Kernel-based portmanteau diagnostic test for ARMA time series models

Esam Mahdi*

Abstract: In this paper, the definition of the Toeplitz autocorrelation matrix is used to derive a kernel-based portmanteau test statistic for ARMA models. Under the null hypothesis of no serial correlation, the distribution of the test statistic is approximated by a standard normal using the kernel-based normalized spectral density estimator, without having to specify any alternative model. Unlike most existing portmanteau test statistics, the proposed test is defined for all lags. Simulation studies are conducted to assess the performance of the proposed test. A real application is given to demonstrate the usefulness of this goodness-of-fit test statistic.

Keywords: diagnostic check; goodness-of-fit; weighted portmanteau test; spectral density; kernel estimation; ARMA models

AMS subject classifications:  62M10; 62G07; 91B84

ABOUT THE AUTHOR

Esam Mahdi obtained his PhD in Statistics from Western University, London Ontario, Canada. He has been working as an assistant professor since 2011 in the Department of Mathematics, Islamic University of Gaza, Palestine. Since September 2015, Dr Mahdi was appointed as an adjunct faculty member at the Department of Statistics and Actuarial Sciences in the University of Western Ontario. His research's interests include time series and regression analysis, multivariate data analysis, and spatio-temporal analysis.

PUBLIC INTEREST STATEMENT

One of the most important stages of constructing a statistical model is that of its diagnosis. If the model is well specified then the residuals (the observable estimate of the unobservable statistical errors) serial autocorrelation should be close to zero. The portmanteau test statistics are used as diagnostic tools to check whether any group of autocorrelations of the residual time series are correlated. Based on this concept, and using the autocorrelation matrix defined from Peña and Rodríguez (2002), we derive a portmanteau test statistic. Our test statistic is essentially the same as the one proposed by Peña and Rodríguez (2006); however, they have different distributions. The distribution of our proposed statistic is approximated by a standard normal using the approach described by Hong (1996a). We support the theoretical results with simulations experiments and provide a real application to demonstrate the usefulness of our test statistic in detecting underfit time series models.
1. Introduction
The autoregressive-moving-average, ARMA($p, q$), models for the univariate time series $Z_t$ can be written as

$$
\Phi_p(B)Z_t = \Theta_q(B)\epsilon_t,
$$

where $\Phi_p(B) = 1 - \sum_{i=1}^{p} \phi_i B^i$ and $\Theta_q(B) = 1 + \sum_{i=1}^{q} \theta_i B^i$ are polynomials in $B$ of degrees $p$ and $q$ respectively; $p$ and $q \geq 0$ are the order of the autoregressive, AR, model and moving average, MA, model respectively; $B$ is the backshift operator on $t$; and $(\epsilon_t)$ is a white noise series that is uncorrelated with a mean zero and a finite variance. It is assumed that the model is stationary, invertible with no common roots between $\Phi_p(B)$ and $\Theta_q(B)$ (Box & Jenkins, 1970, 2005; Brockwell & Davis, 2009).

Denote the variance, autocovariance, and autocorrelation functions of $\epsilon_t$ by $\sigma^2 = \text{VAR}(\epsilon_t)$, $\gamma_\ell = \text{cov}(\epsilon_t, \epsilon_{t-\ell}) = \text{E}(\epsilon_t \epsilon_{t-\ell})$, and $r_\ell = \gamma_\ell / \sigma^2$, for time lag $\ell \in \mathbb{Z}$ respectively, the null and alternative hypothesis are

$$
H_0: r_\ell = 0, \quad \text{for all } \ell \neq 0
$$

$$
H_1: r_\ell \neq 0, \quad \text{for some } \ell \neq 0
$$

The normalized spectral density of the stationary process $(\epsilon_t)$ is

$$
f(\omega) = (2\pi)^{-1} \sum_{\ell \in \mathbb{Z}} r_\ell \cos(\ell \omega), \quad \text{where } \omega \in [-\pi, \pi].
$$

The null hypothesis, $H_0$, is equivalent to the null normalized spectral density $f(\omega) = f_0(\omega) = 1/2\pi$ (Anderson, 1993; Hong, 1996a, 1996b).

After fitting the model in Equation (1) to a series of length $n$ and obtaining the residuals, $\hat{\epsilon}_t$, it is important to conduct some goodness-of-fit tests on these residuals to check the validity of the model assumptions (Li, 2004). Under the null hypothesis which indicates that the model has been correctly identified, we expect the residuals to be approximately white noise. When there is no significant autocorrelation in the residuals, their sample autocorrelations, $r_\ell = \sum_{t=1}^{n-\ell} \hat{\epsilon}_t \hat{\epsilon}_{t-\ell} / \sum_{t=1}^{n} \hat{\epsilon}_t^2$, for $\ell = 1, 2, \ldots, m \leq n - 1$, where $m$ is the largest lag considered for autocorrelation. On the other hand, when there is autocorrelation present, the autocorrelation values should significantly deviate from zero.

