A Powerful Portmanteau Test of Lack of Fit for Time Series

Daniel PEÑA and Julio RODRÍGUEZ

A new portmanteau test for time series, more powerful than the tests of Ljung and Box and Monti, is proposed. The test is based on the mth root of the determinant of the mth autocorrelation matrix. It is shown that the proposed statistic is a function of all of the squared multiple correlation coefficients of the regressions of the residuals on their lags when the number of lags goes from 1 to m. It can also be written as a function of the first m partial autocorrelation coefficients. The asymptotic distribution of the test statistic is a linear combination of chi-squared distributions and can be approximated by a gamma distribution. It is shown, depending on the model and sample size, that this test can be up to 50% more powerful than the Ljung and Box and Monti tests. The test is applied to the detection of several types of nonlinearity by using the autocorrelation matrix of the squared residuals, and it is shown that, in general, the new test is more powerful than the test of McLeod and Li. An example is presented in which this test finds nonlinearity in the residuals of the sunspot series.

KEY WORDS: ARIMA model; Autocorrelation coefficient; Autocorrelation matrix; Goodness of fit; Heteroscedasticity model; Nonlinearity test; Partial autocorrelation.

1. INTRODUCTION

Suppose that a time series $\{X_t\}$ is generated by a stationary and invertible ARMA(p, q) process of the form $\phi(B)X_t = \theta(B)\varepsilon_t$, where $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ and $\phi(B)$ and $\theta(B)$ are polynomials given by $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ and $\theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$, where $B^k X_t = X_{t-k}$. Usually X_t is some transformation of an observed time series such as differencing. Defining $\hat{\theta}(B)$ and $\hat{\phi}(B)$ as the estimated polynomials where the coefficients ϕ_i and θ_j are replaced by the maximum likelihood estimators, $\hat{\phi}_i$ and $\hat{\theta}_j$, the residuals of this model are given by $\hat{\varepsilon}_t = \hat{\theta}^{-1}(B)\hat{\phi}(B)X_t$. Several diagnostic goodness-of-fit tests have been proposed based on the residual autocorrelation coefficients given by $r_k = \sum_{t=k+1}^{n} \hat{\varepsilon}_t \hat{\varepsilon}_{t-k} / \sum_{t=1}^{n} \hat{\varepsilon}_t^2$ for $k = 1, 2, \ldots$.

Box and Pierce (1970) introduced a portmanteau test to check the adequacy of the fitted model, using the statistic

$$Q = n \sum_{k=1}^{m} r_k^2, \tag{1}$$

and they showed that the asymptotic distribution of Q can be approximated by a χ^2 distribution with $m-(p+q)$ degrees of freedom. Ljung and Box (1978) improved this approximation by replacing the autocorrelation coefficients r_k in (1) with their standardized values

$$\tilde{r}_k^2 = \frac{(n+2)}{(n-k)} r_k^2, \tag{2}$$

leading to the statistic

$$Q_{\mathrm{LB}} = n(n+2) \sum_{k=1}^{m} (n-k)^{-1} r_k^2. \tag{3}$$

Ljung (1986) showed that when m is small, the estimated size of Q_{LB} can be improved by using the scaled χ^2 distribution and that computing Q_{LB} with too many residual autocorrelations can reduce the power of the test.

Monti (1994) proposed a portmanteau test of goodness-of-fit based on partial autocorrelations. Let $\hat{\pi}_k$ be the kth residual partial autocorrelation. If the model is correctly specified, $\hat{\pi}_k$ is approximately distributed as a normal random variable with mean zero and variance $(n-k)/(n(n+2))$. Thus, a portmanteau test statistic, similar to the Ljung–Box statistic, can be defined by

$$Q_{\mathrm{MT}} = n(n+2) \sum_{k=1}^{m} (n-k)^{-1} \hat{\pi}_k^2. \tag{4}$$

Under the assumption that the time series has been generated by an ARMA(p, q), the asymptotic distribution of Q_{MT} is χ^2 with $m-(p+q)$ degrees of freedom. Monti (1994) showed by simulation that when the fitted model underestimates the order of the moving average component, Q_{MT} is more powerful than Q_{LB}. Kwan and Wu (1997) examined via Monte Carlo simulation the finite-sample properties of (3) and (4) for data generated with monthly seasonality, finding only small differences between the powers of Q_{MT} and Q_{LB}.

This article proposes a new portmanteau goodness-of-fit test based on a general measure of multivariate dependence and is organized as follows. Section 2 presents the test and its main properties. Section 3 obtains its asymptotic distribution and shows that it can be approximated by a gamma distribution. Section 4 includes a Monte Carlo study of the properties of the test and shows that it is more powerful than the tests proposed by Ljung and Box (1978) and Monti (1994). Section 5 extends the test to check nonlinearity by using the autocorrelations of the squared residuals. It is shown that the proposed test is more powerful than that proposed by McLeod and Li (1983). Section 6 discusses some advantages and limitations of the proposed test.

Daniel Peña and Julio Rodríguez are with the Department of Statistics and Econometrics, University Carlos III of Madrid, Calle Madrid 126, CP 28903, Getafe, Madrid, Spain (E-mail: *dpena@est-econ.uc3m.es* or *puerta@est-econ.uc3m.es*). This research has been sponsored by Ministerio de Ciecia y Techología (Spain) under project BEC-2000-0167 and by the Cátedra Banco BVA de Calidad. The authors are very grateful to Andrés Alonso for many stimulating discussions, to Mike Wiper and Victor Yohai for their help with the final draft, and to an associate editor and a referee for very helpful comments.

Journal of the American Statistical Association
June 2002, Vol. 97, No. 458, Theory and Methods

2. THE PROPOSED TEST

The estimated residuals can be considered a sample of multivariate data from some distribution. We are interested in testing whether or not the covariance (correlation) matrix of their distribution is proportional (equal) to the identity. In multivariate analysis the likelihood ratio test for checking if a set of normal random variables has a scalar covariance matrix is proportional to the determinant of the correlation matrix of the multivariate variables. Thus, it is sensible to explore a test based on this statistic.

For stationary time series data the residual correlation matrix of order m, $\widehat{\mathbf{R}}_m$, is given by

$$\widehat{\mathbf{R}}_m = \begin{bmatrix} 1 & r_1 & \cdots & r_m \\ r_1 & 1 & \cdots & r_{m-1} \\ \vdots & \vdots & \ddots & \vdots \\ r_m & r_{m-1} & \cdots & 1 \end{bmatrix}. \tag{5}$$

We propose to test for autocorrelation in the estimated residuals by using a transformation of $|\widehat{\mathbf{R}}_m|$ that has a simpler distribution under the null hypothesis. The proposed portmanteau statistic is

$$\widehat{D}_m = n[1 - |\widehat{\mathbf{R}}_m|^{1/m}]. \tag{6}$$

This statistic has two interesting interpretations. The first can be obtained by using a recursive expression for the determinant of the matrix (5). If we define $\hat{\mathbf{r}}_{(m)} = (r_1, \ldots, r_m)'$, we can write

$$\widehat{\mathbf{R}}_m = \begin{bmatrix} 1 & \hat{\mathbf{r}}'_{(m)} \\ \hat{\mathbf{r}}_{(m)} & \widehat{\mathbf{R}}_{m-1} \end{bmatrix},$$

and, using the properties of the determinant of a partitioned matrix, we have

$$|\widehat{\mathbf{R}}_m| = |\widehat{\mathbf{R}}_{m-1}|(1 - R_m^2),$$

where $R_m^2 = \hat{\mathbf{r}}'_{(m)}\widehat{\mathbf{R}}_{m-1}^{-1}\hat{\mathbf{r}}_{(m)}$ is the square of the multiple correlation coefficient in the linear fit $\hat{\varepsilon}_t = \sum_{j=1}^{m} b_j \hat{\varepsilon}_{t-j} + u_t$. By recursive use of this expression, we have

$$|\widehat{\mathbf{R}}_m|^{1/m} = \left[\prod_{i=1}^{m}(1 - R_i^2)\right]^{1/m}. \tag{7}$$

Note that $1 - R_i^2$ is a measure of dependence and $|\widehat{\mathbf{R}}_m|^{1/m}$ is the geometric average of these dependence measures. Thus, $1 - |\widehat{\mathbf{R}}_m|^{1/m}$ can be interpreted as an average squared correlation coefficient, obtained when autoregressive models of increasing order are fitted to the residuals of the estimated time series.

