PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 15 (1995), pp. 227–242

INFERENCE FOR MA(1) PROCESSES WITH A ROOT ON OR NEAR THE UNIT CIRCLE

BY

RICHARD A. DAVIS*, MEICHING CHEN* (FORT COLLINS, COLORADO)
AND
WILLIAM T. M. DUNSMUIR (SYDNEY)

Abstract. This paper considers maximum likelihood estimation (MLE) for MA(1) processes when the moving average parameter is on or near the unit circle. The asymptotic theory to be presented allows the use of the generalized likelihood ratio test for testing the null hypothesis of a unit root. The asymptotic distributions of the MLE and the largest local maximizer, the estimator which yields the local maximum closest to the unit circle, are shown to be different. The limit distributions of two estimates provide a very accurate approximation to the finite sample size and power of the tests considered. A comparison is made of the power of four tests of the null hypothesis that the moving average parameter is equal to one versus the alternative that it is less than one. The four tests are based on the MLE, the largest local maximizer, the generalized likelihood ratio test and Tanaka's score type test. The use of the generalized likelihood ratio test is recommended overall since it always dominates the tests based on the MLE and the largest local maximizer and dominates the score type test for close alternatives to the null hypothesis. For alternatives very close to the unit circle the score type test has slightly higher power but this is evident only in the third decimal place.

1. Introduction. In this paper we consider the moving average of order one, or MA(1) for short, generated by

$$Y_t = \varepsilon_t - \theta_0 \varepsilon_{t-1}, \tag{1.1}$$

where $\{\varepsilon_t\} \sim \mathrm{IID}(0, \sigma^2)$ ($\{\varepsilon_t\}$ is a sequence of independent and identically distributed random variables with mean 0 and variance σ^2) with $\mathrm{E}\varepsilon_t^4 < \infty$ and $|\theta_0| \leqslant 1$. We will consider cases in which θ_0 is at or close to 1. Davis and Dunsmuir [4] review the applications in which inference about θ_0 close to or on the unit circle is likely to arise.

* This research supported in part by NSF DMS Grant No. 9100392.

In cases where $|\theta_0| = 1$ the standard asymptotic normal distribution theory (see, e.g., [1]) does not apply. Also when $|\theta_0|$ is near to unity, this standard theory gives a very poor approximation (see [4] for illustration of this). A much better approximation is found by deriving the asymptotic distribution of the MLE for a sequence of true parameter values which converge to the unit circle at rate $1/T$.

We use the parameterization $\theta_T = 1-\beta/T$, where $\beta \geqslant 0$ and T is the sample size. Inference about β, and hence θ, will be based on the observations $Y_1, \ldots, Y_T$ which come from model (1.1) with true parameter $\theta_0 = 1-\gamma/T$, where $\gamma \geqslant 0$.

This paper gives a rigorous derivation of the convergence in distribution of the maximum likelihood estimator $\hat{\theta}_{\mathrm{MLE}}$, the value of θ which maximizes the likelihood over the interval $\theta \in [-1, 1]$. In [4] the asymptotic distribution for the *local maximum* estimate $\hat{\theta}_{\mathrm{LM}}$, defined as the local maximizer of the likelihood closest to $\theta = \pm 1$, is established. Somewhat surprisingly, these two estimators are not equivalent asymptotically.

The main idea in our derivation is to show that the sequence of processes given by $L_T(\beta) - L_T(0) = l_T(1-\beta/T) - l_T(1)$ converges in distribution on $C[0, \infty)$ to a process $Z_\gamma(\beta)$, where $l_T(\theta)$ is the log-likelihood of the observations $Y_1, \ldots, Y_T$. By this result, we show that the maximizer of $L_T(\cdot)$ converges in distribution to the maximizer of $Z_\gamma(\cdot)$, which then furnishes the desired limit distribution of the MLE. These results are presented in Section 2 with the details of the arguments relegated to the appendix.