Since the early 1970s, many research workers including Box and Pierce (1970), Ljung and Box (1978), Monti (1994), Peña and Rodríguez (2002, 2006), Fisher and Gallagher (2012) and Mahdi and McLeod (2012), developed portmanteau test statistics to check the lack of fit of ARMA models; although, the asymptotic distributions of these test statistics were derived based on the assumption that the time-series data are driven by independent and identically distributed (iid) Gaussian innovations, simulation experiments suggested that these tests are asymptotically valid without the Gaussian innovations assumption, taking into account that the zero mean iid and finite fourth moment assumption is kept. In this respect, Hong (1996a, 1996b) implemented the kernel-based normalized spectral density approach to derive three classes of efficient portmanteau test statistics without restrictions on the distribution of the innovations except that they are iid with mean zero and finite fourth moment. Hong (1996a) showed that the distribution of these test statistics are asymptotically normal with better power than the portmanteau tests of Box and Pierce (1970) and Ljung and Box (1978).

In this paper, we devised a portmanteau test statistic for ARMA models using the Toeplitz autocorrelation matrix defined from Peña and Rodríguez (2002). The proposed test statistic is essentially the same to the statistic provided by Peña and Rodríguez (2006); however, they have different distributions. Under the null hypothesis of no serial correlation, we proved that the kernel-based normalized
spectral density estimator described by Hong (1996a) can be used to approximate the distribution of the test statistic by a standard normal, without having to specify any alternative model.

In Section 2, a brief review of commonly univariate portmanteau tests employed for diagnostic checking in ARMA models is given. In Section 3, we propose a kernel-based test statistic using Toeplitz autocorrelation matrix defined from Peña and Rodríguez (2002), where the kernel-based spectral density approach in Hong (1996a) is used to construct its asymptotic distribution. Section 4 provides simulation experiments to demonstrate the behaviour of the proposed test statistic and to compare its power with the Ljung and Box (1978) and Gallagher and Fisher (2015) test statistics. We close this article with Section 5 by introducing an illustrative application of real data demonstrating the usefulness of the proposed test.

2. Portmanteau test statistics for ARMA models
Several works in the time series literature consider the portmanteau tests for diagnosis the adequacy of the fitted ARMA models. In this section, we briefly review the most significant contributed tests.

The well-known portmanteau test statistics are

\[ Q_m = n \sum_{\ell=1}^{m} \hat{r}_\ell^2 \]

\[ Q_m = n(n + 2) \sum_{\ell=1}^{m} (n - \ell)^{-1} \hat{r}_\ell^2 \] (3)

\[ \tilde{Q}_m = n(n + 2) \sum_{\ell=1}^{m} (n - \ell)^{-1} \tilde{r}_\ell^2 \]

where \( Q_m \) was proposed by Box and Pierce (1970), \( Q_m \) was proposed by Ljung and Box (1978), and \( Q_m \) was proposed by Monti (1994), where \( \pi_\ell \) is the residual partial autocorrelation at lag \( \ell \). The asymptotic distribution of each of these three statistics was derived as a chi-squared of \( m - p - q \) degrees of freedom.

Peña and Rodríguez (2002) devised a generalized variance portmanteau test statistic based on the \( m \)th root of the determinant of the Toeplitz Hermitian residual autocorrelation matrix of order \( m + 1 \), \( \tilde{R}_m = [\tilde{r}_{ij}; i, j = 0, 1, \ldots, m] \) where \( \tilde{r}_{ij} = 1 \) for \( i = j \) and \( \tilde{r}_{ij} = \tilde{r}_{ji} \) for \( i \neq j \), i.e. a matrix of the form

\[ \tilde{R}_m = \begin{pmatrix} 1 & \hat{r}_1 & \ldots & \hat{r}_m \\ \hat{r}_1 & 1 & \ldots & \hat{r}_{m-1} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{r}_m & \hat{r}_{m+1} & \ldots & 1 \end{pmatrix} \] (4)

They approximated the distribution of their proposed test statistic by the gamma distribution and provided simulation experiments to demonstrate the improvement of their statistic in comparison with the one given by Ljung and Box (1978). Peña and Rodríguez (2006) modified the previous generalized variance test by taking the log of the \( (m + 1) \)th root of the determinant of \( \tilde{R}_m \) in Equation (4). They approximated the distribution of their modified test using the Gamma and Normal distributions and indicated that the performance of both approximations for diagnostic checking in linear models are similar and more powerful for small sample size than their previous one. The first approximation distributed as a Gamma with parameters \( \alpha = \frac{3(m+1)(m-2p+q)}{2(2m+1)(12m+1)p(q+1)} \) and \( \beta = \frac{3(m+1)(m-2p+q)}{2(2m+1)(12m+1)q+1} \), whereas the second approximation which is distributed as a \( \mathcal{N}(0,1) \) is defined as

\[ (\alpha/\beta)^{1/2}(\lambda/\sqrt{\alpha})[D_m^{1/2} - (\alpha/\beta)^{1/2}(1 - (\lambda - 1)/(2a \lambda^2))] \]
where $D_m^* = \frac{-n}{m+1} \log |\hat{R}_m|$, and $\lambda = \left[ 1 - \frac{2(m/2-(p+q)(q/2)k(m+1)/(6m+1)\cdots(p+q)^2)}{3(m/2m+1)/(6m+1)\cdots(p+q)^2} \right]^{-1}$. Mahdi and McLeod (2012) extended Peña and Rodríguez (2002, 2006) tests to the multivariate time series. The univariate test statistic of Mahdi and McLeod (2012) is

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

(5)

where $\hat{R}_m$ is defined in Equation (4). They demonstrated that the null distribution of $D_m$ is approximately $\chi^2$ with $(3/2)(m(m+1)/(2m+1)) - (p + q)$ degrees of freedom.