The second interpretation is based on the partial autocorrelation coefficients. Note that $1 - R_i^2 = RSS(1, i)/TSS$, where $RSS(1, i)$ is the residual or unexplained sum of squares in the analysis of variance (ANOVA) decomposition of the residual regression $\hat{\varepsilon}_t = \sum_{j=1}^{i} b_j \hat{\varepsilon}_{t-j} + u_t$, and $TSS = \sum \hat{\varepsilon}_t^2$ is the total sum of squares. In the same way, $1 - R_{i-1}^2 = RSS(1, i-1)/TSS$ and

$$\frac{1 - R_i^2}{1 - R_{i-1}^2} = \frac{RSS(1, i)}{RSS(1, i-1)} = (1 - \hat{\pi}_i^2), \tag{8}$$

where $\hat{\pi}_i^2 = [RSS(1, i-1) - RSS(1, i)]/RSS(1, i-1)$ is the ith squared partial autocorrelation coefficient. Thus, using (8) and (7), we obtain

$$|\widehat{\mathbf{R}}_m|^{1/m} = \prod_{i=1}^{m}(1 - \hat{\pi}_i^2)^{(m+1-i)/m}. \tag{9}$$

This expression shows that $|\widehat{\mathbf{R}}_m|^{1/m}$ is a weighted function of the first m partial autocorrelation coefficients of the residuals.

3. DISTRIBUTION OF THE PROPOSED STATISTIC

In this section we derive the asymptotic distribution of the proposed statistic (6) for all m, where m is the number of sample autocorrelations. As this distribution is complicated, we follow Box and Pierce (1970) in obtaining an approximation to this distribution when m is moderately high. Then, we will show by Monte Carlo simulation that this approximation works well in small samples.

3.1 Asymptotic Distribution

The statistic $\widehat{D}_m$ is a continuous function of the sample partial autocorrelations, $\hat{\pi}_i$, as shown in Equation (9). Defining $\hat{\pi}_{(m)} = (\hat{\pi}_1, \ldots, \hat{\pi}_m)'$, and using a result of Monti (1994), we have that $n^{1/2}\hat{\pi}_{(m)}$ is asymptotically multivariate normal with zero mean vector and covariance matrix $(\mathbf{I}_m - \mathbf{Q}_m)$, 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, and $\mathbf{X}_m$ 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 $1/\theta(B) = \sum_{i=0}^{\infty} \theta'_i B^i$ (see Brockwell and Davis, 1991, pp. 296–304). The coefficients ϕ'_i and θ'_i are readily computed using the recursive procedure of Box and Jenkins (1976, pp. 132–134).

Theorem 1. If the model is correctly identified, $\widehat{D}_m$ 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 $\mathbf{W}_m$ is a diagonal matrix with elements $w_i = (m - i + 1)/m$ $(i = 1, \ldots, m)$.

The proof of this theorem 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. However, these eigenvalues are readily calculated for any given ϕ, θ, and m. In the following subsection, we propose an approximation to the distribution $\sum_{i=1}^{m} \lambda_i \chi_{1,i}^2$ for moderately high m.

3.2 An Approximation to the Distribution of $\widehat{D}_m$

The probability $\Pr(\widehat{D}_m > x)$ can be evaluated by inverting the characteristic function of $\sum_{i=1}^{m} \lambda_i \chi_{1,i}^2$ (Imhof, 1961). This procedure requires only one-dimensional numerical integration, but for simplicity, we prefer to use the approximation proposed by Satterthwaite (1941, 1946) and Box (1954). They suggested approximating a distribution of the form $\sum_{i=1}^{m} \lambda_i \chi_{1,i}^2$ by a distribution of the form $a\chi_b^2$, with mean and variance equal to those of the exact distribution, where the degrees of freedom, b, are usually fractional. This implies $a = \sum \lambda_i^2 / \sum \lambda_i$ and $b = (\sum \lambda_i)^2 / \sum \lambda_i^2$. Thus, following this suggestion, we approximate the distribution of $\widehat{D}_m$ by a gamma

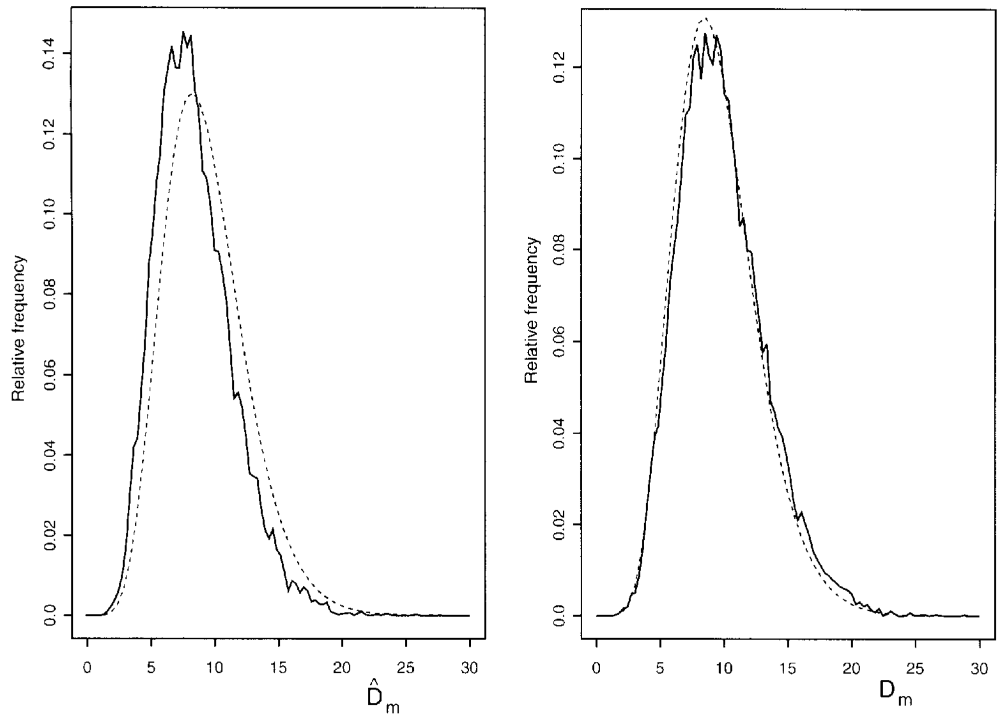

Figure 1. The Solid Line is the Monte Carlo Distribution of $\widehat{D}_m$ (left) and D_m (right) Generated from 10,000 Replications of an AR(1) Model With $\phi_1 = .5$, When $n = 100$ and $m = 20$. The dashed line is the $\mathcal{G}(\alpha, \beta)$ approximation With $\alpha = 7.3$ and $\beta = .77$.

distribution, $\mathcal{G}(\alpha = b/2, \beta = 1/2a)$, where the parameters are defined by

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

and

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

and the distribution has a mean of $\alpha/\beta = (m+1)/2 - (p+q)$ and a variance of $\alpha/\beta^2 = (m+1)(2m+1)/3m - 2(p+q)$. The details of this approximation are given in the Appendix.