Since "the distribution of the MLE of θ under H_0: $\theta = 1$ is unknown even asymptotically," Tanaka [10] was "reluctant to use such tests as likelihood ratio tests or Wald tests." Saikkonen and Luukkonen [9] echo this view and state "the asymptotic distribution of the maximum likelihood estimator of a non-invertible moving average parameter is not known... The development of likelihood ratio tests and Wald tests is therefore intractable." As a consequence these authors propose tests, such as "score type" tests and Lagrange multiplier tests, which only require estimation of the model under the null hypothesis or at a fixed alternative value. The asymptotic results of Section 2 can now be used to develop a test of H_0: $\theta = 1$ based on $\hat{\theta}_{\mathrm{LM}}$, $\hat{\theta}_{\mathrm{MLE}}$, or the generalized likelihood ratio (cutoff values are given in Section 3). Interestingly, the asymptotic operating characteristics of these three tests are different. The test based on the MLE is dominated by the test based on the largest local maximizer which in turn is dominated by the GLR (generalized likelihood ratio) test. In Section 4, the GLR test is compared with Tanaka's LBIU (locally best invariant and unbiased) score type test. While the LBIU test has a slight edge (in the third decimal place) in power for θ's very close to 1 (i.e. $\gamma < 5$), the GLR test dominates the LBIU test by a wide margin for all γ's greater than 5. The desirable operating characteristics of the GLR test, together with its ease of implementation (most time series analysis software packages compute the value of the likelihood for any θ) make the GLR test a desirable test statistic for this problem.

In a more typical situation, Chant [3] considers testing a null hypothesis for a parameter on the boundary of the parameter space when there are nuisance parameters (here σ^2) and following on from Moran [6] demonstrates that when the null hypothesis is simple and concerns a scalar parameter, the maximum likelihood test and optimal $C(\alpha)$ test, as considered by Neyman [7], are equivalent. The use of optimal $C(\alpha)$ tests for testing the null hypothesis that $\theta = 1$ in the present context, while possibly appealing, is not possible since the test statistic is always equal to zero. The essential reason that we cannot derive this test here is that the derivative of the concentrated likelihood is zero under the null hypothesis. For exactly the same reason, the standard score test cannot be applied which led Tanaka to consider a "score type" test based on the second derivative of the log-likelihood. Hence in this moving average problem, the asymptotic equivalence of likelihood tests and related alternative type tests no longer holds.

2. Asymptotic distribution of the MLE. Theorem 2.1 of [4] describes the joint limiting behavior of $L'_T(\beta)$ and $L''_T(\beta)$ and establishes that

(i) $(L'_T(\beta), L''_T(\beta)) \xrightarrow{d} ((\beta/2)Y_\gamma(\beta), (\beta/2)Y'_\gamma(\beta) + \frac{1}{2}Y_\gamma(\beta))$ as $T \to \infty$, where $\xrightarrow{d}$ denotes weak convergence on a subspace S of $C^2[0, \infty)$ and

$$Y_\gamma(\beta) = \sum_{k=1}^{\infty} \frac{4(\pi^2 k^2 + \gamma^2)X_k^2}{(\pi^2 k^2 + \beta^2)^2} - \sum_{k=1}^{\infty} \frac{4}{\pi^2 k^2 + \beta^2}$$

with $\{X_k\} \sim \text{NID}(0, 1)$.

(ii) If $\hat{\beta}_{\text{LM}} = \inf\{\beta \geqslant 0\colon L'_T(\beta) = 0 \text{ and } L''_T(\beta) < 0\}$ (i.e., $\hat{\beta}_{\text{LM}}$ is the local maximum of $L_T(\cdot)$ closest to 0), then

$$\hat{\beta}_{\text{LM}} \xrightarrow{d} \tilde{\beta}_{\text{LM},\gamma},$$

where

$$\tilde{\beta}_{\text{LM},\gamma} = \inf\{\beta \geqslant 0\colon \beta Y_\gamma(\beta) = 0 \text{ and } \beta Y'_\gamma(\beta) + Y_\gamma(\beta) < 0\}.$$