Recently, Fisher and Gallagher (2012) provided a portmanteau statistic consisting of a weighted sum of squared of residual autocorrelation terms as follows

$$Q_w = n \sum_{r=1}^{m} w_r r^2.$$  

(6)

They utilized the gamma approximation similar to Peña and Rodríguez (2002) and derived the limiting distribution of their weighted portmanteau tests as a Gamma distribution. More recently, Gallagher and Fisher (2015) suggested to consider three weighting schemes for the weights in Equation (6). The weighting schemes used in their three statistics were: the squared Daniell kernel-based weights as suggested by Hong (1996a, 1996b), $w_r = (n + 2)(n - r)^{-1}K^2(\ell/m)$, the geometrically decaying weights, $w_r = (p + q)a^{r-1}$, for some $0 < a < 1$, and the data-adaptive weights which gives the data-adaptive weights test as follows

$$Q_d = n(n + 2) \sum_{r=1}^{m_0} (n - r)^{-1} \tilde{r}_r^2 + n \sum_{r=m_0+1}^{m} w_r \tilde{r}_r^2,$$

(7)

where the first $m_0$ terms obtain the standardizing weight $(n + 2)/(n - r)$ from the Ljung-Box statistic, and the remaining weights (from $m_0 + 1$ to $m$) selected from the data to be summable $w_r = -\log(1 - |\tilde{r}_r|)$. In practice we select $m_0 = [\log(n)]$ or $m_0 = [p + q + 1]$ where $[x]$ is the floor function denotes the largest integer less than or equal to $x$. Note that if $m_0 \geq m$ then the data-adaptive weights test in Equation (7) will be the same as the classical Ljung-Box test statistic. Gallagher and Fisher (2015) showed that the weighted portmanteau tests can be more powerful to detect the underfit ARMA models in many situations and less sensitive to the choice of the maximum correlation lag, especially when $m$ depends on $n$ comparing with the other statistics from the literature. They highlighted that the data-adaptive weights test statistic inherit desirable properties of both the Ljung-Box (AR underfit power) and Monti (MA underfit power) statistics.

The problem is that, for large order $p + q$ and small lags $m$ of ARMA($p$, $q$) models, the previous mentioned test statistics are not defined.

Hong (1996a, 1996b) addressed the problem of testing the null hypothesis of no serial correlation of unknown distribution for the residual series without having to specify any alternative model. He proposed three classes of portmanteau test statistics using the kernel-based normalized spectral density estimators of $f(\omega)$, here defined

$$\hat{f}_n(\omega) = \frac{1}{2\pi} \sum_{r=-m_{n}}^{m_{n}} k(\ell/m_n) \tilde{r}_r \cos(\ell \omega), \quad \omega \in [-\pi, \pi],$$

(8)

where $k(\cdot)$ is a nonnegative symmetric kernel function and $m_n$ is the bandwidth that depends on the sample size (Priestley, 1981), where $m_n \to \infty$ and $m_n/n \to 0$. The tests measure the distance (quadratic norm, Hellinger metric, and Kullback-Leibler information criterion) between a kernel-based normalized spectral density estimator and the null normalized spectral density, where each of them was approximated as a standard normal distribution.
3. Asymptotic distribution

In this section, under the null hypothesis of no serial correlation, and without having to specify any alternative model, we use the Toeplitz autocorrelation matrix defined from Peña and Rodríguez (2002) to derive a portmanteau test statistic. The portmanteau test statistic is essentially the same as the one proposed by Peña and Rodríguez (2006) and can be seen as Kullback-Leibler discrimination information test discussed by Hong (1996a). We derive the asymptotic distribution of the portmanteau test statistic as $\mathcal{N}(0, 1)$.

From linear algebra, the determinant of the Hermitian Toeplitz matrix is the same as the product of the eigenvalues. Thus, the logarithmic function of the determinant of $\hat{R}_m$ implies

$$\log |\hat{R}_m|^{(m+1)} = \frac{1}{m+1} \sum_{i=0}^m \log \hat{\lambda}_{m,i} \tag{9}$$

where $\{\hat{\lambda}_{m,i}; i = 0, 1, \ldots, m\}$ are the $m + 1$ eigenvalues of $\hat{R}_m$.

Using Szegö’s theorem (Grenander & Szegö, 1958), which links the asymptotic behavior (as $m$ is large) of the eigenvalues, $\{\hat{\lambda}_{m,i}; i = 0, 1, \ldots, m\}$, to the integral of the spectral density function, we get

$$\frac{1}{m+1} \log |\hat{R}_m| = \frac{1}{m+1} \sum_{i=0}^m \log \hat{\lambda}_{m,i} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \hat{f}(\omega) d\omega \tag{10}$$

where $\hat{f}$ is a positive function associated with positive eigenvalues.

