THE EFFECT OF THE ESTIMATION ON GOODNESS-OF-FIT TESTS IN TIME SERIES MODELS

BY YUE FANG

University of Oregon and China Europe International Business School

First Version received June 2003

Abstract. We analyze, by simulation, the finite-sample properties of goodness-of-fit tests based on residual autocorrelation coefficients (simple and partial) obtained using different estimators frequently used in the analysis of autoregressive moving-average time-series models. The estimators considered are unconditional least squares, maximum likelihood and conditional least squares. The results suggest that although the tests based on these estimators are asymptotically equivalent for particular models and parameter values, their sampling properties for samples of the size commonly found in economic applications can differ substantially, because of differences in both finite-sample estimation efficiencies and residual regeneration methods.

Keywords. Autoregressive-moving average model; conditional least squares; maximum likelihood; goodness-of-fit test; partial autocorrelation; residual autocorrelation; unconditional least squares.

1. INTRODUCTION

Suppose that a time series X_t is generated by the stationary and invertible autoregressive moving-average (ARMA) (p, q) process of the form

$$(1 - \phi_1 B - \cdots - \phi_p B^p)X_t = (1 - \theta_1 B - \cdots - \theta_q B^q)a_t, \quad (1)$$

where $B^j X_t = X_{t-j}$ and a_t is zero-mean Gaussian white noise with variance σ_a^2. Several diagnostic goodness-of-fit tests have been proposed based on the residual autocorrelation coefficients (simple or partial). A popular goodness-of-fit test, proposed by Box and Pierce (1970) and improved by Ljung and Box (1978), uses the statistic

$$Q_{\mathrm{LB}} = n(n+2)\sum_{k=1}^{m}(n-k)^{-1}\hat{r}_k^2, \quad (2)$$

where $\hat{r}_k$ is the kth-order sample autocorrelation and n is the number of observations in the series. Under the hypothesis of correct model specification, Q_{LB} has an asymptotic chi-square (χ^2) distribution with ($m - p - q$) degrees of freedom. It is well known that the empirical size of the Q_{LB} test in finite samples could be much different from those given by the asymptotic theory, and that even for moderate sample sizes, Q_{LB} has low power and often fails to detect model misspecification (Davies and Newbold, 1979).

Alternatively, Monti (1994) proposed a goodness-of-fit test based on residual partial autocorrelations. Let $\hat{\psi}_k$ be the kth-order sample residual partial autocorrelation coefficients. Monti showed that

$$Q_{\mathrm{MT}} = n(n+2)\sum_{k=1}^{m}(n-k)^{-1}\hat{\psi}_k^2 \tag{3}$$

is asymptotically chi-square distributed with $(m - p - q)$ degrees of freedom if the model fitted is appropriate. Monti's simulation results indicated that the empirical sizes of Q_{MT} are adequate in moderate sample sizes and the test is more powerful than Q_{LB} when the fitted model understates the order of the moving-average component. However, Kwan and Wu (1997) found that the size of Q_{MT} can be affected considerably by m and the empirical power of Q_{MT} is similar to that of Q_{LB} if m is properly chosen.

In this paper we analyze, by means of simulation, the effect of the choice of estimation procedures on the finite-sample behaviors of Q_{LB} and Q_{MT} tests. The estimators considered are unconditional least squares (ULS), maximum likelihood (ML), and conditional least squares (CLS). The issue is motivated by the fact that although these estimators are asymptotically equivalent, they differ appreciably from one another in finite samples. Consequently, the finite-sample properties of the goodness-of-fit tests which are based on the residual autocorrelation or partial autocorrelation coefficients are likely to depend on how model parameters are estimated. Furthermore, these estimators use different residual regeneration methods, leading to different residual series and hence, to diverse behaviors of residual autocorrelation or partial autocorrelation coefficients.

2. THREE ESTIMATORS

Given n observations $X_1, X_2, \ldots, X_n$ on such a time series by eqn (1) and the further assumption that the innovations a_t are independent and normally distributed, the ML function can be written as follows:

$$-\frac{1}{2\sigma_a^2}\mathrm{X}'\boldsymbol{\Omega}^{-1}\mathrm{X} - \frac{1}{2}\ln(|\boldsymbol{\Omega}|) - \frac{n}{2}\ln(\sigma_a^2), \tag{4}$$

where $\sigma_a^2\boldsymbol{\Omega}$ is the variance of the vector $\mathrm{X} = \{X_1, X_2, \ldots, X_n\}'$ as a function of the ϕ and θ parameters, and $|\cdot|$ denotes the determinant. The ML estimate of σ_a^2 is

$$\frac{1}{n}\mathrm{X}'\boldsymbol{\Omega}^{-1}\mathrm{X}$$

and the log-likelihood concentrated with respect to σ_a^2 can be taken up to additive constants as

$$-\frac{n}{2}\ln(\mathrm{X}'\boldsymbol{\Omega}^{-1}X)-\frac{1}{2}\ln(|\boldsymbol{\Omega}|). \tag{5}$$

For a stationary, invertible Gaussian process given in eqn (1), the ML estimates are consistent, asymptotically normally distributed (Whittle, 1951) and asymptotically efficient (Aigner, 1971). In general, it is difficult to find closed analytical expressions for ML estimators. In this study, the concentrated log-likelihood function (5) is maximized via nonlinear least squares using Marquardt's method.

