Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Statistical Planning and Inference 136 (2006) 2706–2718

journal of statistical planning and inference

www.elsevier.com/locate/jspi

The log of the determinant of the autocorrelation matrix for testing goodness of fit in time series[☆]

Daniel Peña[a], Julio Rodríguez[b,*]

[a]Department of Statistics, University Carlos III of Madrid, Calle Madrid 126, CP. 28903, Getafe, Madrid, Spain

[b]Laboratorio de Estadística, Universidad Politécnica de Madrid, Calle José Gutiérrez Abascal, 2, CP. 28006, Madrid, Spain

Received 4 February 2004; received in revised form 5 July 2004; accepted 17 October 2004
Available online 24 December 2004

Abstract

A finite sample modification of a test by Peña and Rodríguez is proposed. The new modified test is asymptotically equivalent but it has a more intuitive explanation and it can be 25% more powerful for small sample size than the previous one. The test statistic is the log of the determinant of the mth autocorrelation matrix. We propose two approximations by using the Gamma and the Normal distributions to the asymptotic distribution of the test statistic. It is shown that, depending on the model and sample size, the proposed test can be up to 50% more powerful than the Ljung and Box, Monti and Hong tests, and for finite sample size is always better than the previous Peña–Rodríguez test. This modified test is applied to the detection of several types of nonlinearity by using either the autocorrelation matrix of the squared or the absolute values of the residuals. It is shown that, in general, the new test is more powerful than the one by McLeod and Li.

Keywords: ARIMA model; Autocorrelation coefficients; Heteroscedasticity model; Nonlinearity test; Partial autocorrelation

☆ This research has been sponsored by Ministerio de Ciencia y Tecnología (Spain) under project BEC-2000-0167 and by the Fundación BBVA.

* Corresponding author. Tel.: +34 913363149; fax: +34 913363149.
E-mail address: puerta@etsii.upm.es (J. Rodríguez).

0378-3758/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2004.10.026

1. Introduction

Let X_t be a zero mean process generated by the ARMA(p, q) model $\phi(B) = \theta(B)\varepsilon_t$, where B is the backshift operator, $\phi(B)$ is a polynomial of order p, $\theta(B)$ is a polynomial of order q and ε_t is a white noise process. Let $\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_T$ be the residuals obtained after estimating the model in a sample of size T, and let

$$\hat{r}_j = \sum_{t=j+1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_{t-j} \Bigg/ \sum_{t=1}^{T} \hat{\varepsilon}_t^2 \quad \text{for } j = 1, 2, \ldots, \tag{1}$$

be the estimated residual autocorrelation coefficients. A family of portmanteau goodness of fit tests for the analysis of the independence of the residuals can be obtained by the expression

$$Q = T\left\{\delta \sum_{i=1}^{m} w_i g(\hat{r}_i^2) + (1-\delta)\sum_{i=1}^{m} \omega_i g(\widehat{\pi}_i^2)\right\}, \tag{2}$$

where $\widehat{\pi}_i$ are the estimated residual partial autocorrelation coefficients (see Box and Jenkins, 1976, pp. 64–65), $0 \leqslant \delta \leqslant 1$, $m < T$, $w_i \geqslant 0$, $\omega_i \geqslant 0$, and g is a nondecreasing smooth function with $g(0) = 0$. The partial correlation coefficient $\widehat{\pi}_i$ measures the correlation between $\hat{\varepsilon}_t$ and $\hat{\varepsilon}_{t+i}$ given the values $\hat{\varepsilon}_{t+1}, \ldots, \hat{\varepsilon}_{t+i-1}$. Some well-known members of this class when $g(x) = x$ are the tests proposed by Box and Pierce (1970), where $\delta = 1$ and $w_i = 1$, Ljung and Box (1978), where $\delta = 1$ and $w_i = (T+2)/(T-i)$, and Monti (1994), where $\delta = 0$ and $\omega_i = (T+2)/(T-i)$. This family includes also many tests obtained in the frequency domain by measuring the distance between the spectral density estimator and the one corresponding to a white noise process. The test proposed by Anderson (1993), is a member of this class where $\delta = 1$, $w_i = 1/(\pi i^2)$ and $m = T - 1$, and it was improved by Velilla (1994),who replaced the vector of autocorrelations with a vector of modified autocorrelation which is free of unknown parameters. Finally, Hong (1996) proposed a general class of these statistics where $\delta = 1$ and $w_i = k^2(j/m)$, k is a symmetric function, $k : \mathbb{R} \to [-1, 1]$, that is continuous at zero and at all but a finite number of points, with $k(0) = 1$ and $\int_{-\infty}^{\infty} k^2(z)\,\mathrm{d}z < \infty$. Hong shows that within a suitable class of kernel functions, the Daniell kernel, $(k(z) = \sin(\pi z)/\pi z,\ z \in (-\infty, \infty))$ maximizes the power of the test under both local and global alternatives.

Peña and Rodríguez (2002), based on a general measure of multivariate dependence, the Effective Dependence, see Peña and Rodríguez (2003), proposed a portmanteau test by applying this measure to the autocorrelation matrix, leading to the statistic

$$\hat{D}_m = T[1 - |\widehat{\mathbf{R}}_m|^{1/m}], \tag{3}$$

where $\widehat{\mathbf{R}}_m$ is

$$\widehat{\mathbf{R}}_m = \begin{bmatrix} 1 & \hat{r}_1 & \cdots & \hat{r}_m \\ \hat{r}_1 & 1 & \cdots & \hat{r}_{m-1} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{r}_m & \hat{r}_{m-1} & \cdots & 1 \end{bmatrix}. \tag{4}$$

and $\hat{r}_j$ is estimated by (1). They show that this test is more powerful than the ones proposed by Ljung and Box (1978) and Monti (1994). This test is highly asymmetric with respect to the autocorrelation coefficients and large weights are assigned to lower order lags and smaller weights to higher lags.