This approximation has been checked by a Monte Carlo experiment. As the gamma approximation to the asymptotic distribution improves when the standardized autocorrelation coefficients $\tilde{r}_k$ defined by (2) are used, we will consider both $\widehat{D}_m$, given by (6), and D_m defined by

$$D_m = n[1 - |\widetilde{\mathbf{R}}_m|^{1/m}], \qquad (10)$$

where $\widetilde{\mathbf{R}}_m$ is the correlation matrix built by using the standardized autocorrelation coefficients $\tilde{r}_k$. Under the null hypothesis $\widehat{D}_m$ and D_m are asymptotically equivalent, but obviously, some differences may be expected in their small-sample behaviors. Figure 1 illustrates the accuracy of the approximation of the empirical distribution of $\widehat{D}_m$ and D_m to the gamma distribution, using 10,000 replications of sample size 100 by an AR(1) process with $\phi_1 = .5$ and $m = 20$. We have found similar results in a larger simulation study, and, as the approximation by the standardized autocorrelations is better, we recommend the use of D_m, especially for small sample sizes.

Table 1 shows the percentiles corresponding to $\alpha = .05$ for different values of m and the estimated parameters $(p+q)$ for the proposed approximating distribution. Note that in the approximation $m \geq c(p, q)$, where $c(p, q) \approx 3(p+q-1/2)$ is obtained from the restriction that the variance must be positive.

4. SMALL-SAMPLE SIGNIFICANCE LEVEL AND POWER OF D_m

In this section we present a comparative study of the significance level and power of the three statistics D_m, Q_{LB}, and Q_{MT} given by (10), (3), and (4). The significance levels of Q_{LB} and Q_{MT} have been obtained using the percentiles of the χ^2 distribution, and those of D_m using the approximation obtained in Section 3.2.

These significance levels have been evaluated under both low-order AR and MA models, but here we report only the results for the AR(1). Similar results were found in the other

Table 1. 95% Points of the Recommended Approximation to the Distribution of D_m, for Different Values of m and (p + q)

	$p+q$							
m	0	1	2	3	4	5	6	7
7	8.56	6.70	4.52	—	—	—	—	—
10	10.71	9.00	7.14	4.96	—	—	—	—
12	12.10	10.46	8.71	6.76	4.37	—	—	—
14	13.46	11.87	10.11	8.39	6.35	3.56	—	—
24	19.97	18.52	17.05	15.53	13.96	12.32	10.57	8.63
36	27.42	26.06	24.69	23.30	21.88	20.44	18.96	17.44

Table 2. Significance Levels of D_m, Q_{LB}, and Q_{MT} Under an AR(1) Model

	m = 10			m = 15			m = 20		
ϕ_1	D_m	Q_{LB}	Q_{MT}	D_m	Q_{LB}	Q_{MT}	D_m	Q_{LB}	Q_{MT}
				$\alpha = .05$					
.10	.055	.053	.056	.054	.057	.054	.055	.062	.054
.30	.053	.052	.053	.052	.057	.054	.053	.062	.049
.50	.052	.055	.052	.049	.057	.046	.047	.058	.046
.70	.054	.057	.053	.050	.060	.048	.050	.069	.048
.90	.050	.054	.043	.042	.059	.039	.041	.061	.039
				$\alpha = .01$					
.10	.009	.013	.011	.009	.016	.011	.010	.019	.011
.30	.010	.012	.011	.009	.017	.011	.009	.021	.010
.50	.008	.012	.008	.007	.016	.007	.007	.019	.008
.70	.010	.015	.010	.008	.015	.010	.009	.021	.007
.90	.011	.013	.009	.009	.017	.009	.009	.021	.009

cases, and they are available upon request from the authors. In each case, 10,000 Gaussian series of sample size $n = 100$ were generated. Three values for m, 10, 15, and 20, were considered. Table 2 shows the significance levels of the three statistics for several values of the AR(1) parameter when the nominal levels α are .05 and .01. The observed significance level is close to the nominal level for the three statistics, and the results obtained for Q_{LB} and Q_{MT} are similar to those presented by Monti (1994). In all 30 cases the observed significance level of Q_{LB} is larger than the nominal level, whereas the behavior of Q_{MT} and D_m shows less variability. For $\alpha = .05$, the significance levels of D_m belong to the interval (.041–.055) and are always between those of Q_{LB} (interval .052–.069) and Q_{MT} (interval .039 – .056). For $\alpha = .01$ the observed values for D_m and Q_{MT} are in the interval (.007–.011), whereas those of Q_{LB} are in the range (.012–.021). The significance level of D_m does not seem to be affected by the value of m.

The power of the tests is analyzed for the models proposed by Monti (1994). Twenty-four different ARMA(2, 2) models are considered. Table 3 shows the power of the three tests when, erroneously, an AR(1) or a MA(1) model is fitted to the data. In each case 1,000 series of 100 observations were generated, and the power was computed for $m = 10$ and $m = 20$. The power of the three tests decreases as m increases, as expected, but the loss of power in D_m is relatively small. The test based on D_m is always the most powerful, and the increase in power with respect to the best of the other two statistics, Q_{MT} and Q_{LB}, can be as high as 50% in some cases (see models 1, 11, and 23).

In order to analyze the performance of the test for small sample size, Table 4 presents the same power study with $n = 30$. The values of m are 5 and 10. As before, the power of the three tests decreases as m increases, and the test based on D_m is almost always the most powerful (with the exception of model 22). Again the increase in power can be as high as 75% (see models 5 and 11). As expected, the power of the three tests is generally low, although in some cases (see models 7, 15, and 19) the power of the proposed test can be higher than 70%.

Table 3. Power Levels of the Tests Based on D_m, Q_{LB}, and Q_{MT} When the Data Are Generated From ARMA(2,2) Models and AR(1) and MA(1) Models Are Fitted, $n = 100$

					m = 10			m = 20		
M	ϕ_1	ϕ_2	θ_1	θ_2	D_m	Q_{LB}	Q_{MT}	D_m	Q_{LB}	Q_{MT}
(a) Fitted by AR(1) model										
1	—	—	−.50	—	.415	.234	.278	.299	.189	.178
2	—	—	−.80	—	.987	.751	.959	.972	.590	.855
3	—	—	−.60	.30	.994	.762	.983	.987	.655	.941
4	.10	.30	—	—	.597	.421	.412	.452	.364	.296
5	1.30	−.35	—	—	.807	.620	.605	.649	.540	.415
6	.70	—	−.40	—	.781	.542	.609	.637	.428	.415
7	.70	—	−.90	—	1.000	.982	.998	.998	.905	.992
8	.40	—	−.60	.30	.999	.836	.998	.997	.697	.965
9	.70	—	.70	−.15	.216	.175	.169	.182	.173	.111
10	.70	.20	.50	—	.858	.759	.763	.781	.658	.621
11	.70	.20	−.50	—	.599	.324	.384	.447	.288	.260
12	.90	−.40	1.20	−.30	.988	.694	.971	.970	.555	.893
(b) Fitted by MA(1) model										
13	.50	—	—	—	.366	.295	.265	.287	.243	.189
14	.80	—	—	—	.993	.984	.980	.987	.969	.939
15	1.10	−.35	—	—	1.000	1.000	.999	.999	.986	.988
16	—	—	.80	−.50	.988	.851	.940	.953	.727	.838
17	—	—	−.60	.30	.674	.400	.476	.540	.337	.340
18	.50	—	−.70	—	.957	.888	.876	.888	.801	.736
19	−.5	—	.70	—	.957	.893	.876	.908	.807	.758
20	.30	—	.80	−.50	.859	.583	.743	.765	.468	.556
21	.80	—	−.50	.30	.992	.980	.960	.973	.968	.916
22	1.20	−.50	.90	—	.719	.477	.709	.688	.377	.562
23	.30	−.20	−.70	—	.426	.278	.280	.306	.233	.208
24	.90	−.40	1.20	−.30	.965	.780	.923	.932	.638	.822

