Journal of Statistical Computation and Simulation
Vol. 76, No. 7, July 2006, 593–604

On the power of Portmanteau serial correlation tests

YUE FANG†‡, MICHAEL A. MARTIN*‡, TERENCE J. O'NEILL‡ and STEVEN ROBERTS‡

†Lundquist College of Business, University of Oregon, Eugene, OR 97403, USA
‡School of Finance and Applied Statistics, Faculty of Economics and Commerce, The Australian National University, Canberra, ACT 0200, Australia

(Revised 11 May 2004; in final form 19 January 2005)

This paper studies properties of the portmanteau statistic proposed by Box and Pierce [1] and its modification of Ljung and Box [2]. We show that these portmanteau statistics are feasible analogs to optimal tests for the class of statistics which are linear combinations of consistent estimates of serial correlations. We find, however, that for sample sizes commonly encountered in practice, the efficiency loss in power of portmanteau statistics relative to optimal tests can be substantial, although their size properties are broadly comparable. Our results indicate that tests based on some other non-optimal weighting schemes, including tests with optimal weights constructed from moderately misspecified alternatives, deliver tests with better power than the Box–Pierce or Ljung–Box statistics.

Keywords: Bahadur approximate slope; Local power; Optimal test; Portmanteau statistic; Serial correlation

1. Introduction

Diagnostic checking is fundamental to time series analysis. The common practice of checking whether the fitted model is adequate is to compute the portmanteau statistic of Box and Pierce [1] or its modification of Ljung and Box [2], neither of which requires that a specific alternative be given but rather contemplates general alternatives within several classes of models. By leaving the alternatives so broad and unrestrictive, however, a good deal of power may often be lost.

In this paper, we compare properties of portmanteau statistics of Box and Pierce [1] and Ljung and Box [2] (henceforth Q and Q' statistics) vis-à-vis the optimal tests. The optimal tests are derived with optimality criterion as the maximum local power and the maximum Bahadur approximate slope [3, 4] within a general class of statistics that are linear combinations of consistent estimators of serial correlations. Although neither Q nor Q' belongs to this class literally, it is shown that Q and Q' statistics are feasible analogs to optimal tests, which are

*Corresponding author. Email: michael.martin@anu.edu.au

Journal of Statistical Computation and Simulation
ISSN 0094-9655 print/ISSN 1563-5163 online © 2006 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/10629360500108921

attained by substituting sample serial correlations for population serial correlations in optimal weight functions.

Let $\{X_t\}_1^n$ be a finite sample realization of the stationary real-valued process

$$X_t - \mu = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \tag{1}$$

where $\{Z_t\}$ are uncorrelated identically distributed random variables with mean zero and finite variance σ^2. The aforementioned $\{X_t\}$ series may be residuals in regression analysis or they may arise in a pure time series context. The null hypothesis of interest is that $\{X_t\}$ form a sequence of uncorrelated random variables.

Define the vector of sample serial correlations, $\hat{r}(q) = [\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_q]'$, where $\hat{r}_k$ is a consistent estimator for the kth-order population serial correlation $\hat{r}_k$ of $\{X_t\}$. The Q statistic can be computed as $Q(q) = n\sum_{i=1}^{q} \hat{r}_i^2$, where n is the sample size and q is typically chosen to be small relative to n. Under the null hypothesis of zero serial correlation, the asymptotic distribution of Q is χ^2. However, for relatively small samples, the actual significance levels of Q are known to be considerably different from those predicted by asymptotic theory. A simple modification proposed by Ljung and Box [2], $Q'(q) = n(n+2)\sum_{i=1}^{q}(n-i)^{-1}\hat{r}_i^2$, is more satisfactory in this respect [5, 6].

In section 2, optimal tests are derived and are contrasted to the portmanteau statistics for the general problem of testing for zero serial correlation. In section 3, a simulation study is used to compare finite sample size and power properties of portmanteau statistics vis-à-vis optimal tests. Section 4 gives the conclusion.

2. Optimal tests

2.1 *A synthesis*

Consider a synthesis of statistics based on linear combinations of estimated serial correlations:

$$\mathcal{T}_{q,\mathcal{D}_q}(\hat{r}) \equiv D_q' \hat{r}(q), \tag{2}$$

where $D_q = [d_1, d_2, \ldots, d_q]'$ is a weight vector. To simplify the notation, denote $\mathcal{T}_{q,\mathcal{D}_q}(\hat{r})$ as $\mathcal{T}$ when no confusion arises. Tests which use different (standardized) weights on serial correlations of various lags have the same asymptotic distribution under the null hypothesis but deliver different asymptotic power against various alternatives.

The test (2) can be implemented using the asymptotic distribution of the vector $\hat{r}(q)$. For fixed q, it is possible to derive the asymptotic distribution of $\hat{r}(q)$ under appropriate regularity conditions on the process $\{X_t\}$. One way to put assumptions on X_t is to have constraints on the sequence $\{\psi_j\}$ in equation (1). If we assume that

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty \quad \text{and} \quad \sum_{j=-\infty}^{\infty} |j|\psi_j^2 < \infty, \tag{3}$$

then we have [see, for example, Theorem 7.2.2., ref. 7]

$$\hat{r}(q) \sim N(r(q), n^{-1}\mathcal{W}), \tag{4}$$

where $r(q) = [r_1, r_2, \ldots, r_q]'$ and $\mathcal{W}$ is the covariance matrix whose (i, j)-element is given by Bartlett's formula. For a time series (1) satisfying equation (3), we have

$$\mathcal{T} \sim N(D_q' r(q), n^{-1} D_q' \mathcal{W} D_q). \tag{5}$$

Observing that if $\{X_t\}$ is a sequence of uncorrelated identically distributed random variables, the sample serial correlations will be nearly uncorrelated in large samples and hence, the standardized $\mathcal{T}$ can be used as a test statistic for gauging serial correlation.

