis of the residuals is often performed. The study of the distribution of residual autocorrelations in linear time-series models started with the seminal work of Box and Pierce [2]. If the appropriate model has been chosen, there will be zero autocorrelation in the errors and we use one of the portmanteau tests in time series analysis for testing the adequacy of a fitted linear time series model.

A portmanteau test is proposed to test the goodness of fit of *ARMA* models in time series. This test firstly has been studied by Box and Pierce [2], then it has been improved by Ljung and Box [9], this test is known as the Ljung-Box test. Over decades, this test was improved by many statisticians.

In this study we compare the performance of portmanteau tests through an extensive numerical simulation for different model parameters and sample sizes. These simulations examine the sensitivity of choosing model parameters to different sample sizes. In particular, how do these tests perform for different model parameter specifications and for small, moderate and large sample sizes? In addition, determine the most powerful portmanteau test and study the effect of the lags (m) on the power of these tests based on both simulation study and real data set.

1.2. Literature Review

Portmanteau tests have been studied by many authors, for examples, Ljung [8] examined the properties of the portmanteau test statistic for various choices of m lags autocorrelations of the residuals from the fitted *ARMA*

model. A modification which allows the use of small values of m is shown to result in a more powerful test.

Test of goodness of fit for time series models was proposed by Monti [12] based on the sum of squared residuals partial autocorrelations. The test statistic is asymptotically follows a chi-squared distribution. Small sample performance is studied through a Monte Carlo experiment. It appears sensitive to erroneous specifications especially when the fitted model underestimates the order of the moving average component.

Finite-sample performance of Monti's test was investigated by Kwan and Wu [6], paying special attention to its estimated sizes and empirical powers. their simulation results indicate that (i) the test size can be affected by the choice of the number of residual partial autocorrelations, m , and (ii) the empirical powers of the Monti and the Ljung-Box tests are similar in the cases of both seasonal and non-seasonal data if m is properly chosen.

Chand and Kamal [3] compared the performances of Box-Ljung test and Monti's test under different alternative hypothesis using Monte Carlo experiment. They showed that Monti's test shows better approximation to Chi-squared distribution and is at least as good as that of the Ljung-Box statistic. Monti's test provides stable results over different values of lag " m ".

Lin and McLeod [10] noted several problems with the diagnostic test that has been suggested by Peña and Rodríguez [13] and an improved Monte-Carlo version of this test is suggested. It is shown that quite often the test statistic recommended by Peña and Rodríguez [13] may not exist and their asymptotic distribution of the test does not agree with the suggested gamma approximation very well if the number of lags used by the test is small. It is shown that the convergence of this test statistic to its asymptotic distribution may be quite slow when the series length is less than 1000, and so a Monte-Carlo test is recommended. Simulation experiment suggests the Monte-Carlo test is usually more powerful than the test given by Peña and Rodríguez [13] and often much more powerful than the Ljung Box portmanteau test.

Peña and Rodríguez [14] proposed a finite sample modification of their previous test which is $D = n \left(1 - |\hat{R}_m|^{1/m} \right)$, this statistic is based on the determinant of the m -th residual autocorrelation matrix R_m . The new modified test is asymptotically equivalent to the previous one but it has a more intuitive explanation and it can be 25% more powerful for small sample size. The test statistic is the log of the determinant of the m -th autocorrelation matrix.

Two new statistics were introduced by Fisher [4], First: weighted variations of the common Ljung-Box and second: the less-common, Monti statistics for checking the adequacy of a fitted stationary *ARMA* process. The new test statistics put more emphasis (weight) on the first few

autocorrelations those most likely to deviate from zero and hence is more likely to detect the fitted *ARIMA* model.

This paper is organized as follows: Section 2 focuses on properties of portmanteau tests. Section 3 discusses the power of the portmanteau tests based on Monte Carlo study. In Section 4 a monthly consumption of electricity data is analyzed. Section 5 summaries the results and offers suggestions for future research for the residuals autocorrelation in *ARMA* models using portmanteau tests.

2. Portmanteau Tests for ARMA Models

In this section we present Box-Pierce [2], Ljung-Box [9], Monti [12], Peña and Rodríguez [13 and 14], Mahdi and McLeod Gvtest [11] and Fisher [4] portmanteau tests for univariate and linear time series model extensively for this purpose. The null hypothesis assuming that the fitted model is an adequate model and the residuals behave like white noise series.

Most of portmanteau tests are based on the residual autocorrelation coefficient which is provided by:

$$\hat{r}_k = \frac{\sum_{t=k+1}^n \hat{\varepsilon}_t \hat{\varepsilon}_{t-k}}{\sum_{t=1}^n \hat{\varepsilon}_t^2}, \quad (2.1)$$

where $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_t$ are the residuals obtained after estimating the model in a sample of size n , Peña and Rodríguez [14].

2.1. Box and Pierce Portmanteau Test

The classical portmanteau test statistic is the one proposed by Box and Pierce [2].

Definition 2.1. (Box and Pierce [2]) Box-Pierce \tilde{Q}_{BP} test statistic up to lag m is

$$\tilde{Q}_{BP}(\hat{r}) = n \sum_{k=1}^m \hat{r}_k^2, \quad (2.2)$$

where \hat{r}_k is the sample autocorrelation of order k of the residual, m is the number of lags being tested and n is the sample size.

This statistic is used to test for significant correlation up to lag m . It is well known that for independent and identically distributed data, as $n \rightarrow \infty$ the autocorrelations behave as independent normally distributed random variables, and therefore under the null hypothesis (correctly fitted model) \tilde{Q}_{BP} is shown to be asymptotically distributed Chi-squared random variable with $m-p-q$ degrees of freedom, where p and q are the order of autoregressive and moving average terms estimated in the fitted model, respectively.

Kwan, Sim, and Wu [7] indicated that the normalization procedure used in Box-Pierce test is inappropriate for an independent and identically distributed (*iid*) normal series with an unknown mean. Consequently, the poor empirical performance of the test is not entirely unexpected. On other hand, Arranz [1] showed that in finite samples its distribution falls apart from the asymptotic one.

2.2. Ljung-Box Portmanteau Test

After some discussions about the finite sample distribution of the test statistic proposed by Box and Pierce [2]; Ljung and Box [9] proposed a modified version of that test.

Definition 2.2. (Ljung and Box [9]) Ljung-Box \tilde{Q}_{LP} portmanteau test is

$$\tilde{Q}_{LB}(\hat{r}) = n(n+2) \sum_{k=1}^m \frac{\hat{r}_k^2}{n-k}, \tag{2.3}$$

where n, m, \hat{r}_k as in Definition 2.1.

Ljung and Box [9] showed that their test provides a substantially improved approximation to Chi-squared distribution with $m-p-q$ degrees of freedom that should be adequate for most practical purposes with the same critical region as \tilde{Q}_{BP} . In many applications the value of m has been as high as 20 or 30 even when a simple low-order model has been believed to be appropriate, Ljung [8].