Table 4. Powers of the Tests Based on D_m, Q_{LB}, and Q_{MT} When the Data Are Generated Using ARMA(2, 2) Models and AR(1) and MA(1) Models Are Fitted, n = 30

					m = 5			m = 10		
M	ϕ_1	ϕ_2	θ_1	θ_2	D_m	Q_{LB}	Q_{MT}	D_m	Q_{LB}	Q_{MT}
(a) Fitted by AR(1) model										
1	—	—	−.50	—	.210	.123	.142	.153	.117	.098
2	—	—	−.80	—	.587	.309	.457	.495	.253	.305
3	—	—	−.60	.30	.608	.316	.495	.559	.259	.400
4	.10	.30	—	—	.161	.113	.112	.115	.099	.088
5	1.30	−.35	—	—	.367	.210	.197	.220	.198	.130
6	.70	—	−.40	—	.377	.241	.239	.254	.222	.177
7	.70	—	−.90	—	.864	.498	.736	.754	.447	.596
8	.40	—	−.60	.30	.674	.312	.533	.579	.275	.416
9	.70	—	.70	−.15	.059	.036	.046	.044	.053	.041
10	.70	.20	.50	—	.177	.116	.125	.129	.106	.105
11	.70	.20	−.50	—	.238	.113	.134	.140	.102	.079
12	.90	−.40	1.20	−.30	.430	.222	.386	.435	.204	.310
(b) Fitted by MA(1) model										
13	.50	—	—	—	.112	.087	.088	.086	.084	.072
14	.80	—	—	—	.584	.467	.442	.451	.418	.321
15	1.10	−.35	—	—	.784	.660	.670	.665	.587	.529
16	—	—	.80	−.50	.481	.278	.355	.379	.282	.281
17	—	—	−.60	.30	.311	.127	.200	.243	.142	.172
18	.50	—	−.70	—	.538	.392	.362	.386	.340	.249
19	−.50	—	.70	—	.730	.577	.544	.587	.496	.417
20	.30	—	.80	−.50	.267	.161	.206	.216	.164	.153
21	.80	—	−.50	.30	.655	.547	.491	.514	.504	.400
22	1.20	−.50	.90	—	.133	.131	.163	.165	.109	.144
23	.30	−.20	−.70	—	.236	.146	.163	.166	.146	.122
24	.90	−.40	1.20	−.30	.395	.183	.326	.363	.189	.260

5. CHECKING THE LINEARITY ASSUMPTION

The analysis of time series using nonlinear models has gained much attention in recent years because of the limitations of linear models in capturing some observed real data structures and the advances in computational power. A large number of tests of the linearity assumption have been proposed, see Peña, Tiao, and Tsay (2001, chap. 10). These tests can be classified into two groups. The first group is based on the Volterra expansion (Wiener, 1958, lecture 10) of a stationary time series Y_t, as

$$Y_t = \mu + \sum_{i=-\infty}^{\infty} a_i \varepsilon_{t-i} + \sum_{i,j=-\infty}^{\infty} a_{ij}\varepsilon_{t-i}\varepsilon_{t-j} + \sum_{i,j,k=-\infty}^{\infty} a_{ijk}\varepsilon_{t-i}\varepsilon_{t-j}\varepsilon_{t-k} + \cdots, \qquad (11)$$

where μ is the mean level of Y_t and $\{\varepsilon_t, -\infty < t < \infty\}$ is a strictly stationary process of independent and identically distributed random variables. Obviously, Y_t is nonlinear if any of the higher order coefficients, $\{a_{ij}\}$, $\{a_{ijk}\}, \ldots$, is nonzero. Based on this expansion and on Tukey's one degree of freedom for the nonadditivity test, Keenan (1985) and Tsay (1986), among others, have proposed specific tests for nonlinearity.

A second group of tests is based on the idea, proposed by Granger and Anderson (1978) and Maravall (1983), that looking at the autocorrelation function of the squared values of the time series could be useful for identifying nonlinear time series. If the residuals $\hat{\varepsilon}_t$ are independent then the $\hat{\varepsilon}_t^2$ will also be independent, but if the model is nonlinear and the residuals $\hat{\varepsilon}_t$ are not independent, this feature can appear in the autocorrelation function of $\hat{\varepsilon}_t^2$. McLeod and Li (1983) proposed detecting nonlinearity in time series data, using the statistic

$$Q_{LB}(\hat{\varepsilon}_t^2) = n(n+2)\sum_{k=1}^{m}(n-k)^{-1} r_k^2(\hat{\varepsilon}_t^2), \qquad (12)$$

where $r_k(\hat{\varepsilon}_t^2)$ are the autocorrelation coefficients of the squared residuals computed by

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

where $\hat{\sigma}^2 = \sum \hat{\varepsilon}_t^2/n$. In a similar way, it is possible to build a test similar to that of Monti, using the partial autocorrelation of the squared residuals, $Q_{MT}(\hat{\varepsilon}_t^2)$. We can extend the D_m statistic (10) for testing for nonlinearity to

$$D_m(\hat{\varepsilon}_t^2) = n[1 - |\widetilde{\mathbf{R}}_m(\hat{\varepsilon}_t^2)|^{1/m}],$$

where $\widetilde{\mathbf{R}}_m(\hat{\varepsilon}_t^2)$ is the autocorrelation matrix (5), which is now built using the standardized autocorrelation coefficients $\tilde{r}_k(\hat{\varepsilon}_t^2)$. In the following theorem, we give the asymptotic distribution of the statistic $D_m(\hat{\varepsilon}_t^2)$ under the null hypothesis that the series follows an ARMA model that is correctly identified.

Theorem 2. If the series follows an ARMA model, the statistic $D_m(\hat{\varepsilon}_t^2)$ computed from the squared residuals of the correctly fitted model is asymptotically distributed as $\sum_{i=1}^{m} w_i \chi_{1,i}^2$, where $\chi_{1,i}^2$ $(i = 1, \ldots, m)$ are independent χ_1^2 random variables and $w_i = (m - i + 1)/m$ $(i = 1, \ldots, m)$.

The proof of this theorem is given in the Appendix.

The asymptotic distribution of Theorem 2 can be approximated as before by a Satterthwaite-type approach using a distribution of the form $a\chi_b^2$. Thus, if the series follows an ARMA model, the asymptotic distribution of $D_m(\hat{\varepsilon}_t^2)$ can be approximated by a gamma distribution, $\mathcal{G}(\alpha = b/2, \beta = 1/2a)$, where $\alpha = 3m(m+1)/4(2m+1)$ and $\beta = 3m/2(2m+1)$. The first column of Table 1 shows the percentile corresponding to $\alpha = .05$ for the most usual values of m for this approximated distribution.