The asymptotic distributions of the standardized $\mathcal{T}$ under both null and alternative hypotheses can be derived. Under the null hypothesis,

$$\sqrt{n(D_q' D_q)^{-1}}\mathcal{T} \sim N(0, 1), \tag{6}$$

whereas for a given alternative hypothesis,

$$\sqrt{n(D_q' D_q)^{-1}}\mathcal{T} \sim N\left(\sqrt{n(D_q' D_q)^{-1}}(D_q' r(q)), (D_q' D_q)^{-1} D_q' \mathcal{W} D_q\right). \tag{7}$$

Note that the condition (3) is satisfied by a broad class of time series including stationary autoregressive and moving average (ARMA) processes, models widely used in time series analysis. This condition can be relaxed at the expense of the assumption of the finite fourth moment. Furthermore, the assumption of identical distributions of Z_t's can be replaced by appropriate finite moment conditions and weaker martingale conditions which permit the use of a central limit theorem for non-identically distributed random variables [see, for example, ref. 8].

We also note that the expressions for the asymptotic covariance of $\hat{r}(q)$ involve infinite sums. A more convenient form of $\mathcal{W}$ for computational purposes can be derived using the results that these sums can be interpreted (up to a constant factor) as the autocovariances corresponding to the square of their spectral densities. In particular, the asymptotic distribution of $\hat{r}(q)$ of an ARMA process can be expressed in terms of the autocovariance function of another ARMA process, whose parameters are obtained from the parameters of the initial process by squaring its autoregressive operator, moving average operator and the residual variance [9]. These alternative expressions of $\mathcal{W}$ are useful in the simulation studies in section 3.

2.2 *Optimal tests*

The test $\mathcal{T}$ is consistent for any D_q, provided $D_q' r(q) \neq 0$. We use both local and non-local methods to compare asymptotic performance of the test (2) with different weight functions. The local method considers the limiting relative efficiency of tests against a sequence of alternatives. In the local approach, the sequence of alternatives gets arbitrarily close to the null hypothesis in order to avoid forcing the power of tests to be nearly one. In contrast, the non-local method examines the rate at which the logarithm of the asymptotic marginal significance level of the test decreases as sample size increases, under a given alternative. Two approaches give different insights and optimal tests.

We first focus on the local power analysis. Consider a local alternative hypothesis: $n^{-1/2} r(q)$. The test with the local optimal power against this alternative is the test that maximizes

$$C_{\mathrm{L}} \equiv (D_q' r(q))^2 (D_q' \mathcal{W} D_q)^{-1}. \tag{8}$$

Since we can multiply D_q by any scalar without changing the value of $\mathcal{C}_L$, we shall adopt the normalization that $D_q' r(q) = \gamma$, where γ is a constant. Hence, by fixing q, maximizing $\mathcal{C}_{\mathrm{L}}$ is equivalent to the following minimization problem:

$$\min_{D}: D_q' \mathcal{W} D_q$$

$$\text{subject to: } D_q' r(q) = \gamma.$$

The solution of the above problem is $\gamma(r(q)'\mathcal{W}^{-1}r(q))^{-1}\mathcal{W}^{-1}r(q)$. Substituting this solution for D_q, the test with optimal local power is based on

$$\mathcal{T}_{\mathrm{L}} \equiv \sqrt{n[r(q)'\mathcal{W}^{-2}r(q)]^{-1}}r(q)'\mathcal{W}^{-1}\hat{r}(q). \tag{9}$$

As an alternative approach, the non-local method examines the approximate slope of a test as in ref. [3]. Geweke [4] has shown that if the test statistic's limiting distribution under the null hypothesis is a χ^2 distribution, then the approximate slope of the test equals the probability limit of the statistic divided by sample size n. From equation (6) it follows that $\left(\sqrt{n(D_q'D_q)^{-1}}\mathcal{T}\right)^2 \sim \chi_1^2$ under the null hypothesis. Hence, the approximate slope of the test is $\lim_{n\to\infty} n^{-1}\left(\sqrt{n(D_q'D_q)^{-1}}\mathcal{T}\right)^2$, which has the limit

$$\mathcal{C}_{\mathrm{A}} \equiv (D_q'r(q))^2(D_q'D_q)^{-1}. \tag{10}$$

Note that $\mathcal{C}_{\mathrm{A}}$ is an approximation of $\mathcal{C}_{\mathrm{L}}$ in equation (8) replacing $\mathcal{W}$ by the identity matrix. Therefore, the approximate slope strategy yields a test focusing on maximizing the squared mean and ignoring the variance factor of the asymptotic distribution of the test statistic. The optimal weight based on the approximate slope criteria is $\xi r(q)$, where ξ is a non-zero constant. The test with the maximum approximate slope is based on

$$\mathcal{T}_{\mathrm{A}} \equiv \sqrt{n[r(q)'r(q)]^{-1}}r(q)'\hat{r}(q). \tag{11}$$

It should be noted that there is, in general, no uniformly, asymptotically most powerful test for H_0. The weight function of the test with optimal local power depends explicitly on the serial correlations under the alternative; so too does that of the maximum approximate slope test.

2.3 *Some theoretical properties of optimal and portmanteau tests*

The derived optimal tests in section 2.2 are useful to serve as benchmarks in evaluating properties of Q and Q' statistics (see section 3 for more details on finite-sample power comparisons) and provide an intuitively appealing interpretation of Q and Q' statistics.

The Q and its finite-sample modification, Q', can be viewed as feasible norms of the optimal tests based on the data. In particular, by replacing the optimal weights in $\mathcal{T}_{\mathrm{A}}$ in equation (11) using sample serial correlations, the squared $\mathcal{T}_{\mathrm{A}}$ becomes the Q statistic. The Q statistic can also be constructed by approximating $\mathcal{W}$ with the identity matrix and then replacing the optimal weights in $\mathcal{T}_{\mathrm{L}}$ in equation (9) using sample serial correlations.

