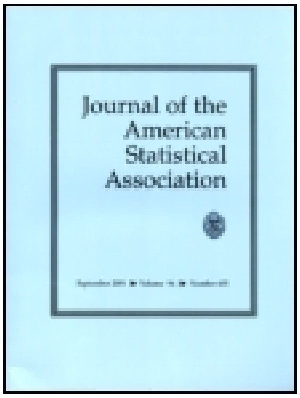

Journal of the American Statistical Association
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uasa20

New Weighted Portmanteau Statistics for Time Series Goodness of Fit Testing

Thomas J. Fisher [a] & Colin M. Gallagher [b]

[a] Department of Mathematics and Statistics , University of Missouri-Kansas City , Kansas City , MO , 64110

[b] Department of Mathematical Sciences , Clemson University , Clemson , SC , 29634

Accepted author version posted online: 16 May 2012.Published online: 24 Jul 2012.

New Weighted Portmanteau Statistics for Time Series Goodness of Fit Testing

Thomas J. FISHER and Colin M. GALLAGHER

We exploit ideas from high-dimensional data analysis to derive new portmanteau tests that are based on the trace of the square of the mth order autocorrelation matrix. The resulting statistics are weighted sums of the squares of the sample autocorrelation coefficients that, unlike many other tests appearing in the literature, are numerically stable even when the number of lags considered is relatively close to the sample size. The statistics behave asymptotically as a linear combination of chi-squared random variables and their asymptotic distribution can be approximated by a gamma distribution. The proposed tests are modified to check for nonlinearity and to check the adequacy of a fitted nonlinear model. Simulation evidence indicates that the proposed goodness of fit tests tend to have higher power than other tests appearing in the literature, particularly in detecting long-memory nonlinear models. The efficacy of the proposed methods is demonstrated by investigating nonlinear effects in Apple, Inc., and Nikkei-300 daily returns during the 2006–2007 calendar years. The supplementary materials for this article are available online.

KEY WORDS: ARMA model; GARCH model; Nonlinear test; Residual diagnostic test.

1. INTRODUCTION

In much of applied time series analysis, parametric models are used to approximate the correlation structure of a stationary process or of a function of that stationary process. Typically, to model the correlation structure of a time series $\{X_t\}$, we use a stationary and invertible autoregressive–moving average (ARMA) process of the form

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} - \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t, \tag{1}$$

where p is the AR order, q is the MA order, and $\{\epsilon_t\}$ is an independent and identically distributed (iid) innovation sequence of zero mean random variables with finite variance. For financial time series, empirical evidence motivates modeling the correlation structure of a nonlinear function of the process. Generalized autoregressive conditional heteroscedasticity (GARCH) processes model the correlation of the squares of the time series and stochastic volatility (SV) models model the correlation structure of the logarithm of the series. Many other nonlinear models have been proposed and typically have error terms that are assumed to be iid. If the fit adequately models the underlying correlation structure, the resulting residuals should be approximately uncorrelated.

To check the fit of an ARMA model, we typically use the sample autocorrelation function of the residuals $\{\hat{\epsilon}_t\}$:

$$\hat{r}_k = \frac{\sum_{k+1}^{n} \hat{\epsilon}_t \hat{\epsilon}_{t-k}}{\sum_{t=1}^{n} \hat{\epsilon}_t^2} \quad k = 1, 2, \ldots, m.$$

If we correctly identify the orders p and q, each of the above correlation coefficients should be approximately equal to zero. However, if the fitted model underestimates the ARMA orders, the values of the autocorrelations should significantly deviate from zero. Most statistical software calculates the portmanteau statistics

$$Q_{\mathrm{BP}} = n \sum_{k=1}^{m} \hat{r}_k^2 \quad \text{and} \quad \widetilde{Q} = n \sum_{k=1}^{m} \frac{n+2}{n-k} \hat{r}_k^2,$$

from Box and Pierce (1970) and Ljung and Box (1978), respectively. Each statistic is asymptotically distributed as a chi-squared random variable with $m - (p+q)$ degrees of freedom. Ljung and Box (1978) demonstrated that the sampling distribution of $\widetilde{Q}$ is closer to the asymptotic chi-squared distribution for smaller sample sizes. Ljung (1986) showed that the distribution of $\widetilde{Q}$ is better approximated by a scaled chi-squared, or gamma, at small m and that power decreases as m increases.

Monti (1994) proposed a test using the residual partial autocorrelations:

$$\widetilde{M} = n(n+2) \sum_{k=1}^{m} \frac{\hat{\pi}_k^2}{n-k}, \tag{2}$$

where $\hat{\pi}_k$ is the residual partial autocorrelation at lag k. If the model is adequately identified, the asymptotic distribution of $\widetilde{M}$ is chi-squared with $m - (p+q)$ degrees of freedom. Simulations demonstrate that $\widetilde{M}$ is more powerful than $\widetilde{Q}$ when the fitted model underestimates the order of the MA component.

Peña and Rodríguez (2002) proposed a statistic based on the determinant of the Toeplitz matrix of sample autocorrelations

$$\widehat{\mathbf{R}}_m = \begin{bmatrix} 1 & \hat{r}_1 & \cdots & \hat{r}_m \\ \hat{r}_1 & 1 & \cdots & \hat{r}_{m-1} \\ \vdots & \cdots & \ddots & \vdots \\ \hat{r}_m & \cdots & \hat{r}_1 & 1 \end{bmatrix}. \tag{3}$$

Thomas J. Fisher is an Assistant Professor of Statistics in the Department of Mathematics and Statistics, University of Missouri-Kansas City, Kansas City, MO 64110 (E-mail: *fishertho@umkc.edu*). Colin M. Gallagher is an Associate Professor of Statistics in the Department of Mathematical Sciences, Clemson University, Clemson, SC 29634 (E-mail: *cgallag@clemson.edu*). Both authors wish to thank the editor, associate editor, and anonymous referees for their thoughtful comments and insights that helped us to improve the article. The first author is grateful for the partial support from the University of Missouri Research Board.

Journal of the American Statistical Association
June 2012, Vol. 107, No. 498, Theory and Methods
DOI: 10.1080/01621459.2012.688465

Their statistic is

$$\widehat{D} = n\left[1 - |\widehat{\mathbf{R}}_m|^{1/m}\right],$$

where $|\widehat{\mathbf{R}}_m|$ is the determinant of matrix $\widehat{\mathbf{R}}_m$. Asymptotically $\widehat{D}$ can be expressed as a linear combination of chi-squared random variables and its distribution can be approximated by a gamma distribution following the methodology of Satterthwaite (1941, 1946) and Box (1954). They recommended replacing $\widehat{\mathbf{R}}_m$ with $\widetilde{\mathbf{R}}_m$, the autocorrelation matrix built by using the standardized autocorrelation coefficients, which are the square roots of the summands in $\widetilde{Q}$. These tests have intuitive appeal, since under the null hypothesis that the model has been correctly identified, the matrix of estimated correlations is close to the identity matrix and checking for an adequately fitted model is equivalent to testing if the determinant is close to unity. However, the estimated matrices can be ill conditioned especially if m is large relative to n. Peña and Rodríguez (2006) demonstrated that $\sum \ln(1 - \hat{\pi}_i^2)$ has the same asymptotic distribution as $\widehat{D}$. Their simulations show an improvement in small sample time series, but the Type I error rates appear to be poor.

Recently Mahdi and McLeod (2012) generalized the results of Peña and Rodríguez (2002, 2006) to the multivariate setting. They suggested the log of the determinant of the Toeplitz matrix of autocorrelations. In the univariate case, they recommended the statistic

$$D_m = \frac{-3n}{2m+1} \log|\widehat{\mathbf{R}}_m|. \tag{4}$$

Using methodology similar to Peña and Rodríguez (2002, 2006) and a key result of Box (1954), they found the asymptotic distribution of D_m as a linear combination of chi-squared random variables. They demonstrated that the null distribution of D_m is approximately chi-squared with $(3/2)m(m+1)/(2m+1) - (p+q)$ degrees of freedom. The degrees of freedom for the chi-squared approximation allow D_m to have conservative Type I errors in practice.

Several other statistics exist and many authors have studied the performance of the stated statistics. Despite a plethora of available methods, practitioners of time series typically rely on the Ljung–Box statistic, as it is easy to understand and to compute, and is included in most statistical software packages.

In this article, we introduce several portmanteau statistics that are asymptotically similar to those in Peña and Rodríguez (2002) but are just as easy to compute as the widely used Ljung–Box statistic. In Section 2, we develop the statistics for fitted ARMA processes. Section 3 addresses the problem of checking for the presence of nonlinearity and the methodology is applied to testing for the adequacy of a fitted autoregressive conditional heteroscedasticity (ARCH) model. Section 4 supplies a computational study and data analysis demonstrating the effectiveness of the proposed test statistics. Section 5 provides concluding remarks. Technical details and proofs appear in the Appendix. Additional details and simulations are provided in the supplementary documents.