In this paper, we propose a modification of this test statistic that has the same asymptotic distribution but better performance for finite samples. Also the new test can be written as a particular case of Eq. (2), whereas $\hat{D}_m$ does not fit exactly into this formulation. The article is organized as follows. Section 2 presents the test as an approximation to the likelihood ratio test and shows that it is a member of the class of tests defined by (2). Section 3 obtains its asymptotic distribution and shows how to approximate it, first, by a gamma distribution and, second, by a normal distribution. Section 4 includes a Monte Carlo study of the properties of the test, both for linear models, by using the residual autocorrelations, and for nonlinear models, by using the autocorrelations of the squared or absolute values of the residuals. For linear models the test is shown to be more powerful than the ones proposed by Ljung and Box (1978), Monti (1994), Peña and Rodríguez (2002) and Hong (1996), whereas for detecting nonlinearity the test is shown to be also more powerful than those proposed by McLeod and Li (1983) and Peña and Rodríguez (2002). In both situations the test is also shown to be more robust to the value of m than other tests considered in the Monte Carlo study.

2. A pseudo-likelihood test

The estimated residuals can be considered as a sample of multivariate data from a distribution, $(\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_T) \sim \mathrm{N}_T(\mathbf{0}, \mathbf{V}_T)$. We are interested in testing whether or not the covariance matrix of the distribution of these residuals is diagonal. In multivariate analysis, see for instance Mardia et al. (1979, pp. 137–138), the likelihood ratio test for checking if a set of p random variables has a diagonal covariance matrix is given in terms of the correlation matrix, $\mathbf{R}_p$, by the statistic

$$-2\ \log \lambda = -T\ \log |\mathbf{R}_p|, \tag{5}$$

which has an asymptotic χ^2 distribution under H_0 with $p(p+1)/2$ degrees of freedom.

The problem when trying to generalize this statistic for goodness of fit in time series is that $p = T$ and thus the likelihood ratio test diverges. In order to find a test based on this criterion for time series we note that

$$|\mathbf{R}_T| = \prod_{i=1}^{T-1} (1 - R_i^2),$$

where $R_i^2 = \widehat{\mathbf{r}}_i' \mathbf{R}_{i-1}^{-1} \hat{\mathbf{r}}_i$ is the squared correlation coefficient in the linear fit, $\hat{\varepsilon}_t = \sum_{j=1}^{i} b_j \hat{\varepsilon}_{t-j} + u_t$. Assuming that the maximum number of nonzero regression coefficients in these regressions is m, and that $R_m^2 = R_{m+1}^2 = \cdots = R_{T-1}^2$, we can decompose the likelihood ratio statistic as

$$-T\ \log |\mathbf{R}_T| = -T\ \log |\mathbf{R}_m| - T(T-m-1)\log(1 - R_m^2).$$

In this equation, the first term converges to a random variable, as we show in Section 3, whereas the second term diverges with T. Thus it seems reasonable to explore a test based on the first component. This leads to the test statistic

$$D_m^* = -\frac{T}{m+1} \log |\hat{\mathbf{R}}_m|, \tag{6}$$

where we have standardized the estimated correlation matrix (4) by its dimension. This statistic can be considered as a modification of D_m as given by (3). We prefer to standardize by the dimension of the matrix instead of using the number of autocorrelation coefficients in order to obtain the following interpretations of this statistic: (i) $D_m^* = -T \log \overline{\delta}$, where $\overline{\delta} = \prod_{i=1}^{m+1} (\delta_i)^{1/(m+1)}$ is the geometric mean of the eigenvalues of $\hat{\mathbf{R}}_m$. (ii) As the eigenvalues of the covariance matrix are approximately equal to the power spectrum ordinates at the frequencies $\delta_i = 2\pi i/m$, (see Hannan, 1970), the statistic is also approximately $D_m^* \approx -T(m+1)^{-1} \sum \log f(\delta_i)$ where $f(\delta_i)$ is the spectral density. (iii) $D_m^* = -T \log(1 - \overline{R}^2)$, where $(1 - \overline{R}^2) = \prod_{i=1}^{m} (1 - R_i^2)^{1/(m+1)}$. (iv) This statistic is a member of the class (2), as shown in Ramsey (1974),

$$|\hat{\mathbf{R}}_m| = \prod_{i=1}^{m} (1 - \hat{\pi}_i^2)^{((m+1-i))},$$

and thus we have

$$D_m^* = -T \sum_{i=1}^{m} \frac{(m+1-i)}{(m+1)} \log(1 - \hat{\pi}_i^2), \tag{7}$$

which is in form (2) with $g(x) = -\log(1-x)$, $\delta = 0$ and $\omega_i = (m+1-i)/(m+1)$. Thus this statistic belongs to class (2), and it is proportional to a weighted average of the squared partial autocorrelation coefficients with larger weights given to low order coefficients and smaller weights to high-order coefficients.

3. Asymptotic distribution

In this section, we obtain the asymptotic distribution of the D_m^* statistic, and propose two approximations of this distribution. The asymptotic distribution is obtained by a straightforward extension of the one obtained in Peña and Rodríguez (2002). The first approximation is similar to the one presented in that paper, whereas the second one is new.