Before comparing the power of Q and Q' vis-à-vis optimal tests, we examine $\mathcal{T}_{\mathrm{L}}$ and $\mathcal{T}_{\mathrm{A}}$ in some detail for two simple autocorrelation structures: AR(1) and MA(1) processes.

First, consider the AR(1) process: $X_t = \phi X_{t-1} + Z_t$ with $|\phi| < 1$. Using results that $r_0 = 1$ and $r_k = \phi^{|k|}$ for $k = \pm 1, \pm 2, \ldots$, it can be verified algebraically that the diagonal element of $\mathcal{W}$ has form

$$w_{ii} = \phi^{2i}\sum_{l=1}^{i}(\phi^{-l} - \phi^{l})^2 + (\phi^{-i} - \phi^{i})^2\sum_{l=i+1}^{\infty}\phi^{2l} = (1-\phi^{2i})(1+\phi^2)(1-\phi^2)^{-1} - 2i\phi^{2i}.$$

For $i \neq j$, the (i, j)th element of $\mathcal{W}$ is

$$w_{ij} = \phi^{i+j}\sum_{l=1}^{j}(\phi^{-l} - \phi^{l})^2 + \phi^{i}(\phi^{-j} - \phi^{j})\sum_{l=j+1}^{i}(1 - \phi^{2l}) + (\phi^{-i} - \phi^{i})(\phi^{-j} - \phi^{j})\sum_{l=i+1}^{\infty}\phi^{2l}$$
$$= \{(\phi^2 + 1) + (1 - \phi^2)(i - j)\}(\phi^{i-j} - \phi^{i+j})(1 - \phi^2)^{-1} - 2j\phi^{i+j}.$$

Figure 1 presents the optimal weight functions of both $\mathcal{T}_{\mathrm{L}}$ and $\mathcal{T}_{\mathrm{A}}$ for several selected ϕ values for sample size 50. For easy comparison, the weights are standardized by having a unit (L^2-) norm and the weighting coefficient of $\hat{r}_1$ non-negative.

Figure 1 shows that serial correlations with relatively small lags play more important roles than those with relatively large lags in $\mathcal{T}_{\mathrm{A}}$, whose weight on lag k, proportional to $\phi^{|k|}$, decays rapidly as k increases. In contrast, the weights for $\mathcal{T}_{\mathrm{L}}$ are not necessarily, in general, a monotone decreasing function of lag, as indicated by the cases of $|\phi| = 0.9$ [figure 1(e) and (f)]. Note that optimal weights of $\mathcal{T}_{\mathrm{L}}$ depend not only intimately on the magnitude of ϕ but also on its sign. For example, the optimal weights alternate around zero when ϕ is negative; the fluctuation becomes much more visible for alternatives with ϕ close to -1.

As in the AR(1) case, we can obtain a closed form for $\mathcal{W}$ for MA(1) alternative: $X_t = Z_t - \theta Z_{t-1}$. Specifically, it can be shown that for the MA(1) process, $\mathcal{W} = \mathcal{W}_1 + \mathcal{W}_2$, where $\mathcal{W}_1 = (w_{ij}^{(1)})$ is a symmetric $q \times q$ matrix given below:

$$w_{ij}^{(1)} = \begin{cases} -5r_1^2 + 4r_1^4, & i = j = 1 \\ -2r_1^3, & i = 1,\ \ j = 2 \quad \text{or} \quad i = 2, j = 1 \\ 0, & \text{otherwise,} \end{cases}$$

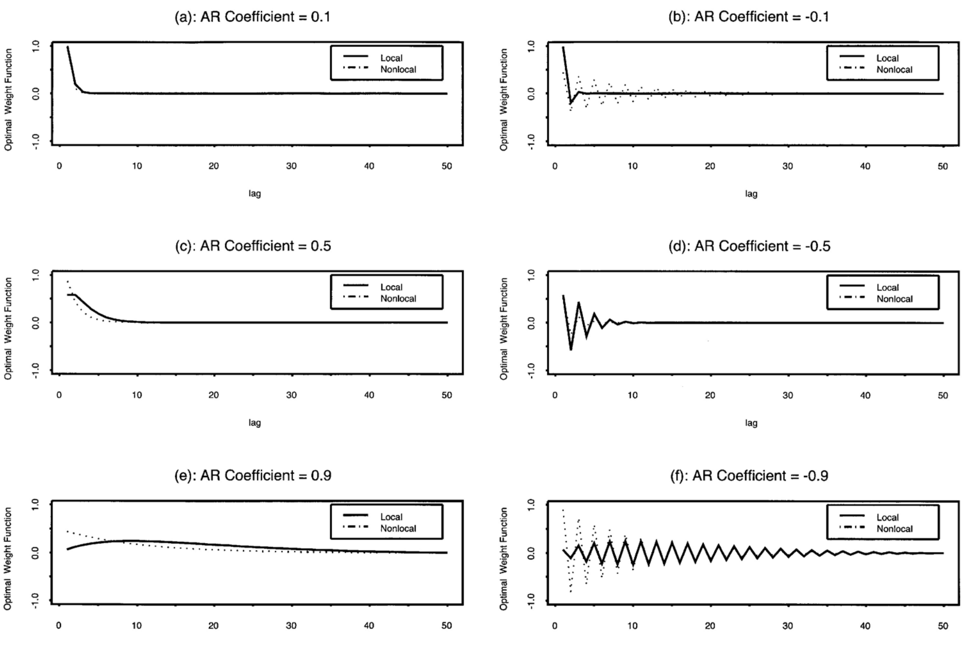

Figure 1. Weight functions of optimal tests for AR(1) processes.

and $\mathcal{W}_2 = (w_{ij}^{(2)})$, a Toeplitz matrix, by

$$w_{ij}^{(2)} = \begin{cases} 1 + 2r_1^2, & i = j \\ 2r_1, & |i-j| = 1 \\ r_1^2, & |i-j| = 2 \\ 0, & \text{otherwise,} \end{cases}$$

where $r_1 = -\theta/(1+\theta^2)$. The weights of $\mathcal{T}_{\mathrm{L}}$ and $\mathcal{T}_{\mathrm{A}}$ for the MA(1) alternatives with various values of θ are shown in figure 2.