2. TESTING THE ADEQUACY OF A FITTED ARMA PROCESS

The statistics proposed in Peña and Rodríguez (2002, 2006) and Mahdi and McLeod (2012) have an asymmetric structure that helps explain why they work well compared with the Ljung–Box and the Monti statistics. Intuitively, the sample autocorrelation at lag 1 will be the most accurate since its numerator is based on $n-1$ data points, whereas the autocorrelation at lag m is estimated with only $n-m$ observations. The statistics from Peña and Rodríguez (2002, 2006) and Mahdi and McLeod (2012) put more emphasis on the more accurate autocorrelations. However, they are not without their own set of potential drawbacks. As described in section 7.2 of Brockwell and Davis (1991), the sample autocorrelation matrix $\widehat{\mathbf{R}}_m$ is nonnegative definite, but when the autocorrelations are replaced by their standardized counterparts, the matrix $\widetilde{\mathbf{R}}_m$ is not necessarily nonnegative definite. Lin and McLeod (2006) highlighted this deficiency as the recommended statistic from Peña and Rodríguez (2002) is frequently undefined. They suggest using the statistic $\widehat{D}$ and improving on its performance by using a Monte Carlo method to determine the distribution.

The computational instability of the determinant of $\widetilde{\mathbf{R}}_m$ is addressed by the statistic from Peña and Rodríguez (2006), which does not require the computation of a determinant. However, this statistic still requires logarithmic calculation on m partial autocorrelations that could lead to other computational stability and accuracy problems. Mahdi and McLeod (2012) used $\widehat{\mathbf{R}}_m$ which is nonsingular, however when constructed at large lags, it too may become ill conditioned. Bickel and Levina (2008) proposed a consistent regularization procedure that bands or tapers the autocorrelation matrix under a Gaussian white noise process; their simulations are promising and this method might provide a more numerically stable statistic. In this article, we consider statistics based on the trace of the square of estimated correlation matrices. These statistics share the same asymptotic distribution as Peña and Rodríguez (2002, 2006), but are simpler to calculate and are computationally stable. Also, the proposed statistics seem to be less sensitive to the choice of m.

2.1 Proposed Statistics for ARMA Processes

In the field of multivariate analysis, many authors have proposed methods for testing on the covariance matrix in high-dimensions; see Schott (2005), Srivastava (2005), Fisher, Sun, and Gallagher (2010), and Srivastava, Kollo, and von Rosen (2011) to name a few. In particular, when the dimensionality exceeds the sample size, the likelihood ratio method degenerates as the sample covariance matrix becomes singular. Due to some of the computational concerns with the matrix $\widehat{\mathbf{R}}_m$, similar methodology appears well justified.

Consider the problem discussed in Srivastava (2005) for testing for an identity covariance matrix. Using his approach, let $\mathbf{\Sigma}_m$ denote the probability limit of matrix $\widehat{\mathbf{R}}_m$. To test for an iid correlation structure, we test $H_0 : \mathbf{\Sigma}_m = \mathbf{I}$, which is equivalent to testing if each eigenvalue of $\mathbf{\Sigma}_m$ is one. In practice, we can use the matrix $\widehat{\mathbf{R}}_m$ and its eigenvalues a_i, $i = 1, \ldots, m+1$ to test H_0. The likelihood ratio criterion uses the geometric mean of the eigenvalues. If $\widehat{\mathbf{R}}_m$ is singular or close to singular, a statistic based on the arithmetic mean of the eigenvalues would be preferred. Note that

$$\frac{1}{m+1}\sum_{k=1}^{m+1}(a_k - 1)^2 \geq 0$$

with equality when each $a_k = 1$. An equivalent statement after some algebra is

$$\frac{1}{m+1}\operatorname{tr}\left(\widehat{\mathbf{R}}_m^2\right) - 1 \geq 0.$$

From calculating the trace of the matrix $\widehat{\mathbf{R}}_m^2$, we may consider testing the hypothesis based on the inequality

$$\sum_{k=1}^{m} \frac{2(m-k+1)}{m+1}\hat{r}_k^2 \geq 0.$$

Noting the inequality will not change if each side is multiplied by $(m+1)/2m$, and replacing the squared autocorrelations with their standardized residual counterparts leads to the suggested statistic

$$\widetilde{Q}_W = n(n+2)\sum_{k=1}^{m} \frac{(m-k+1)}{m}\frac{\hat{r}_k^2}{n-k}. \tag{5}$$

The statistic $\widetilde{Q}_W$ can be interpreted as a Weighted Ljung–Box statistic. The residual at lag 1 is given the most weight, 1, while the residual at lag m is given the least weight, $1/m$. The statistic is easy to implement and is computationally stable. A similar derivation using the matrix of partial autocorrelations leads to a Weighted Monti statistic

$$\widetilde{M}_W = n(n+2)\sum_{k=1}^{m} \frac{(m-k+1)}{m}\frac{\hat{\pi}_k^2}{n-k}. \tag{6}$$

2.2 Asymptotic Distribution

Define $\hat{\mathbf{r}}' = (\hat{r}_1, \ldots, \hat{r}_m)$ and $\hat{\boldsymbol{\pi}}' = (\hat{\pi}_1, \ldots, \hat{\pi}_m)$ and using the results of Ljung and Box (1978) and Monti (1994) that $\sqrt{n}\hat{\mathbf{r}}$ and $\sqrt{n}\hat{\boldsymbol{\pi}}$ are asymptotically multivariate normal with mean zero vector and covariance matrix $(\mathbf{I} - \mathbf{Q})$, where $\mathbf{Q} = \mathbf{X}\boldsymbol{\Psi}^{-1}\mathbf{X}'$, $\boldsymbol{\Psi}$ is the information matrix for parameters $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$, and $\mathbf{X}$ is an $m \times (p+q)$ matrix with elements $\boldsymbol{\phi}'$ and $\boldsymbol{\theta}'$ defined by $1/\phi(B) = \sum \phi_i' B^i$ and $1/\theta(B) = \sum \theta_i' B^i$.

Theorem 1. Under the null hypothesis of an adequately fitted model, the statistics $\widetilde{Q}_W$ and $\widetilde{M}_W$ are asymptotically distributed as $\sum_{k=1}^m \lambda_k \chi_k^2$ where $\{\chi_k^2\}$ are independent chi-squared random variables with one degree of freedom and λ_k $(k = 1, \ldots, m)$ are the eigenvalues of $(\mathbf{I} - \mathbf{Q})\mathbf{W}$, where $\mathbf{W}$ is a diagonal matrix with elements $w_{ii} = (m - i + 1)/m$ $(i = 1, \ldots, m)$.

We note that the statistics $\widetilde{Q}_W$ and $\widetilde{M}_W$ follow the same null asymptotic distribution as that of Peña and Rodríguez (2002, 2006) and this distribution is similar to that in Mahdi and McLeod (2012). The density is difficult to write in explicit form but approximating the distribution has been discussed by many authors. Imhof (1961) had shown that probabilities can be found through numerical integration. Peña and Rodríguez (2006) suggested a normal approximation using a generalization of the Wilson-Hilferty cube root transformation of a chi-squared random variable. Lin and McLeod (2006) and Mahdi and McLeod (2012) suggested a Monte Carlo method to determine the critical points and p-values. For computational ease, like that in Peña and Rodríguez (2002), we recommend an approximation using a gamma distribution described in Satterthwaite (1941, 1946) and Box (1954) with shape and scale parameters

$$\alpha = \frac{3}{4}\frac{m(m+1)^2}{2m^2 + 3m + 1 - 6m(p+q)}, \tag{7}$$

and

$$\beta = \frac{2}{3}\frac{2m^2 + 3m + 1 - 6m(p+q)}{m(m+1)}, \tag{8}$$

respectively. The derivation of α and β follows similar methodology as in Peña and Rodríguez (2002) except our recommended shape and scale are conservative to improve the approximation. The details of the derivation appear in the Appendix.

It may be natural to let m depend on the sample size n, particularly to detect serial correlation at higher lags. The following theorem, which is similar to the results of Bhansali, Giraitis, and Kokoszka (2007), provides the asymptotic distribution of our statistic as $(n, m) \to \infty$. Similar results can be derived for the statistics by Peña and Rodríguez (2002) and Mahdi and McLeod (2012), but in practice if m is large relative to n, their statistics can become numerically unstable.

Theorem 2. Let $\{X_t\}$ be an ARMA(p, q) process with innovation sequence $\{\epsilon_t\}$. If $m = O(n^\delta)$ for some $0 < \delta < 1$ such that $m/n \to 0$ and $E[\epsilon^8] < \infty$, then under the null hypothesis as $n \to \infty$, $(Q_W - K_1)/\sqrt{K_2}$ converges to a standard normal random variable, where K_1 and K_2 are given in (A.2) and (A.3), respectively.