The ULS method is also referred to as the exact least-squares (ELS) method. For ULS, the estimates minimize

$$\sum_{t=1}^{n}\tilde{a}_t^2=\sum_{t=1}^{n}(X_t-\mathrm{C}_t\mathrm{V}_t^{-1}(X_1,X_2,\ldots,X_{t-1})')^2=\mathrm{X}'\boldsymbol{\Omega}^{-1}\mathrm{X}, \tag{6}$$

where C_t is the covariance matrix of X_t and $(X_1, X_2, \ldots, X_{t-1})$, and V_t is the variance matrix of $(X_1, X_2, \ldots, X_{t-1})$. Therefore, the ULS estimates are obtained by minimizing the sum of squared residuals rather than using the log-likelihood (5) as the criterion function. Note that for large n, the quantity (5) will be dominated by $\mathrm{X}'\boldsymbol{\Omega}^{-1}\mathrm{X}$.

The CLS estimates are conditional on the assumption that the past unobserved errors are equal to their expected value. The series (1) can be represented in terms of the previous observations:

$$X_t=a_t+\sum_{i=1}^{\infty}\pi_i X_{t-i}.$$

The π weights are computed from the ratio of the ϕ and θ polynomials, as follows:

$$\frac{\phi(B)}{\theta(B)}=1-\sum_{i=1}^{\infty}\pi_i B^i.$$

The CLS method produces estimates minimizing

$$\sum_{t=1}^{n}\hat{a}_t^2=\sum_{t=1}^{n}\left(X_t-\sum_{i=1}^{\infty}\hat{\pi}_i X_{t-i}\right)^2, \tag{7}$$

where the unobserved past values of X_t are set to zero. Note that there are n values of a_t in eqn (7), in contrast to that only $n-p$ values of a_t contributes to the sum of squares in the CLS method of residual regeneration described in Section 7.1 by Box *et al.* (1994).

It is well known that these three estimators are asymptotically equivalent. In general, ML estimates are more expensive to compute than ULS and especially, than CLS estimates. However, ML estimates may be preferable in some cases (Ansley and Newbold, 1980).

By contrasting eqns (6) and (7), we can see that the residuals regenerated from the CLS method are based on infinite memory forecasts (also called conditional forecasts), which may be different from those obtained from ULS or ML methods in which finite-memory forecasts (or unconditional forecasts) are used. Although the difference in residual regeneration methods is of less consequence for the estimation, it has serious size and power implications on the tests and will be discussed more completely in Section 3.

3. MONTE CARLO EXPERIMENTS AND SIMULATION RESULTS

An exhaustive simulation study is carried out to explore the finite-sample properties of Q_{LB} and Q_{MT} based on residuals obtained using different estimators. Both the accuracy of the chi-square approximation to the distributions of Q_{LB} and Q_{MT} and the powers of the tests under a number of alternative models are considered. We have chosen model parameterizations of $\{X_t\}$ not only to conform with other previously published studies such as Davies and Newbold (1979) and Koreisha and Fang (1999, 2001), but also to provide a representative set of examples of possible autocorrelated error structures. Without loss of generality, we assume that $\sigma_a^2 = 1$.

Our computations were performed on a Dell Pentium III computer using SAS Window release 8.1 programs. All simulations are based on 10,000 replications and the pseudo-random number generator, *rannor*, is used to generate independent Gaussian deviates. Although we simulate tens of thousands of trials, for brevity, we report results only for tests with the 5% nominal size based on $n = 50$ with $m = 5$, 10 and 20. For illustrative purposes, however, we also provide results on $n = 100$ and 200 for selected models. The general conclusions reached for the results presented here extend to tests with other nominal values and wider class of models with other parameters also.

3.1. The sizes of the tests

Tables I–IV contrast the theoretical and empirical values of the size, mean, standard deviation and skewness of the Q_{LB} and Q_{MT} tests based on three estimators. The tables correspond to the AR(1) and MA(1), AR(2) and MA(2), ARMA(1,1), and ARMA(1,2) and ARMA(2,1) models, respectively, with selected parameter values.

The results show that the behaviours of both Q_{LB} and Q_{MT} depend on the choice of the estimators, in addition to levels of m and model structures and parameterizations. In general, for models with parameters not close to the boundary of the invertibility or stationarity region, the impact of estimators is relatively negligible, as exemplified by cases of the AR(1) process with $\phi_1 = -0.5$ and the MA(1) process with $\theta_1 = 0.5$ (Table I), little differentiation can be made

TABLE I

THEORETICAL AND EMPIRICAL VALUES OF THE SIZE, MEAN, STANDARD DEVIATION, AND SKEWNESS OF THE 5% Q_{LB} AND Q_{MT} TESTS UNDER THE AR(1) OR MA(1) NULL HYPOTHESIS ($n = 50$)