As shown in figure 2, for the MA(1) alternative hypothesis, optimal local weights do not cut off after lag 1. Any other weight functions would make the resulting test statistics asymptotically inefficient under the local optimality criteria. The Bahadur approach delivers a different weighting scheme for which $\mathcal{T}_{\mathrm{A}}$ puts non-zero weights only on the first-order sample serial correlation regardless of the value of q.

Although one may expect that the (asymptotic) behaviours of Q or Q' should be very similar to those of optimal tests due to the close connection between their weight functions, the two groups of tests have, however, remarkably different (asymptotic) power properties. For example, the value of q has different implications on the power of tests. We note that the optimality for tests $\mathcal{T}_{\mathrm{L}}$ and $\mathcal{T}_{\mathrm{A}}$ is established assuming that the lag parameter q is predetermined. If we allow q to vary, maximizing $\mathcal{C}_{\mathrm{L}}$ (or $\mathcal{C}_{\mathrm{A}}$) may not be feasible with a finite range of q for some alternatives. Since the optimal approximate slope will not decrease with q, it becomes apparent that large values of q are preferred for $\mathcal{T}_{\mathrm{L}}$ and $\mathcal{T}_{\mathrm{A}}$. We anticipate that under a given alternative, inclusion of the serial correlations with higher lags in $\mathcal{T}_{\mathrm{L}}$ and $\mathcal{T}_{\mathrm{A}}$ should not, in general, decrease the power of these tests (more finite-sample evidence is given in section 3).

In contrast, similar to other test statistics distributed asymptotically as χ^2 (see ref. [10] for discussion on specification tests based on generalized method of moments), the power

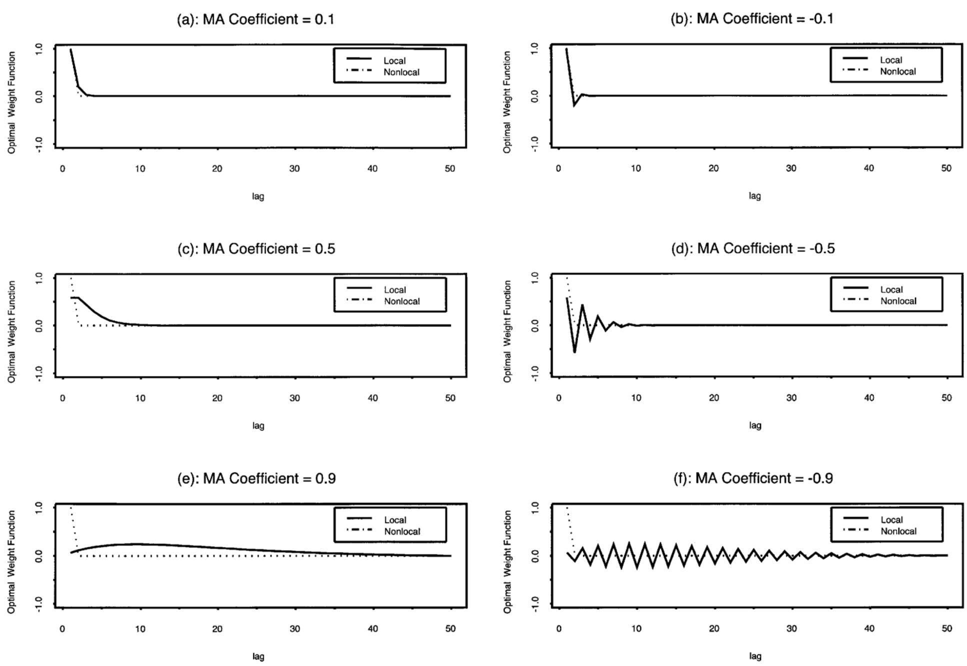

Figure 2. Weight functions of optimal tests for MA(1) processes.

function of Q' (or Q) depends on the lag parameter q in a more complicated way. For example, it can be shown that although additional serial correlations are used in Q' with higher values of q, this does not necessarily imply that large values of q are generally more desirable even asymptotically. Since $\mathcal{W}$ is non-singular, one can by means of a non-singular linear transformation [11] express Q' in the form

$$Q' \sim n \sum_{i=1}^{q} \lambda_i \chi_1^2(\delta_i^2), \tag{12}$$

where the λ_i's are the characteristic roots of $\mathcal{W}$, the δ_i's are certain linear combinations of components of $r(q)$ and the $\chi_1^2(\delta_i^2)$'s are independent χ^2-variables with one degree of freedom and non-centrality parameter δ_i^2. From equation (12), we infer that increasing q has two opposing effects: a tendency for power to increase as δ_i^2 tends to increase as q does; and a tendency for power to decrease since the variance of $\sum_{i=1}^{q} \lambda_i \chi_1^2(\delta_i^2)(= \sum_{i=1}^{q} \lambda_i^2(2 + 4\delta_i^2))$ may increase with q. The extent to which these effects compensate for each other in practice is difficult to predict.

Simulation is an obvious and useful way to proceed here.

3. Power comparison

Monte Carlo simulation experiments were carried out to examine finite sample performance of the Q' statistic in comparison to $\mathcal{T}_{\mathrm{L}}$ and $\mathcal{T}_{\mathrm{A}}$ (results on Q are omitted for brevity.) Due to the unreliability of asymptotic theory under the null when qn^{-1} is large, we have chosen q to be no more than 20% of the sample size, n, set to 100, 200 and 400. All simulations are based on 10,000 replications and performed using the R statistical programming environment (version 2.0.1). The critical values of all test statistics are empirically determined by simulation under the IID Gaussian null.

Although the results obtained in the previous section apply to a broad class of time series, for illustrative purposes, we will focus solely on ARMA models and report results only for alternative hypotheses based on several selected model specifications, which are simple but can reveal main aspects of the results in our analysis.