2.3. Monti Portmanteau Test

The tests in Definitions 2.1 and 2.2 are based on residual autocorrelation; alternatively Monti [12] proposed a test statistic based on residual partial autocorrelation.

Definition 2.3 (Monti [12]) Let $\hat{\pi}_k$ be the k th residual partial autocorrelation, then Monti portmanteau test up to lag m is provided by:

$$\tilde{Q}_M(\hat{\pi}) = n(n+2) \sum_{k=1}^m \frac{\hat{\pi}_k^2}{n-k}, \tag{2.4}$$

where n is the length of the time series.

Monti [12] proved that if the model is correctly identified, \tilde{Q}_M is asymptotically distributed as a χ_{m-p-q}^2 random variable. If the model is correctly specified, $\hat{\pi}_k$ is approximately distributed as normal with mean zero and variance $\frac{n-k}{n(n+2)}$, Kwan and Wu [6].

One can alternatively uses the statistic:

$$Q_M^*(\hat{\pi}) = n \sum_{k=1}^m \hat{\pi}_k^2 \tag{2.5}$$

where Q_M^* is asymptotically equivalent to \tilde{Q}_M . The difference between Q_M^* and \tilde{Q}_M is the same as between

the Box-Pierce test \tilde{Q}_{BP} and Ljung-Box test \tilde{Q}_{LP} . In both cases, the approximation to the small-sample distribution by a Chi-squared is more accurate for the latter versions, which is therefore recommended, Monti [12].

In addition, Monti [12] showed by simulations that the performance of \tilde{Q}_M is comparable to that of \tilde{Q}_{LB} and better if the order of the moving average is understated. On the other hand, \tilde{Q}_{LP} performs better if the order of the autoregressive part is understated.

2.4. Peña and Rodríguez Portmanteau Test (2002)

The estimated residuals can be considered as a sample of multivariate data from some distribution, Peña and Rodríguez [13] interested in testing the adequacy for the ARMA models on a statistic based on the determinant of the residual autocorrelation matrix:

$$\hat{R}_m = \begin{pmatrix} 1 & \hat{r}_1 & \cdots & \hat{r}_m \\ \hat{r}_1 & 1 & \cdots & \hat{r}_{m-1} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{r}_m & \cdots & \hat{r}_1 & 1 \end{pmatrix}, \tag{2.6}$$

where \hat{r}_k is the sample autocorrelation of order k of the residual. So, under the null hypothesis; testing for model adequacy is equivalent to testing if \hat{R}_m is approximately the identity matrix. Thus, it is sensible to explore a test based on this statistic.

Definition 2.4. (Peña and Rodríguez [13]) For stationary time series data a portmanteau diagnostic test statistic up to the lag m is

$$D = n \left(1 - \left| \hat{R}_m \right|^{\frac{1}{n}} \right), \tag{2.7}$$

where n is the length of the time series.

Peña and Rodríguez [13] showed that if the model is correctly identified D is asymptotically distributed as a linear combination of Chi-squared random variables and is approximately a Gamma distributed random variable for large values of m with parameters α and β , where

$$f(x; \alpha, \beta) = (\Gamma(\alpha))^{-1} \beta^{-\alpha} x^{\alpha-1} \exp(-x\beta^{-1}), 0 < x < \infty, \tag{2.8}$$

where,

$$\alpha = \frac{3m \left[(m+1) - 2(p+q) \right]^2}{2 \left[2(m+1)(2m+1) - 12m(p+q) \right]} \tag{2.9}$$

and

$$\beta = \frac{3m \left[(m+1) - 2(p+q) \right]}{2(m+1)(2m+1) - 12m(p+q)} \tag{2.10}$$

In practice, they recommended the matrix $|\hat{R}_m|$ can be constructed using the standardized residuals as this improves the Gamma distribution approximation by replacing \hat{r}_k with \ddot{r}_k , where $\ddot{r}_k = [(n+2)/(n-k)]^{1/2} \hat{r}_k$. Hence $|\hat{R}_m|$ will be replaced with $|\hat{R}_m|$, Lin and McLeod [10].

As pointed out in Lin and McLeod [10], the statistic D constructed using the standardized residuals are frequently does not exist in practice since the matrix \ddot{R}_m is not always positive definite, so they recommended to concentrate on the original D statistic. They showed that the test based on the Gamma approximation is not conservative, despite the fact that as shown by Peña and Rodríguez [13] the small sample performance is acceptable in some cases, the more general use of tests based on the gamma approximation cannot be recommended. D test statistic may be difficult to implement since it involves calculating the determinant of a matrix, Fisher [4]. For finite sample size, Monti test \tilde{Q}_M is always better than D test, Peña and Rodríguez [14].

In order to improve the properties of D test, Peña and Rodríguez [14] proposed a new test statistic for diagnostics as follows.

2.5. Peña and Rodríguez Portmanteau Test (2006)

A finite sample modification of a test by Peña and Rodríguez [14] is proposed. The new modified test has a more intuitive explanation than the Peña and Rodríguez [13].

Definition 2.5. (Peña and Rodríguez [14]) For stationary time series data a new portmanteau diagnostic test statistic is

$$D^* = -\frac{n}{m+1} \log |\hat{R}_m|, \quad (2.11)$$

the notation as outlined in Definition 2.4.

They showed that the test statistic D^* is asymptotically distributed as a linear combination of Chi-squared random variables and proposed two different approximations to the asymptotic distribution of that test statistic: The first one is based on the Gamma distribution. The test statistic follows asymptotically a Gamma distribution with parameters α and β , where

$$\alpha = \frac{3(m+1)[m-2(p+q)]^2}{2[2m(2m+1)-12(m+1)(p+q)]} \quad (2.12)$$

and

$$\beta = \frac{3(m+1)[m-2(p+q)]}{2m(2m+1)-12(m+1)(p+q)} \quad (2.13)$$

The distribution has mean $\alpha\beta^{-1} = \frac{m}{2} - (p+q)$ and variance $\alpha\beta^{-2} = \frac{m(2m+1)}{3(m+1)} - 2(p+q)$. They denoted this

first approximation by GD^* which is distributed as $G(\alpha, \beta)$

The second approximation is based on Normal distribution. They suggested a power transformation which reduces the skewness in order to improve the normal approximation. The test statistic is

$$ND^* = (\alpha\beta^{-1})^{-\lambda^{-1}} (\alpha^{-\lambda}) \left[D^{*\lambda^{-1}} - (\alpha\beta^{-1})^{\lambda^{-1}} \left(1 - \frac{1}{2\alpha} \left(\frac{\lambda-1}{\lambda^2} \right) \right) \right] \quad (2.14)$$

where

$$\lambda = \left[1 - \frac{2 \left[\frac{m}{2} - (p+q) \right] \left[m^2 / (4(m+1)) - (p+q) \right]}{3 \left[m(2m+1) / (6(m+1)) - (p+q) \right]^2} \right]^{-1} \quad (2.15)$$

For m moderately large we get, $\lambda \approx 4$ and α and β are the values as shown in (2.12) and (2.13). The statistic ND^* is the second approximation which is distributed as $N(0,1)$.