Model	m		Theoretical	Q_{LB}			Q_{MT}		
				ULS	ML	CLS	ULS	ML	CLS
AR(1)	5								
$\phi_1 = -0.5$		Test size	5%	5.19%	5.06%	5.02%	5.71%	5.72%	5.68%
		Mean	4.000	4.117	4.081	4.068	4.259	4.252	4.237
		St. dev.	2.828	2.835	2.818	2.818	2.913	2.915	2.915
		Skewness	1.414	1.511	1.520	1.518	1.369	1.369	1.369
	10								
		Test size	5%	5.78%	5.63%	5.63%	6.01%	5.99%	5.95%
		Mean	9.000	9.064	9.020	9.003	9.430	9.421	9.406
		St. dev.	4.243	4.465	4.442	4.441	4.306	4.307	4.309
		Skewness	0.943	1.305	1.301	1.304	0.906	0.905	0.906
	20								
		Test size	5%	7.20%	6.96%	7.05%	4.36%	4.37%	4.36%
		Mean	19.000	18.970	18.906	18.886	19.223	19.214	19.198
		St. dev.	6.164	7.166	7.128	7.131	5.889	5.891	5.893
		Skewness	0.649	1.203	1.196	1.202	0.510	0.510	0.511
AR(1)	5								
$\phi_1 = 0.9$		Test size	5%	4.67%	4.71%	7.04%	5.11%	5.07%	7.34%
		Mean	4.000	4.078	4.060	4.496	4.164	4.154	4.570
		St. dev.	2.828	2.772	2.765	3.096	2.755	2.755	3.017
		Skewness	1.414	1.631	1.549	1.588	1.307	1.316	1.251
	10								
		Test size	5%	5.36%	4.94%	7.58%	4.36%	4.23%	6.62%
		Mean	9.000	8.893	8.771	9.550	8.984	8.941	9.672
		St. dev.	4.243	4.458	4.406	4.812	4.038	4.050	4.334
		Skewness	0.943	1.399	1.399	1.442	0.855	0.864	0.830
	20								
		Test size	5%	6.44%	5.98%	8.56%	3.10%	2.97%	4.52%
		Mean	19.000	18.643	18.478	19.643	18.489	18.419	19.479
		St. dev.	6.164	7.082	7.032	7.530	5.637	5.680	5.861
		Skewness	0.649	1.356	1.334	1.396	0.508	0.509	0.472
MA(1)	5								
$\theta_1 = 0.5$		Test size	5%	4.92%	4.69%	4.67%	5.37%	4.91%	4.97%
		Mean	4.000	3.998	3.956	3.930	4.118	4.042	4.020
		St. dev.	2.828	2.767	2.762	2.757	2.806	2.769	2.769
		Skewness	1.414	1.523	1.536	1.559	1.305	1.306	1.333
	10								
		Test size	5%	5.50%	5.14%	5.22%	5.60%	5.21%	5.27%
		Mean	9.000	8.864	8.826	8.793	9.176	9.112	9.075
		St. dev.	4.243	4.366	4.356	4.345	4.164	4.145	4.141
		Skewness	0.943	1.311	1.320	1.333	0.842	0.840	0.852
	20								
		Test size	5%	6.52%	6.35%	6.49%	3.96%	3.70%	3.98%
		Mean	19.000	18.713	18.645	18.623	18.959	18.897	18.843
		St. dev.	6.164	6.937	6.928	6.926	5.743	5.735	5.745
		Skewness	0.649	1.195	1.209	1.204	0.465	0.464	0.479
MA(1)	5								
$\theta_1 = -0.9$		Test size	5%	6.35%	6.45%	5.77%	7.14%	6.85%	6.18%
		Mean	4.000	4.486	4.443	4.266	4.681	4.579	4.411
		St. dev.	2.828	2.917	2.974	2.913	2.997	2.977	2.941
		Skewness	1.414	1.485	1.539	1.558	1.284	1.302	1.337

TABLE I

CONTINUED

Model	m		Theoretical	Q_{LB} ULS	Q_{LB} ML	Q_{LB} CLS	Q_{MT} ULS	Q_{MT} ML	Q_{MT} CLS
	10								
		Test size	5%	6.32%	6.79%	6.03%	6.69%	6.35%	5.66%
		Mean	9.000	9.289	9.301	9.056	9.670	9.567	9.323
		St. dev.	4.243	4.527	4.604	4.548	4.306	4.301	4.255
		Skewness	0.943	1.303	1.325	1.345	0.861	0.873	0.900
	20								
		Test size	5%	7.49%	7.82%	7.46%	4.58%	4.51%	4.02%
		Mean	19.000	19.109	19.180	18.866	19.340	19.225	18.915
		St. dev.	6.164	7.247	7.355	7.308	5.903	5.896	5.840
		Skewness	0.649	1.219	1.244	1.244	0.512	0.516	0.520

TABLE II

THEORETICAL AND EMPIRICAL VALUES OF THE SIZE, MEAN, STANDARD DEVIATION, AND SKEWNESS OF THE 5% Q_{LB} AND Q_{MT} TESTS UNDER THE AR(2) OR MA(2) HYPOTHESIS ($n = 50$)

Model	m		Theoretical	Q_{LB} ULS	Q_{LB} ML	Q_{LB} CLS	Q_{MT} ULS	Q_{MT} ML	Q_{MT} CLS
AR(2)	5								
$(\phi_1, \phi_2) =$		Test size	5%	5.42%	4.31%	7.25%	4.27%	3.97%	6.90%
(1.42, −0.73)		Mean	3.000	3.269	3.047	3.424	3.207	3.076	3.395
		St. dev.	2.449	2.430	2.264	2.813	2.221	2.172	2.601
		Skewness	1.633	1.868	1.859	2.371	1.426	1.469	1.883
	10								
		Test size	5%	3.76%	2.88%	5.33%	2.96%	2.84%	5.77%
		Mean	8.000	7.344	7.021	7.548	7.491	7.371	7.933
		St. dev.	4.000	3.845	3.645	4.342	3.562	3.553	4.162
		Skewness	1.000	1.569	1.522	2.153	0.971	0.990	1.285
	20								
		Test size	5%	3.52%	3.00%	5.09%	2.54%	2.44%	3.08%
		Mean	18.000	15.833	15.408	16.188	16.031	15.925	16.625
		St. dev.	6.000	6.143	5.908	6.732	5.151	5.166	5.576
		Skewness	0.667	1.354	1.283	1.789	0.881	0.881	0.702
AR(2)	5								
$(\phi_1, \phi_2) =$		Test size	5%	5.34%	4.37%	10.67%	4.66%	4.21%	10.41%
(1.60, −0.64)		Mean	3.000	3.199	3.027	4.151	3.194	3.088	4.110
		St. dev.	2.449	2.392	2.242	2.949	2.265	2.215	2.816
		Skewness	1.633	2.024	1.989	1.614	1.495	1.551	1.325
	10								
		Test size	5%	4.13%	4.05%	8.86%	3.46%	3.38%	8.27%
		Mean	8.000	6.710	6.528	9.067	6.821	6.725	9.201
		St. dev.	4.000	3.844	3.654	4.581	3.531	3.526	4.224
		Skewness	1.000	1.559	1.526	1.338	0.998	1.025	0.840
	20								
		Test size	5%	3.91%	3.82%	8.97%	3.15%	3.18%	5.38%
		Mean	18.000	14.794	14.456	19.001	15.902	15.809	19.079
		St. dev.	6.000	6.548	6.330	7.249	5.717	5.747	5.997
		Skewness	0.667	1.192	1.147	1.235	0.727	0.754	0.471