3.1 *Portmanteau statistics vis-à-vis optimal tests based on correctly specified alternatives*

The results in this subsection are based on the optimal tests resulting from use of the correctly specified alternative. In the first set of Monte Carlo experiments, we consider alternative hypotheses in which serial correlations at relatively high lags are not significant. Table 1 presents empirical powers of Q' vis-à-vis $\mathcal{T}_{\mathrm{L}}$ and $\mathcal{T}_{\mathrm{A}}$ for six alternatives: AR(1), MA(1), AR(2), MA(2), ARMA(1,1) and ARMA(1,2) with selected parameterizations. We also give the ratios of the power of Q' relative to those of optimal tests (quoted in percentages). The results show that regardless of sample size, the test Q' is dominated by both $\mathcal{T}_{\mathrm{L}}$ and $\mathcal{T}_{\mathrm{A}}$ in terms of power for all alternatives considered. The relative efficiency loss of Q' in power can be considerable: as the case of 200 observations demonstrates, the power of the 5% Q' test against the AR(1) model is ~70% relative to $\mathcal{T}_{\mathrm{A}}$ for $q = 5$, and falls to as low as 40% when q is taken to be 40.

We note that although the relative efficiency loss of Q' vis-à-vis $\mathcal{T}_{\mathrm{L}}$ and $\mathcal{T}_{A}$ depends upon the alternative hypothesis, the results seem to be consistent, at least qualitatively, across all alternatives considered previously and many other low-order ARMA models (data not shown).

Table 1. Empirical power of Q' relative to that of $\mathcal{T}_A$ for non-seasonal models. (Results are quoted in percentages.)

		Estimated size of nominal 5% test			Estimated power of nominal 5% test				
n	q	Q'	$\mathcal{T}_L$	$\mathcal{T}_A$	Q'	$\mathcal{T}_L$	$\mathcal{T}_A$	$Q'/\mathcal{T}_L$	$Q'/\mathcal{T}_A$
AR(1) alternative: $X_t = \phi X_{t-1} + Z_t$ with $\phi = 0.2$									
100	2	4.80	4.36	4.36	37.02	42.40	45.81	87.31	80.81
	5	5.11	4.36	4.36	26.86	42.40	45.31	63.35	59.28
	10	6.08	4.36	4.36	22.27	42.40	45.29	52.52	49.17
	20	7.29	4.36	4.36	20.71	42.40	45.29	48.84	45.73
200	2	5.05	5.27	5.27	69.76	76.74	77.88	90.90	89.57
	5	5.23	5.27	5.27	54.41	76.74	77.40	70.90	70.30
	10	5.73	5.27	5.27	43.31	76.74	77.40	56.44	55.96
	20	6.42	5.27	5.27	35.04	76.74	77.40	45.66	45.27
	40	7.92	5.27	5.27	30.57	76.74	77.40	39.84	39.50
400	2	5.01	4.76	4.76	95.51	97.31	97.41	98.15	98.05
	5	4.78	4.76	4.76	88.97	97.31	97.39	91.43	91.35
	10	4.66	4.76	4.76	79.20	97.31	97.39	81.39	81.32
	20	5.16	4.76	4.76	66.39	97.31	97.39	68.23	68.17
	40	6.32	4.76	4.76	54.93	97.31	97.39	56.45	56.40
	80	7.62	4.76	4.76	46.38	97.31	97.39	47.66	47.62
MA(1) alternative: $X_t = Z_t - \theta Z_{t-1}$ with $\theta = 0.2$									
100	2	5.26	4.89	4.89	39.66	46.86	49.98	84.64	79.35
	5	5.12	4.89	4.89	28.23	45.82	49.98	61.61	56.48
	10	5.94	4.89	4.89	22.62	45.83	49.98	49.36	45.26
	20	7.37	4.89	4.89	20.50	45.83	49.98	44.73	41.02
200	2	4.73	4.45	4.45	71.73	79.24	80.00	90.52	89.66
	5	5.04	4.45	4.45	55.22	78.24	80.00	70.58	69.03
	10	5.56	4.45	4.45	42.88	78.25	80.00	54.80	53.60
	20	6.34	4.45	4.45	33.89	78.25	80.00	43.31	42.36
	40	7.77	4.45	4.45	29.20	78.25	80.00	37.32	36.50
400	2	4.74	4.81	4.81	95.54	97.38	97.75	98.11	97.74
	5	4.93	4.81	4.81	88.07	97.28	97.75	90.53	90.10
	10	5.23	4.81	4.81	77.67	97.29	97.75	79.83	79.46
	20	5.71	4.81	4.81	63.38	97.29	97.75	65.15	64.84
	40	6.80	4.81	4.81	50.98	97.29	97.75	52.40	52.15
	80	8.04	4.81	4.81	43.50	97.29	97.75	44.71	44.50
ARMA(1,1) alternative: $X_t = \phi X_{t-1} + Z_t - \theta Z_{t-1}$ with $\theta = 0.65$ and $\theta = 0.5$									
100	2	4.90	4.76	4.76	32.61	39.83	40.75	81.87	80.02
	5	5.21	4.76	4.76	26.77	40.30	42.01	66.43	63.72
	10	6.17	4.76	4.76	22.56	40.40	41.00	55.84	55.02
	20	7.42	4.76	4.76	20.88	40.42	40.85	51.66	51.11
200	2	4.83	4.88	4.88	60.03	67.92	69.24	88.38	86.70
	5	5.41	4.88	4.88	51.48	69.16	70.73	74.44	72.78
	10	5.38	4.88	4.88	42.17	69.23	70.04	60.91	60.21
	20	6.60	4.88	4.88	36.05	69.29	69.94	52.03	51.54
	40	8.13	4.88	4.88	31.67	69.29	69.94	45.71	45.28
400	2	5.04	5.18	5.18	90.16	93.78	94.02	96.14	95.89
	5	5.19	5.18	5.18	84.84	94.03	94.53	90.23	89.75
	10	5.26	5.18	5.18	77.16	94.08	94.41	82.02	81.73
	20	5.77	5.18	5.18	67.23	94.07	94.38	71.47	71.23
	40	6.87	5.18	5.18	56.88	94.07	94.38	60.47	60.27
	80	8.41	5.18	5.18	49.36	94.07	94.38	52.47	52.30
AR(2) alternative: $X_t = \phi X_{t-1} + \phi_2 X_{t-2} + Z_t$ with $\theta = 0.15$ and $\phi_2 = -0.1$									
100	2	4.86	4.32	4.32	24.49	30.64	30.63	79.93	79.95
	5	5.55	4.32	4.32	18.19	26.91	32.79	67.60	55.47
	10	5.56	4.32	4.32	15.65	26.91	32.77	58.16	47.76
	20	6.65	4.32	4.32	15.27	26.91	32.77	56.74	46.60