Peña and Rodríguez [14] found that the performance of both approximations, GD^* or ND^* for checking for goodness of fit in linear models is similar. By Monte Carlo study they showed that the new test can be up to 50% more powerful than the Ljung-Box and Monti tests, and for finite sample size is always better than previous one, D .

The new tests do not seem to be affected by the value of m . The statistics D and Ljung-Box, \tilde{Q}_{LB} , have a good size performance but they are much more sensitive to the value of m .

Their simulations showed an improvement in small sample time series, but the Type I error rates appear to be poor. So Mahdi and McLeod [11] improved D^* test such that the degrees of freedom for the Chi-squared approximation allow the improved one to have conservative Type I errors in practice, Fisher and Gallagher [5].

Mahdi and McLeod [11] generalized the results of Peña and Rodríguez [13 and 14] to the multivariate setting. In the univariate case, they recommended the statistic

$$D_m^* = -\frac{3n}{2m+1} \log |\hat{R}_m| \quad (2.16)$$

The null distribution is approximately Chi-squared with $[1.5m(m+1)(2m+1)^{-1} - p - q]$ degrees of freedom and it is implemented in the R function `gvttest()`, so this modified statistic is called as `Gvttest`, since it's a generalized variance portmanteau test based on the determinant matrix, hence we will not include Peña and Rodríguez tests in our

comparisons because of generalization by Mahdi and McLeod [11].

2.6. Fisher Portmanteau Tests

Fisher [4] introduced two new statistics that are weighted variations of the common Ljung-Box and, the less-common, Monti statistics and these two new statistics are easy to implemented.

Definition 2.6. (Fisher [4]) Weighted portmanteau tests are provided by

$$\tilde{Q}_{WL}(\hat{r}) = n(n+2) \sum_{k=1}^m \frac{\hat{r}_k^2 (m-k-1)}{m(n-k)} \tag{2.17}$$

and

$$\tilde{Q}_{WM}(\hat{\pi}) = n(n+2) \sum_{k=1}^m \frac{\hat{\pi}_k^2 (m-k-1)}{m(n-k)}, \tag{2.18}$$

where \hat{r}_k and $\hat{\pi}_k$ as defined in 2.1 and 2.3 respectively.

The two statistics look similar to the Ljung-Box and Monti statistics with the exception a weight $\frac{m-k-1}{m}$ on each autocorrelation or partial autocorrelation. The weights are derived using multivariate analysis techniques on the matrix of autocorrelations or matrix of partial autocorrelations.

Fisher [4] noted that the sample autocorrelation at lag 1, \hat{r}_1 , is given weight $\frac{m}{m} = 1$. The sample autocorrelation at lag 2, \hat{r}_2 , is given weight $\frac{m-1}{m} < 1$. We can interpret the weights as putting more emphasis on the first autocorrelation, and the least emphasis on the autocorrelation at lag m (corresponding weight $\frac{1}{m}$). This matches the intuition about statistical estimators. The first autocorrelation \hat{r}_1 is calculated using information from all n observations. The second autocorrelation \hat{r}_2 is based on $n-1$ observations, and the m -th autocorrelation is based on $n-m$ observations. Intuitively, it makes sense to put more emphasis on the first autocorrelation as it should be the most accurate. This idea also holds true for the partial autocorrelations.

The two statistics are asymptotically distributed as a linear combination of Chi-squared random variables. This is the same asymptotic distribution as the statistics in Peña and Rodríguez [13 and 14]. The weighted Ljung-Box \tilde{Q}_{WL} and weighted Monti \tilde{Q}_{WM} statistics are asymptotically equivalent to D but have the added benefit of easy calculation and computational stability. When a small number of parameters have been fit under the null hypothesis of an adequate model, the statistics \tilde{Q}_{WL} and

\tilde{Q}_{WM} are approximately distributed as Gamma random variables with shape parameter,

$$\gamma = \frac{3[m^2 + m - 2(m-1)(p+q)]^2}{4[2m^3 + 3m^2 + m - 6(m^2 - 2m - 1)(p+q)]}, \tag{2.19}$$

and scale parameter

$$\lambda = \frac{2[2m^3 + 3m^2 + m - 6(m^2 - 2m - 1)(p+q)]}{3[m(m^2 + m - 2(m-1)(p+q))]} \tag{2.20}$$

The Gamma approximation is constructed to have the same theoretical mean and variance as the true asymptotic distribution.

3. Simulation Study

In this section, we consider the robustness of various portmanteau tests. We compare the power among some of the portmanteau tests such as Box-Pierce \tilde{Q}_{BP} , Ljung-Box \tilde{Q}_{LB} , Monti \tilde{Q}_M , Gvtest D_m^* , Weighted Ljung-Box \tilde{Q}_{WL} , and Weighted Monti \tilde{Q}_{WM} . Monte Carlo simulations will be conducted to examine the sensitivity of the selected portmanteau tests to model diagnostics. In particular, which is the most appropriate test for examining the adequacy of linear and nonseasonal *ARMA* models. This section displays the results of simulation study.

Three finite sample sizes (50, 200, and 500) are generated from different *ARMA(p,q)* models with different values of the model parameters, where p and $q \leq 2$ and *ARIMA(1,0,0)* is fitted for each model. In each case 1000 of Monte-Carlo simulations with 1000 replications were generated by R statistical software package and the power of the tests were computed for selected lags m for 0.05 nominal α level.

3.1. Simulation Study for Small Data

We generated small sample size data (N=50) by R statistical software and the power of the tests are computed for $m = 5, 10, 15, 20$. The Results are presented in Table 3.1 for $m = 5$ and 20. The other choices of m are shown in Appendix A, Table A.1.

The simulation results reveal that all portmanteau tests are sensitive to the choice of lag m and reach its maximum at lag $m = 5$. For an example, in model 3 *MA(1)*, there are deficiencies of the power of the portmanteau test \tilde{Q}_{WM} from lags 5 to lag 10 which equals to $\frac{0.293 - 0.245}{0.293} \times 100\% = 16.3\%$ and similar calculations for other models. Then, the averages of the power decrease with respect to m from lag 5 to 10 by 21.2%, 21.1%, 19.8%,

13.3%, 18.6% and 15.1% for \tilde{Q}_{BP} , \tilde{Q}_{LB} , \tilde{Q}_M , D_m^* , \tilde{Q}_{WL} and \tilde{Q}_{WM} , respectively with exception in model 5.

For $AR(p)$ and $MA(q)$, the simulation results show that the performance of \tilde{Q}_M is better than \tilde{Q}_{LB} if the order of the moving average component is underestimated (see models 1, 2, 3, 6 and 7 for all selected lags). On the other hand, \tilde{Q}_{LB} performs better than \tilde{Q}_M if the order of the autoregressive component is underestimated (see models 4 and 5 for all lags).