TABLE II

CONTINUED

Model	m		Theoretical	Q_{LB} ULS	Q_{LB} ML	Q_{LB} CLS	Q_{MT} ULS	Q_{MT} ML	Q_{MT} CLS
MA(2)	5								
$(\theta_1, \theta_2) =$		Test size	5%	6.29%	6.11%	4.96%	6.98%	5.86%	4.94%
(1.42, −0.73)		Mean	3.000	3.551	3.373	3.048	3.671	3.390	3.078
		St. dev.	2.449	2.431	2.424	2.360	2.502	2.377	2.357
		Skewness	1.633	1.647	1.892	1.861	1.497	1.643	1.732
	10								
		Test size	5%	4.93%	4.27%	3.77%	4.27%	3.79%	3.36%
		Mean	8.000	7.475	7.375	7.034	7.927	7.604	7.255
		St. dev.	4.000	3.717	3.799	3.775	3.705	3.623	3.619
		Skewness	1.000	1.554	1.751	1.617	0.959	1.014	1.075
	20								
		Test size	5%	3.86%	4.31%	3.81%	2.56%	2.38%	1.77%
		Mean	18.000	15.856	15.855	15.389	16.547	16.233	15.683
		St. dev.	6.000	6.023	6.144	6.149	5.224	5.166	5.144
		Skewness	0.667	1.323	1.346	1.424	0.584	0.597	0.608
MA(2)	5								
$(\theta_1, \theta_2) =$		Test size	5%	3.60%	3.27%	2.93%	4.73%	4.16%	3.98%
(0.3, −0.5)		Mean	3.000	2.850	2.763	2.710	3.071	2.907	2.857
		St. dev.	2.449	2.107	2.080	2.054	2.336	2.231	2.219
		Skewness	1.633	1.352	1.446	1.404	1.340	1.382	1.386
	10								
		Test size	5%	2.84%	2.70%	2.59%	3.96%	3.58%	3.37%
		Mean	8.000	7.017	7.025	6.958	7.652	7.520	7.468
		St. dev.	4.000	3.428	3.481	3.437	3.678	3.638	3.628
		Skewness	1.000	1.090	1.149	1.087	0.877	0.890	0.886
	20								
		Test size	5%	3.13%	3.04%	2.90%	2.59%	2.20%	2.19%
		Mean	18.000	15.679	15.734	15.661	16.634	16.518	16.460
		St. dev.	6.000	5.641	5.723	5.667	5.220	5.194	5.195
		Skewness	0.667	1.028	1.072	1.003	0.522	0.524	0.521

between estimators. In these two simple cases, regardless of the choice of the estimators, the size of Q_{LB} is close to the nominal level if m is chosen to be relatively small but may be well above the nominal level for relatively large m and appears to increase monotonically as m increases. The size of Q_{MT} first increases slightly above and then falls back below the nominal level as m goes from 5 to 20. The empirical mean and standard deviations of both Q_{LB} and Q_{MT} are close to those of the theoretical values based on the asymptotic chi-square distribution. The empirical skewness of Q_{LB} (Q_{MT}) exhibits, however, a persistent tendency to be larger (smaller) than its theoretical counterpart. We note that the empirical skewness of Q_{LB} can be considerably larger than the theoretical value if m is relatively large, while the empirical skewness of Q_{MT} has somewhat better agreement with that of chi-square distribution across all levels of m examined.

More substantial differences emerge in models with parameters close to the boundary of the invertibility or stationarity region. For example, consider the AR(2) process with $(\phi_1, \phi_2) = (1.60, -0.64)$ (Table II) and the ARMA(2,1) process with $(\phi_1, \phi_2, \theta_1) = (1.40, -0.60, -0.80)$ (Table IV). In both cases, tests

TABLE III

THEORETICAL AND EMPIRICAL VALUES OF THE SIZE, MEAN, STANDARD DEVIATION, AND SKEWNESS OF THE 5% Q_{LB} AND Q_{MT} TESTS UNDER THE ARMA(1,1) NULL HYPOTHESIS