(*continued*)

Table 1. Continued

		Estimated size of nominal 5% test			Estimated power of nominal 5% test				
n	q	Q'	$\mathcal{T}_{\mathrm{L}}$	$\mathcal{T}_{\mathrm{A}}$	Q'	$\mathcal{T}_{\mathrm{L}}$	$\mathcal{T}_{\mathrm{A}}$	$Q'/\mathcal{T}_{\mathrm{L}}$	$Q'/\mathcal{T}_{\mathrm{A}}$
200	2	5.16	4.70	4.70	51.62	62.52	62.53	82.57	82.55
	5	5.22	4.70	4.70	36.20	59.55	63.13	60.79	57.34
	10	5.47	4.70	4.70	27.58	59.55	63.13	46.31	43.69
	20	6.25	4.70	4.70	22.18	59.55	63.13	37.25	35.13
	40	7.79	4.70	4.70	20.75	59.55	63.13	34.84	32.87
400	2	4.94	5.18	5.18	85.81	92.07	92.15	93.20	93.12
	5	4.85	5.18	5.18	70.86	91.06	91.78	77.82	77.21
	10	5.51	5.18	5.18	55.50	91.06	91.76	60.95	60.48
	20	5.51	5.18	5.18	42.68	91.06	91.77	46.87	46.51
	40	6.09	5.18	5.18	33.37	91.06	91.77	36.65	36.36
	80	7.87	5.18	5.18	29.50	91.06	91.77	32.40	32.15
MA(2) alternative: $X_t = Z_t - \theta_1 Z_{t-1} - \theta_2 Z_{t-2}$ with $\theta_1 = -0.1$ and $\theta_2 = 0.15$									
100	2	4.99	4.38	4.38	27.87	34.11	36.43	81.71	76.50
	5	5.11	4.38	4.38	19.31	25.40	36.43	76.02	53.01
	10	6.15	4.38	4.38	16.56	24.34	36.43	68.04	45.46
	20	7.63	4.38	4.38	15.92	24.35	36.43	65.38	43.70
200	2	4.80	4.60	4.60	57.16	66.48	68.37	85.98	83.60
	5	4.88	4.60	4.60	40.29	61.75	68.37	65.25	58.93
	10	5.48	4.60	4.60	30.48	60.96	68.37	50.00	44.58
	20	6.28	4.60	4.60	24.47	60.94	68.37	40.15	35.79
	40	7.40	4.60	4.60	23.38	60.94	68.37	38.37	34.20
400	2	5.00	4.95	4.95	88.87	93.87	94.05	94.67	94.49
	5	5.06	4.95	4.95	75.85	93.04	94.05	81.52	80.65
	10	5.23	4.95	4.95	61.61	92.80	94.05	66.39	65.51
	20	5.71	4.95	4.95	47.84	92.80	94.05	51.55	50.87
	40	6.46	4.95	4.95	37.79	92.80	94.05	40.72	40.18
	80	7.84	4.95	4.95	33.11	92.80	94.05	35.68	35.20
ARMA(1,2) alternative: $X_t = \phi X_{t-1} + Z_t - \theta_1 Z_{t-1} - \theta_2 Z_{t-2}$ with $\theta = 0.4$, $\theta_1 = 0.3$, $\theta_2 = 0.2$									
100	2	4.61	4.78	4.78	32.62	41.43	42.83	78.74	76.16
	5	5.05	4.78	4.78	26.01	42.37	49.51	61.39	52.53
	10	5.72	4.78	4.78	21.68	43.20	49.69	50.19	43.63
	20	6.72	4.78	4.78	19.87	43.37	49.71	45.82	39.97
200	2	5.09	4.97	4.97	63.31	73.61	74.65	86.01	84.81
	5	5.17	4.97	4.97	52.42	75.88	79.38	69.08	66.04
	10	5.60	4.97	4.97	40.49	76.32	79.40	53.05	50.99
	20	6.14	4.97	4.97	32.39	76.37	79.40	42.41	40.79
	40	7.61	4.97	4.97	28.80	76.36	79.40	37.72	36.27
400	2	4.97	4.87	4.87	93.66	96.71	96.82	96.85	96.74
	5	5.06	4.87	4.87	88.05	97.13	97.71	90.65	90.11
	10	5.43	4.87	4.87	77.21	97.33	97.72	79.33	79.01
	20	5.75	4.87	4.87	63.32	97.32	97.72	65.06	64.80
	40	6.56	4.87	4.87	50.14	97.32	97.72	51.52	51.31
	80	8.20	4.87	4.87	42.98	97.32	97.72	44.16	43.98

The power of two optimal tests are comparable for all sample sizes regardless of the values of q. In most cases, when q is taken to be relatively large compared with n, the power of optimal tests do not decrease. However, insignificance of serial correlations at high lags causes low power which, in turn, leads to substantial loss in relative efficiency of Q'. The results are generally consistent with the asymptotic properties presented in section 2.3.