It follows from Hadamard’s inequality for the determinant of a positive definite matrix, $\log |\hat{R}_m| \leq 0$. When there is no significant autocorrelation in the residuals, $\hat{r}_i = O(n^{-1/2})$, so $\hat{R}_m$ is approximately identity matrix and hence $\log |\hat{R}_m| \approx 0$. On the other hand, when there is autocorrelation present, $\log |\hat{R}_m|$ will be expected to be less than 0. i.e. under the null hypothesis $H_0$, $\hat{R}_m \approx R_m = \{r_{ij}; i, j = 0, 1, \ldots, 1\}$ where $r_{ij} = 1$ for all $i = j$ and 0 otherwise. Hence the difference between $(m+1)^{-1} \log |\hat{R}_m|$ and $(m+1)^{-1} \log |R_m|$ will be

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \log \hat{f}(\omega) d\omega - \frac{1}{2\pi} \int_{-\pi}^{\pi} \log f_0(\omega) d\omega = \int_{-\pi}^{\pi} \log \left[ \frac{\hat{f}(\omega)}{f_0(\omega)} \right] f_0(\omega) d\omega \tag{11}$$

where $f_0(\omega) = 1/2\pi$ is the null spectral density.

Note that the integral in Equation (11) is the same as Kullback-Leibler discrimination information (Kullback, 1951; Kullback & Leibler, 1959). Note also that Hong (1996a) used the Kullback-Leibler information criterion as an entropy-based test to measure the divergence between the two spectral densities $\hat{f}(\omega)$ and $f_0(\omega)$ as follows

$$I(\hat{f}(\omega); f_0(\omega)) = -\int_{-\pi}^{\pi} \log \left[ \frac{\hat{f}(\omega)}{f_0(\omega)} \right] f_0(\omega) d\omega.$$

Under the iid assumption of $\varepsilon_i$, where $m_n \rightarrow \infty$, $m_n/n \rightarrow 0$, Hong (1996a) showed that

$$\frac{nI(\hat{f}_n(\omega); f_0(\omega)) - C_n(k)}{\sqrt{2D_n(k)}} \xrightarrow{d} \mathcal{N}(0, 1), \tag{12}$$

where $\hat{f}_n$ is the kernel spectral density estimates of $\hat{f}$ and defined from Equation (8), $C_n(k) = \sum_{\ell=0}^{m_n} \sum_{\ell=0}^{m_n} (1 - \ell/n)k^2(\ell/m_n)$, $D_n(k) = \sum_{\ell=0}^{m_n} (1 - \ell/n)(1 - (\ell + 1)/n)k^4(\ell/m_n)$, the lags of order $\ell > m_n$ receive zero weight, and $\Rightarrow \mathcal{N}(0, 1)$ denotes convergence to a standard normal distribution. Hong (1996a) indicated that the distribution given in Equation (12) still hold if we replaced...
\( C_n(k) \) by \( m_nC(k) \) and \( D_n(k) \) by \( m_nD(k) \), where \( C(k) = \int_{-\infty}^{\infty} k^2(u)du \) and \( D(k) = \int_{-\infty}^{\infty} k^4(u)du \). Within a suitable class of kernel functions, Hong (1996a, 1996b) demonstrated that the Daniell kernel
\[
k(u) = \sin(\pi u)/\pi u \text{ for } u \in (-\infty, \infty),
\] (13)
is the optimal that maximizes the power of the proposed tests under alternative hypotheses. He showed that the Daniell kernel tests have better power than the portmanteau tests of Box and Pierce (1970) and Ljung and Box (1978). The kernel \( k(\cdot) \) satisfies the following conditions:

Assumption 1. The kernel function \( k: \mathbb{R} \to [-1, 1] \) is symmetric differentiable except at a finite number of points, with \( k(0) = 1 \) and \( \int_{-\infty}^{\infty} k'(u)du < \infty \).

Assumption 2. \( \int_{-\infty}^{\infty} |k(u)|du < \infty \) and for \( \omega \in (-\infty, \infty) \), the Fourier transform \( K(\omega) \) for \( k(u) \) exist and defined as
\[
K(\omega) = (2\pi)^{-1} \int_{-\infty}^{\infty} k(u)e^{-iu\omega}du \geq 0.
\]

Assumption 3. \( \{a_i\} \) is a mean zero fourth order stationary process with \( \sum_{l=-\infty}^{\infty} \gamma^2_l < \infty \) and \( \sum_i \sum_j \sum_l |k^4(i,j,l)| < \infty \) where \( k^4(i,j,l) \) is the fourth joint cumulant of the distribution \( \{a_i, a_{t+i}, a_{t+j}, a_{t+l}\} \) and defined as
\[
k^4(i,j,l) = E(a_ia_{t+i}a_{t+j}a_{t+l}) - E(a_ia_{t+i})E(a_{t+j}a_{t+l}),
\]
where \( \{a_i\} \) is a Gaussian sequence with the same mean and covariance as \( \{a_j\} \), i.e. \( \{a_j\} \) is a fourth order stationary linear process with absolute summable coefficients and innovations whose fourth moments exist, the cumulant condition also holds.