Model	m		Theoretical	Q_{LB} ULS	Q_{LB} ML	Q_{LB} CLS	Q_{MT} ULS	Q_{MT} ML	Q_{MT} CLS
ARMA(1,1)	$n=50$								
$(\phi_1, \theta_1) =$	5								
(0.8, 0.7)		Test size	5%	6.51%	5.78%	6.18%	6.81%	6.12%	6.04%
		Mean	3.000	3.285	3.252	3.253	3.347	3.320	3.321
		St. dev.	2.449	2.497	2.471	2.480	2.539	2.525	2.533
		Skewness	1.633	1.607	1.590	1.598	1.534	1.544	1.553
	10								
		Test size	5%	6.10%	5.42%	5.80%	6.53%	5.46%	6.33%
		Mean	8.000	8.015	7.969	7.971	8.349	8.315	8.314
		St. dev.	4.000	4.043	4.013	4.023	4.012	3.999	4.008
		Skewness	1.000	1.230	1.217	1.222	0.877	0.878	0.886
	20								
		Test size	5%	6.02%	5.36%	5.74%	5.48%	4.92%	5.38%
		Mean	18.000	17.643	17.581	17.575	18.127	18.089	18.073
		St. dev.	6.000	6.489	6.461	6.458	5.654	5.655	5.646
		Skewness	0.667	1.060	1.054	1.053	0.487	0.487	0.494
ARMA(1,1)	$n=50$								
$(\phi_1, \theta_1) =$	5								
(− 0.8, 0.7)		Test size	5%	4.65%	4.25%	3.52%	5.72%	5.07%	4.27%
		Mean	3.000	3.308	3.142	2.837	3.399	3.254	2.938
		St. dev.	2.449	2.263	2.202	2.158	2.343	2.304	2.258
		Skewness	1.633	1.473	1.531	1.573	1.415	1.460	1.509
	10								
		Test size	5%	4.03%	3.72%	3.36%	3.96%	3.73%	3.35%
		Mean	8.000	7.612	7.449	7.072	7.867	7.742	7.317
		St. dev.	4.000	3.759	3.716	3.729	3.714	3.699	3.684
		Skewness	1.000	1.227	1.211	1.253	0.964	0.963	0.977
	20								
		Test size	5%	4.59%	4.31%	4.19%	2.84%	2.72%	2.07%
		Mean	18.000	16.822	16.561	15.993	17.145	17.022	16.254
		St. dev.	6.000	6.310	6.211	6.273	5.454	5.446	5.461
		Skewness	0.667	1.122	1.087	1.121	0.557	0.555	0.561
	$n=100$								
	5								
		Test size	5%	5.40%	4.76%	4.33%	5.52%	5.18%	4.71%
		Mean	3.000	3.240	3.130	2.933	3.280	3.184	2.988
		St. dev.	2.449	2.405	2.323	2.293	2.430	2.375	2.349
		Skewness	1.633	1.612	1.587	1.643	1.516	1.526	1.618
	10								
		Test size	5%	4.16%	4.01%	3.65%	4.12%	3.98%	3.53%
		Mean	8.000	7.713	7.614	7.374	7.833	7.754	7.501
		St. dev.	4.000	3.887	3.829	3.832	3.807	3.775	3.771
		Skewness	1.000	1.295	1.219	1.238	0.992	0.985	1.027
	20								
		Test size	5%	4.84%	4.53%	4.25%	3.85%	3.82%	3.36%
		Mean	18.000	16.982	16.862	16.496	17.310	17.242	16.816
		St. dev.	6.000	6.244	6.180	6.211	5.763	5.743	5.743
		Skewness	0.667	1.108	1.100	1.109	0.702	0.695	0.709

TABLE III

CONTINUED

Model	m		Theoretical	Q_{LB} ULS	Q_{LB} ML	Q_{LB} CLS	Q_{MT} ULS	Q_{MT} ML	Q_{MT} CLS
	$n = 200$								
	5								
		Test size	5%	4.73%	4.64%	4.48%	5.12%	5.06%	4.78%
		Mean	3.000	3.160	3.130	3.025	3.193	3.165	3.058
		St. dev.	2.449	2.385	2.372	2.348	2.415	2.407	2.378
		Skewness	1.633	1.593	1.599	1.637	1.573	1.580	1.611
	10								
		Test size	5%	4.55%	4.45%	4.29%	4.64%	4.64%	4.52%
		Mean	8.000	7.844	7.811	7.662	7.922	7.899	7.745
		St. dev.	4.000	3.905	3.897	3.885	3.906	3.902	3.878
		Skewness	1.000	1.069	1.073	1.100	0.999	1.001	1.009
	20								
		Test size	5%	4.54%	4.45%	4.36%	4.25%	4.21%	3.89%
		Mean	18.000	17.355	17.315	17.051	17.582	17.563	17.273
		St. dev.	6.000	6.028	6.018	6.017	5.832	5.831	5.800
		Skewness	0.667	0.916	0.915	0.980	0.667	0.667	0.683

TABLE IV

THEORETICAL AND EMPIRICAL VALUES OF THE SIZE, MEAN, STANDARD DEVIATION, AND SKEWNESS OF THE 5% Q_{LB} AND Q_{MT} TESTS UNDER THE ARMA(1,2) OR ARMA(2,1) NULL HYPOTHESIS ($n = 50$)

Model	m		Theoretical	Q_{LB} ULS	Q_{LB} ML	Q_{LB} CLS	Q_{MT} ULS	Q_{MT} ML	Q_{MT} CLS
ARMA(1,2)	5								
$(\phi, \theta_1, \theta_2) =$		Test size	5%	8.99%	7.54%	4.04%	10.85%	8.23%	4.91%
$(-0.8, 1.4, -0.6)$		Mean	2.000	3.301	3.068	1.946	3.405	3.153	2.017
		St. dev.	2.000	2.437	2.288	1.949	2.608	2.467	2.153
		Skewness	2.000	1.862	1.970	2.616	2.266	2.506	3.335
	10								
		Test size	5%	4.71%	3.75%	2.24%	4.79%	3.90%	2.33%
		Mean	7.000	6.892	6.626	5.438	7.172	6.873	5.635
		St. dev.	3.742	3.701	3.554	3.396	3.725	3.586	3.470
		Skewness	1.069	1.557	1.582	1.686	1.394	1.443	1.648
	20								
		Test size	5%	3.66%	3.20%	2.30%	2.72%	2.36%	1.34%
		Mean	17.000	14.951	14.596	12.889	15.465	15.190	13.200
		St. dev.	5.831	6.140	5.948	5.870	5.594	5.467	5.330
		Skewness	0.686	1.167	1.203	1.327	0.694	0.709	0.884
ARMA(1,2)	5								
$(\phi, \theta_1, \theta_2) =$		Test size	5%	8.15%	7.99%	8.21%	9.53%	8.84%	9.33%
$(0.3, -0.5, 0.3)$		Mean	2.000	3.193	3.068	2.966	3.321	3.152	3.056
		St. dev.	2.000	2.493	2.421	2.430	2.708	2.572	2.595
		Skewness	2.000	1.650	1.638	1.746	1.933	1.889	2.011
	10								
		Test size	5%	7.36%	7.50%	6.40%	8.39%	8.08%	7.71%
		Mean	7.000	7.777	7.636	7.514	8.268	8.029	7.907
		St. dev.	3.742	3.956	3.909	3.928	4.175	4.066	4.078
		Skewness	1.069	1.226	1.208	1.242	1.085	1.075	1.098