Theorem 1. *If the model is correctly identified, $\hat{D}_m^*$ is asymptotically distributed as $\sum_{i=1}^{m} \lambda_i \chi_{1,i}^2$, where $\chi_{1,i}^2$ $(i = 1, \ldots, m)$ are independent χ_1^2 random variables and λ_i $(i = 1, \ldots, m)$ are the eigenvalues of $(\mathbf{I}_m - \mathbf{Q}_m)\mathbf{W}_m$, where $\mathbf{Q}_m = \mathbf{X}_m \mathbf{V}^{-1} \mathbf{X}_m'$, $\mathbf{V}$ is the information matrix for the parameters ϕ and θ, $\mathbf{X}_m$ is an $m \times (p+q)$ matrix, with elements ϕ' and θ' defined by $1/\phi(B) = \sum_{i=0}^{\infty} \phi_i' B^i$ and $1/\theta(B) = \sum_{i=0}^{\infty} \theta_i' B^i$, and $\mathbf{W}_m$ is a diagonal matrix with elementsw$_i = (m-i+1)/(m+1)$, $(i = 1, \ldots, m)$.*

The proof of Theorem 1 is given in the Appendix.

For a general ARMA model the expression for the eigenvalues of $(\mathbf{I}_m - \mathbf{Q}_m)\mathbf{W}_m$ is complicated. Box and Pierce (1970) assumed that $m = \mathrm{O}(T^{1/2})$ when $T \to \infty$ and obtained that the matrix $\mathbf{Q}_m$ can be approximated by the projection matrix $\mathbf{Q}_m = \mathbf{X}_m(\mathbf{X}'_\mathbf{m}\mathbf{X}_\mathbf{m})^{-1}\mathbf{X}'_m$ when m is moderately high. Based on this result, we present here two approximations of the percentiles of the distribution $\sum_{i=1}^m \lambda_i \chi^2_{1,i}$. The first is the one proposed by Peña and Rodríguez (2002), where this distribution is approximated by a distribution of the form $a\chi^2_b$, a Gamma distribution with mean and variance equal to those of the exact distribution. This implies $a = \sum \lambda_i^2 / \sum \lambda_i$ and $b = (\sum \lambda_i)^2 / \sum \lambda_i^2$. Thus, we can approximate the distribution of D_m^* by a gamma distribution, $\mathscr{G}(\alpha = b/2, \beta = 1/2a)$ where the parameters are defined by

$$\alpha = \frac{3(m+1)\{m-2(p+q)\}^2}{2\{2m(2m+1)-12(m+1)(p+q)\}} \tag{8}$$

and

$$\beta = \frac{3(m+1)\{m-2(p+q)\}}{2m(2m+1)-12(m+1)(p+q)} \tag{9}$$

and the distribution has mean $\alpha/\beta = m/2-(p+q)$ and variance, $\alpha/\beta^2 = m(2m+1)/(3(m+1))-2(p+q)$. We denote by GD_m^* this first approximation which is distributed as a $\mathscr{G}(\alpha, \beta)$.

The second approximation is a generalization of the Wilson–Hilferty cube root transformation of a χ^2 random variable proposed by Chen and Deo (2004). They suggested a power transformation which reduces the skewness in order to improve the normal approximation. In our case this approximation is

$$\mathrm{ND}_m^* = (\alpha/\beta)^{-1/\lambda}(\lambda/\sqrt{\alpha})\left((D_m^*)^{1/\lambda} - (\alpha/\beta)^{1/\lambda}\left(1 - \frac{1}{2\alpha}\left(\frac{\lambda-1}{\lambda^2}\right)\right)\right) \tag{10}$$

and

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

where for m moderately large $\lambda \simeq 4$ and α and β are the values obtained in (8) and (9). The statistic ND_m^* is the second approximation which is distributed as a $\mathrm{N}(0, 1)$.

We have checked in a Monte Carlo experiment that both approximations to the asymptotic distribution improve by using the standardized autocorrelation coefficients $\tilde{r}_j^2 = \hat{r}_j^2(T+2)/(T-j)$. Thus we recommend their use instead of $\tilde{r}_j^2$, specially for small sample sizes.

4. Monte Carlo study

In this section we present a comparative study of the significance level and power of the tests based on the two approximations of the statistic $(\mathrm{ND}_m^*, \mathrm{GD}_m^*)$, and compare them to other tests. We first present the results for linear models and, second, for nonlinear models. The Matlab codes that implement the tests, as described in this article, are

Table 1
Significance levels of ND_m^*, GD_m^*, D_m and Q_{LB} under an AR(1) Model when $\alpha = 0.05$

T	ϕ	$m=10$				$m=20$				$m=30$			
		ND_m^*	GD_m^*	D_m	Q_{LB}	ND_m^*	GD_m^*	D_m	Q_{LB}	ND_m^*	GD_m^*	D_m	Q_{LB}
100	0.1	0.060	0.062	0.050	0.051	0.062	0.063	0.040	0.061	0.061	0.062	0.031	0.072
	0.5	0.060	0.062	0.052	0.055	0.060	0.062	0.040	0.064	0.061	0.062	0.029	0.075
	0.9	0.065	0.067	0.054	0.055	0.055	0.057	0.036	0.063	0.051	0.052	0.024	0.072
500	0.1	0.050	0.052	0.049	0.049	0.052	0.054	0.049	0.053	0.053	0.054	0.046	0.056
	0.5	0.053	0.055	0.052	0.050	0.054	0.055	0.049	0.055	0.053	0.054	0.046	0.057
	0.9	0.066	0.068	0.065	0.052	0.056	0.057	0.051	0.054	0.053	0.054	0.046	0.057

available for download at http://www.etsii.upm.es/ingor/estadistica/investigacion/pubpuerta.htm