Following the kernel-based normalized density approach in Hong (1996c), with the above mentioned assumption, one can show that \( -n(m+1)^{-1} \log |\hat{R}_{m}| \) is equal to \( nI(f_n(\omega); f_0(\omega)) \). In this respect, we propose the kernel-based portmanteau test statistic
\[
\mathcal{D}_m^* = -n(m+1)^{-1} \log |\hat{R}_{m}| - C_n(k) \quad \sqrt{D_n(k)} 
\]
\( \Rightarrow \mathcal{N}(0, 1), \) (14)

where \( C_n(k) = \sum_{r=1}^{m-1} (1 - \ell/r)k^2(\ell/m) \), \( D_n(k) = \sum_{r=1}^{m-1} (1 - \ell/r)(1 - (\ell+1)/r)k^4(\ell/m) \), where \( k(\cdot) \) is the estimated kernel which can be computed from Equation (13).

It is worth noting that the test statistic \( \mathcal{D}_m^* \) is defined for all lags \( m < n \) regardless the ARMA order \( p + q \).

4. Simulation study
In this section, we conduct numerical simulation experiments to explore the performance of the proposed portmanteau test statistic in finite samples. We investigate the accuracy of its approximation distribution in estimating the empirical type I error rates. We also conduct comparison studies between the empirical power of the classical Ljung-Box test, \( Q_n \), the data-adaptive weights test of Gallagher-Fisher, \( Q_{Dp} \), and the proposed test, \( \mathcal{D}_m^* \), against some ARMA(p, q) models. The main objective of our simulation is to show that the proposed test has a good performance as \( Q_n \) and \( Q_{Dp} \) and more powerful in many cases. In our simulation, we consider the Daniell kernel-based weights suggested by Hong (1996a, 1996b). The simulations were run on a quad-core personal computer using the \( \text{R} \) packages portes (Mahdi, Xiao, & McLeod, 2014) and WeightedPortTest (Fisher & Gallagher, 2012) that are available from the CRAN website (R Development Core Team, 2015).
4.1. Empirical size

Here we investigate the accuracy of the $\mathcal{D}_m^\ast$ statistic by comparing the empirical significance levels with the corresponding nominal levels at several lags $m$ and sample sizes $n$. The type I error rates at nominal levels 1, 5, and 10% have been evaluated under the fitted AR(1) models with different AR coefficients $\phi$ using the approximation distribution significance test based on $10^4$ iterations. We also evaluate the empirical size of the proposed statistic as the lag increases in the case of a properly fitted ARMA(1, 1) model of sample size $n = 150$ based on $10^4$ iterations. We use the Gaussian and uniform distributions to generate the sequence of innovations. The results are summarized in Figures 1–3 and Tables 1 and 2.

Figure 1 shows the empirical rejection rate using the distribution of $\mathcal{D}_m^\ast$. The AR(1) with parameter $\phi = 0.7$ is simulated $10^4$ times for Gaussian series of lengths $n = 50, 100, 200$, and $500$ and the empirical distribution of $\mathcal{D}_m^\ast$ at lag $m = 10$ is compared with its asymptotic distribution using Q-Q plot with corresponding levels $0.01$ and $0.05 \times i$ where $i = 1, 2, \ldots, 19$. Figure 1 shows that the empirical rate gets more closer to its nominal 5% level as $n$ increases. In Figure 2, the empirical rejection rates and their 95% confidence intervals are shown for the lags $m = 10, 15, 20$, and $25$ with $10^4$ simulations based on the non-Gaussian (uniform) innovation series. The 95% confidence intervals are indicated by the vertical lines in Figure 2. From the plots in this figure, we see the empirical rate varies between 0.039 and 0.060 so that the test is not conservative. It can also be seen from Figure 2, that the empirical rate gets a little closer to its nominal 5% level as $n$ increases. Although
simulation results suggest that empirical size of the $\mathcal{D}_m^*$ statistic based on the uniform innovations distribution worsens as $m$ increases, it is not significantly different from the corresponding nominal level based on the Gaussian innovations distribution in most cases (Tables 1 and 2). This suggests that the proposed test statistic is robust under strong assumptions on the noise (iid innovations but not necessarily Gaussian).

Figure 3 displays the empirical size at $\alpha = 0.05$ of the statistic $\mathcal{D}_m^*$ as $m$ increases for a simulated Gaussian and Uniform (non-Gaussian) series of length $n = 150$ in the case of a true fitted ARMA(1, 1) model where $\phi = 0.7$ and $\theta = -0.5$ based on $10^4$ replicates. We see from the two size simulations that the proposed statistic become very conservative as the lag increases.