3. NONLINEAR MODELS

Due to the limitations of linear models in capturing some observed data features, much attention has been focused on the analysis of time series using nonlinear models. It may be the case that the innovation sequence is uncorrelated, but not iid. In this case, some function of the residuals may be autocorrelated. Many authors have considered nonlinear parametric models of the form

$$\epsilon_t = g(h_t)\eta_t, \tag{9}$$

where η_t are iid with mean zero and variance 1 and $\{h_t\}$ follows an ARMA type recursion. Two special cases are the GARCH process of Bollerslev (1986), which takes $g(x) = \sqrt{x}$, and the SV model of Taylor (1986), which assumes $g(x) = \exp(x)$. For these models, a nonlinear function of $\{\epsilon_t\}$ inherits the correlation structure of an ARMA process. Two problems of interest are studied; (1) using a portmanteau statistic to detect nonlinear processes and (2) determining the goodness of fit of a nonlinear process.

3.1 Detecting Nonlinear Processes

Many authors have studied detecting the presence of a nonlinear innovation process by using transformed residuals of a fitted model. Pérez and Ruiz (2003) provided a nice review of the topic. McLeod and Li (1983) first showed that if the innovation sequence is iid, then the vector of autocorrelation coefficients based on the squared residuals of a fitted ARMA model is asymptotically normally distributed with mean zero and unit covariance matrix. In particular, the distribution of the autocorrelations based on the squared residuals is not influenced by the number of fitted ARMA parameters. Similarly, Peña and

Rodríguez (2006) and several other authors had suggested using the absolute residuals or the log of the squared residuals to detect GARCH or SV structures. Define

$$\hat{r}_k(*) = \frac{\sum_{t=k+1}^{n}(g(\hat{\epsilon}_t) - \bar{g}(\hat{\epsilon}_t))(g(\hat{\epsilon}_{t-k}) - \bar{g}(\hat{\epsilon}_t))}{\sum_{t=1}^{n}(g(\hat{\epsilon}_t) - \bar{g}(\hat{\epsilon}_t))^2}, \tag{10}$$

where $\bar{g}(\hat{\epsilon}_t) = \sum g(\hat{\epsilon}_t)/n$, as the autocorrelation function based on the transformation $g(\cdot)$ of the residuals. To simplify the notation, we use the $*$ character to indicate either a squared, absolute, or log-squared transformation of the residuals. McLeod and Li (1983) suggested a Ljung–Box type test statistic,

$$\widetilde{Q}^* = n(n+2)\sum_{k=1}^{m}\frac{\hat{r}_k^2(*)}{n-k}, \tag{11}$$

where the $*$ represents the squared residuals in their work, for detecting nonlinear processes in ARMA innovations. They showed that $\widetilde{Q}^*$ is asymptotically distributed as chi-squared with m degrees of freedom. Peña and Rodríguez (2002) provided an analogous version of their statistics in which the matrix $\widetilde{\mathbf{R}}_m$ is constructed based on the standardized squared residuals. The statistic in Mahdi and McLeod (2012) can easily be extended to use any of these transformations of the residuals.

As pointed out in Peña and Rodríguez (2002), it is possible to build a test based on the partial autocorrelations, $\hat{\pi}_k(*)$. From this, the Weighted Ljung–Box and the Weighted Monti test can easily be extended for checking the linearity assumption in a time series. Consider the following statistics for checking the linearity assumption

$$\widetilde{Q}_W^* = n(n+2)\sum_{k=1}^{m}\frac{(m-k+1)}{m}\frac{\hat{r}_k^2(*)}{n-k}, \tag{12}$$

and

$$\widetilde{M}_W^* = n(n+2)\sum_{k=1}^{m}\frac{(m-k+1)}{m}\frac{\hat{\pi}_k^2(*)}{n-k}, \tag{13}$$

where $\hat{r}_k(*)$ is the autocorrelation at lag k based on $\hat{\epsilon}^2$, $|\hat{\epsilon}|$ or $\log(\hat{\epsilon}^2)$ and $\hat{\pi}_k(*)$ are the corresponding partial autocorrelations. $\widetilde{Q}_W^*$ can be considered as a Weighted McLeod–Li type statistic and $\widetilde{M}_W^*$ is a Weighted Monti type statistic.

Theorem 3. If the series follows a stationary ARMA process, the statistics $\widetilde{Q}_W^*$ and $\widetilde{M}_W^*$ computed from the squared residuals are asymptotically distributed as $\sum_{k=1}^{m} w_k\chi_k^2$ where $\{\chi_k^2\}$ are independent chi-squared random variables with one degree of freedom and w_k $(k = 1, \ldots, m)$ are the weights $w_k = (m-k+1)/m$.

We note that a similar argument can be made for the statistics using the autocorrelations based on $|\hat{\epsilon}|$ and $\log(\hat{\epsilon}^2)$. As before, the asymptotic distribution can be approximated using a gamma distribution. If the series follows an ARMA process, the approximate distribution of $\widetilde{Q}_W^*$ and $\widetilde{M}_W^*$ is gamma with shape $\alpha = 3m(m+1)/(8m+4)$ and scale $\beta = 2(2m+1)/3m$.

3.2 GARCH and ARCH Goodness of Fit

With advances in the area of macroeconomics and financial time series, the ARCH models of Engle (1982) and GARCH models of Bollerslev (1986) have gained in popularity. The GARCH(b, a) model is given by $g(x) = \sqrt{x}$ in (9) where $\{\eta_t\}$ is iid with zero mean and variance of unity and

$$h_t = \omega + \sum_{i=1}^{b}\alpha_i\epsilon_{t-i}^2 + \sum_{j=1}^{a}\beta_j h_{t-j}.$$

Much like fitting an ARMA process, after fitting the parameters for the GARCH process, checking for the adequacy of that model should follow. Li and Mak (1994) showed that the Box–Pierce type statistic suggested by Higgins and Bera (1992) does not converge to a chi-squared distribution asymptotically when constructed with the squared residuals. They proposed the statistic for a fitted GARCH process based on the autocorrelation function developed by the standardized sample squared residuals,

$$\hat{r}_k\left(\epsilon_t^2/h_t\right) = \frac{\sum_{t=k+1}^{n}\left(\hat{\epsilon}_t^2/\hat{h}_t - \bar{\epsilon}\right)\left(\hat{\epsilon}_{t-k}^2/\hat{h}_{t-k} - \bar{\epsilon}\right)}{\sum_{t=1}^{n}\left(\hat{\epsilon}_t^2/\hat{h}_t - \bar{\epsilon}\right)^2},$$

where $\bar{\epsilon} = (1/n)\sum\hat{\epsilon}_t^2/\hat{h}_t$ and $\hat{h}_t$ are the sample conditional variances. They showed

$$L(m) = n\hat{\mathbf{r}}\left(\hat{\epsilon}_t^2/\hat{h}_t\right)'\hat{\mathbf{V}}^{-1}\hat{\mathbf{r}}\left(\hat{\epsilon}_t^2/\hat{h}_t\right) \tag{14}$$

will be asymptotically distributed as a chi-squared random variable with m degrees of freedom. The matrix $\hat{\mathbf{V}}$ is a consistent estimator for the covariance matrix of $\hat{\mathbf{r}}(\hat{\epsilon}_t^2/\hat{h}_t)$.

In addition, Li and Mak (1994) showed that for an ARCH(b) model (GARCH$(b, 0)$), the residual autocorrelations $\hat{r}_k(\hat{\epsilon}_t^2/\hat{h}_t)$ for $k = b+1, \ldots, m$ are asymptotically iid standard normal. They proposed the modified statistic

$$L(b, m) = n\sum_{k=b+1}^{m}\hat{r}_k^2\left(\hat{\epsilon}_t^2/\hat{h}_t\right). \tag{15}$$

Their simulation study shows only modest improvement using $L(m)$ compared with $L(b, m)$ when the data follow an ARCH process. They suggested that a practitioner may prefer $L(b, m)$ for its simplicity in checking the adequacy of a fitted ARCH model.

Tse (2002) proposed a statistic by modifying the ordinary least squares procedure that includes lagged squared standardized residuals. The statistic can be calculated using the recursive method in Tse (2000). Tsui (2004) provided a nice review of these statistics and a simulation study showing that the statistics of Li and Mak (1994) and Tse (2002) are comparable and more powerful than the general regression-based diagnostic provided in Wooldridge (1991). Li and Li (2005) derived the distribution of $\hat{r}_k(\hat{\epsilon}_t^2/\hat{h}_t)$ after an absolute deviations approach for fitting the GARCH process. They suggested a similar statistic to $L(m)$ in (14).