4.1. Linear models

The significance level of the two approximations are compared with those of the previous test (D_m), and of the Ljung–Box statistic (Q_{LB}) for several values of the AR(1) parameters when the nominal level α is 0.05 (Table 1). In each case 50,000 Gaussian time series of sample sizes $T = 100$ and 500 were generated. Three values for m, 10, 20 and 30 were considered. The new tests for sample size 100 and 500 have reasonable size at the 5% level and do not seem to be affected by the value of m. The statistics D_m and Q_{LB} have a good size performance but they are clearly much more sensitive to the value of m. This robustness to the value of m is an advantage of this statistic compared to previous proposals.

Power of the tests are analyzed for the models proposed by Monti (1994). Table 2 presents the power study for 24 ARMA(2,2) models when AR(1) or MA(1) are fitted, for sample size $T = 100$. In each case 1000 series were generated and the power is computed for $m = 10$ and 20. We have included the tests proposed by Hong (1996) in this Monte Carlo study, with the Daniell kernel, H_n, and the test proposed by Monti (1994), Q_{MT}. The modified test D_m^* proposed in this article is almost always the most powerful (with the exception of the models 5, 14, 19 and 21 for $m = 20$, where the decrease in power with respect to Hong's statistic is as an average of only 0.6%). The differences in power of the test D_m and GD_m^* or ND_m^* are small: the average increase in power over the 24 models considered are 2.7% for $m = 10$ and 9.1%, for $m = 20$, respectively. However, the two implementations of the modified statistics are always more powerful than D_m and in some particular models (see models 1, 9, 11, 13, 23 for $m = 20$) the power increase is over 20%. The increase in power of the modified statistics with respect to Q_{LB}, Q_{MT} and H_n tests are as an average of 34.2%, 23.6% and 16.1% for $m = 10$ and 41.2%, 23.6% and 13.7% for $m = 20$.

Table 3 presents the results for $m = 5$ and 10 of the same tests for small sample size ($n = 30$). The proposed tests are always the most powerful in all models. The increase in power with respect to D_m, Q_{LB}, Q_{MT} and H_n tests is in average 26.1%, 82.6%, 48.5% and 414.7% for $m = 5$ and 26.1%, 78.9%, 73.9% and 133.8% for $m = 10$. We conclude that

Table 2
Powers of the tests based on ND_m^*, GD_m^*, D_m, Q_{LB}, Q_{MT} and H_n when the data are generated using ARMA(2,2) models and AR(1) and MA(1) models are fitted, $T=100$ and $\alpha=0.05$

M	ϕ_1	ϕ_2	θ_1	θ_2	$m=10$						$m=20$					
					ND_m^*	GD_m^*	D_m	Q_{LB}	Q_{MT}	H_n	ND_m^*	GD_m^*	D_m	Q_{LB}	Q_{MT}	H_n
(a) Fitted by AR(1) model																
1	—	—	−0.5	—	0.474	0.477	0.445	0.275	0.296	0.337	0.381	0.386	0.306	0.249	0.197	0.292
2	—	—	−0.8	—	0.985	0.985	0.985	0.772	0.956	0.930	0.964	0.965	0.953	0.582	0.839	0.835
3	—	—	−0.6	0.30	0.996	0.996	0.996	0.782	0.992	0.947	0.997	0.997	0.994	0.652	0.939	0.870
4	0.1	0.30	—	—	0.624	0.629	0.598	0.424	0.408	0.562	0.508	0.509	0.440	0.361	0.288	0.485
5	1.3	−0.35	—	—	0.795	0.796	0.776	0.599	0.545	0.770	0.647	0.650	0.589	0.505	0.390	0.656
6	0.7	—	−0.4	—	0.792	0.797	0.771	0.536	0.569	0.677	0.676	0.683	0.612	0.422	0.421	0.592
7	0.7	—	−0.9	—	0.999	0.999	0.999	0.987	0.999	0.999	0.999	0.999	0.999	0.910	0.988	0.995
8	0.4	—	−0.6	0.30	1.000	1.000	1.000	0.831	0.998	0.984	1.000	1.000	0.997	0.688	0.970	0.912
9	0.7	—	0.7	−0.15	0.260	0.264	0.239	0.200	0.186	0.181	0.220	0.226	0.173	0.205	0.139	0.199
10	0.7	0.2	0.5	—	0.872	0.878	0.860	0.738	0.747	0.817	0.799	0.804	0.759	0.601	0.580	0.776
11	0.7	0.2	−0.5	—	0.637	0.641	0.610	0.306	0.368	0.476	0.484	0.489	0.402	0.279	0.242	0.354
12	0.9	−0.4	1.2	−0.30	0.986	0.987	0.984	0.725	0.973	0.863	0.978	0.978	0.969	0.565	0.902	0.804
(b) Fitted by MA(1) model																
13	0.5	—	—	—	0.414	0.416	0.386	0.283	0.259	0.345	0.324	0.328	0.266	0.249	0.189	0.295
14	0.8	—	—	—	0.994	0.994	0.992	0.980	0.975	0.992	0.984	0.984	0.976	0.963	0.940	0.986
15	1.1	−0.35	—	—	0.999	0.999	0.999	0.997	0.996	0.999	0.998	0.999	0.996	0.988	0.987	0.998
16	—	—	0.8	−0.5	0.970	0.970	0.968	0.823	0.928	0.928	0.945	0.947	0.928	0.704	0.810	0.882
17	—	—	−0.6	0.3	0.695	0.700	0.674	0.395	0.466	0.546	0.581	0.585	0.505	0.332	0.344	0.429
18	0.5	—	−0.7	—	0.966	0.968	0.954	0.880	0.867	0.957	0.924	0.924	0.889	0.798	0.720	0.921
19	−0.5	—	0.7	—	0.975	0.978	0.972	0.912	0.888	0.976	0.939	0.940	0.907	0.832	0.767	0.946
20	0.3	—	0.8	−0.5	0.888	0.891	0.877	0.636	0.757	0.766	0.805	0.808	0.750	0.511	0.592	0.691
21	0.8	—	−0.5	0.3	0.992	0.992	0.991	0.983	0.975	0.992	0.982	0.982	0.976	0.963	0.920	0.988
22	1.2	−0.50	0.9	—	0.743	0.746	0.720	0.474	0.703	0.411	0.748	0.750	0.693	0.370	0.578	0.478
23	0.3	−0.20	−0.7	—	0.454	0.462	0.426	0.285	0.295	0.342	0.365	0.369	0.295	0.247	0.196	0.304
24	0.9	−0.40	1.2	−0.3	0.970	0.973	0.967	0.794	0.934	0.891	0.950	0.952	0.933	0.621	0.819	0.852