4.2. Power comparisons

It is worth noting that, for ARMA models, Gallagher and Fisher (2015) demonstrated that the power of data-adaptive weights test statistic, $Q_D$, is always larger than the power of the classical test statistic of Box and Pierce (1970) and in most cases it has a better performance than Ljung and Box (1978), Peña and Rodríguez (2002) and Peña and Rodríguez (2006). In this respect, we conduct a comparison study between the empirical power of the proposed test statistic and the empirical power of the Ljung-Box ($Q_m$) and the data-adaptive weights ($Q_D$) test statistics, against some ARMA($p$, $q$) models. We use the same models originally used by Monti (1994) and which appears in several later articles (Fisher & Gallagher, 2012; Gallagher & Fisher, 2015; Peña & Rodríguez, 2002, 2006). Tables 3 and 4 provide the empirical powers for a nominal 5% level of $Q_m$, $Q_D$, and $\mathcal{D}_m^*$ tests.
Table 1. The empirical sizes of nominal 1, 5 and 10% levels of $\Phi^m$ test for the fitted AR(1) models

| $\Phi$ | $n = 100$ | $n = 200$ | $n = 100$ | $n = 200$ | $n = 100$ | $n = 200$ |
|--------|-----------|-----------|-----------|-----------|-----------|-----------|
| $\alpha = 0.01$ | | | | | | |
| 0.1 | 0.012 | 0.015 | 0.010 | 0.016 | 0.009 | 0.016 |
| 0.3 | 0.013 | 0.018 | 0.012 | 0.008 | 0.011 | 0.015 |
| 0.5 | 0.011 | 0.017 | 0.009 | 0.017 | 0.010 | 0.015 |
| 0.7 | 0.015 | 0.015 | 0.012 | 0.012 | 0.010 | 0.011 |
| 0.9 | 0.018 | 0.009 | 0.013 | 0.010 | 0.009 | 0.010 |
| $\alpha = 0.05$ | | | | | | |
| 0.1 | 0.044 | 0.036 | 0.040 | 0.041 | 0.035 | 0.050 |
| 0.3 | 0.054 | 0.039 | 0.046 | 0.048 | 0.041 | 0.062 |
| 0.5 | 0.054 | 0.047 | 0.044 | 0.044 | 0.041 | 0.059 |
| 0.7 | 0.059 | 0.042 | 0.047 | 0.041 | 0.042 | 0.055 |
| 0.9 | 0.061 | 0.045 | 0.053 | 0.038 | 0.049 | 0.045 |
| $\alpha = 0.10$ | | | | | | |
| 0.1 | 0.095 | 0.089 | 0.090 | 0.083 | 0.080 | 0.088 |
| 0.3 | 0.101 | 0.089 | 0.091 | 0.082 | 0.082 | 0.100 |
| 0.5 | 0.106 | 0.082 | 0.087 | 0.080 | 0.082 | 0.098 |
| 0.7 | 0.115 | 0.085 | 0.097 | 0.078 | 0.082 | 0.086 |
| 0.9 | 0.144 | 0.094 | 0.150 | 0.088 | 0.100 | 0.098 |

Note: Data were generated using Gaussian iid innovations.

Table 2. The empirical sizes of nominal 1, 5 and 10% levels of $\Phi^m$ test for the fitted AR(1) models

| $\Phi$ | $n = 100$ | $n = 200$ | $n = 100$ | $n = 200$ | $n = 100$ | $n = 200$ |
|--------|-----------|-----------|-----------|-----------|-----------|-----------|
| $\alpha = 0.01$ | | | | | | |
| 0.1 | 0.008 | 0.008 | 0.009 | 0.010 | 0.009 | 0.010 |
| 0.3 | 0.014 | 0.017 | 0.014 | 0.015 | 0.010 | 0.019 |
| 0.5 | 0.018 | 0.016 | 0.009 | 0.016 | 0.009 | 0.016 |
| 0.7 | 0.018 | 0.016 | 0.012 | 0.019 | 0.016 | 0.014 |
| 0.9 | 0.016 | 0.009 | 0.017 | 0.015 | 0.009 | 0.015 |
| $\alpha = 0.05$ | | | | | | |
| 0.1 | 0.039 | 0.037 | 0.044 | 0.040 | 0.037 | 0.043 |
| 0.3 | 0.059 | 0.062 | 0.060 | 0.070 | 0.066 | 0.061 |
| 0.5 | 0.060 | 0.059 | 0.070 | 0.064 | 0.060 | 0.069 |
| 0.7 | 0.059 | 0.056 | 0.057 | 0.049 | 0.041 | 0.055 |
| 0.9 | 0.071 | 0.041 | 0.050 | 0.036 | 0.049 | 0.052 |
| $\alpha = 0.10$ | | | | | | |
| 0.1 | 0.085 | 0.077 | 0.083 | 0.080 | 0.090 | 0.083 |
| 0.3 | 0.111 | 0.079 | 0.090 | 0.086 | 0.077 | 0.110 |
| 0.5 | 0.105 | 0.087 | 0.080 | 0.085 | 0.089 | 0.099 |
| 0.7 | 0.105 | 0.095 | 0.0902 | 0.088 | 0.094 | 0.082 |
| 0.9 | 0.141 | 0.073 | 0.184 | 0.086 | 0.102 | 0.094 |

Note: Data were generated using Uniform (non-Gaussian) iid innovations.
when a series of 12 different ARMA(2,2) processes (each of size \(n = 100\)) are inadequately fitted as an AR(1) and tested at lag \(m = 10\). The results from Table 3 are obtained from data driven by Gaussian \(iid\) innovations, whereas the results of Table 4 are obtained from data with Uniform noise.