3.3 Proposed Statistics

Weighted versions of the statistics from Li and Mak (1994) can easily be constructed. For a fitted GARCH(b, a) process, the vector of autocorrelation coefficients on the standardized squared residuals is asymptotically distributed as multivariate normal with covariance matrix $\mathbf{V}$. Their statistic $L(m)$ is a standard quadratic form where $\hat{\mathbf{V}}$ is estimated with a consistent estimator. A Weighted Li–Mak test could be constructed by considering the quadratic form $n\hat{\mathbf{r}}(\hat{\epsilon}_t^2/\hat{h}_t)'\mathbf{W}\hat{\mathbf{r}}(\hat{\epsilon}_t^2/\hat{h}_t)$ where $\mathbf{W}$ is a

Table 1. Empirical size at 5% for AR(1) and MA(1) models, $n = 100$, $m = 20$

	Fitted AR(1)					Fitted MA(1)				
ϕ/θ	$\widetilde{Q}_W$	$\widetilde{M}_W$	D_m	$\widetilde{Q}$	$\widetilde{M}$	$\widetilde{Q}_W$	$\widetilde{M}_W$	D_m	$\widetilde{Q}$	$\widetilde{M}$
0.10	0.042	0.040	0.030	0.059	0.050	0.032	0.026	0.015	0.049	0.035
0.30	0.059	0.043	0.033	0.077	0.059	0.035	0.029	0.020	0.056	0.048
0.50	0.053	0.040	0.034	0.070	0.050	0.045	0.038	0.026	0.064	0.057
0.70	0.031	0.024	0.014	0.045	0.034	0.056	0.039	0.027	0.075	0.056
0.90	0.045	0.038	0.025	0.056	0.047	0.062	0.043	0.032	0.078	0.059

matrix of weights described in the Appendix. The distribution of that statistic would follow the previous theorems. Unfortunately, the covariance matrix $\mathbf{V}$ depends on the asymptotic covariance matrix of the vector of estimated GARCH parameters, which is typically estimated by the Hessian of the maximized likelihood and tends to be numerically unstable. We recommend building a test based on the simpler Li–Mak test for fitted ARCH(b) processes. Consider the statistic

$$L_W(b, m) = n \sum_{k=b+1}^{m} \frac{m - k + (b+1)}{m} \hat{r}_k^2 \left(\hat{\epsilon}_t^2/\hat{h}_t\right). \quad (16)$$

Theorem 4. Under the null hypothesis of an adequately fitted ARCH(b) model, the statistic $L_W(b, m)$ is asymptotically distributed as $\sum_{k=1}^{m} w_k \chi_k^2$ where each χ_k^2 is an independent chi-squared random variable with one degree of freedom and w_k ($k = 1, \ldots, m$) are the weights $w_k = (m - k + (b+1))/m$.

As with the other proposed statistics in this article, the distribution can be approximated with a gamma distribution with shape and scale parameters:

$$\alpha = \frac{3}{4} \frac{(m+b+1)^2(m-b)}{2m^2 + 3m + 2mb + 2b^2 + 3b + 1},$$

and

$$\beta = \frac{2}{3} \frac{2m^2 + 3m + 2mb + 2b^2 + 3b + 1}{m(m+b+1)}.$$

4. COMPUTATIONAL STUDY

As aforementioned, the proposed statistics are easy to implement in any available statistical language. Source code for the proposed statistics are provided in the functions Weighted.Box.test and Weighted.LM.test in the "WeightedPortTest" package of the GNU-licensed R-Project. In this simulation study, the proposed statistics are compared with the statistic D_m from Mahdi and McLeod (2012), $\widetilde{Q}$ from Ljung and Box (1978), $\widetilde{M}$ from Monti (1994), and $L(b, m)$ from Li and Mak (1994). The statistic D_m was chosen over those in Peña and Rodríguez (2002, 2006) since it has conservative Type-I performance and is implemented in the "portes" package in R. Ljung–Box statistic is implemented in the Box.test function and Monti statistic is easy to implement; both were chosen as they are common statistics used by practitioners. To assist the reader, in each included power study, the best performing statistic is noted in boldface.

4.1 Study on Fitted ARMA Models

This section presents a comparative study of the empirical significance level and simulated power for the proposed statistics against those in the literature. This study is similar to that provided in Peña and Rodríguez (2002). Table 1 provides the empirical size of the statistics under low-order AR and MA models. In each case, 1000 replicates were performed where an AR(1) or MA(1) process of size $n = 100$ was generated using several ϕ or θ values. The model was properly fit by an AR(1) or MA(1) process and the test statistics were computed. The values show the empirical significance levels when the statistics were calculated at lag $m = 20$. We see in Table 1 that the statistic of Mahdi and McLeod (2012) tends to be conservative. The Ljung and Box (1978) statistic tends to be liberal, whereas the proposed statistics and that from Monti (1994) tend to be close to the nominal level of 0.05. At higher ordered models, the proposed statistics tend to be conservative as seen in the additional simulations in the supplementary documentation.

A study similar to that of Monti (1994) and Peña and Rodríguez (2002, 2006) indicates the relative power of the statistics in detecting 24 different under fit ARMA(2,2) models. Table 2 shows the power of the five statistics when, erroneously, an AR(1) or MA(1) model is fitted to the ARMA(2,2) process. In each case, 1000 replicates of $n = 100$ observations were generated and the power was calculated at lag $m = 20$. We see in Table 2 that one of the two proposed statistics is always the most powerful, or tied for most powerful. Apparently, the weighting improves power.

Overall, we conclude that the proposed statistics are comparable with those in Mahdi and McLeod (2012) (and likewise Peña and Rodríguez 2002, 2006) and both tend to outperform the commonly used Ljung and Box (1978) and Monti (1994).

4.2 Detecting Nonlinear Processes

The study in Pérez and Ruiz (2003) suggests that statistics based on the squared and absolute residuals are more powerful than those based on the log of the squared residuals, particularly when the true model contains some long persistence. Based on their result, we restrict our study to the case of squared and absolute residuals.

Consider four nonlinear models proposed in Keenan (1985):

NL-1: $Y_t = \epsilon_t - 0.4\epsilon_{t-1} + 0.3\epsilon_{t-2} + 0.5\epsilon_t\epsilon_{t-2}$;
NL-2: $Y_t = \epsilon_t - 0.3\epsilon_{t-1} + 0.2\epsilon_{t-2} + 0.4\epsilon_{t-1}\epsilon_{t-2} - 0.25\epsilon_{t-2}^2$;
NL-3: $Y_t = 0.4Y_{t-1} - 0.3Y_{t-2} + 0.5Y_{t-1}\epsilon_{t-1} + \epsilon_t$;
NL-4: $Y_t = 0.4Y_{t-1} - 0.3Y_{t-2} + 0.5Y_{t-1}\epsilon_{t-1} + 0.8\epsilon_{t-1} + \epsilon_t$.

Table 2. Power levels based on 5% significance when data are generated from ARMA(2,2) models and AR(1) or MA(1) models are fitted, $n = 100$, $m = 20$

Model #	ϕ_1	ϕ_2	θ_1	θ_2	$\widetilde{Q}_W$	$\widetilde{M}_W$	D_m	$\widetilde{Q}$	$\widetilde{M}$
				Fitted with AR(1) model					
1	–	–	−0.50	–	0.275	**0.286**	0.250	0.223	0.210
2	–	–	−0.80	–	0.792	**0.966**	0.956	0.624	0.856
3	–	–	−0.60	0.30	0.805	**0.990**	0.986	0.632	0.937
4	0.10	0.30	–	–	**0.463**	0.421	0.394	0.375	0.295
5	1.30	−0.35	–	–	**0.745**	0.721	0.704	0.615	0.517
6	0.70	–	−0.40	–	0.573	**0.645**	0.618	0.442	0.455
7	0.70	–	−0.90	–	0.997	**1.000**	**1.000**	0.927	**1.000**
8	0.40	–	−0.60	0.30	0.879	**0.998**	**0.998**	0.707	0.979
9	0.70	–	0.70	−0.15	**0.156**	0.140	0.114	0.138	0.111
10	0.70	0.20	0.50	–	**0.759**	0.755	0.729	0.650	0.613
11	0.70	0.20	−0.50	–	0.369	**0.465**	0.422	0.293	0.271
12	0.90	−0.40	1.20	−0.30	0.738	**0.963**	0.953	0.579	0.887
				Fitted with MA(1) model					
13	0.50	–	–	–	**0.323**	0.290	0.251	0.274	0.207
14	0.80	–	–	–	**0.986**	0.974	0.974	0.962	0.926
15	1.10	−0.35	–	–	0.996	**0.997**	**0.997**	0.986	0.986
16	–	–	0.80	−0.50	0.860	**0.937**	0.924	0.696	0.800
17	–	–	−0.60	0.30	0.420	**0.473**	0.441	0.319	0.306
18	0.50	–	−0.70	–	**0.897**	0.869	0.854	0.777	0.695
19	−0.50	–	0.70	–	**0.901**	0.882	0.876	0.796	0.740
20	0.30	–	0.80	−0.50	0.640	**0.762**	0.727	0.498	0.590
21	0.80	–	−0.50	0.30	**0.986**	0.978	0.977	0.966	0.929
22	1.20	−0.50	0.90	–	0.439	**0.677**	0.617	0.383	0.596
23	0.30	−0.20	−0.70	–	0.238	**0.263**	0.230	0.228	0.207
24	0.90	−0.40	1.20	−0.30	0.799	**0.939**	0.929	0.615	0.828

NOTE: The best performing statistic is noted in bold.