Table 3
Powers of the tests based on ND_m^*, GD_m^*, D_m, Q_{LB}, Q_{MT} and H_n when the data are generated using ARMA(2,2) models and AR(1) and MA(1) models are fitted, $T = 30$ and $\alpha = 0.05$

M	ϕ_1	ϕ_2	θ_1	θ_2	$m = 5$						$m = 10$					
					ND_m^*	GD_m^*	D_m	Q_{LB}	Q_{MT}	H_n	ND_m^*	GD_m^*	D_m	Q_{LB}	Q_{MT}	H_n
(a) Fitted by AR(1) model																
1	—	—	−0.5	—	0.252	0.256	0.226	0.118	0.164	0.052	0.224	0.226	0.172	0.118	0.116	0.082
2	—	—	−0.8	—	0.618	0.621	0.588	0.306	0.455	0.219	0.567	0.569	0.494	0.269	0.328	0.253
3	—	—	−0.6	0.30	0.650	0.653	0.611	0.305	0.488	0.200	0.612	0.614	0.553	0.238	0.370	0.244
4	0.1	0.30	—	—	0.183	0.185	0.159	0.093	0.114	0.040	0.158	0.161	0.113	0.103	0.080	0.054
5	1.3	−0.35	—	—	0.382	0.386	0.345	0.212	0.193	0.254	0.283	0.288	0.222	0.194	0.140	0.203
6	0.7	—	−0.4	—	0.407	0.410	0.363	0.220	0.216	0.173	0.324	0.329	0.238	0.186	0.154	0.169
7	0.7	—	−0.9	—	0.879	0.881	0.857	0.478	0.726	0.588	0.799	0.801	0.748	0.419	0.569	0.477
8	0.4	—	−0.6	0.30	0.721	0.721	0.696	0.331	0.558	0.266	0.676	0.678	0.608	0.280	0.425	0.286
9	0.7	—	0.7	−0.15	0.088	0.088	0.078	0.060	0.065	0.019	0.096	0.098	0.065	0.065	0.061	0.031
10	0.7	0.20	0.5	—	0.220	0.224	0.193	0.125	0.125	0.080	0.178	0.181	0.120	0.105	0.093	0.091
11	0.7	0.20	−0.5	—	0.287	0.290	0.260	0.132	0.145	0.118	0.206	0.209	0.152	0.113	0.088	0.102
12	0.9	−0.40	1.2	−0.30	0.465	0.466	0.420	0.224	0.404	0.086	0.504	0.508	0.451	0.215	0.316	0.171
(b) Fitted by MA(1) model																
13	0.5	—	—	—	0.134	0.135	0.115	0.085	0.097	0.042	0.143	0.145	0.102	0.103	0.096	0.055
14	0.8	—	—	—	0.631	0.636	0.603	0.493	0.476	0.503	0.555	0.559	0.483	0.443	0.364	0.488
15	1.1	−0.35	—	—	0.812	0.813	0.790	0.659	0.679	0.692	0.751	0.751	0.685	0.591	0.531	0.664
16	—	—	0.8	−0.50	0.505	0.508	0.457	0.279	0.339	0.265	0.456	0.461	0.370	0.275	0.288	0.275
17	—	—	−0.6	0.30	0.377	0.381	0.344	0.130	0.215	0.078	0.312	0.318	0.240	0.132	0.164	0.086
18	0.5	—	−0.7	—	0.582	0.585	0.545	0.387	0.368	0.447	0.484	0.491	0.399	0.349	0.272	0.386
19	−0.5	—	0.7	—	0.735	0.735	0.709	0.572	0.524	0.642	0.661	0.664	0.588	0.474	0.400	0.592
20	0.3	—	0.8	−0.50	0.307	0.313	0.273	0.161	0.206	0.091	0.295	0.299	0.221	0.165	0.172	0.117
21	0.8	—	−0.5	0.30	0.722	0.723	0.691	0.582	0.520	0.641	0.618	0.620	0.552	0.513	0.398	0.589
22	1.2	−0.50	0.9	—	0.166	0.170	0.140	0.135	0.181	0.003	0.237	0.242	0.193	0.129	0.158	0.041
23	0.3	−0.20	−0.7	—	0.292	0.296	0.259	0.166	0.176	0.088	0.252	0.253	0.184	0.155	0.150	0.112
24	0.9	−0.40	1.2	−0.30	0.468	0.471	0.427	0.198	0.353	0.095	0.450	0.454	0.390	0.192	0.272	0.127

the performance of both approximations, GD_m^* or ND_m^*, for checking for goodness of fit in linear models is similar, and that the new test seems to be more powerful than the previous one.