This theorem can be used directly to establish the limit

$$L_T(\beta) - L_T(0) \xrightarrow{d} Z_\gamma(\beta) = \int_0^\beta (\tau/2) Y_\gamma(\tau) d\tau$$

on $C[0, \infty)$. The integral can be evaluated in the closed form from which it follows that

$$Z_\gamma(\beta) = \sum_{k=1}^{\infty} \frac{\beta^2(\pi^2 k^2 + \gamma^2)X_k^2}{(\pi^2 k^2 + \beta^2)\pi^2 k^2} + \sum_{k=1}^{\infty} \ln\left(\frac{\pi^2 k^2}{\pi^2 k^2 + \beta^2}\right).$$

The above convergence suggests that the global maximizer of L_T converges in distribution to the global maximizer of Z_γ. Of course, in general, convergence on $C[0, \infty)$ does not necessarily imply convergence of the corresponding maximizers. However, as shown below, the maximum likelihood estimator converges in distribution to the value maximizing Z_γ. This is the content of the following theorem:

THEOREM 2.1. *Suppose $Y_1, \ldots, Y_T$ are observations from model* (1.1) *with $\theta_0 = 1-\gamma/T$ for some $\gamma \geqslant 0$. Then*

$$L_T(\beta)-L_T(0) \xrightarrow{d} Z_\gamma(\beta)$$

on $C[0, \infty)$ and

$$T(\hat{\theta}_{\mathrm{MLE}}-1) \xrightarrow{d} -\tilde{\beta}_{\mathrm{MLE},\gamma},$$

where $\tilde{\beta}_{\mathrm{MLE},\gamma}$ is the global maximizer of $Z_\gamma(\cdot)$.

Proof of Theorem 2.1. The first convergence follows immediately from Theorem 2.1 in [4]. Now, set $\hat{\beta}_T = T(1-\hat{\theta}_{\mathrm{MLE}})$, and let $\hat{\beta}_{T,M}$ and $\tilde{\beta}_{M,\gamma}$ be the values of β which maximize the likelihood $L_T(\beta)$ and the limit process $Z_\gamma(\beta)$, respectively, over the interval $[0, M]$. Then

$$|P(\hat{\beta}_T \leqslant x)-P(\tilde{\beta}_{\mathrm{MLE},\gamma} \leqslant x)| \leqslant R_1(x)+R_2(x)+R_3(x),$$

where

$$R_1(x) = |P(\hat{\beta}_T \leqslant x)-P(\hat{\beta}_{T,M} \leqslant x)|, \quad R_2(x) = |P(\hat{\beta}_{T,M} \leqslant x)-P(\tilde{\beta}_{M,\gamma} \leqslant x)|,$$

$$R_3(x) = |P(\tilde{\beta}_{M,\gamma} \leqslant x)-P(\tilde{\beta}_{\mathrm{MLE},\gamma} \leqslant x)|.$$

By Theorem 2.1 in [4], it follows that for each $M > 0$, $\hat{\beta}_{T,M}$ converges in distribution to $\tilde{\beta}_{M,\gamma}$, and hence $R_2(x)$ converges to zero as $T \to \infty$. In the Appendix, we show that $\tilde{\beta}_{\mathrm{MLE},\gamma} < \infty$ a.s. so that $\tilde{\beta}_{M,\gamma} \xrightarrow{\text{a.s.}} \tilde{\beta}_{\mathrm{MLE},\gamma}$ as $M \to \infty$. It follows that $R_3(x) \to 0$ as $M \to \infty$. Moreover, since the sequence $\{\hat{\beta}_T\}$ is tight (see the Appendix), we conclude that

$$\limsup_{T\to\infty} R_1(x) = \limsup_{T\to\infty} P(\hat{\beta}_T \geqslant M) \to 0 \quad \text{as } M \to \infty.$$

Since $T(\hat{\theta}_{\mathrm{MLE}}-1) = -\hat{\beta}_T$, the result now follows. ∎