For each model, 1000 replicates of sample size $n = 204$ are generated and an AR(p) is fitted to the data, where p is selected by the Akaike Information Criterion (AIC) with $p \in \{1, 2, 3, 4\}$. This study is analogous to those performed in Peña and Rodríguez (2002, 2006). The simulation results in Table 3 show that the statistics based on the squared residuals are more powerful for the first three nonlinear models, whereas in NL-4, the statistics based on absolute residuals are the most powerful. In general, the proposed Weighted McLeod–Li type test appears to be the most powerful in detecting the nonlinear process, with a few exceptions in which the Weighted Monti is most powerful or ties Mahdi–McLeod as the most powerful. As with the results in Tsay (1986) and Peña and Rodríguez (2002, 2006), we see that the statistics have difficulty in detecting the nonlinearity of NL-1. Empirical sizes for these nonlinear detection statistics are provided in the supplementary documents.

The study in Peña and Rodríguez (2002) suggests that the statistic based on the likelihood ratio criterion is less powerful than the traditional McLeod–Li or Monti type statistics when detecting a heteroscedastic process with long persistence. Rodríguez and Ruiz (2005) proposed a new statistic using similar methodology as McLeod and Li (1983) attempting to improve the results by including additional information in the statistic to detect possible patterns among the sample

Table 3. Power levels of the $\widetilde{Q}^*_W$, $\widetilde{M}^*_W$, D^*_m, $\widetilde{Q}^*$, and $\widetilde{M}^*$ tests at 5% significance for four nonlinear models and the fitted model is an AR(p), $n = 204$

		$m = 7$					$m = 24$				
Model		$\widetilde{Q}^*_W$	$\widetilde{M}^*_W$	D^*_m	$\widetilde{Q}^*$	$\widetilde{M}^*$	$\widetilde{Q}^*_W$	$\widetilde{M}^*_W$	D^*_m	$\widetilde{Q}^*$	$\widetilde{M}^*$
NL-1	ϵ^2	0.159	**0.162**	**0.162**	.132	0.145	**0.120**	0.118	0.108	0.079	0.082
	$\|\epsilon\|$	**0.073**	0.068	0.065	**0.073**	**0.073**	**0.069**	0.068	0.055	0.062	0.065
NL-2	ϵ^2	**0.615**	0.607	0.601	0.504	0.501	**0.459**	0.448	0.423	0.321	0.295
	$\|\epsilon\|$	**0.484**	0.467	0.462	0.391	0.358	**0.373**	0.351	0.325	0.273	0.223
NL-3	ϵ^2	**0.970**	0.965	0.962	0.914	0.915	**0.898**	0.892	0.887	0.757	0.737
	$\|\epsilon\|$	0.932	**0.936**	0.935	0.858	0.849	**0.851**	0.834	0.824	0.732	0.659
NL-4	ϵ^2	**0.929**	0.926	0.926	0.864	0.841	**0.825**	0.797	0.788	0.675	0.611
	$\|\epsilon\|$	**0.963**	0.959	0.959	0.925	0.902	**0.904**	0.885	0.880	0.801	0.766

NOTE: The best performing statistic is noted in bold.

Table 4. Power levels of the $\widetilde{Q}^*_W$, $\widetilde{M}^*_W$, D^*_m, $\widetilde{Q}^*$, and $\widetilde{M}^*$ tests at 5% significance for detecting GARCH(1,1) models of financial time series

			$m = 12$					$m = 24$				
n	Model		$\widetilde{Q}^*_W$	$\widetilde{M}^*_W$	D^*_m	$\widetilde{Q}^*$	$\widetilde{M}^*$	$\widetilde{Q}^*_W$	$\widetilde{M}^*_W$	D^*_m	$\widetilde{Q}^*$	$\widetilde{M}^*$
250	GF-1	ϵ^2	**0.317**	0.280	0.271	0.285	0.251	**0.309**	0.270	0.253	0.276	0.222
		$\lvert\epsilon\rvert$	**0.246**	0.202	0.192	0.230	0.161	**0.245**	0.177	0.166	0.217	0.141
	GF-2	ϵ^2	**0.877**	0.844	0.839	0.847	0.794	**0.861**	0.822	0.808	0.802	0.723
		$\lvert\epsilon\rvert$	**0.846**	0.802	0.796	0.813	0.744	**0.836**	0.781	0.762	0.782	0.668
500	GF-1	ϵ^2	**0.561**	0.519	0.514	0.556	0.490	**0.571**	0.515	0.504	0.515	0.441
		$\lvert\epsilon\rvert$	0.457	0.400	0.395	**0.458**	0.373	**0.470**	0.400	0.382	0.428	0.300
	GF-2	ϵ^2	**0.985**	0.981	0.981	0.982	0.973	**0.982**	0.979	0.979	0.973	0.952
		$\lvert\epsilon\rvert$	**0.979**	0.973	0.973	0.973	0.957	**0.975**	0.966	0.964	0.965	0.934

NOTE: The best performing statistic is noted in bold.

autocorrelation. Their statistic has an embedded structure and is a generalization of the McLeod–Li statistic. We omit it from our study, as it is outside the scope and aim of this article.

We perform the study from Peña and Rodríguez (2002) involving the financial models from Carnero, Peña, and Ruiz (2001). A GARCH(1,1) process (9) is generated with two sets of parameters: GF-1: $(\omega, \alpha, \beta) = (1, 0.05, 0.90)$ and GF-2: $(\omega, \alpha, \beta) = (1, 0.15, 0.80)$. One thousand replicates of two different sample sizes are generated, the statistics based on the squared and absolute residuals are computed, and the empirical power is found at lags $m = 12$ and $m = 24$. Table 4 shows that in general, the statistics based on the squared residuals are more powerful. Furthermore, the proposed Weighted McLeod–Li type test tends to be the most powerful, with the exception of one instance when it is comparable to the most powerful McLeod–Li type test. Generally, the McLeod–Li type test outperforms that based on the likelihood ratio criterion. As with the previous study, there is a decrease in power as m increases, although the decrease appears to be less pronounced for the Weighted McLeod–Li type statistic. A similar study based on the results of Tol (1996) is provided in the supplementary documents.

The final study for detecting heteroscedastic processes studies the long memory stochastic volatility (LMSV) models found in Pérez and Ruiz (2003),

$$Y_t = \exp(h_t/2)\epsilon_t$$
$$(1 - \phi B)(1 - B)^d h_t = \eta_t,$$

Table 5. Power levels of the $\widetilde{Q}^*_W$, $\widetilde{M}^*_W$, D^*_m, $\widetilde{Q}^*$, and $\widetilde{M}^*$ tests at 5% significance for long memory stochastic volatility processes

			$m = 10$					$m = 50$				
n	Model		$\widetilde{Q}^*_W$	$\widetilde{M}^*_W$	D^*_m	$\widetilde{Q}^*$	$\widetilde{M}^*$	$\widetilde{Q}^*_W$	$\widetilde{M}^*_W$	D^*_m	$\widetilde{Q}^*$	$\widetilde{M}^*$
256	LM-1	ϵ^2	**0.907**	0.881	0.878	0.895	0.853	**0.878**	0.823	0.796	0.815	0.696
		$\lvert\epsilon\rvert$	0.937	0.922	0.919	**0.941**	0.908	**0.930**	0.877	0.845	0.894	0.748
	LM-2	ϵ^2	**0.287**	0.252	0.247	0.267	0.247	**0.266**	0.218	0.160	0.197	0.123
		$\lvert\epsilon\rvert$	**0.236**	0.195	0.189	0.235	0.180	**0.232**	0.150	0.103	0.192	0.104
	LM-3	ϵ^2	**0.934**	0.914	0.912	0.927	0.892	**0.893**	0.834	0.811	0.796	0.633
		$\lvert\epsilon\rvert$	**0.987**	0.981	0.981	0.985	0.971	**0.973**	0.943	0.935	0.950	0.841
	LM-4	ϵ^2	**0.402**	0.373	0.369	0.396	0.348	**0.402**	0.330	0.273	0.312	0.203
		$\lvert\epsilon\rvert$	**0.390**	0.341	0.332	0.378	0.313	**0.378**	0.258	0.211	0.315	0.162
	LM-5	ϵ^2	**0.825**	0.795	0.791	0.822	0.774	**0.805**	0.730	0.691	0.724	0.574
		$\lvert\epsilon\rvert$	**0.843**	0.802	0.797	0.839	0.772	**0.821**	0.714	0.676	0.757	0.574
	LM-6	ϵ^2	**0.201**	0.189	0.185	0.192	0.180	**0.180**	0.153	0.112	0.136	0.103
		$\lvert\epsilon\rvert$	0.135	0.122	0.117	**0.138**	0.109	**0.135**	0.095	0.068	0.121	0.070
512	LM-1	ϵ^2	0.994	0.994	0.994	**0.995**	0.992	**0.994**	0.990	0.989	0.987	0.967
		$\lvert\epsilon\rvert$	**0.999**	**0.999**	**0.999**	0.998	0.997	**0.998**	0.997	0.997	**0.998**	0.984
	LM-2	ϵ^2	0.480	0.439	0.437	**0.484**	0.431	**0.490**	0.411	0.385	0.405	0.313
		$\lvert\epsilon\rvert$	**0.430**	0.384	0.374	0.429	0.353	**0.425**	0.300	0.276	0.340	0.217
	LM-3	ϵ^2	**0.993**	0.991	0.991	**0.993**	0.988	**0.987**	0.975	0.972	0.967	0.925
		$\lvert\epsilon\rvert$	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	0.998
	LM-4	ϵ^2	**0.738**	0.701	0.699	0.737	0.675	**0.739**	0.674	0.643	0.663	0.545
		$\lvert\epsilon\rvert$	**0.709**	0.643	0.638	0.699	0.623	**0.701**	0.591	0.549	0.624	0.432
	LM-5	ϵ^2	0.989	0.982	0.982	**0.992**	0.984	**0.991**	0.978	0.978	0.977	0.922
		$\lvert\epsilon\rvert$	**0.993**	0.991	0.991	0.991	0.983	**0.992**	0.978	0.977	0.983	0.950
	LM-6	ϵ^2	**0.297**	0.275	0.272	0.293	0.265	**0.310**	0.266	0.242	0.263	0.205
		$\lvert\epsilon\rvert$	**0.265**	0.226	0.220	0.256	0.217	**0.283**	0.207	0.164	0.237	0.131