4.2. Checking the linearity assumption

McLeod and Li (1983) proposed detecting nonlinearity in time series data by replacing the auto-correlation coefficients in the statistic Q_{LB} by the autocorrelation coefficients of the squared residuals. The asymptotic distribution of this statistic does not depend on the order of the ARMA model fitted to the data. This fact makes it different from the Ljung–Box statistic. Peña and Rodríguez (2002) used a similar idea in generalizing D_m for testing for nonlinearity and Pérez and Ruiz (2003) compared the size and power of the Q_{LB} and D_m tests when they are applied by using squared residuals, the absolute values of residuals and the log of one minus the squared residuals. They obtained that the test based on the absolute values of the residuals is more powerful in finding nonlinearity in heteroscedastic models (as GARCH and Stochastic Volatility models). An additional advantage when using the absolute values of the residuals is that in order to obtain the asymptotic distribution we need to assume only that the fourth order moments of $|\hat{\varepsilon}_t|$ exist, whereas for the squared residuals we required the existence of the eighth order moments of $\hat{\varepsilon}_t^2$, see Hannan (1970). The test applied to the absolute value of the residuals is

$$\hat{D}_m^*(|\hat{\varepsilon}_t|) = -\frac{T}{m+1} \log |\tilde{\mathbf{R}}_m(|\hat{\varepsilon}_t|)|,$$

where $\tilde{\mathbf{R}}_m(|\hat{\varepsilon}_t|)$ is the autocorrelation matrix (4) which is now built using the standardized autocorrelation coefficients $\tilde{r}_k(|\hat{\varepsilon}_t|)$, given by

$$r_k(|\hat{\varepsilon}_t|) = \frac{T+2}{T-k} \frac{\sum_{t=k+1}^{n} (|\hat{\varepsilon}_t| - |\hat{\sigma}|)(|\hat{\varepsilon}_{t-k}| - |\hat{\sigma}|)}{\sum_{t=1}^{n} (|\hat{\varepsilon}_t| - |\hat{\sigma}|)^2} \quad (k = 1, 2, \ldots, m),$$

where $|\hat{\sigma}| = \sum |\hat{\varepsilon}_t|/T$. The asymptotic distribution of $\hat{D}_m^*(|\hat{\varepsilon}_t|)$ is similar to the one shown in Theorem 1, the weights are $\lambda_i = w_i$, and the asymptotic distribution does not depend on the population parameters. We can approximate this distribution, in a similar way to the $\hat{D}_m^*$, by either a Gamma or a Normal distribution where now the degrees of freedom do not depend on the order of the ARMA model fitted and therefore both p and q are equal to zero in expressions (8), (9) and (10).

In the next Monte Carlo study, we compare the power of the statistics D_m^*, D_m, Q_{LB} and Q_{MT} for testing for linearity for the six nonlinear models indicated in Table 4. The first four models were analyzed by Keenan (1985), whereas the last two models have changing conditional variance. In M5 the parameters are taken from real financial time series (see Carnero et al., 2001), whereas in M6 they correspond to environmental data (see Tol, 1996). The e_t's in the six models are independent N(0, 1).

Table 5 summarizes the power results. For each model 1000 replications of sample size $T = 204$ were generated. An AR(p) model was fitted to the data, where p was selected by the AIC criterion (Akaike, 1974) with $p \in \{1, 2, 3, 4\}$. The power of the proposed test D_m^* when using both approximations, GD_m^* and ND_m^*, is slightly better than the D_m test and

Table 4
Six nonlinear models: the first four models were proposed by Keenan and the last two are models with changing conditional variance

M1:	$Y_t = e_t - 0\cdot 4e_{t-1} + 0\cdot 3e_{t-2} + 0\cdot 5e_t e_{t-2}$@
M2:	$Y_t = e_t - 0\cdot 3e_{t-1} + 0\cdot 2e_{t-2} + 0\cdot 4e_{t-1}e_{t-2} - 0\cdot 25e^2_{t-2}$
M3:	$Y_t = 0\cdot 4Y_{t-1} - 0\cdot 3Y_{t-2} + 0\cdot 5Y_{t-1}e_{t-1} + e_t$
M4:	$Y_t = 0\cdot 4Y_{t-1} - 0\cdot 3Y_{t-2} + 0\cdot 5Y_{t-1}e_{t-1} + 0\cdot 8e_{t-1} + e_t$
M5:	$y_t = e_t\sigma_t, \quad \sigma_t^2 = 1\cdot 21 + 0\cdot 404y^2_{t-1} + 0\cdot 153\sigma^2_{t-1}$
M6:	$y_t = 0\cdot 025e_t\sigma_t, \quad \log\sigma_t^2 = 0\cdot 9\log\sigma^2_{t-1} + \eta_t, \quad \eta_t \sim N(0, 0\cdot 363)$