3. Comparison of tests based on the MLE, local maximum estimators and the generalized likelihood ratio. In this and the following section we use simulation to derive type I error probabilities and power functions for testing the null hypothesis that H_0: $\theta = 1$ versus H_A: $\theta < 1$. Tests based on $\hat{\beta}_{\mathrm{MLE}}$, $\hat{\beta}_{\mathrm{LM}}$ and the likelihood ratio are considered in this section. In the next section the likelihood ratio test (which is shown here to have superior power performance than the tests based on $\hat{\beta}_{\mathrm{MLE}}$ and $\hat{\beta}_{\mathrm{LM}}$) is compared with the score type test of Tanaka [10].

The asymptotic theory of Section 2 also allows us to approximate the nominal power of the above test against local alternatives of the form

$$H_A\colon\ \theta = \theta_A, \quad \text{where } \theta_A = 1-\gamma/T.$$

The tests considered all have asymptotic power equal to 1 against any fixed local alternative. This property is not shared by the procedure proposed in [2].

To describe the test based on the generalized likelihood ratio let $Z_T(\beta) = L_T(\beta) - L_T(0)$ be the $-2\log$ of the likelihood ratio. Define the generalized likelihood ratio statistic as $\hat{Z}_T = Z_T(\hat{\beta}_{\text{MLE}})$. Also, let $\tilde{Z}_\gamma = Z_\gamma(\tilde{\beta}_{\text{MLE},\gamma})$ denote the limit random variable of $\hat{Z}_T$ when γ is the true value. The $(1-\alpha)$-th asymptotic quantile $b_{\text{GLR}}(\alpha)$ is defined as

$$P(\tilde{Z}_0 > b_{\text{GLR}}(\alpha)) = \alpha.$$

In the results to follow the tests are defined using these asymptotic quantiles to define the critical region.

To describe the tests based on the MLE and LM point estimates we define the following asymptotic quantile. Let $b_{\text{LM}}(\alpha)$ and $b_{\text{MLE}}(\alpha)$ be the $(1-\alpha)$-th quantiles defined as

$$P(\tilde{\beta}_{\text{MLE},0} > b_{\text{MLE}}(\alpha)) = \alpha, \qquad P(\tilde{\beta}_{\text{LM},0} > b_{\text{LM}}(\alpha)) = \alpha.$$

In order to find the values of $b_{\text{GLR}}(\alpha)$, $b_{\text{LM}}(\alpha)$ and $b_{\text{MLE}}(\alpha)$ using the asymptotic results of Section 2 the following simulation method was used. The infinite sums required in $Z_\gamma(\beta)$ and $Y_\gamma(\beta)$ are approximated by truncating them at $k = 1000$. For all results reported below, 100,000 replications were used when $\gamma = 0$ and 10,000 replications were used when $\gamma > 0$. For each replicate the three statistics were evaluated thereby reducing the between replicate variability as a component in the comparison of the three methods. For finite sample results, further details on the methods used to compute the likelihood based estimates are given in [4]. Probabilities reported below are accurate to ± 0.0032 when $\gamma = 0$ and to ± 0.01 when $\gamma > 0$ with 95% confidence. Since the p-th quantile estimate $\hat{\xi}_p$ is approximately $N(\xi_p,\ p(1-p)/(nf^2(\xi_p)))$, where $f(\cdot)$ is the pdf and n is the number of replications, the function DENSITY in Splus was used to estimate $f(\xi_p)$ which provides a standard error estimate of $\hat{\xi}_p$.