NOTE: The best performing statistic is noted in bold.

that is, the log volatility follows an AR fractionally integrated MA, or ARFIMA(p, d, q). We note that when $d = 0$, we have a short memory autoregressive stochastic volatility model (ARSV). Table 5 provides the power of the statistics based on the squared and absolute residuals after 1000 replicates of two sample sizes were calculated and tested at lags $m = 10$ and $m = 50$. The six different long memory stochastic volatility models considered are LM-1: $(\phi, d, \sigma_\eta^2) = (0.98, 0, 0.05)$, LM-2: $(\phi, d, \sigma_\eta^2) = (0.9, 0.2, 0.01)$, LM-3: $(\phi, d, \sigma_\eta^2) = (0.9, 0.2, 0.1)$, LM-4: $(\phi, d, \sigma_\eta^2) = (0.8, 0.45, 0.01)$, LM-5: $(\phi, d, \sigma_\eta^2) = (0.9, 0.45, 0.01)$, and LM-6: $(\phi, d, \sigma_\eta^2) = (0.0, 0.45, 0.1)$.

Table 5 shows that on models LM-1, LM-3, and LM-5, the statistics based on the absolute residuals tend to be the most powerful. On models LM-2, LM-4, and LM-6, the statistics based on the squared residuals are the most powerful; this follows the results in Pérez and Ruiz (2003). We also note that for lag $m = 10$, the proposed Weighted McLeod–Li type and original McLeod–Li type tests are comparable and the most powerful in every case. When testing at the larger lag of $m = 50$, we see the proposed Weighted McLeod–Li type test is always the most powerful and its decrease in power appears less pronounced than that of the McLeod–Li type.

4.3 Fitted ARCH Processes

In Section 3.2, the versatility of the methodology is shown as it is extended to check for the adequacy of a GARCH or ARCH fit. As discussed in Section 3.2, a Weighted Li and Mak or Li and Li statistic can be constructed but the exact distribution (and approximation) under the null hypothesis will depend on a consistent estimate for the covariance of the autocorrelations. We omit that study here, although an example is provided in the supplementary documents.

Table 6 provides the empirical Type I errors where the data are generated, and properly fit, by two different AR(1)-ARCH(b) models suggested in Li and Mak (1994): ARCH-1: $(\phi, \omega, \alpha_1) = (0.2, 0.2, 0.2)$ and ARCH-2: $(\phi, \omega, \alpha_1, \alpha_2) = (0.2, 0.2, 0.2, 0.2)$ where ARCH-1 has $b = 1$ and ARCH-2 has $b = 2$. Four different sample sizes are considered and 1000 replicates are performed. We see in Table 6 that the simulated size is comparable to that found in Li and Mak (1994). We extend their study by comparing the statistics at two lags, $m = 6$ and $m = 12$. All of the statistics are close to the nominal size of 0.05 or conservative.

Table 6. Empirical size at size 5% for ARCH-1 and ARCH-2 models

	$m = 6$		$m = 12$	
	$L_W(b, m)$	$L(b, m)$	$L_W(b, m)$	$L(b, m)$
	Model ARCH-1, $b = 1$			
$n = 100$	0.022	0.020	0.017	0.014
$n = 200$	0.039	0.037	0.038	0.037
$n = 300$	0.036	0.040	0.033	0.040
$n = 400$	0.038	0.033	0.033	0.041
	Model ARCH-2, $b = 2$			
$n = 100$	0.024	0.016	0.019	0.016
$n = 200$	0.038	0.027	0.034	0.024
$n = 300$	0.050	0.051	0.042	0.033
$n = 400$	0.047	0.042	0.039	0.043

Table 7. Power levels at size 0.05 for ARCH(1) and ARCH(2) models

	$m = 6$		$m = 12$	
	$L_W(b, m)$	$L(b, m)$	$L_W(b, m)$	$L(b, m)$
	AR(1)-ARCH(1) fitted to ARCH-2, $b = 1$			
$n = 100$	**0.185**	0.134	**0.139**	0.081
$n = 200$	**0.393**	0.330	**0.338**	0.233
$n = 300$	**0.579**	0.491	**0.511**	0.399
$n = 400$	**0.753**	0.662	**0.676**	0.553
	AR(1)-ARCH(2) fitted to ARCH-3, $b = 2$			
$n = 100$	**0.088**	0.084	**0.079**	0.062
$n = 200$	**0.207**	0.197	**0.193**	0.158
$n = 300$	**0.343**	0.323	**0.331**	0.260
$n = 400$	**0.471**	0.454	**0.450**	0.355

NOTE: The best performing statistic is noted in bold.

Table 7 supplies a power study analogous to that in Li and Mak (1994). The data are generated from models ARCH-2 and model ARCH-3: $(\phi, \omega, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5) = (0.2, 0.2, 0.2, 0.2, 0.1, 0.05, 0.05)$. An inadequate AR(1)-ARCH(1) or AR(1)-ARCH(2) are fit, respectively. One thousand replicates were performed on four different sample sizes and the power levels are reported. Table 7 shows that the proposed statistic is more powerful in detecting an inadequate ARCH fit. We see an improvement of upwards of 10% for $n = 300$ when an AR(1)-ARCH(1) is fit to a true model of AR(1)-ARCH(2).

4.4 Monte Carlo Methods and Computational Efficiency

Lin and McLeod (2006) and Mahdi and McLeod (2012) recommended a Monte Carlo method to obtain the distribution and p-values for the statistic based on the determinant of the autocorrelation matrix $\widehat{\mathbf{R}}_m$. With advances in modern computing power, particularly with parallel processing, this procedure can generally be handled quite easily. However, in some applications, the use of a Monte Carlo method may be too expensive due to the computational efficiency of the statistic. In practice, calculation of the determinant of an $m \times m$ matrix requires on the order of $O(m^2)$ operations, where our proposed statistics only require $O(m)$ operations. Further studies are included in the supplemental documents. If a practitioner wishes to use a Monte Carlo method, the more computationally efficient proposed statistics may be preferred.

4.5 Data Analysis

To demonstrate the effectiveness of the proposed methods, consider the analysis of two return series. We examine the log daily closing return of Apple, Inc., stock (AAPL) and the Nikkei-300 index (N300) on the market days from May 1, 2006, through October 31, 2007. This period essentially accounts for the year and a half before the very volatile recession that began in late 2007. The two series can be seen in Figure 1.

When analyzing the Apple, Inc., log returns, using any of the statistics, no ARMA process is detected. Furthermore, through analysis of the squared series, no nonlinear process is detected using any method. Table 8 provides the p-values of the five

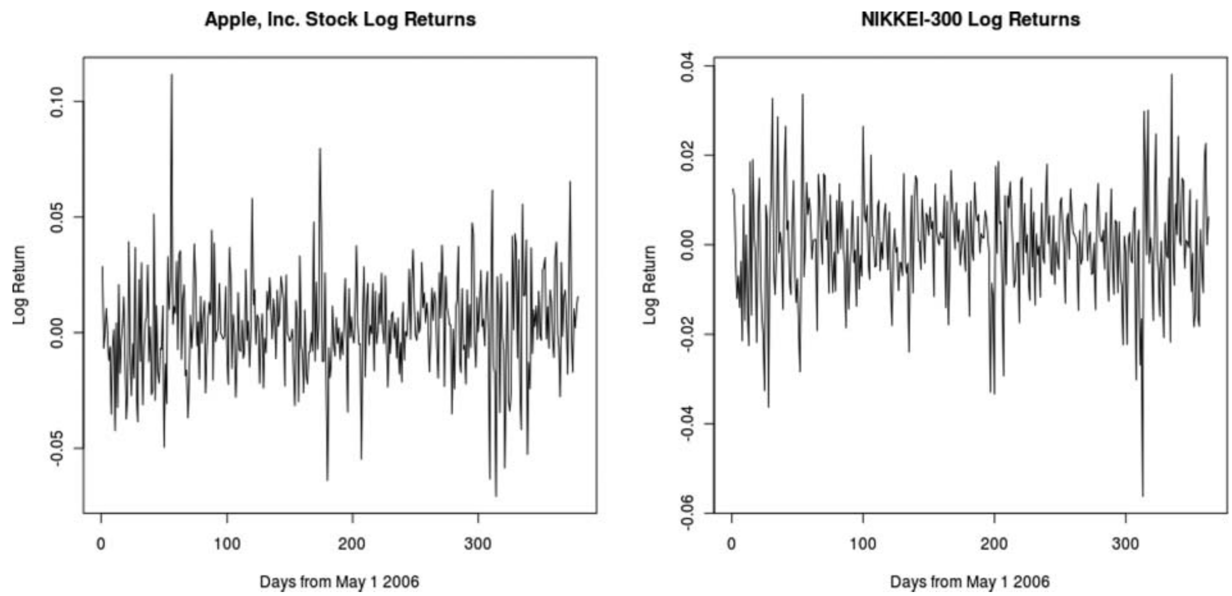

Figure 1. Apple, Inc., and Nikkei-300 return May 1, 2006, through October 31, 2007.