Table 5
Powers of the tests based on ND^*_m, GD^*_m, D_m, Q_{LB} and Q_{MT} when the data are generated by four nonlinear models, and the fitted model is an AR(p) and $\alpha = 0.05$

	$m=7$					$m=12$					$m=24$				
	ND^*_m	GD^*_m	D_m	Q_{LB}	Q_{MT}	ND^*_m	GD^*_m	D_m	Q_{LB}	Q_{MT}	ND^*_m	GD^*_m	D_m	Q_{LB}	Q_{MT}
*M*1															
ε^2	0.143	0.145	0.140	0.124	0.128	0.142	0.143	0.134	0.099	0.101	0.119	0.120	0.100	0.084	0.066
$\|\varepsilon\|$	0.076	0.078	0.071	0.068	0.057	0.074	0.076	0.068	0.067	0.058	0.066	0.066	0.051	0.077	0.059
*M*2															
ε^2	0.615	0.616	0.611	0.517	0.506	0.567	0.572	0.543	0.416	0.419	0.469	0.472	0.425	0.315	0.276
$\|\varepsilon\|$	0.451	0.456	0.442	0.350	0.340	0.414	0.416	0.391	0.318	0.292	0.333	0.334	0.296	0.252	0.214
*M*3															
ε^2	0.965	0.966	0.963	0.933	0.921	0.952	0.953	0.949	0.879	0.872	0.906	0.908	0.885	0.775	0.727
$\|\varepsilon\|$	0.938	0.939	0.936	0.896	0.883	0.919	0.920	0.907	0.820	0.797	0.857	0.858	0.836	0.723	0.664
*M*4															
ε^2	0.930	0.932	0.928	0.867	0.852	0.888	0.889	0.880	0.796	0.761	0.812	0.814	0.794	0.677	0.612
$\|\varepsilon\|$	0.963	0.964	0.963	0.928	0.914	0.943	0.946	0.940	0.898	0.865	0.900	0.903	0.888	0.831	0.768
*M*5															
ε^2	0.864	0.865	0.860	0.792	0.781	0.823	0.825	0.816	0.731	0.708	0.748	0.750	0.731	0.638	0.574
$\|\varepsilon\|$	0.856	0.858	0.853	0.798	0.780	0.834	0.836	0.822	0.745	0.713	0.770	0.772	0.743	0.666	0.587
*M*6															
ε^2	0.781	0.784	0.777	0.775	0.738	0.766	0.766	0.756	0.730	0.693	0.712	0.715	0.694	0.625	0.573
$\|\varepsilon\|$	0.966	0.966	0.963	0.959	0.940	0.955	0.956	0.953	0.945	0.908	0.927	0.929	0.916	0.915	0.826

the advantage increases with m. The limited experience presented in this study shows that the most convenient transformation for the residuals depends on the model: the squared residuals lead to more powerful tests for M1, M2 and M3, whereas the absolute values of the residuals leads to more powerful tests for M4, M5 and M6.

Table 6
Average decrease in power of the tests compared when m is increased from $m = 10$ to 20 in the models of Table 2

	ND^*_m	GD^*_m	D_m	Q_{LB}	Q_{MT}	H_n
Models 1–12	10.2	10.0	15.1	14.5	19.0	9.0
Models 13–24	6.8	6.8	10.8	11.8	16.0	4.8

Table 7
Average decrease in power of the tests compared when m is increased

m		ND^*_m	GD^*_m	D_m	Q_{LB}	Q_{MT}
7–12	ε^2	4.297	4.005	4.884	9.404	9.728
7–12	$\|\varepsilon\|$	3.396	3.252	4.147	5.792	8.242
12–24	ε^2	9.858	9.278	10.882	15.633	21.316
12–24	$\|\varepsilon\|$	8.873	8.295	9.652	10.765	16.263

In the first two rows m is increased from $m = 7$ to 12, and in the last two rows from $m = 12$ to 24. The models considered are the ones in Table 4.

4.3. The effect of m

In the derivation of the asymptotic results about the distribution of the test statistic it is assumed that m is of order $T^{1/2}$ and this is the usual value for m recommended in applications. If the value for m is chosen too large the power of the test of the form (2) is expected to decrease. See Battaglia (1990) for a study of the effect of m on the Ljung–Box–Pierce statistics. Table 6 shows the average decrease in power when going from $m = 10$ to 20 for the models in Table 2. It can be seen that all the tests suffered a decrease in power but the two tests which are more robust to the value of m are the one by Hong and the test proposed in this article. The small sensitivity of Hong's test appears because in this test larger autocorrelations have a very small weight and thus the test statistic changes very little with the increase of m. However, this structure makes this test less powerful when the information is not concentrated in the low lags (see for instance models 11, 17 and 22 in Table 2). The sensitivity to the value of m of the proposed test with respect to D_m, Q_{LB}, and Q_{MT} is much smaller.

Table 7 shows that the average decrease in power of the proposed test in the nonlinear case is also smaller than for the other test compared. Thus we conclude that the proposed test is less sensitive to the value of m than its competitors.

Acknowledgements

The authors are very grateful to Fernando Tusell for helpful comments.