To further investigate the power of the portmanteau tests, the simulation results indicate that the power of the test is increased -on average- by 23.6% when using \tilde{Q}_{WL} , which is proposed by Fisher [4], instead of \tilde{Q}_{LB} . In addition, the power of the test is increased -on average- by 25.0% when using \tilde{Q}_{WM} instead of \tilde{Q}_M .

Based on the simulation results, the difference of the power test between D_m^* and \tilde{Q}_{WM} is very small (approximately 0.009 only in average) and is smaller than the difference of the power test between D_m^* and \tilde{Q}_{WL} tests (approximately 0.05 only in average), with exception in model 5.

Furthermore, the simulation results reveal that Gvtest D_m^* is the most powerful test for the most selected models for large lags ($m = 15$ and 20), whereas Weighted Ljung-Box \tilde{Q}_{WL} and Weighted Monti \tilde{Q}_{WM} outperform and more powerful than D_m^* for moderate lags ($m=5$ and 10).

The Gvtest D_m^* seems to be sensitive for model 5, $AR(2)$ with parameters 1.2 and -0.73 for all lags in small data ($N=50$). The test statistic with the highest power for any particular model is shown in bold font.

Table 3.1. Powers of portmanteau tests for $N = 50$, $m=5$ and 20 , and $\alpha = 0.05$

Model	ϕ_1	ϕ_2	θ_1	θ_2	\tilde{Q}_{BP}	\tilde{Q}_{LP}	\tilde{Q}_M	D_m^*	\tilde{Q}_{WL}	\tilde{Q}_{WM}
<i>m=5</i>										
1			0.7		0.428	0.417	0.548	0.685	0.571	0.682
2			0.4		0.140	0.136	0.138	0.196	0.172	0.193
3			-0.5		0.200	0.197	0.209	0.294	0.260	0.293
4	0.6	0.3			0.252	0.244	0.231	0.300	0.325	0.300
5	1.2	-0.73			0.997	0.997	0.995	0.143	0.999	0.999
6			1	-0.6	0.726	0.712	0.755	0.816	0.817	0.816
7			0.24	0.1	0.117	0.115	0.127	0.177	0.154	0.163
8	0.8		0.4		0.095	0.088	0.093	0.116	0.114	0.115
9	0.5		-0.7		0.700	0.676	0.879	0.944	0.864	0.943
10	-0.2		-0.6		0.186	0.183	0.193	0.293	0.233	0.296
11	0.7	0.2	-0.5		0.234	0.210	0.272	0.406	0.336	0.402
12	1.3	-0.35	0.1		0.287	0.295	0.276	0.414	0.406	0.404
13	0.4		-0.6	0.3	0.580	0.530	0.821	0.881	0.766	0.882
14	0.9	-0.3	1.3	-0.5	0.172	0.155	0.201	0.281	0.237	0.271
<i>m=20</i>										
1			0.7		0.274	0.241	0.327	0.525	0.319	0.490
2			0.4		0.112	0.105	0.106	0.137	0.125	0.122
3			-0.5		0.157	0.149	0.117	0.205	0.179	0.187
4	0.6	0.3			0.156	0.142	0.117	0.204	0.190	0.185
5	1.2	-0.73			0.974	0.943	0.930	0.116	0.986	0.988
6			1	-0.6	0.496	0.441	0.484	0.716	0.622	0.696
7			0.24	0.1	0.096	0.085	0.086	0.120	0.106	0.108
8	0.8		0.4		0.071	0.074	0.067	0.076	0.070	0.071
9	0.5		-0.7		0.444	0.382	0.596	0.859	0.655	0.829
10	-0.2		-0.6		0.133	0.116	0.108	0.197	0.146	0.154
11	0.7	0.2	-0.5		0.167	0.142	0.139	0.272	0.186	0.238
12	1.3	-0.35	0.1		0.212	0.202	0.168	0.276	0.250	0.250
13	0.4		-0.6	0.3	0.353	0.310	0.529	0.792	0.442	0.769
14	0.9	-0.3	1.3	-0.5	0.133	0.102	0.100	0.175	0.124	0.149

3.2. Simulation Study for Moderate Data

We generated moderate sample size data ($N=200$) by R statistical software and the power of the tests are computed for $m = 5, 10, 15, 20$. The Results are presented in Table

3.2 for $m = 5$ and 15 . The other choices of m are shown in Appendix A, Table A.2.

The simulation study shows that the power of the portmanteau tests is more powerful for $N=200$ than $N = 50$. The averages of the power decrease with respect to m from lag 5 to 10 by 13.3%, 14%, 11.2%, 4.9%, 6.7% and 5.3%

for \tilde{Q}_{BP} , \tilde{Q}_{LB} , \tilde{Q}_M , D_m^* , \tilde{Q}_{WL} and \tilde{Q}_{WM} , respectively with some exceptions in 5, 6, 9 and 13, where all tests have the same highest power for most of lags with approximately 100%.

Table 3.2 shows that \tilde{Q}_{LB} tends to perform better than \tilde{Q}_M when the fitted model underestimates the order of the autoregressive component (see model 4 for all lags in Tables 3.2 and A.2). Conversely, if the fitted model underestimates the order of the moving average component then \tilde{Q}_M does better than \tilde{Q}_{LB} (see models 1, 2, 3, 6 and 7 for all lags in Tables 3.2 and A.2). The results reveal that the Weighted Ljung-Box \tilde{Q}_{WL} test raises the power of the

classical one \tilde{Q}_{LB} test by 25.0% and Weighted Monti \tilde{Q}_{WM} test raises the power of Monti \tilde{Q}_M test by 25.1% with some exceptions in models 5, 6, 9, and 13, where all tests have the same highest power for most of lags with approximately 100%.

Furthermore, we notice that the highest power of the tests for other models varies between D_m^* , \tilde{Q}_{WL} and \tilde{Q}_{WM} for lags 5 and 10 such that the power differences between these tests are small (approximately 0.008 only in average). In general, the test statistic D_m^* is the most powerful test for lags $m=15$ and 20. The test statistic with the highest power for any particular model is shown in bold font.