TABLE IV

CONTINUED

Model	m		Theoretical	Q_{LB} ULS	Q_{LB} ML	Q_{LB} CLS	Q_{MT} ULS	Q_{MT} ML	Q_{MT} CLS
	20								
		Test size	5%	6.77%	7.30%	6.06%	5.86%	5.99%	5.10%
		Mean	17.000	17.200	17.052	16.890	17.891	17.604	17.463
		St. dev.	5.831	6.442	6.372	6.408	5.774	5.683	5.699
		Skewness	0.686	1.052	1.045	1.050	0.582	0.575	0.585
ARMA(2,1)	5								
$(\phi_1, \phi_2, \theta_1) =$		Test size	5%	8.43%	6.44%	14.49%	8.83%	6.95%	16.30%
(1.4, −0.6, −0.8)		Mean	2.000	3.062	2.783	4.008	3.093	2.846	4.218
		St. dev.	2.000	2.361	2.080	2.987	2.333	2.120	3.141
		Skewness	2.000	2.180	1.827	1.671	1.764	1.723	1.437
	10								
		Test size	5%	5.31%	4.36%	11.65%	5.94%	4.96%	14.11%
		Mean	7.000	6.862	6.535	8.700	7.174	6.889	9.708
		St. dev.	3.742	3.659	3.545	4.560	3.639	3.465	4.694
		Skewness	1.069	1.824	1.349	1.468	1.077	1.065	0.854
	20								
		Test size	5%	4.55%	4.08%	10.26%	4.82%	4.47%	10.00%
		Mean	17.000	14.959	14.516	18.595	15.660	15.307	19.502
		St. dev.	5.831	6.028	5.696	7.160	5.563	5.434	6.149
		Skewness	0.686	1.293	1.071	1.229	0.593	0.591	0.526
ARMA(2,1)	5								
$(\phi_1, \phi_2, \theta_1) =$		Test size	5%	10.88%	10.10%	10.48%	12.29%	11.73%	12.12%
(0.3, −0.5, 0.3)		Mean	2.000	2.784	2.708	2.689	2.908	2.833	2.818
		St. dev.	2.000	2.592	2.512	2.550	2.864	2.741	2.807
		Skewness	2.000	2.537	2.414	2.493	3.196	2.778	3.053
	10								
		Test size	5%	7.39%	6.87%	7.12%	9.87%	8.33%	9.59%
		Mean	7.000	7.352	7.255	7.227	7.783	7.700	7.663
		St. dev.	3.742	4.050	3.946	4.004	4.235	4.145	4.182
		Skewness	1.069	1.714	1.576	1.690	1.471	1.312	1.414
	20								
		Test size	5%	7.06%	6.43%	6.74%	6.51%	6.20%	6.41%
		Mean	17.000	16.743	16.619	16.584	17.489	17.413	17.349
		St. dev.	5.831	6.459	6.315	6.417	5.841	5.782	5.812
		Skewness	0.686	1.371	1.216	1.396	0.751	0.678	0.715

based on either ULS or ML estimator perform generally quite well for appropriately selected values of m. We note that some differences in sizes of the tests based on ULS and ML estimators are noticeable. This can be partly explained by the fact that in such cases, it becomes much more difficult to obtain accurate estimates. Consequently, the properties of test statistics are sensitive to the estimation methods which are subject to more tangible variations.

On the contrary, tests based on CLS do not appear to be reliable in the circumstances that model parameters are close to the boundary of the invertibility or stationarity region. As the cases of the AR(2) process with $(\phi_1, \phi_2) = (1.60, -0.64)$ and the ARMA(2,1) process with $(\phi_1, \phi_2, \theta_1) = (1.40, -0.60, -0.80)$ indicate, tests based on CLS tend to reject the null hypothesis too frequently regardless of the values of m, except in the cases of Q_{MT} with relatively large m for

TABLE V

ESTIMATED COECIENT BIASES AND MEAN SQUARED ERRORS FOR ARMA(1,1) MODELS ($n = 50$)

Bias $\hat{\phi}_1$			MSE $\hat{\phi}_1$			Bias $\hat{\theta}_1$			MSE $\hat{\theta}_1$		
ULS	ML	CLS	ULS	ML	CLS	ULS	ML	CLS	ULS	ML	CLS
(a) $(\phi_1, \theta_1) = (-0.8, 0.7)$											
0.040	0.039	0.039	0.013	0.012	0.012	0.137	0.108	−0.040	0.046	0.038	0.034
(b) $(\phi_1, \theta_1) = (0.8, 0.7)$											
−0.653	−0.653	−0.655	0.813	0.798	0.803	−0.631	−0.629	−0.631	0.784	0.767	0.768

the AR(2) process. In contrast to the earlier cases [i.e. the AR(1) process with $\phi_1 = -0.5$ and the MA(1) process with $\theta_1 = 0.5$], the empirical mean, standard deviation and skewness of the two tests are sensitive to the choice of the estimators.