Appendix

Proof of Theorem 1. Suppose that under the null hypothesis, $\hat{D}_m^*$ is asymptotically distributed as the random variable X. To find the distribution of (7), suppose that $(T\hat{\pi}_1^2, T\hat{\pi}_2^2, \dots, T\hat{\pi}_m^2)$ is asymptotically distributed as Y. Then, applying the multivariate δ-method (v.g. in Arnold, 1990) to $g(\hat{\pi}_1^2, \dots, \hat{\pi}_m^2) = -\sum_{i=1}^m \left(\frac{m-i+1}{m+1}\right) \ln(1-\hat{\pi}_i^2)$, it follows that

$$-T\sum_{i=1}^{m}\left(\frac{m-i+1}{m+1}\right)\ln(1-\hat{\pi}_i^2) \to \left(\frac{m}{m+1}, \frac{m-1}{m+1}, \dots, \frac{1}{m+1}\right) Y, \tag{A.1}$$

where $\to$ stands for convergence in distribution. From the Cramer–Wold theorem (v.g. in Arnold, 1990), it follows that

$$\begin{aligned}&\left(\frac{m}{m+1}, \frac{m-1}{m+1}, \dots, \frac{1}{m+1}\right)(T\hat{\pi}_1^2, T\hat{\pi}_2^2, \dots, T\hat{\pi}_m^2)' \\ &\quad \to \left(\frac{m}{m+1}, \frac{m-1}{m+1}, \dots, \frac{1}{m+1}\right) Y.\end{aligned} \tag{A.2}$$

Using the fact that $T^{1/2}\hat{\pi}_{(m)}$ is asymptotically distributed as $\mathrm{N}(\mathbf{0}, \mathbf{I}_m - \mathbf{Q}_m)$, see Monti (1994), and from the theorem on quadratic forms given by Box (1954), it follows that

$$\begin{aligned}&\left(\frac{m}{m+1}, \frac{m-1}{m+1}, \dots, \frac{1}{m+1}\right)(T\hat{\pi}_1^2, T\hat{\pi}_2^2, \dots, T\hat{\pi}_m^2)' \\ &\quad = T\hat{\pi}_{(m)}'\mathbf{W}\hat{\pi}_{(m)} \to \sum_{i=1}^{m} \lambda_i \chi_{1,i}^2.\end{aligned} \tag{A.3}$$

Finally, from (A.2) and (A.3), $(\frac{m}{m+1}, \frac{m-1}{m+1}, \dots, \frac{1}{m+1})Y \to \sum_{i=1}^m \lambda_i \chi_{1,i}^2$, and from (A.1), $\hat{D}_m^* \to \sum_{i=1}^m \lambda_i \chi_{1,i}^2$. □

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Automat. Control AC 19, 203–217.

Anderson, T.W., 1993. Goodness of fit tests for spectral distributions. Ann. Statist. 21, 830–847.

Arnold, S.F., 1990. Mathematical Statistics. Prentice-Hall International, Inc., Englewood cliffs, NJ.

Battaglia, F., 1990. Approximate power of Portmanteau tests for time series. Statist. Probab. Lett. 9, 337–341.

Box, G.E.P., 1954. Some theorems on quadratic forms applied in the study of analysis of variance problems I: effect of the inequality of variance in the one-way classification. Ann. Math. Statist. 25, 290–302.

Box, G.E.P., Jenkins, G.M., 1976. Time Series Analysis, Forecasting and Control. Holden-Day, San Francisco.

Box, G.E.P., Pierce, D.A., 1970. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. J. Amer. Statist. Assoc. 65, 1509–1526.

Carnero, M.A., Peña, D., Ruiz, E., 2001. Outliers and conditional autoregressive heteroscedasticity in time series. Estadística 53, 143–213.

Chen, W.W., Deo, R.S., 2004. Power transformations to induce normality and their applications. J. Roy. Statist. Soc. Ser. B 66, 117–130.

Hannan, E., 1970. Multiple Time Series. Wiley, New York.

Hong, Y., 1996. Consistent testing for serial correlation of unknown form. Econometrica 64, 837–864.

Keenan, D.M., 1985. A Tukey nonadditivity-type test for time series nonlinearity. Biometrika 72, 39–44.

Ljung, G.M., Box, G.E.P., 1978. On a measure of lack of fit in time series models. Biometrika 65, 297–303.

Mardia, K.V., Kent, J.T., Bibby, J.M., 1979. Multivariate Analysis. Academic Press, New York.

McLeod, A.I., Li, W.K., 1983. Diagnostic checking ARMA time series models using squared-residual autocorrelations. J. Time Ser. Anal. 4, 269–273.

Monti, A.C., 1994. A proposal for residual autocorrelation test in linear models. Biometrika 81, 776–780.

Peña, D., Rodríguez, J., 2002. A powerful Portmanteau test of lack of fit for time series. J. Amer. Statist. Assoc. 97, 601–610.

Peña, D., Rodríguez, J., 2003. Descriptive measures of multivariate scatter and linear dependence. J. Multivariate Anal. 85, 361–374.

Pérez, A., Ruiz, E., 2003. Properties of the sample autocorrelations of non-linear transformations in long memory stochastic volatility models. J. Fin. Econ. 1, 420–444.

Ramsey, F.L., 1974. Characterization of the partial autocorrelation function. Ann. Statist. 6, 1296–1301.

Tol, R.S.J., 1996. Autoregressive conditional heteroscedasticity in daily temperatures measurements. Environmetrics 7, 67–75.

Velilla, S., 1994. A goodness-of-fit test for autoregressive moving-average models based on the standardized sample spectral distribution of the residuals. J. Time Ser. Anal. 15, 637–647.