Table 3.2. Powers of portmanteau tests for $N = 200$, $m=5$ and 15, and $\alpha = 0.05$

Model	ϕ_1	ϕ_2	θ_1	θ_2	\tilde{Q}_{BP}	\tilde{Q}_{LP}	\tilde{Q}_M	D_m^*	\tilde{Q}_{WL}	\tilde{Q}_{WM}
<i>m=5</i>										
1			0.7		0.997	0.997	0.999	1	1	0.999
2			0.4		0.356	0.349	0.381	0.497	0.477	0.497
3			-0.5		0.666	0.665	0.721	0.840	0.812	0.839
4	0.6	0.3			0.920	0.919	0.914	0.956	0.962	0.956
5	1.2	-0.73			1	1	1	1	1	1
6			1	-0.6	1	1	1	1	1	1
7			0.24	0.1	0.373	0.372	0.385	0.503	0.492	0.509
8	0.8		0.4		0.570	0.568	0.602	0.660	0.643	0.660
9	0.5		-0.7		1	1	1	1	1	1
10	-0.2		-0.6		0.663	0.661	0.707	0.834	0.788	0.814
11	0.7	0.2	-0.5		0.836	0.851	0.891	0.942	0.927	0.948
12	1.3	-0.35	0.1		0.824	0.819	0.823	0.930	0.922	0.928
13	0.4		-0.6	0.3	1	1	1	1	1	1
14	0.9	-0.3	1.3	-0.5	0.830	0.822	0.830	0.932	0.920	0.931
<i>m=15</i>										
1			0.7		0.897	0.885	0.989	1	0.993	0.999
2			0.4		0.212	0.204	0.277	0.363	0.332	0.335
3			-0.5		0.422	0.407	0.471	0.684	0.619	0.679
4	0.6	0.3			0.761	0.751	0.734	0.916	0.907	0.911
5	1.2	-0.73			1	1	1	1	1	1
6			1	-0.6	0.998	0.997	1	1	1	1
7			0.24	0.1	0.220	0.212	0.233	0.366	0.335	0.356
8	0.8		0.4		0.346	0.335	0.348	0.562	0.529	0.553
9	0.5		-0.7		1	1	1	1	1	1
10	-0.2		-0.6		0.399	0.387	0.457	0.689	0.604	0.679
11	0.7	0.2	-0.5		0.595	0.573	0.701	0.867	0.796	0.886
12	1.3	-0.35	0.1		0.667	0.656	0.627	0.854	0.83	0.820
13	0.4		-0.6	0.3	0.997	0.995	1	1	1	1
14	0.9	-0.3	1.3	-0.5	0.609	0.578	0.621	0.844	0.795	0.839

3.3. Simulation Study for Large Data

We generated moderate sample size data (N=500) by R statistical software and the power of the tests are computed for $m = 5, 10, 15, 20$. The Results are presented in Table 3.3 for $m = 5$ and 10. The other choices of m are shown in Appendix A, Table A.3.

Table 3.3 shows that all tests appear sensitivity to the choice of the value m and reach its maximum at lag $m = 5$. The averages of the power decrease with respect to m from

lag 5 to 10 by 3.6%, 3.6%, 2.8%, 1.1%, 1% and 1% for \tilde{Q}_{BP} , \tilde{Q}_{LB} , \tilde{Q}_M , D_m^* , \tilde{Q}_{WL} and \tilde{Q}_{WM} , respectively.

For $AR(p)$ and $MA(q)$, the performance of \tilde{Q}_M is better than \tilde{Q}_{LB} if the order of the moving average component is underestimated. Contrariwise, \tilde{Q}_{LB} performs better than \tilde{Q}_M if the order of the autoregressive component is

underestimated. However, \tilde{Q}_{LB} and \tilde{Q}_M are comparable for models 1, 4, 5, and 6 for lags from 5 to 15.

In addition, \tilde{Q}_{WL} test increases the power only by 7.9% only in average when we replaced \tilde{Q}_{LB} by \tilde{Q}_{WL} , and 6.9% when replacing \tilde{Q}_M by \tilde{Q}_{WM} , with exceptions in models 2 and 7 where the power slightly greater than 6.9%. According to Monti Carlo study the differences in power

between D_m^* , \tilde{Q}_{WL} and \tilde{Q}_{WM} tests are very small around 0.003 only unlike previous samples. In general, D_m^* , \tilde{Q}_{WL} and \tilde{Q}_{WM} statistics have the most powerful tests. The test statistic with the highest power for any particular model is shown in bold font.

Table 3.3. Powers of portmanteau tests for $N = 500$, $m=5$ and 10, and $\alpha = 0.05$

Model	ϕ_1	ϕ_2	θ_1	θ_2	\tilde{Q}_{BP}	\tilde{Q}_{LP}	\tilde{Q}_M	D_m^*	\tilde{Q}_{WL}	\tilde{Q}_{WM}
<i>m=5</i>										
1			0.7		1	1	1	1	1	1
2			0.4		0.747	0.747	0.782	0.873	0.841	0.864
3			-0.5		0.989	0.989	0.992	0.998	0.995	0.996
4	0.6	0.3			1	1	1	1	1	1
5	1.2	-0.73			1	1	1	1	1	1
6			1	-0.6	1	1	1	1	1	1
7			0.24	0.1	0.791	0.790	0.803	0.895	0.885	0.895
8	0.8		0.4		0.977	0.967	0.978	0.979	0.991	0.979
9	0.5		-0.7		1	1	1	1	1	1
10	-0.2		-0.6		0.988	0.988	0.991	0.999	0.996	0.998
11	0.7	0.2	-0.5		1	1	1	1	1	1
12	1.3	-0.35	0.1		0.998	0.997	0.998	0.999	1	1
13	0.4		-0.6	0.3	1	1	1	1	1	1
14	0.9	-0.3	1.3	-0.5	0.999	0.999	1	1	1	1
<i>m=10</i>										
1			0.7		1	1	1	1	1	1
2			0.4		0.598	0.590	0.682	0.814	0.800	0.810
3			-0.5		0.957	0.957	0.969	0.993	0.988	0.992
4	0.6	0.3			1	1	1	1	1	1
5	1.2	-0.73			1	1	1	1	1	1
6			1	-0.6	1	1	1	1	1	1
7			0.24	0.1	0.659	0.655	0.664	0.832	0.821	0.831
8	0.8		0.4		0.928	0.927	0.942	0.975	0.987	0.977
9	0.5		-0.7		1	1	1	1	1	1
10	-0.2		-0.6		0.947	0.947	0.957	0.994	0.99	0.994
11	0.7	0.2	-0.5		0.996	0.997	0.999	1	1	1
12	1.3	-0.35	0.1		0.995	0.993	0.991	0.998	1	1
13	0.4		-0.6	0.3	1	1	1	1	1	1
14	0.9	-0.3	1.3	-0.5	0.992	0.994	0.998	0.999	0.999	0.999

4. Illustrative Example: Consumption of Electricity Data

To validate the simulation results, we apply the portmanteau tests for testing the adequacy of the fitted model on real data set. We consider the monthly consumption of electricity (in kilowatt-hours, KWH) in

Khan Younis city, Palestine, from April 2009 through May 2013. R-statistical software is used for fitting ARMA model for the time series.

4.1. Data Exploration

Figure 4.1 displays the time series plot. The series displays considerable fluctuations over time, and a stationary model does not seem to be reasonable. The

higher values display considerably more variation than the lower values.