Table 3.1 compares the limit probabilities that the MLE and LM es-

TABLE 3.1. Probabilities of $\tilde{\beta}_{\text{LM},\gamma}$ and $\tilde{\beta}_{\text{MLE},\gamma}$ being zero

γ	$P(\tilde{\beta}_{\text{LM},\gamma} = 0)$	$P(\tilde{\beta}_{\text{MLE},\gamma} = 0)$
0.00	0.6574	0.6518
0.50	0.6516	0.6461
1.00	0.6336	0.6281
1.25	0.6198	0.6150
2.00	0.5667	0.5612
2.50	0.5312	0.5268
3.75	0.4319	0.4264
5.00	0.3454	0.3401
6.25	0.2750	0.2695
7.50	0.2214	0.2145
8.75	0.1765	0.1706
10.00	0.1437	0.1362
12.50	0.0934	0.0865
15.00	0.0633	0.0571

timates of β are equal to zero (or equivalently that MLE and LM estimates of θ are 1).

It is clear from this table that the probability of the MLE being at $\theta = 1$ is about 0.006 smaller than the corresponding probability for the local maximizer. However, for the purposes of using these two estimates for hypothesis testing a comparison of the $(1-\alpha)$ quantiles of the distributions of the two estimates is relevant. Table 3.2 provides the required quantiles $b_{\mathrm{GLR}}(\alpha)$, $b_{\mathrm{LM}}(\alpha)$ and $b_{\mathrm{MLE}}(\alpha)$ for the three tests against selected values of α (estimated standard deviations given in parentheses).

TABLE 3.2. $(1-\alpha)$ quantiles for the distribution of $\tilde{\beta}_{\mathrm{LM},0}$, $\tilde{\beta}_{\mathrm{MLE},0}$ and $\tilde{Z}_0$

Quantiles	α			
	0.01	0.025	0.05	0.1
$b_{\mathrm{LM}}(\alpha)$	11.25 (0.100)	8.55 (0.056)	6.52 (0.037)	4.75 (0.024)
$b_{\mathrm{MLE}}(\alpha)$	11.93 (0.103)	8.97 (0.062)	6.80 (0.042)	4.90 (0.026)
$b_{\mathrm{GLR}}(\alpha)$	4.41 (0.048)	2.95 (0.030)	1.94 (0.020)	1.00 (0.012)

Table 3.2 illustrates the difference between the asymptotic behavior of the local and global maximizers. As is expected the quantiles of $\tilde{\beta}_{\mathrm{MLE},0}$ are larger than those for $\tilde{\beta}_{\mathrm{LM},0}$ by a few percent which increases further out in the tail of the distributions.

It is clear from this table that substantial differences exist between the asymptotic quantile for the two estimators and that it is not safe to assume that the quantiles for the largest local maximizer provide adequate approximations to those of the MLE.

The next table indicates that the use of asymptotic quantiles gives very accurate values of the size of the test for sample size $T = 50$.

TABLE 3.3. Achieved significance levels using the asymptotic quantiles of Table 3.2 for $T = 50$

Finite sample power of LM, MLE and GLR tests	α			
	0.01	0.025	0.05	0.1
$P(\hat{\beta}_{\mathrm{LM}} > b_{\mathrm{LM}}(\alpha) \mid \gamma = 0)$	0.010	0.024	0.050	0.099
$P(\hat{\beta}_{\mathrm{MLE}} > b_{\mathrm{MLE}}(\alpha) \mid \gamma = 0)$	0.010	0.025	0.049	0.098
$P(\hat{Z}_T > b_{\mathrm{GLR}}(\alpha) \mid \gamma = 0)$	0.011	0.025	0.051	0.102

Figure 1 compares the finite sample power ($T = 50$) for various test statistics. Quite clearly, the maximum likelihood estimator is dominated by the local maximizer, and the generalized likelihood ratio test dominates both of these. This is also the case for the asymptotic values shown in Figure 2.