aforementioned statistics using absolute values at various lags m. The simulation study suggested the proposed Weighted McLeod–Li type statistic appeared to be the most powerful in detecting long memory processes, particularly when testing at larger lags. As seen in Table 8, in general, only the proposed Weighted McLeod–Li type and the McLeod–Li type statistic suggest any presence of a nonlinear process. It is also interesting to note that the proposed statistic tends to have more stable p-values. That is, once the nonlinear process is detected, it continues to be detected, even when testing at larger lags, whereas the McLeod–Li type statistic becomes insignificant.

A similar phenomenon is seen with the Weighted Li–Mak statistic. When analyzing the Nikkei-300 index, no ARMA process is detected. Every statistic, using both the squared series and absolute series, suggests the presence of a nonlinear process. Following the work of Li and Mak (1994), we fit increasing orders of ARCH processes to the index series. After fitting an ARCH(3) model to the series, Table 9 provides the p-values of the proposed statistics and that of Li and Mak (1994) at various lags. The proposed Weighted Li–Mak test appears to suggest that an ARCH(3) is inadequate, while the conclusion of the Li and Mak (1994) test is not consistent across m. The proposed method tends to be more stable as the lag m increases. Even at $m = 50$ when the statistic is insignificant, it still suggests weak significance while that of Li–Mak has a p-value greater than 0.30.

Table 8. The p-values in detecting nonlinear effects in Apple, Inc., returns using absolute values

m	$\widetilde{Q}_W^*$	$\widetilde{M}_W^*$	D_m^*	$\widetilde{Q}^*$	$\widetilde{M}^*$
10	0.1033	0.2089	0.2170	0.1480	0.3454
20	0.0148	0.0707	0.0816	0.0037	0.0266
30	0.0079	0.0482	0.0610	0.0133	0.0665
40	0.0066	0.0541	0.0735	0.0231	0.1422
50	0.0092	0.0773	0.1095	0.0368	0.2029
60	0.0123	0.1043	0.1546	0.0878	0.3853

5. DISCUSSION

This article introduces new time series goodness of fit tests, which can be derived by considering traces of squares of estimated correlation matrices. Using the trace of $\widehat{\mathbf{R}}_m^2$ rather than the determinant results in a test statistic that is much more stable in terms of m. The resulting weighted portmanteau tests are easy to compute and numerically stable, but still tend to be more powerful than the usual Ljung–Box tests in detecting under fit ARMA processes. In many cases, the proposed methods are more effective in detecting nonlinear effects compared with those in the literature. The simulations and data analysis considered in this article demonstrate that the weighting scheme also results in test statistics that are less sensitive to the choice of m; that is, the conclusion of the test is more consistent as m varies.

There are several open questions pointed out by referees. In what sense is the weighting scheme optimal? It may be possible to analytically show that the large sample distributional approximation is optimal if the weights decrease in the lag. A second notion of optimality is in terms of power of detecting under fit models. Since for ARMA processes the autocorrelations (eventually) decay exponentially in the lag, it may be the case that in the class of ARMA processes, a decreasing sequence of weights can optimize power in certain settings. It has been known for some time that partial correlations pick up missing MA components, while autocorrelations are better at detecting missing

Table 9. The p-values of statistics after ARCH(3) has been fit to Nikkei-300 index series

m	$L_W(b, m)$	$L(b, m)$
10	0.0298	0.0115
20	0.0335	0.0883
30	0.0184	0.0244
40	0.0341	0.1724
50	0.0690	0.3227

autoregressive terms. Can a combination of $\widetilde{Q}_W$ and $\widetilde{M}_W$ be used to improve performance in detecting under fit ARMA models?

APPENDIX: TECHNICAL DETAILS

A.1 Testing on ARMA Models

Proof of Theorem 1. Both $\widetilde{Q}_W$ and $\widetilde{M}_W$ can be expressed as quadratic forms. Define

$$\mathbf{W} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & \dfrac{m-1}{m} & \cdots & 0 \\ \vdots & \cdots & \ddots & \vdots \\ 0 & \cdots & 0 & \dfrac{1}{m} \end{bmatrix}.$$

as the diagonal matrix of our weights. Then $\widetilde{Q}_W$ and $\widetilde{M}_W$ are asymptotically expressed as the quadratic forms

$$\widetilde{Q}_W \simeq n\hat{\mathbf{r}}'\mathbf{W}\hat{\mathbf{r}} \quad \text{and} \quad \widetilde{M}_W \simeq n\hat{\boldsymbol{\pi}}'\mathbf{W}\hat{\boldsymbol{\pi}} \quad \text{as } n \to \infty,$$

respectively, where $\mathbf{A}'$ denotes the transpose operation on vector/matrix $\mathbf{A}$ and $\hat{\mathbf{r}}$ ($\hat{\boldsymbol{\pi}}$) is the $m \times 1$ vector of the autocorrelations (partial autocorrelations) from lag 1 to m. From the results in Box (1954), both quadratic forms will be distributed as

$$\sum_{k=1}^{m} \lambda_k \chi_k^2, \tag{A.1}$$

where each χ_k^2 are independently distributed chi-squared random variables with one degree of freedom, and the λ_k are the m real nonzero characteristic roots of the matrix $(\mathbf{I}-\mathbf{Q})\mathbf{W}$ where $\mathbf{I}-\mathbf{Q}$ is the $m \times m$ covariance matrix of both $\sqrt{n}\hat{\mathbf{r}}$ and $\sqrt{n}\hat{\boldsymbol{\pi}}$. Box and Pierce (1970) and McLeod (1978) approximated the matrix $\mathbf{Q}$ by the projection matrix $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ when m is moderately high.

Gamma Approximation Box (1954) and Satterthwaite (1941, 1946) demonstrated that the distribution of this particular quadratic form can be well approximated by $c\chi_k^2$, or gamma distribution. The parameters are chosen so that the distribution has the same first two cumulants as the quadratic form. In particular, consider the shape and scale parameters for a gamma distribution:

$$\alpha = \frac{K_1^2}{K_2} \quad \text{and} \quad \beta = \frac{K_2}{K_1},$$

where K_i is the ith cumulant of the distribution (A.1). Box (1954) provided a formula (theorem 2.2 in Box 1954) for the cumulants and a simple algebraic manipulation of the results in Peña and Rodríguez (2002) provides

$$K_1 = \sum_{i=1}^{m} \lambda_i = \mathrm{tr}((\mathbf{I}-\mathbf{Q})\mathbf{W}) = \frac{m+1}{2} - (p+q) + \frac{1}{m}\sum_{i=2}^{m}(i-1)q_{ii},$$

and

$$\begin{aligned} K_2 = 2\sum_{i=1}^{m}\lambda_i^2 &= 2\mathrm{tr}\,((\mathbf{I}-\mathbf{Q})\mathbf{W}(\mathbf{I}-\mathbf{Q})\mathbf{W}) \\ &= \frac{1}{3m}(m+1)(2m+1) - 2(p+q) \\ &\quad + \frac{4}{m}\sum_{i=2}^{m}(i-1)q_{ii} - \frac{4}{m^2}\sum_{i=2}^{m}(i-1)^2 q_{ii} \\ &\quad + \frac{2}{m^2}\sum_{i=2}^{m}\sum_{j=2}^{m}(i-1)(j-1)q_{ij}^2, \end{aligned}$$

where q_{ij} is the element of the matrix $\mathbf{Q}$ on row i, column j. Peña and Rodríguez (2002) demonstrated that as m grows large, the $O(m^{-1})$ and $O(m^{-2})$ terms tend to zero. We recommend using an upper bound on K_1 as it improves the approximation under the null distribution and the limiting argument for K_2 from Peña and Rodríguez (2002):

$$K_1 = \frac{m+1}{2}, \tag{A.2}$$

and

$$K_2 = \frac{(m+1)(2m+1)}{3m} - 2(p+q). \tag{A.3}$$

Proof of Theorem 3. From the result in McLeod and Li (1983), the autocorrelations of the squared residuals are asymptotically normally distribution as $\sqrt{n}\hat{\mathbf{r}}(\hat{\epsilon}_t^2) \to N(0, \mathbf{I}_m)$ where $\mathbf{I}_m$ is an $m \times m$ identity matrix. Applying this result to the results in Monti (1994) provides the asymptotic distribution $\sqrt{n}\hat{\boldsymbol{\pi}}(\hat{\epsilon}_t^2) \to N(0, \mathbf{I}_m)$. Both $\widetilde{Q}_W(\hat{\epsilon}_t^2)$ and $\widetilde{M}_W(\hat{\epsilon}_t^2)$ can be expressed as quadratic forms as in Theorem 1 and the result follows. We note that the gamma approximation follows the previous result with the exception that the cumulants are easier to find since the covariance matrix of the autocorrelations and partial autocorrelations is an identity matrix.