It is interesting to note that the imprecision of parameter estimates does not necessarily imply a poor approximation of the test statistics to the chi-square distribution, as demonstrated in the case of the ARMA(1,1) process with $\phi_1 = -0.8$ and $\theta_1 = 0.7$ (Table III), where ML estimates are preferable (see the biases and mean-squared errors of the estimators in Table Va) but both Q_{LB} and Q_{MT} based on ULS perform better in terms of the empirical size for $n = 50$ and all m values considered. We note that as the sample size n increases, the differences appear to diminish by contrasting the tests based on ULS and ML estimators for various sample sizes ranging from 50 up to 200 in Table III.

This point is further reinforced by the ARMA(1,1) process with $\phi_1 = 0.8$ and $\theta_1 = 0.7$ in Table III, where the autoregressive and moving-average operators are nearly cancelled out, so that there is near parameter redundancy in the model. When there is near cancellation in the autoregressive and moving-average operators, all three estimators tend to be severely biased towards zero, with unacceptably large mean-squared errors (Table Vb). However, both Q_{LB} and Q_{MT} generally perform rather well in terms of size (Table III). Note that the approximation of $\hat{r}_k(\hat{\psi}_k)$ to $r_k(\psi_k)$ depends on parameter estimates via mainly the term $X(\hat{\beta} - \beta)$, where $\beta = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)$ and X is an $m \times (p + q)$ matrix with entries given by eqn (16) in McLeod (1978). In the ARMA(1,1) case:

$$X = \begin{pmatrix} -1 & 1 \\ -\phi_1 & \theta_1 \\ \vdots & \vdots \\ -\phi_1^{m-1} & \theta_1^{m-1} \end{pmatrix}.$$

If ϕ_1 is close to θ_1, although one experiences difficulties in obtaining accurate parameter estimates, the magnitude of the product of X and $(\hat{\beta} - \beta)$ may not be any great cause for concern.

Table V indicates that the differences in estimation efficiencies between CLS and ULS (or ML) may be small. This suggests that the substantial differences in the empirical sizes of the tests between CLS and ULS (or ML) reported earlier can not be wholly due to the difference in estimation efficiencies. Recall from

Section 2, CLS adopts a residual regeneration method based on AR representation with infinite memory. On the contrary, ULS and ML use finite-memory forecasts to regenerate residuals. The difference in residual regeneration methods may have an effect in the performance of those test statistics. To examine this issue, Table VI reports the sizes of both Q_{LB} and Q_{MT} based on the true model parameter values (after restoring the full degrees of freedom for assessing the statistics) for two ARMA(1,1) processes. In order to compare the results with those based on estimated model parameters, we also present (in parentheses) the corresponding empirical sizes of tests based on the estimated model parameters.

By eliminating the estimation aspect, it becomes evident from Table VI that if the π weights of AR representation decay very quickly as lag increases [as in the case of the ARMA(1,1) process with $\phi_1 = 0.8$ and $\theta_1 = 0.7$ in Table VIb], the sizes of the tests based on CLS are very similar to those based on ULS or ML (ULS and ML use the same residual regeneration method and hence yield the identical residuals when the true parameter values are used). With respect to test

TABLE VII

EMPIRICAL POWERS OF THE 5% Q_{LB} AND Q_{MT} TESTS

Actual model	m	Q_{LB} ULS	Q_{LB} ML	Q_{LB} CLS	Q_{MT} ULS	Q_{MT} ML	Q_{MT} CLS
(a) Fitting AR(1) model							
MA(1) $\theta_1 = 0.5$	$n = 50$						
	5	19.40	19.12	19.11	24.53	24.50	24.53
	10	16.44	16.30	16.29	17.82	17.81	17.78
	20	17.14	16.84	16.86	11.45	11.41	11.43
	$n = 100$						
	5	35.30	35.15	35.10	42.88	42.91	42.85
	10	26.06	25.86	25.88	30.59	30.59	30.57
	20	22.60	22.44	22.47	20.38	20.38	20.37
	$n = 200$						
	5	68.39	68.27	68.24	74.68	74.69	74.67
	10	51.61	51.47	51.42	59.30	59.33	59.33
	20	38.61	38.52	38.48	41.78	41.77	41.77
MA(2)	$n = 50$						
$(\theta_1, \theta_2) = (1.42, -0.73)$	5	96.55	96.22	95.99	99.09	99.02	98.95
	10	84.36	83.21	83.11	95.09	95.00	94.85
	20	70.99	69.87	69.87	80.79	80.81	80.72
ARMA(1,1)	$n = 50$						
$(\phi_1, \theta_1) = (-0.8, 0.7)$	5	68.67	68.25	67.62	85.63	85.88	85.47
	10	55.17	53.67	53.79	71.95	71.80	71.27
	20	48.29	47.01	47.34	49.92	49.74	49.41
ARMA(2,1)	$n = 50$						
$(\phi_1, \phi_2, \theta_1) = (0.3, -0.5, 0.3)$	5	82.41	82.41	82.40	83.96	83.95	83.96
	10	71.13	71.14	71.14	68.42	68.43	68.42
	20	61.18	61.25	61.22	44.04	44.04	44.04
(b) Fitting MA(1) model							
AR(1) $\phi_1 = 0.5$	$n = 50$						
	5	15.06	15.37	15.19	16.68	16.65	16.61
	10	13.45	13.61	13.56	11.98	12.02	12.06
	20	14.12	14.41	14.52	7.84	7.87	7.82
AR(2)	$n = 50$						
$(\phi_1, \phi_2) = (0.3, -0.5)$	5	49.87	51.15	51.23	62.81	62.54	62.62
	10	36.76	37.96	38.02	42.90	42.75	42.75
	20	31.49	32.58	32.61	23.50	23.40	23.46
	$n = 100$						
	5	89.51	89.95	89.96	93.83	93.81	93.77
	10	76.94	77.48	77.52	84.00	83.95	83.94
	20	63.82	64.55	64.56	67.10	67.01	66.98
	$n = 200$						
	5	99.91	99.92	99.92	99.95	99.95	99.95
	10	99.35	99.35	99.35	99.70	99.70	99.70
	20	96.60	96.72	96.73	98.11	98.11	98.11
ARMA(1,1)	$n = 50$						
$(\phi_1, \theta_1) = (0.8, 0.7)$	5	5.74	5.81	5.79	6.00	5.92	5.97
	10	5.54	5.58	5.54	5.77	5.76	5.69
	20	6.83	6.85	6.82	4.17	4.15	4.18
ARMA(2,1)	$n = 50$						
$(\phi_1, \theta_1, \theta_2) = (0.3, -0.5, 0.3)$	5	73.85	75.12	75.09	83.59	83.48	83.40
	10	58.69	60.20	60.33	69.18	68.99	68.93
	20	50.88	52.01	52.49	49.10	48.76	49.14