5.1 Power Study

In this section we compare the powers of the statistics $D_m(\hat{\varepsilon}_t^2)$, $Q_{\mathrm{LB}}(\hat{\varepsilon}_t^2)$, and $Q_{\mathrm{MT}}(\hat{\varepsilon}_t^2)$ for testing for linearity. We will calculate the test powers for the four nonlinear models used by Keenan (1985),

Model 1, $Y_t = e_t - .4e_{t-1} + .3e_{t-2} + .5e_t e_{t-2}$;

Model 2, $Y_t = e_t - .3e_{t-1} + .2e_{t-2} + .4e_{t-1}e_{t-2} - .25e_{t-2}^2$;

Model 3, $Y_t = .4Y_{t-1} - .3Y_{t-2} + .5Y_{t-1}e_{t-1} + e_t$;

Model 4, $Y_t = .4Y_{t-1} - .3Y_{t-2} + .5Y_{t-1}e_{t-1} + .8e_{t-1} + e_t$;

where the e_t's are independent $N(0, 1)$.

Table 5 summarizes the power results. For each model 1,000 replications of sample size $n = 204$ were generated. An AR(p) model is fitted to the data, where p is selected by the Akaike information criterion (AIC) (Akaike, 1974) with $p \in \{1, 2, 3, 4\}$. The power of the proposed test, $D_m(\hat{\varepsilon}_t^2)$, is broadly between 6% (Model 4, $m = 7$) and 40% (Model 2, $m = 24$) higher than the powers of the tests based on $Q_{\mathrm{LB}}(\hat{\varepsilon}_t^2)$ and $Q_{\mathrm{MT}}(\hat{\varepsilon}_t^2)$. None of these tests is powerful in handling Model 1, which contains a concurrent nonlinear term $e_t e_{t-2}$. The difficulties in observing the nonlinearity of Model 1 are seen in the study by Tsay (1986), which compares some statistics for checking nonlinearity over the same four models. From the comparison of Table 5 and the results obtained by Tsay (1986), we conclude that the proposed test $D_m(\hat{\varepsilon}_t^2)$ is better than that proposed by Keenan for all lags, for $m = 7$, 12, and 24, and for all models except Model 2. However, the proposed portmanteau test is slightly worse than the test proposed by Tsay (1986).

We are interested in checking the behavior of the proposed statistics in the detection of nonlinearity in Threshold Autoregressive (TAR) models, which are among the most popular nonlinear time series models in applied research. The simplest, two-regime TAR(1) model is given by

$$y_t = \phi_0^{(1)} + \phi_1^{(1)} y_{t-1} + \varepsilon_{1t}, \qquad \text{if } y_{t-1} \le c,$$
$$y_t = \phi_0^{(2)} + \phi_1^{(2)} y_{t-1} + \varepsilon_{2t}, \qquad \text{if } y_{t-1} > c.$$

With this aim, 1,000 replications were generated with sample sizes 100, 200, and 500 from the TAR(2) model with four regimes proposed by Tiao and Tsay (1994). This model is an alternative to the AR(2) model for capturing the structure of the U.S. real GNP series from the first quarter of 1947 to the first quarter of 1991 with a total of 177 observations. The regimes and their economic interpretation are described in Tiao and Tsay (1994).

Table 6 shows the power of the three statistics D_m, Q_{LB}, and Q_{MT} applied to the residual autocorrelations after fitting an AR(2) model. The values shown in the table are close to the nominal significance level used, (.05), confirming the lack of power of these statistics to detect threshold nonlinear structure.

Table 7 displays the power of the statistics based on the autocorrelations of the squared residuals of the AR(2) fit, $D_m(\hat{\varepsilon}_t^2)$, $Q_{\mathrm{LB}}(\hat{\varepsilon}_t^2)$, and $Q_{\mathrm{MT}}(\hat{\varepsilon}_t^2)$. The power of the three statistics is now much larger, especially in large samples, and again the statistic $D_m(\hat{\varepsilon}_t^2)$ is the most powerful for all sample sizes and lags. The increase in power of $D_m(\hat{\varepsilon}_t^2)$ with respect to the best of Q_{LB} and Q_{MT} for $m = 12$ goes from 22.9% to 40.8%.

Finally we analyze an important class of nonlinear time series models with changing conditional variance. Engle (1982) proposes Autoregressive Conditional Heteroscedasticity (ARCH) models, which have been generalized to the Autoregressive Stochastic Volatility (ARSV(1)) models, proposed by Taylor (1986), and the Generalized Autoregressive Conditional Heteroscedasticity (GARCH(1,1)) models, proposed by Bollerslev (1986). These models have been used for analyzing financial time series (see Carnero, Peña, and Ruiz (2001), for a review of their applications in finance) and for

Table 5. Powers of the Tests Based on D_m, Q_{LB}, and Q_{MT} When the Data Are Generated by Four Nonlinear Models, and the Fitted Model Is an AR(p), $\alpha = .05$

	$m = 7$			$m = 12$			$m = 24$		
Model	D_m	Q_{LB}	Q_{MT}	D_m	Q_{LB}	Q_{MT}	D_m	Q_{LB}	Q_{MT}
1	.157	.114	.120	.126	.100	.091	.091	.094	.089
2	.566	.497	.480	.511	[illegible]	[illegible]	406	.289	.277
3	.960	.900	.914	[illegible]	[illegible]	[illegible]	.870	.740	.70
4	.914	.860	.830	[illegible]	[illegible]	[illegible]	.788	.686	.614

Table 6. Powers of the Tests Based on D_m, Q_{LB}, and Q_{MT} Give Data Generated From a TAR(2) Model When AR(2) Is Fitted, $\alpha = .05$

	$m = 12$			$m = 24$			$m = 36$		
n	D_m	Q_{LB}	Q_{MT}	D_m	Q_{LB}	Q_{MT}	D_m	Q_{LB}	Q_{MT}
100	.045	.036	.046	.043	.044	.045	.025	.056	.032
200	.054	.051	.053	.045	.050	.056	.037	.059	.044
500	.054	.048	.050	.046	.054	.046	.040	.059	.051

Table 7. Powers of the Tests Based on $D_m(\hat{\varepsilon}_t^2)$, $Q_{\mathrm{LB}}(\hat{\varepsilon}_t^2)$, and $Q_{\mathrm{MT}}(\hat{\varepsilon}_t^2)$ Give Data Generated From a TAR(2) Model When AR(2) Is Fitted, $\alpha = .05$

	$m = 12$			$m = 24$			$m = 36$		
n	D_m	Q_{LB}	Q_{MT}	D_m	Q_{LB}	Q_{MT}	D_m	Q_{LB}	Q_{MT}
100	.162	.115	.109	.089	.076	.077	.075	.075	.055
200	.296	.228	.213	.216	.17	.178	.18	.146	.122
500	.648	.527	.524	.556	.442	.415	.499	.369	.342

Table 8. Powers of the $D_m(\hat{\varepsilon}_t^2)$, $Q_{LB}(\hat{\varepsilon}_t^2)$, and $Q_{MT}(\hat{\varepsilon}_t^2)$ Tests for Two ARCH(p) Models

		m = 7			m = 12			m = 24		
p	n	D_m	Q_{LB}	Q_{MT}	D_m	Q_{LB}	Q_{MT}	D_m	Q_{LB}	Q_{MT}
1	100	.567	.484	.469	.504	.427	.411	.409	.341	.298
	250	.787	.733	.731	.755	.707	.691	.699	.644	.623
	500	.854	.835	.840	.843	.818	.814	.825	.794	.788
	1000	.921	.904	.903	.912	.894	.893	.898	.886	.882
3	100	.446	.421	.405	.417	.370	.350	.334	.313	.265
	250	.724	.707	.692	.709	.685	.657	.670	.628	.579
	500	.796	.778	.773	.788	.754	.754	.765	.734	.708
	1000	.882	.873	.870	.875	.850	.842	.854	.830	.814

modeling environmental variables (see Tol, 1996, for a meteorological application). The test most often used for checking this type of nonlinearity is $Q_{LB}(\hat{\varepsilon}_t^2)$ as given by (12).