Table 1 also reports empirical sizes of the tests considered in order to assess whether the disparity in power properties is compensated by favourable size properties for the Q' test. The results show that the estimated sizes of the optimal tests are very close to the nominal 5% level. The estimated sizes of the Q' tests were also close to the nominal level of 5%, particularly so for smaller values of q. At larger values of q, the size of the Q' tests tended to exceed the nominal level, but not to a serious extent. Most likely, the insignificance of serial correlations at high lags is again responsible for the negative impact on empirical sizes for the Q' tests for larger values of q.

To gauge the efficiency of Q' over the optimal tests when serial correlations for some relatively large lags are significant, in the second set of Monte Carlo experiments, we consider two multiplicative seasonal models: ARMA(0, 1) $\times$ ARMA$(0, 1)_4$ and ARMA(0, 1) $\times$ ARMA$(0, 1)_{12}$ processes. The results are reported in table 2. As in the case of non-seasonal alternatives, $\mathcal{T}_\mathrm{L}$ and $\mathcal{T}_\mathrm{A}$ outperform the Q' statistic. However, the power of Q' and, hence, its relative efficiency to the optimal tests do not, in general, decline monotonically with q. For example, with a sample size of 200, the power of the 5% Q' test starts at 6.30% when $q = 2$ and the ARMA(0, 1) $\times$ ARMA$(0, 1)_4$ alternative is considered (table 2). The power rises to 55.52% at $q = 5$. As q is increased further, the power declines. Again, this rise and fall in power with q is as anticipated because a test Q' with q at least 4 is desirable to detect the significant serial correlations in the alternative, but increased sampling variation of any additional serial correlations leads to the decline in power. Table 2 reports similar results for an ARMA(0, 1) $\times$ ARMA$(0, 1)_{12}$ alternative. The power of Q' falls first with q and then rises after $q \geq 12$. Overall, the relative efficiency of Q' to optimal tests has the similar rise-and-fall

Table 2. Empirical power of Q' relative to that of $\mathcal{T}_\mathrm{A}$ for seasonal models. (Results are quoted in percentages.)

		Estimated power of nominal 5% test				
n	q	Q'	$\mathcal{T}_\mathrm{L}$	$\mathcal{T}_\mathrm{A}$	$Q'/\mathcal{T}_\mathrm{L}$	$Q'/\mathcal{T}_\mathrm{A}$
ARMA(0, 1) $\times$ $(0, 1)_4$ alternative: $X_t = (1 - \theta B)(1 - \Theta B^4)Z_t$ with $\theta = -0.05$ and $\Theta = 0.1$						
100	2	4.85	7.87	7.87	61.63	61.63
	5	26.27	48.37	48.26	54.31	54.43
	10	22.41	49.58	49.25	45.20	45.50
	20	21.37	49.51	49.54	43.16	43.14
200	2	6.30	10.09	10.31	62.44	61.11
	5	55.52	80.34	80.13	69.11	69.29
	10	43.20	80.82	80.11	53.45	53.93
	20	35.37	81.82	80.91	43.23	43.72
	40	31.17	80.73	79.67	38.61	39.12
400	2	10.80	16.57	16.73	65.18	64.55
	5	88.81	97.97	97.87	90.65	90.74
	10	79.20	98.27	98.00	80.59	80.82
	20	66.38	98.34	97.95	67.50	67.77
	40	53.55	98.31	98.01	54.47	54.64
	80	40.55	98.56	98.21	41.14	41.29
ARMA(0, 1) $\times$ $(0, 1)_{12}$ alternative: $X_t = (1 - \theta B)(1 - \Theta B^{12})Z_t$ with $\theta = -0.05$ and $\Theta = 0.1$						

or fall-and-rise pattern as the power of Q' does, depending primarily on the size of the highest lag of the significant serial correlations relative to the value of q used. Empirical sizes of tests are not reported in table 2 as the null hypothesis is the same as for table 1, so results would be the same as those presented in table 1 up to simulation error.

Overall, the results, for both size and power, suggest that the Q' tests perform best at small to moderate values of q for alternative models of the type considered here. Obviously, for models with serial correlation at relatively high lags, the value of q needs to be selected accordingly to accommodate the significance of sample autocorrelations at high lags.

Results for tests at nominal levels 1% and 10% are qualitatively similar to those reported for 5% tests, and so are not reported here.

3.2 *Portmanteau statistics vis-à-vis optimal tests based on misspecified alternatives*

In reality, the true underlying structure of $\{X_t\}$ is rarely known. From a practical standpoint, a more relevant comparison is how well Q' performs relative to tests based on statistics resulting from use of misspecified alternatives.

To investigate this issue, consider the situation in which the true alternative is the ARMA(1, 1) process with $(\phi, \theta) = (0.65, 0.5)$, but one uses the optimal test based on an AR(1) model, $X_t = \phi^* X_{t-1} + Z_t^*$, denoted as $\mathcal{T}_A^*$ to distinguish it from $\mathcal{T}_A$. This kind of misspecification is common in practice. Since the first-order autocorrelation of the misspecified AR(1) model is given by $(\phi - \theta)(1 - \phi\theta)/(1 + \theta^2 - 2\phi\theta) = 0.16875$, which is positive, it is reasonable to consider $\mathcal{T}_A^*$ based on a positive pre-determined ϕ^*. Negative values of ϕ^* are unlikely to be preferred in practice. We consider this value of ϕ^* as well as the values 0.25, 0.5 and 0.75 to exemplify possible realistic values in its full positive range.

Table 3 reports the power of Q' relative to that of $\mathcal{T}_A^*$ for sample sizes of 100, 200 and 400. As for tables 1 and 2, we present results for 5% tests only. Again, all simulations are based on 10,000 replications.

As far as power is concerned, the test $\mathcal{T}_A$ clearly dominates Q' by substantial margins, especially when q is relatively large. Observe that although the power of $\mathcal{T}_A$ depends on the

Table 3. Empirical power of Q' relative to that of $\mathcal{T}_A^*$ based on misspecified alternatives for nominal 5% tests. (Results are quoted in percentages.)