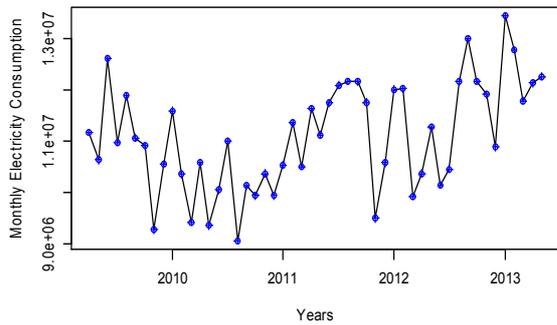


Figure 4.1. Monthly Electricity Consumption, April 2009 through May 2013

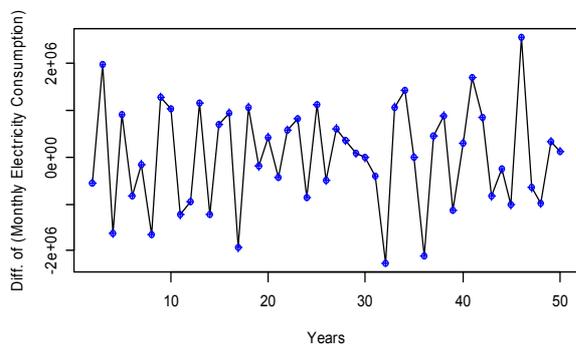


Figure 4.2. The Difference Series of the Monthly Electricity Consumption

In addition, software implementation of the ADF test for stationarity is applied to the original consumption leads to a test statistic of -2.4231 and a p -value of 0.4046 . With nonstationarity as the null hypothesis, this provides strong evidence supporting the nonstationarity and the appropriateness of taking a difference of the original series.

The differences of the electricity values are displayed in Figure 4.2. The differenced series looks much more stationary when compared with the original time series shown in Figure 4.1. On the basis of this plot, we might well consider a stationary model as appropriate.

Table 4.1. P-values of the portmanteau tests of the residuals for ARIMA (1,1,0) Model

Model	m	\tilde{Q}_{BP}	\tilde{Q}_{LP}	\tilde{Q}_M	D_m^*	\tilde{Q}_{WL}	\tilde{Q}_{WM}
ARIMA(1,1,0)	5	0.0363	0.0248	0.0126	0.0044	0.0051	0.0013
	10	0.2002	0.1330	0.0738	0.0189	0.0430	0.0165
	15	0.1519	0.0486	0.0148	0.0241	0.0582	0.0154
	20	0.3675	0.1421	0.0556	0.0357	0.0692	0.0176

4.3. Fitting an Appropriate Model

The sample EACF computed on the first differences of the electricity consumption series is shown in Table 4.2. In this table, an $ARMA(p,q)$ process will have a theoretical

ADF test is applied to the differenced series leads to a test statistic of -5.0478 and a p -value of 0.01 . That is, we reject the null hypothesis of nonstationarity.

4.2. Fitting an Inappropriate ARIMA(1,1,0) Model

Now, suppose the difference of the electricity consumption time series data is fitted wrongly by $ARIMA(1,1,0)$ model. We perform the aforementioned tests to the $ARIMA(1,1,0)$ fitted model. Table 4.1 shows the p -values of the portmanteau tests for selected lags, $m=5, 10, 15,$ and 20 with $\alpha = 0.05$. Clearly the results indicate that for lag 5, all portmanteau tests reject the null hypothesis of adequacy model, i.e. all portmanteau tests have the same result for autocorrelation diagnostics.

In addition, for more clarification if one uses Box-Pierce \tilde{Q}_{BP} test for model adequacy diagnostics, although for $m = 5$, \tilde{Q}_{BP} test shows that the fitted $ARIMA(1,1,0)$ model appears to be inadequate, however for $m = 10, 15$ and 20 the \tilde{Q}_{BP} test indicates significant evidence to support the null hypothesis of model adequacy. While the classic Ljung-Box \tilde{Q}_{LB} and Monti \tilde{Q}_M tests give unstable p -values across m , since for $m = 5$ and 15 , \tilde{Q}_{LB} and \tilde{Q}_M show that the fitted $ARIMA(1,1,0)$ model appears to be inadequate model, whereas for $m = 10$ and 20 , these tests support the model adequacy. Likewise, Weighted Ljung-Box test \tilde{Q}_{WL} suggests an inadequate model for $m = 5$ and 10 and an adequate model for $m = 15$ and 20 .

Furthermore, the analysis shows that Gvtest D_m^* and Weighted Monti \tilde{Q}_{WM} portmanteau tests tend to be the most powerful statistics in detecting inadequacy fitted model, the corresponding p -values for all selected lags are small enough to reject the null hypothesis of adequacy model. Therefore, Gvtest D_m^* and Weighted Monti \tilde{Q}_{WM} portmanteau tests are recommended for autocorrelation diagnostics.

pattern of a triangle of zeroes, with the upper left-hand vertex corresponding to the $ARMA$ orders.

Table 4.2 displays the schematic pattern for an $MA(1)$ model. The upper left-hand vertex of the triangle of zeros is marked with the symbol 0^* and is located in the $p = 0$ row and $q = 1$ column, an indication of an $MA(1)$ model.

The model for the original electricity consumption series would then be a nonstationary $ARIMA(0,1,1)$ model.

We perform the aforementioned portmanteau tests to the $ARIMA(0,1,1)$ model. Table 4.3 shows the p -values of the portmanteau tests for selected lags, $m=5, 10, 15,$ and 20

with $\alpha = 0.05$. Clearly the result indicates that for all lags, all portmanteau tests have insufficient evidence to reject the null hypothesis of adequacy model, i.e. all portmanteau have the same result for autocorrelation diagnostics and detect the fitted model correctly.

Table 4.2. EACF for Difference of Electricity Consumption Series

		MA													
		0	1	2	3	4	5	6	7	8	9	10	11	12	13
AR	0	X	0*	0	0	0	0	0	0	0	0	0	0	0	0
	1	X	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	X	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	0	X	0	0	0	0	0	0	0	0	0	0	0	0
	4	X	X	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	X	0	0	0	0	0	0	0	0	0	0	0	0
	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	7	X	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 4.3. P-value of the portmanteau tests of the residuals for $ARIMA(0,1,1)$ Model

Model	m	\tilde{Q}_{BP}	\tilde{Q}_{LP}	\tilde{Q}_M	D_m^*	\tilde{Q}_{WL}	\tilde{Q}_{WM}
$ARIMA(0,1,1)$	5	0.5851	0.5369	0.5999	0.5099	0.7073	0.7418
	10	0.6392	0.4886	0.4291	0.7531	0.8245	0.8358
	15	0.6216	0.3595	0.2226	0.6589	0.6839	0.6119
	20	0.6329	0.5617	0.4097	0.6615	0.6515	0.5221

Overall, in case of choosing an inappropriate and appropriate fitted models, Gvtest and Weighted Monti portmanteau tests detect the model correctly in both cases for all selected lags. Therefore, for all circumstances Gvtest and Weighted Monti portmanteau tests are recommended to use for model diagnostic phase.

5. Summary and Future Research

In this section, we introduce summary of the results and offer suggestions for future research for the residuals autocorrelation diagnostics in $ARMA$ models using portmanteau tests.