Note: All Values are Percentages.

The results suggest that the choice of the criterion function used in ULS and ML estimators [namely, eqn (5) or (6)] has a little effect on the asymptotic approximations of Q_{LB} and Q_{MT} to the chi-square distribution, especially when the model parameters are not close to the boundary of the invertibility or stationarity region. Further analysis also reveals that the substantial differences of empirical sizes of tests between CLS and ULS (or ML) are due to differences in both estimation efficiencies and residual regeneration methods, and could be primarily due to the difference in residual regeneration methods if the π weights of AR representation of the process decay relatively slowly as lag increases.

3.2. *The power of the tests*

Table VII reports the empirical powers of Q_{LB} and Q_{MT} when, erroneously, an AR(1) or a MA(1) model is fitted to the data. Four alternative ARMA processes are considered in each case. In general, there do not appear to have vast differences between powers of tests based on three estimators.

The evidence in Table VII suggests that both Q_{LB} and Q_{MT} have reasonable powers against a wide range of alternatives, excluding the case of the ARMA(1,1) process with $\phi_1 = 0.8$ and $\theta_1 = 0.7$ (in Table VIIb). In this case, regardless of the choice of m and estimation methods, both Q_{LB} and Q_{MT} do not distinguish between the ARMA(1,1) and MA(1) because this ARMA(1,1) process is close to the white noise due to the near cancellation in the autoregressive and moving average operators.

For various alternatives examined here, it is clear that Q_{MT} is generally more powerful than Q_{LB} if m is relatively small, while Q_{LB} may be preferable for large values of m. On the whole, the powers of the tests appear to improve significantly as n increases but decline as m increases for fixed n.

Finally, Monti (1994) suggested that when the fitted model underestimates the order of the moving average component, Q_{MT} is more powerful than Q_{LB}. The results from Table VII, however, present rather a mixed picture. The powers of Q_{LB} and Q_{MT} depend greatly on the value of m (as pointed out by Kwan and Wu, 1997) and the estimator.

4. CONCLUDING REMARKS

Although ULS, ML and CLS estimators are asymptotically equivalent, it is found that, for particular models and parameter values in the null and alternative hypotheses, the size and power of tests based on these estimators can differ substantially. The implication of the results is that the choice of estimators will matter in finite samples whether one has a sensible goodness-of-fit testing strategy.

In general, we would recommend that CLS not be used, especially in situations involving processes that the π weights of AR representation decay relatively slowly as lag increases. Tests based on either ULS or ML generally yield reliable inferences. It is well known that the ML estimator is generally preferable in terms of the bias and mean squares error in parameter estimates. However, our results show that it does not necessarily imply that tests based on ML should perform best. At least there seems to be no justification that tests based on ML are markedly superior to tests based on ULS.

ACKNOWLEDGEMENTS

The author would like to thank an anonymous referee for helpful comments leading to a significantly improved article.

NOTE

Corresponding author: Yue Fang, Lundquist College of Business, University of Oregon, Eugene, OR 97403, USA. Tel.: (541) 346-3265; E-mail: yfang@darkwing.uoregon.edu

REFERENCES

Aigner, D. J. (1971) A compendium on estimation of the autoregressive-moving average model from time series data. *International Economic Review* 12, 348–71.

Ansley, C. F. and Newbold, P. (1980) Finite sample properties of estimators for autoregressive moving average models. *Journal of Econometrics* 13, 159–83.

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, 1509–26.

Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (1994) *Time Series Analysis*. Princeton, NJ: Prentice-Hall, Inc.

Davies, N. and Newbold, P. (1979) Some power studies of a portmanteau test of time series model specification. *Biometrika* 66, 153–5.

Koreisha, S. and Fang, Y. (1999) The impact of measurement errors on ARMA prediction. *Journal of Forecasting* 18, 95–109.

Koreisha, S. and Fang, Y. (2001) GLS estimation of regression models with misspecified serial correlation structures. *Journal of the Royal Statistical Society, Series* B 63, 515–31.

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, 733–6.

Ljung, G. M. and Box, G. E. P. (1978) On a measure of lack of fit in time series models. *Biometrika* 65, 67–72.

McLeod, A. I. (1978) On the distribution of residual autocorrelations in Box–Jenkins models. *Journal of the Royal Statistical Society, Series* B 40, 296–302.

Monti, A. C. (1994) A proposal for a residual autocorrelation test in linear models. *Biometrika* 81, 776–80.

Whittle, P. (1951) *Hypothesis Testing in Time Series*. Uppsala: Almqvist and Wiksells.