A.2 Testing on ARCH Models

Proof of Theorem 4. From the result in Li and Mak (1994), for an ARCH(b) model, the autocorrelations of the standardized squared residuals are asymptotically normally distributed as $N(0, \mathbf{I}_{m-(b+1)})$ where $\mathbf{I}_{m-(b+1)}$ is an $(m-(b+1)) \times (m-(b+1))$ identity matrix. $L_W(b, m)$ can be expressed as a quadratic form and the result follows that in A.1.

Cumulants for the Gamma Approximation. We note that only $m-(b+1)$ terms are in the above statistic, but the weights are still based on m. Only a submatrix of $\mathbf{W}$ from above is used, the cumulants are

$$K_1 = \frac{m(m+1) - b(b+1)}{2m},$$

and

$$K_2 = \frac{m(m+1)(2m+1) - b(b+1)(2b+1)}{3m^2}.$$

SUPPLEMENTARY MATERIALS

Proof of Theorem 2: Outline of a proof to Theorem 2 in Section 2.2. (mtoinfty.pdf, Portable Document Format file)

Additional Simulations: Extension of Section 4 including additional simulation studies. (simulations.pdf, Portable Document Format file)

Data for analysis: Apple, Inc., (AAPL) stock and Nikkei-300 index (N300) closing prices from May 1, 2006, through October 31, 2007. (apple0607.csv and n300-0607.csv, comma-separated values file)

WeightedPortTest R-package: R-package contaning source code for the weighted portmanteau tests introduced in this article. (WeightedPortTest_1.0.tar.gz, GNU zipped tar file)

[Received August 2011. Revised January 2012.]

REFERENCES

Bhansali, R. J., Giraitis, L., and Kokoszka, P. S. (2007), "Convergence of Quadratic Forms With Nonvanishing Diagonal," *Statistics and Probability Letters*, 77 (7), 726–734. [779]

Bickel, P. J., and Levina, E. (2008), "Regularized Estimation of Large Covariance Matrices," *The Annals of Statistics*, 36 (1), 199–227. [778]

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31 (3), 307–327. [779,780]

Box, G. E. P. (1954), "Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems. I. Effect of Inequality of Variance in the One-Way Classification," *The Annals of Mathematical Statistics*, 25, 290–302. [778,779,786]

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*, 65, 1509–1526. [777,786]

Brockwell, P. J., and Davis, R. A. (1991), *Time Series: Theory and Methods* (Springer Series in Statistics, 2nd ed.), New York: Springer-Verlag. [778]

Carnero, M. A., Peña, D., and Ruiz, E. (2001), "Outliers and Conditional Autoregressive Heteroscedasticity in Time Series," *Estadística*, 53, 143–213. [783]

Engle, R. F. (1982), "Autoregressive Conditional Heteroscedasticity With Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50 (4), 987–1007. [780]

Fisher, T. J., Sun, X., and Gallagher, C. M. (2010), "A New Test for Sphericity of the Covariance Matrix for High Dimensional Data," *Journal of Multivariate Analysis*, 101 (10), 2554–2570. [778]

Higgins, M. L., and Bera, A. K. (1992), "A Class of Nonlinear ARCH Models," *International Economic Review*, 33 (1), 137–158. [780]

Imhof, J. P. (1961), "Computing the Distribution of Quadratic Forms in Normal Variables," *Biometrika*, 48, 419–426. [779]

Keenan, D. M. (1985), "A Tukey Nonadditivity-Type Test for Time Series Nonlinearity," *Biometrika*, 72, 39–44. [781]

Li, G., and Li, W. K. (2005), "Diagnostic Checking for Time Series Models With Conditional Heteroscedasticity Estimated by the Least Absolute Deviation Approach," *Biometrika*, 92 (3), 691–701. [780]

Li, W. K., and Mak, T. K. (1994), "On the Squared Residual Autocorrelations in Non-Linear Time Series With Conditional Heteroskedasticity," *Journal of Time Series Analysis*, 15 (6), 627–636. [780,781,784,785,786]

Lin, J.-W., and McLeod, A. I. (2006), "Improved Peña-Rodriguez Portmanteau Test," *Computational Statistics and Data Analysis*, 51 (3), 1731–1738. [778,779,784]

Ljung, G. M. (1986), "Diagnostic Testing of Univariate Time Series Models," *Biometrika*, 73 (3), 725–730. [777]

Ljung, G. M., and Box, G. E. P. (1978), "On a Measure of Lack of Fit in Time Series Models," *Biometrika*, 65 (2), 297–303. [777,779,781]

Mahdi, E., and McLeod, I. A. (2012), "Improved Multivariate Portmanteau Test," *Journal of Time Series Analysis*, 33 (2), 211–222. [778,779,780,781,784]

McLeod, A. I. (1978), "On the Distribution of Residual Autocorrelations in Box-Jenkins Models," *Journal of the Royal Statistical Society,* Series B, 40 (3), 296–302. [786]

McLeod, A. I., and Li, W. K. (1983), "Diagnostic Checking ARMA Time Series Models Using Squared-Residual Autocorrelations," *Journal of Time Series Analysis*, 4 (4), 269–273. [779,780,782,786]

Monti, A. C. (1994), "A Proposal for a Residual Autocorrelation Test in Linear Models," *Biometrika*, 81 (4), 776–780. [777,779,781,786]

Peña, D., and Rodríguez, J. (2002), "A Powerful Portmanteau Test of Lack of Fit for Time Series," *Journal of the American Statistical Association*, 97 (458), 601–610. [777,778,779,780,781,782,783,786]

——— (2006), "The Log of the Determinant of the Autocorrelation Matrix for Testing Goodness of Fit in Time Series," *Journal of Statistical Planning and Inference*, 136 (8), 2706–2718. [778,779,781]

Pérez, A., and Ruiz, E. (2003), "Properties of the Sample Autocorrelations of Nonlinear Transformations in Long-Memory Stochastic Volatility Models," *Journal of Financial Econometrics*, 1 (3), 420–444. [779,781,783,784]

Rodríguez, J., and Ruiz, E. (2005), "A Powerful Test for Conditional Heteroscedasticity for Financial Time Series With Highly Persistent Volatilities," *Statistica Sinica*, 15 (2), 505–525. [782]

Satterthwaite, F. E. (1941), "Synthesis of Variance," *Psychometrika*, 6, 309–316. [778,779,786]

——— (1946), "An Approximate Distribution of Estimates of Variance Components," *Biometrics Bulletin*, 2 (6), 110–114. [778,779,786]

Schott, J. R. (2005), "Testing for Complete Independence in High Dimensions," *Biometrika*, 92 (4), 951–956. [778]

Srivastava, M. S. (2005), "Some Tests Concerning the Covariance Matrix in High Dimensional Data," *Journal of Japan Statistical Society*, 35 (2), 251–272. [778]

Srivastava, M. S., Kollo, T., and von Rosen, D. (2011), "Some Tests for the Covariance Matrix With Fewer Observations Than the Dimension Under Non-Normality," *Journal of Multivariate Analysis*, 102 (6), 1090–1103. [778]

Taylor, S. J. (1986), *Modelling Financial Time Series*, New York: Wiley. [779]

Tol, R. S. (1996), "Autoregressive Conditional Heteroscedasticity in Daily Temperature Measurements," *Environmetrics*, 7 (1), 67–75. [783]

Tsay, R. S. (1986), "Nonlinearity Tests for Time Series," *Biometrika*, 73 (2), 461–466. [781]

Tse, Y. K. (2000), "A Test for Constant Correlations in a Multivariate GARCH Model," *Journal of Econometrics*, 98 (1), 107–127. [780]

——— (2002), "Residual-Based Diagnostics for Conditional Heteroscedasticity Models," *Econometrics Journal*, 5 (2), 358–374. [780]

Tsui, A. K. (2004), "Diagnostics for Conditional Heteroscedasticity Models: Some Simulation Results," *Mathematics and Computation Simulation*, 64 (1), 113–119. First special issue: selected papers of the MSSANZ/IMACS 14th Biennial Conference on Modelling and Simulation (Canberra, 2001). [780]

Wooldridge, J. M. (1991), "On the Application of Robust, Regression-Based Diagnostics to Models of Conditional Means and Conditional Variances," *Journal of Econometrics*, 47 (1), 5–46. [780]