5.1. Summary

We compare six portmanteau tests for goodness-of-fit for $ARMA$ time series models. Using Monti Carlo simulation, we found that these test have the highest values of power for large sample data ($N=500$) comparing to small and moderate samples ($N = 50,$ and 200).

The study concluded that these tests are sensitive to the m values, such that there are loss in the power with respect to m ranging from 5 to 20. The power loss reaches its smallest values for large sample data comparing to small and moderate samples. For $N = 50$ and $200,$ Gvtest

and Weighted Monti test were not affected by the m values as large as the other tests and the same behavior for Weighted Monti and Weighted Box for $N = 500$.

For $AR(p)$ and $MA(q),$ the performance of \tilde{Q}_M test is better than \tilde{Q}_{LB} test if the order of the moving average component is underestimated, whereas, \tilde{Q}_{LB} performs better if the order of the autoregressive component is underestimated. Similar behavior of \tilde{Q}_M and \tilde{Q}_{LB} tests was also noticed by Monti [12].

Fisher and Gallagher [5] compared Fisher tests [4] and Gvtest D_m^* for $N=100$ at lag $m = 20$ and showed that one of the two Weighted statistics is always produce the most powerful test, or tied for most powerful test.

For $N=50$ and $200,$ we showed that the highest power of the tests varies between Gvtest D_m^* and Fisher tests for lags $m=5$ and $10.$

In general Gvtest D_m^* statistic outperforms and preferable for lags $m=15$ and $20.$ On other hand, when $N=500$ all tests are almost identical, however, Gvtest D_m^* and Fisher tests outperform for some time series models.

It's interesting to note from real data analysis that Gvtest D_m^* and Weighted Monti tests tend to have more stable p -

values across m and to be the most powerful tests in detecting the goodness of fit for $ARMA$ time series models, hence these tests are strongly recommended to be used.

5.2. Future Research

Many opportunities of future research are available. The plan of the future research on portmanteau tests for the residuals autocorrelation diagnostics in $ARMA$ models can be split into four main areas. First, extension of the portmanteau tests for seasonal, Generalized Autoregressive Conditional Heteroskedasticity ($GARCH$) and Threshold Autoregressive (TAR) models. Second, extend the research to examine the relationship between $ARMA(p,q)$ coefficients and the power of the tests for testing the goodness-of- fit tests in time series. Third, one may study a portmanteau test that combines Monti \tilde{Q}_M and Ljung–Box \tilde{Q}_{LB} tests for detecting adequacy fitting for $ARMA(p,q)$ model since \tilde{Q}_{LB} tends to perform better than \tilde{Q}_M when the fitted model underestimates the order of the

autoregressive component and vice versa. Finally, determine the appropriate range for lag m to each portmanteau test to increase the ability for model misspecification detection.

Acknowledgment

We would like to express the deepest appreciation to Professor Thomas J. Fisher from Department of Mathematics at University of Missouri, Kansas City, USA for his comments and guidance for the simulation study. Also we would also like to express our thanks to Mr. Iyad Ashshamy from Department of Information Technology at the Islamic University of Gaza for his help in MPICH2 programming software in order to accelerate simulation's time. Finally, we deeply grateful to the referees for their valuable comments and suggestions on earlier preliminary of this paper.

Appendix A

Table A.1. Powers of portmanteau tests for $N = 50, m=10$ and 15 , and $\alpha = 0.05$

Model	ϕ_1	ϕ_2	θ_1	θ_2	\tilde{Q}_{BP}	\tilde{Q}_{LP}	\tilde{Q}_M	D_m^*	\tilde{Q}_{WL}	\tilde{Q}_{WM}
$m=10$										
1			0.7		0.298	0.279	0.423	0.606	0.444	0.597
2			0.4		0.117	0.110	0.118	0.156	0.148	0.155
3			-0.5		0.162	0.146	0.166	0.248	0.213	0.245
4	0.6	0.3			0.110	0.183	0.145	0.255	0.256	0.234
5	1.2	-0.73			0.987	0.983	0.981	0.134	0.997	0.996
6			1	-0.6	0.572	0.537	0.651	0.786	0.744	0.784
7			0.24	0.1	0.105	0.104	0.107	0.147	0.122	0.136
8	0.8		0.4		0.072	0.069	0.074	0.092	0.092	0.091
9	0.5		-0.7		0.558	0.510	0.753	0.919	0.735	0.910
10	-0.2		-0.6		0.148	0.147	0.173	0.244	0.175	0.211
11	0.7	0.2	-0.5		0.192	0.168	0.192	0.335	0.234	0.320
12	1.3	-0.35	0.1		0.232	0.235	0.199	0.340	0.322	0.317
13	0.4		-0.6	0.3	0.428	0.386	0.71	0.866	0.614	0.864
14	0.9	-0.3	1.3	-0.5	0.146	0.117	0.128	0.226	0.177	0.206
$m=15$										
1			0.7		0.282	0.257	0.346	0.560	0.358	0.535
2			0.4		0.116	0.107	0.110	0.144	0.139	0.133
3			-0.5		0.149	0.136	0.140	0.221	0.199	0.208
4	0.6	0.3			0.159	0.145	0.125	0.222	0.221	0.206
5	1.2	-0.73			0.975	0.964	0.956	0.123	0.992	0.992
6			1	-0.6	0.519	0.478	0.531	0.749	0.670	0.738
7			0.24	0.1	0.102	0.094	0.096	0.129	0.109	0.108
8	0.8		0.4		0.069	0.069	0.066	0.083	0.073	0.078
9	0.5		-0.7		0.475	0.421	0.647	0.880	0.643	0.867
10	-0.2		-0.6		0.133	0.122	0.113	0.209	0.158	0.171
11	0.7	0.2	-0.5		0.177	0.150	0.150	0.301	0.201	0.265
12	1.3	-0.35	0.1		0.211	0.208	0.177	0.301	0.283	0.272
13	0.4		-0.6	0.3	0.363	0.325	0.607	0.822	0.506	0.805
14	0.9	-0.3	1.3	-0.5	0.133	0.100	0.115	0.200	0.144	0.170

Table A.2. Powers of portmanteau tests for $N = 200$, $m=10$ and 20 , and $\alpha = 0.05$