We see in the results of Tables 3 and 4 that the statistics \(Q_m\) and \(\mathcal{D}_m^\star\) are always the most powerful in comparison with the Ljung-Box test statistic. Generally, the simulation results show that the proposed test has as good performance as \(Q_m\) and it provides more power in some situations. Furthermore, the results suggest that the powers associated with data driven by Gaussian innovations are not significantly different from the corresponding powers associated with data driven by non-Gaussian innovations.
5. An illustrative application

Here, we are not interested in selecting the best fitted model, but the main objective of this application is to demonstrate that the proposed test statistic can be useful for investigating whether the autocorrelations of the residual ARMA model are different from zero.

In this section, we make use of the monthly Recruitment (number of new fish) series. Data is available from the R package astsa with the name rec ranging over the years 1950–1987 with 453 observations (Shumway & Stoffer, 2011). Although, Figure 4 suggests that the Recruitment series tend to exhibit repetitive behavior every 12 months, (Shumway & Stoffer, 2011, p. 108) indicated that the second-order ($p = 2$) autoregressive model might provide a good fit.

Following the recommendation of Shumway and Stoffer (2011), we purposely fit the AR(2) model for the Recruitment series, and then we investigate how the different portmanteau statistics detect the underfit at several lags (see Table 5). Using the 5% significance level and considering the lags $m = 5, 10, 15, 20, 25,$ and $30$, the resulted $p$-values of the classical Box-Pierce and Ljung-Box test statistics on the estimated residual series suggest that significant correlations remain in the original series, and hence, both suggest model inadequacy for all lags $m \geq 20$. By lags $m \geq 25$, the Mahdi-McLeod and the proposed test statistics support the results obtained from the Box-Pierce and Ljung-Box tests, whereas both fail to detect a significant correlation for lags $m \leq 20$. On the other hand, the Gallagher-Fisher test statistic fails to detect significant correlation, and hence, it indicates that the model is adequate.

### Table 4. The empirical power for a nominal 5% level test comparing the Ljung-Box ($Q_m$) and the data-adaptive weights test ($Q_D$) statistics with the proposed one ($\hat{Q}_m$) based on $10^3$ simulations, where the AR(1) models are fitted to data generated from different ARMA(2, 2) models, $n = 100$ and $m = 20$

| Model number | $\phi_1$ | $\phi_2$ | $\theta_1$ | $\theta_2$ | $Q_{m}$ | $Q_{D}$ | $\hat{Q}_{m}$ |
|--------------|----------|----------|------------|------------|---------|---------|--------------|
| 1            | *        | *        | 0.5        | *          | 0.1871  | 0.2116  | 0.2000       |
| 2            | *        | *        | 0.8        | *          | 0.6001  | 0.9091  | 0.9187       |
| 3            | *        | *        | 0.6        | -0.3       | 0.6103  | 0.8722  | 0.8991       |
| 4            | 0.1      | 0.3      | *          | *          | 0.3028  | 0.4978  | 0.5001       |
| 5            | 1.1      | -0.35    | *          | *          | 0.5998  | 0.8016  | 0.7709       |
| 6            | 0.7      | *        | 0.4        | *          | 0.4071  | 0.6500  | 0.7015       |
| 7            | 0.7      | *        | 0.9        | *          | 0.8977  | 0.9293  | 0.9270       |
| 8            | 0.4      | *        | 0.6        | -0.3       | 0.6670  | 1.0000  | 1.0000       |
| 9            | 0.7      | *        | -0.7       | 0.15       | 0.1162  | 0.1876  | 0.1776       |
| 10           | 0.7      | 0.2      | -0.5       | *          | 0.6488  | 0.8116  | 0.7899       |
| 11           | 0.7      | 0.2      | 0.5        | *          | 0.3003  | 0.3999  | 0.3970       |
| 12           | 0.9      | -0.4     | -1.2       | 0.3        | 0.5622  | 0.8409  | 0.8900       |

Note: Data were driven by Uniform (non-Gaussian) iid innovations.

### Table 5. The $p$-values at lags $m = 5, 10, 15, 20, 25$ and $30$ for the portmanteau tests for the fitted AR(2) model on the Recruitment (number of new fish) series, 1950–1987

| Lag $m$ | Box-Pierce | Ljung-Box | Mahdi-McLeod | Gallagher-Fisher | Kernel-based test |
|---------|------------|-----------|--------------|-----------------|------------------|
| 5       | 0.190      | 0.185     | 0.193        | 0.185           | 0.409            |
| 10      | 0.500      | 0.487     | 0.329        | 0.518           | 0.448            |
| 15      | 0.108      | 0.094     | 0.216        | 0.401           | 0.247            |
| 20      | 0.009      | 0.006     | 0.102        | 0.288           | 0.079            |
| 25      | 0.001      | $3 \times 10^{-3}$ | 0.048       | 0.151           | 0.020            |
| 30      | 0.001      | 0.001     | 0.021        | 0.141           | 0.004            |
6. Conclusion