To analyze the power of the new proposed test for these models, we generate first 1,000 series of sizes $n = 100$, 250, 500, and 1,000 for the covariance stationary process ARCH(p),

$$y_t = \varepsilon_t \sigma_t$$
$$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \cdots + \alpha_p y_{t-p}^2,$$

where $\alpha_0 > 0$ and $\sum_{i=1}^{p} \alpha_i < 1$. We consider the cases $p = 1$ and $p = 3$. The parameters α_i have been sampled from a uniform $U(0, 1)$ and are rescaled by an auxiliary variable, s, from a uniform distribution $U(0, 1)$ so that $\sum_{i=1}^{p} \alpha_i = s$.

Table 8 shows the power of the tests based on $D_m(\hat{\varepsilon}_t^2)$, $Q_{LB}(\hat{\varepsilon}_t^2)$, and $Q_{MT}(\hat{\varepsilon}_t^2)$ for the ARCH(p) models, with $m = 7$, 12, and 24. The test based on $D_m(\hat{\varepsilon}_t^2)$ is again the most powerful for all sample sizes and lags, but the advantage of this test over the other two is lower here than with other nonlinear models. For the ARCH(3) model the differences in power of the three tests are small, and the increase in power of $D_m(\hat{\varepsilon}_t^2)$ goes from 1.0% to 12.7%.

Tables 9 and 10 show the power of these tests for four covariance stationary GARCH(1,1) models,

$$y_t = \varepsilon_t \sigma_t,$$
$$\sigma_t^2 = 1 + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2,$$

where $\alpha \geq 0$, $\beta \geq 0$, and $\alpha + \beta < 1$. Again 1,000 replicates have been made. The parameters in the first two models (see Table 9), are taken from real financial time series (see Carnero et al., 2001), whereas those in Table 10 are taken from environmental data (see Tol, 1996). An interesting result is that the power of the proposed test depends very much on the parameter values: it is slightly less than the one for $Q_{LB}(\hat{\varepsilon}_t^2)$ (although greater than the one for $Q_{MT}(\hat{\varepsilon}_t^2)$) in financial data with long persistence, whereas it is the most powerful of the three in Table 10 with parameter values obtained from the environmental data. A possible reason for this different behavior is given in the following section.

Table 11 shows the power for 1,000 replicates from an ARSV(1) model,

$$y_t = \sigma_* \varepsilon_t \sigma_t,$$
$$\log \sigma_t^2 = \phi \log \sigma_{t-1}^2 + \eta_t,$$

where $\sigma_* > 0$, $|\phi| < 1$, and ε_t and η_t are assumed to be mutually independent and normally distributed white noise processes with zero mean and variances 1 and σ_η^2, respectively. The simulation experiment follows the design of Sandmann and Koopman (1998). The autoregressive parameters are $\phi = \{.7, .9, .98\}$, and the parameter σ_η is selected so that the coefficient of variation of $\log \sigma_t^2$ takes the values 10, 1, and .01. High values of the ratio of volatility variance to its squared mean indicate pronounced relative strength of a stochastic volatility process.

We can observe that the proposed test is more powerful than the other tests for $m = 24$ and $m = 36$. However, for $m = 12$ and $\phi \geq .9$ the power of $D_m(\hat{\varepsilon}_t^2)$ is similar to or slightly slower than the test based on Q_{LB}. These results are consistent with the behavior observed in Table 9, since both models, GARCH(1,1) and ARSV(1), are used to fit financial time series with long persistence.

5.2 An Example with Real Data

In this section we apply the $D_m(\hat{\varepsilon}_t^2)$ statistic to find nonlinearities in the well-known sunspot time series. This time series has been studied previously by various authors who have applied both linear and nonlinear models. A detailed description of some of the different models proposed can be found in Priestley (1989). Following Priestley (1989, p. 882), we have fitted an AR(9) model to the sample of 246 observations corresponding to the first 246 observations from 1,700. The order of the linear model was determined by the AIC. The model is estimated by maximum likelihood, and no structure is found in the residuals using the statistics D_m, Q_{LB},

Table 9. Powers of the $D_m(\hat{\varepsilon}_t^2)$, $Q_{LB}(\hat{\varepsilon}_t^2)$, and $Q_{MT}(\hat{\varepsilon}_t^2)$ Tests for Two GARCH(1,1) Models of Financial Time Series

			m = 12			m = 24			m = 32		
n	α	β	D_m	Q_{LB}	Q_{MT}	D_m	Q_{LB}	Q_{MT}	D_m	Q_{LB}	Q_{MT}
250	.05	.90	.268	.301	.256	.244	.279	.221	.213	.246	.182
	.15	.80	.821	.825	.778	.782	.775	.695	.731	.732	.623
500	.05	.90	.522	.553	.494	.510	.527	.441	.479	.486	.388
	.15	.80	.986	.986	.976	.981	.979	.959	.973	.971	.937
1,000	.05	.90	.807	.837	.786	.802	.809	.728	.776	.771	.672
	.15	.80	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999

Table 10. Powers of the $D_m(\hat{\varepsilon}_t^2)$, $Q_{LB}(\hat{\varepsilon}_t^2)$, and $Q_{MT}(\hat{\varepsilon}_t^2)$ Tests for Two GARCH(1,1) Models of Meteorological Time Series

				m = 7			m = 12			m = 24		
n	ω	α	β	D_m	Q_{LB}	Q_{MT}	D_m	Q_{LB}	Q_{MT}	D_m	Q_{LB}	Q_{MT}
90	1.21	.404	.153	.506	.433	.425	.454	.369	.337	.328	.290	.220
	1.58	.55	.105	.631	.546	.523	.553	.479	.435	.426	.369	.309
180	1.21	.404	.153	.825	.751	.731	.781	.703	.671	.697	.611	.536
	1.58	.55	.105	.905	.867	.854	.875	.818	.794	.810	.739	.675

and Q_{MT}, with lags $m = 12$ and 24 and $\alpha = .05$. However, when the statistic $D_m(\hat{\varepsilon}_t^2)$ is used, a clear indication of nonlinear structure appears. Table 12 presents the values of the three portmanteau statistics applied to the squared residuals. To facilitate the comparison, in addition to the value of each statistic the ratio between the statistic and the percentile corresponding to $\alpha = .05$ for each distribution is also given (see columns 3, 5, and 7). The statistic $D_m(\hat{\varepsilon}_t^2)$ clearly suggests a nonlinear structure in the residuals of the sunspot series, whereas the results from the Ljung–Box and Monti statistics on the squared residuals are not decisive.