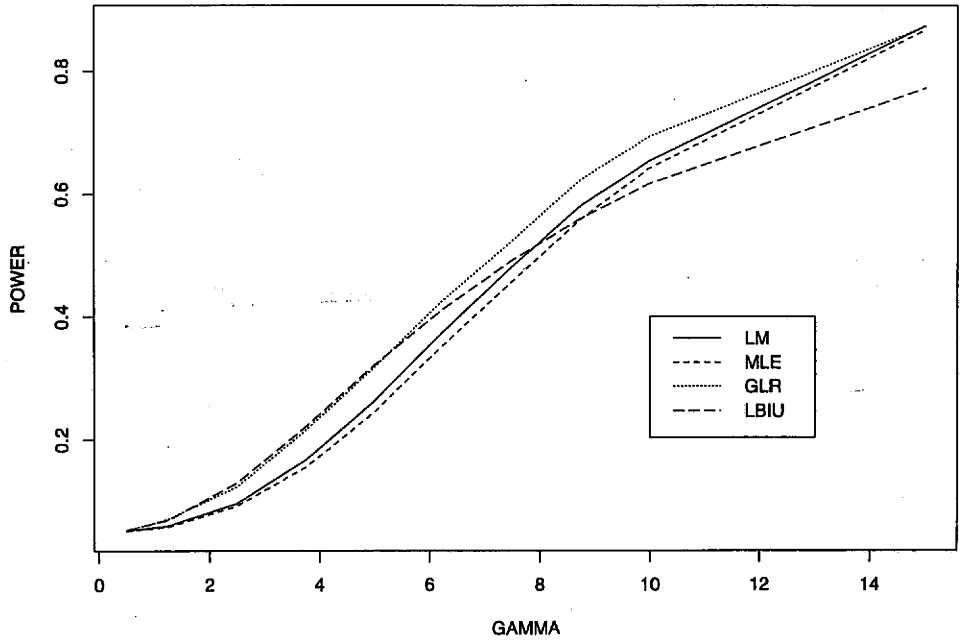

Fig. 1. Power curves based on LM, MLE, GLR and LBIU tests for $T = 50$

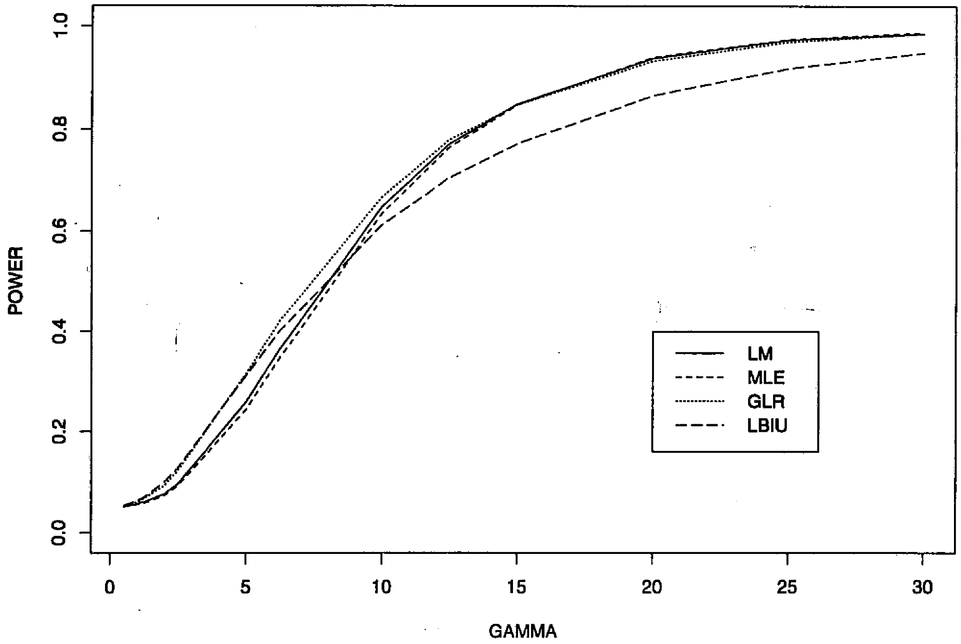

Fig. 2. Limiting power curves based on LM, MLE, GLR, and LBIU tests

For more detailed reference these power values are also presented in Tables 3.4 and 3.5.

It is clear from Table 3.5 that the use of the local maximizer yields a test with greater power than that based on the MLE. This result also holds (results not given here) when the type I error probability varies between $\alpha = 0.01$ and $\alpha = 0.1$.