Model	ϕ_1	ϕ_2	θ_1	θ_2	\tilde{Q}_{BP}	\tilde{Q}_{LP}	\tilde{Q}_M	D_m^*	\tilde{Q}_{WL}	\tilde{Q}_{WM}
$m=10$										
1			0.7		0.964	0.962	0.994	1	0.997	0.999
2			0.4		0.262	0.255	0.329	0.412	0.393	0.411
3			-0.5		0.501	0.493	0.561	0.751	0.693	0.751
4	0.6	0.3			0.841	0.839	0.837	0.944	0.943	0.942
5	1.2	-0.73			1	1	1	1	1	1
6			1	-0.6	0.999	0.999	1	1	1	1
7			0.24	0.1	0.269	0.261	0.292	0.437	0.412	0.434
8	0.8		0.4		0.432	0.425	0.446	0.619	0.590	0.622
9	0.5		-0.7		1	1	1	1	1	1
10	-0.2		-0.6		0.4999	0.482	0.562	0.768	0.678	0.735
11	0.7	0.2	-0.5		0.678	0.684	0.763	0.911	0.867	0.913
12	1.3	-0.35	0.1		0.726	0.721	0.702	0.883	0.860	0.860
13	0.4		-0.6	0.3	1	1	1	1	1	1
14	0.9	-0.3	1.3	-0.5	0.701	0.680	0.724	0.891	0.848	0.889
$m=20$										
1			0.7		0.849	0.820	0.996	0.997	0.985	0.998
2			0.4		0.190	0.185	0.220	0.324	0.297	0.314
3			-0.5		0.381	0.371	0.407	0.639	0.557	0.626
4	0.6	0.3			0.733	0.696	0.674	0.886	0.866	0.878
5	1.2	-0.73			1	1	1	1	1	1
6			1	-0.6	0.996	0.996	0.998	1	1	1
7			0.24	0.1	0.179	0.170	0.188	0.319	0.289	0.307
8	0.8		0.4		0.320	0.304	0.302	0.507	0.472	0.500
9	0.5		-0.7		0.995	0.991	0.997	1	1	1
10	-0.2		-0.6		0.355	0.334	0.378	0.650	0.531	0.618
11	0.7	0.2	-0.5		0.524	0.504	0.605	0.828	0.742	0.857
12	1.3	-0.35	0.1		0.607	0.595	0.565	0.808	0.78	0.784
13	0.4		-0.6	0.3	0.981	0.976	1	1	1	1
14	0.9	-0.3	1.3	-0.5	0.544	0.512	0.561	0.808	0.739	0.797

Table A.3. Powers of portmanteau tests for $N = 500$, $m=15$ and 20 , and $\alpha = 0.05$

Model	ϕ_1	ϕ_2	θ_1	θ_2	\tilde{Q}_{BP}	\tilde{Q}_{LP}	\tilde{Q}_M	D_m^*	\tilde{Q}_{WL}	\tilde{Q}_{WM}
$m=15$										
1			0.7		1	1	1	1	1	1
2			0.4		0.493	0.484	0.574	0.742	0.712	0.729
3			-0.5		0.895	0.890	0.736	0.990	0.980	0.985
4	0.6	0.3			1	0.999	1	1	1	1
5	1.2	-0.73			1	1	1	1	1	1
6			1	-0.6	1	1	1	1	1	1
7			0.24	0.1	0.551	0.548	0.573	0.788	0.765	0.784
8	0.8		0.4		0.867	0.863	0.905	0.960	0.974	0.967
9	0.5		-0.7		1	1	1	1	1	1
10	-0.2		-0.6		0.906	0.904	0.919	0.981	0.971	0.984
11	0.7	0.2	-0.5		0.985	0.985	0.996	1	1	1
12	1.3	-0.35	0.1		0.992	0.992	0.986	0.998	1	1
13	0.4		-0.6	0.3	1	1	1	1	1	1
14	0.9	-0.3	1.3	-0.5	0.986	0.989	0.992	0.999	0.999	0.999
$m=20$										
1			0.7		1	1	1	1	1	1
2			0.4		0.444	0.431	0.521	0.691	0.662	0.683
3			-0.5		0.840	0.832	0.891	0.981	0.969	0.978
4	0.6	0.3			0.996	0.998	0.997	1	1	1
5	1.2	-0.73			1	1	1	1	1	1
6			1	-0.6	1	1	1	1	1	1
7			0.24	0.1	0.468	0.468	0.492	0.744	0.722	0.737
8	0.8		0.4		0.813	0.808	0.862	0.951	0.966	0.957
9	0.5		-0.7		1	1	1	1	1	1
10	-0.2		-0.6		0.858	0.842	0.878	0.973	0.955	0.977
11	0.7	0.2	-0.5		0.974	0.973	0.989	1	0.998	1
12	1.3	-0.35	0.1		0.980	0.981	0.973	0.998	0.997	0.998
13	0.4		-0.6	0.3	1	1	1	1	1	1
14	0.9	-0.3	1.3	-0.5	0.974	0.975	0.984	0.999	0.996	0.999

References

- [1] Arranz, M.A. (2005). "Portmanteau test statistics in Time Series". Tol-Project.
- [2] Box, G.E.P. and Pierce, D.A. (1970). "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models". *Journal of the American Statistical Association*, Vol. 65(332), pp. 1509-1526.
- [3] Chand, S. and Kamal, S. (2006). "A Comparative Study of Portmanteau Tests for Univariate Time Series Models", *Pakistan Journal of Statistics and Operation Research*, Vol. 2(2), pp. 111-114.
- [4] Fisher, T. J. (2011). "Testing the Adequacy of ARMA Models using a Weighted Portmanteau Test on Residual Autocorrelations". *Contributed Paper 327*, 2011 SAS Global Forum, Las Vegas, NV.
- [5] Fisher, T. J. and Gallagher, C. M. (2012). "New Weighted Portmanteau Statistics for Time Series Goodness of Fit Testing". *Journal of the American Statistical Association*, Vol. 107(498), pp. 777-787.
- [6] Kwan, A.C.C. and Wu, Y. (1997). "Further results on the finite sample distribution of Monti's portmanteau test for the adequacy of an ARMA (p, q) model". *Biometrika*, Vol. 84(3), pp. 733-736.
- [7] Kwan, A.C.C., Sim, A. and Wu, Y. (2005). "A comparative study of finite-sample performance of some portmanteau tests for randomness of a time series", *Computational Statistics and Data Analysis*, Vol. 48, pp. 391-413.
- [8] Ljung, G.M. (1986). "Diagnostic testing of univariate time series models". *Biometrika*, Vol. 73(3), pp. 725-30.
- [9] Ljung, G.M. and Box, G.E.P. (1978). "On a measure of lack of fit in time series models". *Biometrika*, Vol. 65(2), pp. 297-303.
- [10] Lin, J.W. and McLeod, A.I. (2006). "Improved Peña-Rodríguez portmanteau test". *Computational Statistics and Data Analysis*, Vol. 51(3), pp. 1731-1738.
- [11] Mahdi, E. and McLeod, A.I. (2012). "Improved Multivariate Portmanteau Test". *Journal of Time Series Analysis*, Vol. 33(2), pp. 211-222.
- [12] Monti, A.C. (1994). "A Proposal for a Residual Autocorrelation Test in Linear Models". *Biometrika*, Vol. 81(4), pp. 776-790.
- [13] Peña, D. and Rodríguez, J. (2002). "A Powerful Portmanteau Test of Lack of Fit for Time Series". *Journal of the American Statistical Association*, Vol. 97(458), pp. 601-610.
- [14] Peña, D. and Rodríguez, J. (2006). "The log of the determinant of the autocorrelation matrix for testing goodness of fit in time series". *Journal of Statistical Planning and Inference*, Vol. 136(8), pp. 2706-2718.