6. ADVANTAGES AND LIMITATIONS OF THE PROPOSED TEST

The proposed test has an interesting property that may explain why it works so well in some cases compared with other portmanteau goodness-of-fit tests. Note that, for large n, both the Ljung–Box and Monti tests are symmetric on the autocorrelation coefficients; that is, given a sequence of autocorrelation coefficients $(r_1, \dots, r_m)$, a permutation of this sequence does not affect the test. However, the proposed test is highly asymmetric with respect to the autocorrelation coefficients. In fact, $\widehat{\mathbf{R}}_m$ includes $2(m-i+1)$ times the ith autocorrelation coefficient, which implies that the first autocorrelation coefficient appears $2m$ times in the matrix, whereas the mth autocorrelation coefficient appears only twice. One would expect that this asymmetry will imply a larger sensitivity of $|\widehat{\mathbf{R}}_m|$ to changes in the low-order coefficients, and this seems to be the case. We have checked that for $m = 1, 2, 3, 4, 5$, we have

$$\left.\frac{d|\widehat{\mathbf{R}}_m|}{dr_j}\right|_{\mathbf{r}_{(m)}=\varepsilon\mathbf{1}'} = -2\varepsilon(m+1-j)(1-\varepsilon)^{m-1}, \qquad j = 1, \dots, m,$$

where the derivative of the determinant with respect to the jth autocorrelation coefficient is evaluated at $\mathbf{r}_{(m)} = (\varepsilon, \varepsilon, \dots, \varepsilon)'$. We can see that, in these particular cases, the effect of a change in the jth autocorrelation coefficient decreases with the lag. Thus, we can conclude that the proposed test is more sensitive to the low-order autocorrelation coefficients than to the large order ones, in contrast to Q_{LB} and Q_{MT}, which have the same sensitivity for all order autocorrelation coefficients.

This same property applies to the partial autocorrelation coefficients. In the proposed statistic they also have a weight that decreases with the lags. From (9) it is easy to see that

$$\frac{d|\widehat{\mathbf{R}}_m|}{d\pi_j} = -2(m+1-j)\left(\frac{\hat{\pi}_j}{1-\hat{\pi}_j^2}\right)|\widehat{\mathbf{R}}_m|, \qquad j = 1, \dots, m,$$

and evaluating this function at $\pi_{(m)} = (\varepsilon, \varepsilon, \dots, \varepsilon)'$,

$$\left.\frac{d|\widehat{\mathbf{R}}_m|}{d\pi_j}\right|_{\pi_{(m)}=\varepsilon\mathbf{1}'} = -2(m+1-j)\varepsilon(1-\varepsilon^2)^{m(m+1)/2-1},$$

we conclude that the effect on the determinant of a change in the jth partial autocorrelation coefficient decreases linearly with the lag.

When the information about the lack of fit is mainly included in the low lag autocorrelation coefficients (simple or partial), D_m seems to be more powerful because it gives more weight to these coefficients than do both Q_{LB} and Q_{MT}. On the other hand, when this is not the case and the information is spread over a long number of lags, this advantage may disappear and D_m has a power similar to those of Q_{LB} and Q_{MT}. For instance, for heteroscedastic data with long persistence (see Tables 9 and 11) we have seen that the performances of

Table 11. Power of the $D_m(\hat{\varepsilon}_t^2)$, $Q_{LB}(\hat{\varepsilon}_t^2)$, and $Q_{MT}(\hat{\varepsilon}_t^2)$ Tests for ARSV(1) Models ($n = 250$)

				m = 12			m = 24			m = 36		
ϕ	σ_n	σ_*	CV	D_m	Q_{LB}	Q_{MT}	D_m	Q_{LB}	Q_{MT}	D_m	Q_{LB}	Q_{MT}
.70	1.239	.014	10	.503	.425	.396	.410	.324	.315	.342	.274	.257
.70	.666	.041	1	.468	.389	.362	.376	.299	.275	.315	.263	.228
.70	.247	.029	0.1	.297	.237	.226	.232	.179	.155	.179	.164	.117
.90	.675	.016	10	.837	.818	.787	.787	.733	.687	.735	.662	.606
.90	.363	.025	1	.868	.832	.790	.795	.767	.703	.751	.709	.607
.90	.135	.029	0.1	.688	.695	.631	.647	.604	.525	.566	.532	.435
.98	.308	.016	10	.959	.959	.939	.935	.927	.888	.910	.892	.847
.98	.166	.025	1	.954	.954	.934	.937	.936	.898	.917	.907	.839
.98	.061	.029	0.1	.905	.925	.886	.885	.885	.830	.854	.852	.763

Table 12. Nonlinearity Tests on the Residuals of the Sunspot Data When an AR(9) Model Is Fitted

m	$D_m(\hat{\varepsilon}_t^2)$	$D_m(\hat{\varepsilon}_t^2)/d_C$	$Q_{LB}(\hat{\varepsilon}_t^2)$	$Q_{LB}(\hat{\varepsilon}_t^2)/d_C$	$Q_{MT}(\hat{\varepsilon}_t^2)$	$Q_{MT}(\hat{\varepsilon}_t^2)/d_C$
7	17.35	1.93	21.13	1.50	19.15	1.36
12	18.52	1.53	22.9	1.08	21.17	1.01
24	21.01	1.05	27.07	.74	26.56	.73

the three tests are similar. We have checked that for seasonal models the performance of D_m is still better in general (and sometimes equal) to the performances of Q_{LB} and Q_{MT}. However, it is possible that if the relevant information for model checking is mainly given by the high-order autocorrelation coefficients, D_m might be less powerful than Q_{LB} and Q_{MT}.

APPENDIX: PROOFS OF THEOREMS AND APPROXIMATIONS TO THE DISTRIBUTIONS

Proof of Theorem 1

Suppose that under the null hypothesis, $\widehat{D}_m$ is asymptotically distributed as the random variable X. Then, applying the δ-method (e.g., in Arnold, 1990) to $g(x) = \log(1-x)$, it follows that $-n\log|\mathbf{R}_m|^{1/m}$ is also asymptotically distributed as X. From (9) we obtain the equivalent expression,

$$-n\log(|\widehat{\mathbf{R}}_m|^{1/m}) = -n\sum_{i=1}^{m}\frac{m-i+1}{m}\log(1-\hat{\pi}_i^2). \tag{A.1}$$

To find the distribution of (A.1), suppose that $(n\hat{\pi}_1^2, n\hat{\pi}_2^2, \ldots, n\hat{\pi}_m^2)$ is asymptotically distributed as Y. Then, applying the multivariate δ-method (e.g., in Arnold, 1990) to $g(\hat{\pi}_1^2, \ldots, \hat{\pi}_m^2) = -\sum_{i=1}^m((m-i+1)/m)\log(1-\hat{\pi}_i^2)$, it follows that

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

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

$$\left(1, \frac{m-1}{m}, \ldots, \frac{1}{m}\right)(n\hat{\pi}_1^2, n\hat{\pi}_2^2, \ldots, n\hat{\pi}_m^2)' \to \left(1, \frac{m-1}{m}, \ldots, \frac{1}{m}\right)Y. \tag{A.3}$$

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

$$\left(1, \frac{m-1}{m}, \ldots, \frac{1}{m}\right)(n\hat{\pi}_1^2, n\hat{\pi}_2^2, \ldots, n\hat{\pi}_m^2)' = n\hat{\pi}_{(m)}'\mathbf{W}\hat{\pi}_{(m)} \to \sum_{i=1}^{m}\lambda_i\chi_{1,i}^2. \tag{A.4}$$

Finally, from (A.3) and (A.4),

$$\left(1, \frac{m-1}{m}, \ldots, \frac{1}{m}\right)Y \to \sum_{i=1}^{m}\lambda_i\chi_{1,i}^2,$$

and from (A.2), $\widehat{D}_m \to \sum_{i=1}^m \lambda_i\chi_{1,i}^2$.