Most existing portmanteau test statistics are not consistent against serial correlation of unknown form. Moreover, these tests for ARMA($p$, $q$) models are not defined in many cases where $p + q$ is large and $m$ is small. In this paper, we proposed a kernel-based consistent portmanteau test statistic for serial correlation of unknown form for the residual (i.e. when no prior information about the true alternative is available). Regardless of the ARMA order $p + q$, the proposed test is defined for all lags $m < n$, where $n$ is series length. We derived the test statistic from the Hermitian Toeplitz autocorrelation matrix defined from Peña and Rodríguez (2002). Essentially, the proposed test statistic is the same as Peña and Rodríguez (2006); however, they have different distributions. The asymptotic distribution of the proposed statistic is derived as a normal distribution. We conducted simulation studies to demonstrate the efficacy in the power of the proposed statistic in finite samples. The simulation results suggest that the proposed test statistic is robust under strong assumptions on the noise (iid innovations but not necessarily Gaussian). Perhaps, the idea and procedures in this paper can be extended in several ways to check goodness-of-fit of other univariate and multivariate time series models. For instance, the methods of Franço and Raïssi (2007) and Franço, Roy, and Zakoïan (2005) could be modified using the ideas in this article to provide new portmanteau test statistics with higher powers to detect the underfit time series models when the underlying noise process is assumed to be uncorrelated rather than independent (weakly dependent white noise). It may be possible to implement the idea of the Toeplitz autocorrelation matrix, as defined by Peña and Rodríguez (2002), and the kernel-based normalized density method, as described by Hong (1996a), to derive a new portmanteau test statistic to check the adequacy of a fitted detecting long-memory nonlinear model as in McLeod and Li (1983).
Acknowledgements
The author is most grateful to the Editor and three anonymous referees for their constructive suggestions and references on the earlier versions, which helped to improve the presentation of this article considerably. All errors are attributed solely to the author.

Funding
The author received no direct funding for this research.

Author details
Esam Mahdi
E-mail: emahdi@iugaza.edu.ps
ORCID ID: http://orcid.org/0000-0002-3447-0639

1 Department of Mathematics, Islamic University of Gaza, P.O. Box 108, Gaza, Palestine.

Citation information
Cite this article as: Kernel-based portmanteau diagnostic test for ARMA time series models, Esam Mahdi, Cogent Mathematics (2017), 4: 1296327.

References
Anderson, T. W. (1993). Goodness of fit tests for spectral distributions. The Annals of Statistics, 21, 830–847.
Box, G. E. P., & Jenkins, G. M. (1970). Time series analysis: Forecasting and control. San Francisco, CA: Holden-Day.
Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2005). Time series analysis: Forecasting and control (4th ed.). New York, NY: Wiley.
Box, G. E. P., & Pierce, D. A. (1970). Distribution of residual autocorrelation in autoregressive-integrated moving average time series models. Journal of the American Statistical Association, 65, 1509–1526.
Brockwell, P. J., & Davis, R. A. (2009). Time series: Theory and methods (2nd ed.). New York, NY: Springer.
Fisher, T. J., & Gallagher, C. M. (2002). New weighted portmanteau statistics for time series goodness of fit testing. Journal of the American Statistical Association, 107, 777–787.
França, C., & Raisi, H. (2007). Multivariate portmanteau test for autoregressive models with correlated but nonindependent errors. Journal of Time Series Analysis, 28, 454–470.
França, C., Roy, R., & Zakoian, J.-M. (2003). Diagnostic checking in arma models with uncorrelated errors. Journal of the American Statistical Association, 100, 532–544.
Gallagher, C. M., & Fisher, T. J. (2015). On weighted portmanteau tests for time-series goodness-of-fit. Journal of Time Series Analysis, 36, 67–83.
Grenander, U., & Szegö, G. (1958). Toeplitz forms and their applications. Berkeley and Los Angeles, CA: University of California Press.
Hong, Y. (1996a). Consistent testing for serial correlation of unknown form. Econometrica, 64, 837–864.
Hong, Y. (1996b). Testing for independence between two covariance stationary time series. Biometrika, 83(3), 615–625.
Kullback, S. (1959). Information theory and statistics. New York, NY: Wiley.
Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. Biometrika, 65, 297–303.
Li, W. K. (2004). Diagnostic checks in time series. New York, NY: Chapman and Hall/CRC.
Monti, A. C. (1994). A proposal for a residual autocorrelation test in linear models. Biometrika, 81, 776–780.
Peña, D., & Rodriguez, J. (2000). A powerful portmanteau test of lack of test for time series. Journal of the American Statistical Association, 97, 601–610.
Peña, D., & Rodríguez, J. (2006). The log of the determinant of the autocovariance matrix for testing goodness of fit in time series. Journal of Statistical Planning and Inference, 136, 2706–2718.
Priestley, M. B. (1981). Spectral analysis and time series (Vol. 1). London: Academic Press.
R Development Core Team. (2015). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.
Shumway, R. H., & Stoffer, D. S. (2011). Time series analysis and its applications with R applications (3rd ed.). New York, NY: Springer.

© 2017 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.
You are free to:
Share — copy and redistribute the material in any medium or format
Adapt — remix, transform, and build upon the material for any purpose, even commercially.
The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:
Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made.
You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
No additional restrictions
You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.