			$\phi^* = 0.16875$		$\phi^* = 0.25$		$\phi^* = 0.5$		$\phi^* = 0.75$	
n	q	Q'	$\mathcal{T}_A^*$	$Q'/\mathcal{T}_A^*$	$\mathcal{T}_A^*$	$Q'/\mathcal{T}_A^*$	$\mathcal{T}_A^*$	$Q'/\mathcal{T}_A^*$	$\mathcal{T}_A^*$	$Q'/\mathcal{T}_A^*$
100	2	33.27	37.10	89.68	38.79	85.77	41.04	81.07	40.90	81.34
	5	26.33	37.55	70.12	39.77	66.21	43.38	60.70	40.86	64.44
	10	22.64	37.54	60.31	39.77	56.93	43.24	52.36	37.94	59.67
	20	21.25	37.54	56.61	39.77	53.43	43.25	49.13	37.10	57.28
200	2	60.13	64.81	92.78	66.90	89.88	68.96	87.20	68.96	87.20
	5	52.35	65.58	79.83	68.04	76.94	71.93	72.78	69.31	75.53
	10	43.41	65.59	66.18	68.07	63.77	71.94	60.34	66.84	64.95
	20	36.84	65.59	56.17	68.07	54.12	71.90	51.24	66.12	55.72
	40	32.05	65.59	48.86	68.07	47.08	71.90	44.58	66.06	48.52
400	2	89.95	91.87	97.91	92.91	96.81	93.95	95.74	93.58	96.12
	5	84.54	92.25	91.64	93.58	90.34	94.90	89.08	93.00	90.90
	10	76.34	92.25	82.75	93.58	81.58	94.98	80.37	91.87	83.10
	20	66.72	92.25	72.33	93.58	71.30	94.99	70.24	91.68	72.77
	40	56.65	92.25	61.41	93.58	60.54	94.99	59.64	91.67	61.80
	80	49.44	92.25	53.59	93.58	52.83	94.99	52.05	91.67	53.93

The true model is the ARMA(1,1) process: $X_t = 0.65X_{t-1} - 0.5Z_{t-1} + Z_t$. $\mathcal{T}_A^*$ is based on the AR(1) process: $X_t = \phi^* X_{t-1} + Z_t$.

pre-determined AR coefficient in the misspecified AR(1) process, the difference in power of $\mathcal{T}_{\mathrm{A}}^*$s with different ϕ^*'s is small. In fact, the optimal approximate slopes of $\mathcal{T}_{\mathrm{A}}^*$ compare favorably with that of $\mathcal{T}_{\mathrm{A}}$. For example, when $q = 10$, the optimal approximate slopes of $\mathcal{T}_{\mathrm{A}}^*$ for the values of ϕ^*, 0.25, 0.5 and 0.75, are 0.038, 0.047 and 0.048, respectively, each of which is just slightly lower than $\mathcal{C}_{\mathrm{A}}$, 0.049.

The previously mentioned results hold generally for $\mathcal{T}_{\mathrm{A}}^*$ based on other moderately misspecified alternatives, unless the misspecified alternative has a considerably different correlation structure to that of the true underlying process. In such a case, $\mathcal{T}_{\mathrm{A}}^*$ may be insensitive against the alternative in comparison with the Q' statistic.

4. Concluding remarks

A fundamental distinction between Box-Pierce and Ljung-Box portmanteau statistics and optimal tests lies in whether the weighting scheme of the serial correlations in forming the test statistic is exactly known or has to be estimated from data. Our results indicate that the efficiency loss in estimating the weight function can be substantial. In particular, caution must be exercised when q is large relative to the sample size, since the influence of significant serial correlations at low lags may be diluted by serial correlations at high lags which are not significant but subject to sampling variation. The results emphasize dramatically the fact that portmanteau statistics may not achieve a high level of success against many commonly used alternatives and a more sensible testing strategy can often deliver much higher power than portmanteau statistics do. Nevertheless, the portmanteau tests have size properties similar to those of the optimal tests, though the lack of power remains a concern.

Although optimal tests, resulting from use of correctly or even some incorrectly specified alternatives, may have advantages over the portmanteau tests against some alternatives, there are of course other situations (for example, when the form of the alternative is completely vague or when a whole class of models are of interest) in which portmanteau statistics may possess more desirable properties. Nevertheless, the results in this paper suggest that the optimal tests may provide a useful basis for evaluating the power performance of portmanteau statistics. In cases that the loss of efficiency in power of portmanteau statistics is substantial for the most relevant alternative, empirical researchers should consider the use of tests for the specified alternative(s) rather than portmanteau statistics often applied.

References

[1] Box, G. and Pierce, D., 1970, Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, **65**, 1509–1526.

[2] Ljung, G., and Box, G., 1978, On a measure of lack of fit in time series models. *Biometrika*, **65**, 297–303.

[3] Bahadur, R., 1967, Rates of convergence of estimates and test statistics. *Annals of Mathematical Statistics*, **38**, 303–324.

[4] Geweke, J., 1981, The approximate slopes of econometric tests. *Econometrica*, **49**, 1427–1442.

[5] Godfrey, L.G., 1979, Testing the adequacy of a time series model. *Biometrika*, **66**, 67–72.

[6] Davies, N. and Newbold, P., 1979, Some power studies of a portmanteau test of time series model specification. *Biometrika*, **66**, 153–155.

[7] Brockwell, P. and Davis, R., 1991, *Time Series: Theory and Methods* (New York: Springer-Verlag).

[8] Hannan, E. and Heyde, C., 1972, On limit theorems for quadratic functions of discrete time series. *Annals of Mathematical Statistics*, **43**, 2058–2066.

[9] Boshnakov, G., 1989, Bartlett's formulae – closed forms and recurrent equations. *Annals of the Institute of Statistical Mathematics*, **48**, 49–59.

[10] Fang, Y., 2003, GMM tests for the Katz family of distributions. *Journal of Statistical Planning and Inference*, **110**, 55–73.

[11] Scheffè, H., 1959, *The Analysis of Variance* (New York: John Wiley and Sons).