Approximation to the Distribution of D_m

Box and Pierce (1970) and McLeod (1978) approximate the matrix $\mathbf{Q}_m = \mathbf{X}_m\mathbf{V}^{-1}\mathbf{X}_m'$ by the projection matrix $\mathbf{Q}_m = \mathbf{X}_m(\mathbf{X}_m'\mathbf{X}_m)^{-1}\mathbf{X}_m'$ when m is moderately high. This approximation is useful for computing an expression for a and b that does not depend on the ARMA parameters ϕ and θ. We have

$$\begin{aligned}\sum_{i=1}^{m}\lambda_i &= \operatorname{tr}((\mathbf{I}_m - \mathbf{Q}_m)\mathbf{W}_m) \\ &= \operatorname{tr}(\mathbf{W}_m) - \operatorname{tr}(\mathbf{Q}_m) + (1/m)\operatorname{tr}(\mathbf{Q}_m\mathbf{C}_m),\end{aligned} \tag{A.5}$$

where $\mathbf{C}_m$ is a diagonal matrix with elements $c_i = i$, $i = 0, \ldots, (m-1)$, and

$$\begin{aligned}\sum_{i=1}^{m}\lambda_i^2 &= \operatorname{tr}((\mathbf{I}_m - \mathbf{Q}_m)\mathbf{W}_m(\mathbf{I}_m - \mathbf{Q}_m)\mathbf{W}_m) \\ &= \operatorname{tr}(\mathbf{W}_m^2) - \operatorname{tr}(\mathbf{Q}_m) + (2/m)\operatorname{tr}(\mathbf{Q}_m\mathbf{C}_m) \\ &\quad - (2/m^2)\operatorname{tr}(\mathbf{Q}_m\mathbf{C}_m^2) + (1/m^2)\operatorname{tr}(\mathbf{Q}_m\mathbf{C}_m\mathbf{Q}_m\mathbf{C}_m).\end{aligned} \tag{A.6}$$

An alternative expression for $\sum\lambda_i$ and $\sum\lambda_i^2$ can be obtained using the Cholesky decomposition of the matrix $(\mathbf{I}_m - \mathbf{Q}_m)$ (see Velilla (1994)). As $\mathbf{Q}_m$ is an idempotent matrix with rank $p+q$, (A.5) and (A.6) can be written as a function of p, q, m, q_{ii}, and q_{ij}, where the q_{ij} are the elements of $\mathbf{Q}_m$,

$$\sum_{i=1}^{m}\lambda_i = \frac{m+1}{2} - (p+q) + \frac{1}{m}\sum_{i=2}^{m}(i-1)q_{ii}, \tag{A.7}$$

$$\begin{aligned}\sum_{i=1}^{m}\lambda_i^2 &= \frac{1}{6m}(m+1)(2m+1) - (p+q) \\ &\quad + \frac{2}{m}\sum_{i=2}^{m}(i-1)q_{ii} - \frac{2}{m^2}\sum_{i=2}^{m}(i-1)^2q_{ii} \\ &\quad + \frac{1}{m^2}\sum_{i=2}^{m}\sum_{j=2}^{m}(i-1)(j-1)q_{ij}^2.\end{aligned} \tag{A.8}$$

Now we will show that the terms in (A.8) which depend on q_{ij} tend to zero when m increases. Consider the sequences $a_i = i$ and $b_i = (i-1)q_{ii}$. Then $\sum_{i=1}^m(i-1)q_{ii}/i \le \sum_{i=1}^m q_{ii} = p+q < \infty$ as $m \to \infty$, and by Kronecker's lemma (Davidson, 1997, pp. 34–35) we obtain that $(2/m)\sum_{i=1}^m(i-1)q_{ii} \to 0$. A similar argument and the property of idempotent matrices, $q_{ii} = q_{ii}^2 + \sum_{i=1}^m q_{ij}^2$, is used to show that $(2/m^2)\sum_{i=1}^m(i-1)^2q_{ii} \to 0$ and $(1/m^2)\sum_{i=2}^m\sum_{j=2}^m(i-1)(j-1)q_{ij}^2 \to 0$. Thus, for large m we can approximate Equations (A.7) and (A.8) by

$$\sum_{i=1}^{m}\lambda_i = \frac{m+1}{2} - (p+q), \tag{A.9}$$

$$\sum_{i=1}^{m}\lambda_i^2 = \frac{1}{6m}(m+1)(2m+1) - (p+q). \tag{A.10}$$

Proof of Theorem 2

This is based on the result by McLeod and Li (1983) for the asymptotic distribution of $n^{1/2}\mathbf{r}_{(m)}(\hat{\varepsilon}_t^2)$, which is $N(\mathbf{0}, \mathbf{I}_m)$. Applying this result to the one obtained by Monti (1994), the asymptotic distribution of $n^{1/2}\hat{\pi}_{(m)}(\hat{\varepsilon}_t^2)$ is $N(\mathbf{0}, \mathbf{I}_m)$. Following the same reasoning as in Theorem 1, the asymptotic distribution for $D_m(\hat{\varepsilon}_t^2)$ is obtained.

[Received December 2000. Revised August 2001.]

REFERENCES

Akaike, H. (1974), "A New Look at the Statistical Model Identification," *IEEE Trans. Automat. Control AC*, 19, 203–217.

Arnold, S. F. (1990), *Mathematical Statistics*, Englewood Cliffs, NJ: Prentice-Hall International.

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

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," *Annals of Mathematical Statistics*, 25, 290–302.

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

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.

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

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

Davidson, J. (1997), *Stochastic Limit Theory*, New York: Oxford University Press.

Engle, R. F. (1982), "Autoregressive Conditional Heteroskedasticity with Estimates of the Variances," *Econometrica*, 50, 987–1007.

Granger, C. W. J., and Anderson, A. P. (1978), *An Introduction to Bilinear Time Series Models*, Göttingen: Vandenhoeck & Ruprecht.

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

Keenan, D. M. (1985), "A Tukey Nonadditivity-Type Test for Time Series Nonlinearity," *Biometrika*, 48, 39–44.

Kwan, A. C. C., and Wu, Y. (1997), "Further Results on the Finite-Sample Distribution of Monti's Portmanteau Test for the Adequacy of an ARMA(p, q) Model," *Biometrika*, 84, 733–736.

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

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

Maravall, A. (1983), "An Application of Nonlinear Time Series Forecasting," *Journal of Business & Economics Statistics*, 1, 66–74.

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

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

Monti, A. C. (1994), "A Proposal for Residual Autocorrelation Test in Linear Models," *Biometrika*, 81, 776–780.

Peña, D., Tiao, G. C., and Tsay, R. S. (2001), *A Course in Time Series Analysis*, New York: Wiley.

Priestley, M. B. (1989), *Spectral Analysis and Time Series* (Vol. 1), San Diego: Academic Press.

Sandmann, G., and Koopman, S. J. (1998), "Estimation of Stochastic Volatility Models via Monte Carlo Maximun Likelihood," *Journal of Econometrics*, 87, 271–301.

Satterthwaite, F. E. (1941), "Synthesis of Variance," *Psychometrica*, 6, 309–316.

——— (1946), "An Approximate Distribution of Estimates of Variance Components," *Biometrics Bulletin*, 2, 110–114.

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

Tiao, G. C., and Tsay, R. S. (1994), "Some Advances in Non-linear and Adaptive Modeling in Time Series," *Journal of Forecasting*, 13, 109–131.

Tol, R. S. J. (1996), "Autoregressive Conditional Heteroscedasticity in Daily Temperature Measurements," *Envirometrics*, 7, 67–75.

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

Velilla, S. (1994), "A Goodness-of-Fit Test for Autoregressive Moving-Average Models Based on the Standardized Sample Spectral Distribution of the Residuals," *Journal of Time Series Analysis*, 15, 637–647.

Wiener, N. (1953), *Non-linear Problems in Random Theory*, Cambridge, MA: MIT Press.