Also evident from Table 3.5 (and similar results for $\alpha = 0.01$ to $\alpha = 0.1$) is the better power performance of the generalized likelihood ratio test. As a consequence, the GLR test is recommended. Interestingly, although the asymptotic distributions of $\hat{\beta}_{\text{MLE}}$ and $\hat{\beta}_{\text{LM}}$ differ somewhat (see below), the asymptotic powers,

TABLE 3.4. Comparison of the finite sample ($T = 50$) power of LM and MLE tests with the GLR test for $\alpha = 0.05$

γ	$P(\hat{\beta}_{\text{LM}} > b_{\text{LM}}(\alpha)\mid\gamma)$	$P(\hat{\beta}_{\text{MLE}} > b_{\text{MLE}}(\alpha)\mid\gamma)$	$P(\hat{Z}_T > b_{\text{GLR}}(\alpha)\mid\gamma)$
0.50	0.052	0.051	0.053
1.25	0.060	0.057	0.070
2.50	0.096	0.092	0.123
3.75	0.167	0.155	0.215
5.00	0.262	0.244	0.317
6.25	0.375	0.353	0.427
7.50	0.482	0.458	0.524
8.75	0.581	0.559	0.625
10.0	0.655	0.643	0.695
15.0	0.872	0.866	0.872

TABLE 3.5. Comparison of the limiting power of LM, MLE and GLR tests for $\alpha = 0.05$

γ	$P(\tilde{\beta}_{\text{LM},\gamma} > b_{\text{LM}}(\alpha))$	$P(\tilde{\beta}_{\text{MLE},\gamma} > b_{\text{MLE}}(\alpha))$	$P(\tilde{Z}_\gamma > b_{\text{GLR}}(\alpha))$
0.50	0.050	0.050	0.052
1.00	0.056	0.055	0.060
1.25	0.060	0.057	0.067
2.00	0.076	0.073	0.092
2.50	0.096	0.092	0.122
5.00	0.258	0.243	0.315
6.25	0.365	0.348	0.422
10.0	0.647	0.634	0.666
12.5	0.771	0.765	0.779
15.0	0.849	0.847	0.847
20.0	0.938	0.940	0.933
25.0	0.974	0.975	0.970
30.0	0.987	0.989	0.986
40.0	0.997	0.998	0.997
50.0	0.999	0.999	0.999
60.0	1.000	1.000	1.000

of the GLR test using either estimate are almost identical (results not presented). This implies that the calculation of asymptotic power of the GLR test can be done reasonably accurately (to the third decimal place) using the value of $\tilde{\beta}_{\mathrm{LM},\gamma}$ rather than $\tilde{\beta}_{\mathrm{MLE},\gamma}$. However, in the next section the simulated distribution of $Z(\tilde{\beta}_{\mathrm{MLE},\gamma})$ is used as the basis for comparing the asymptotic performance of the GLR test with Tanaka's S_T statistic.

Power estimates against various alternative values of θ using the asymptotic theory are compared with the exact values in the next table. Both the asymptotic estimates and the finite sample results use the asymptotic quantiles presented in Table 3.2. Clearly, the asymptotic theory provides very accurate approximations for alternatives as low as $\theta = 0.9$. For values of $\theta < 0.9$ the asymptotic results continue to provide very useable approximations.

TABLE 3.6. Comparison of the limiting power and exact power of the GLR test using quantiles from Table 3.2. Exact values are computed via simulation

θ_A		α 0.01	0.025	0.05	0.10
0.7	exact	0.797	0.840	0.872	0.900
	limit	0.767	0.810	0.845	0.886
0.8	exact	0.566	0.636	0.695	0.761
	limit	0.538	0.610	0.666	0.734
0.9	exact	0.176	0.247	0.317	0.412
	limit	0.176	0.246	0.315	0.411
0.95	exact	0.041	0.076	0.123	0.201
	limit	0.041	0.076	0.122	0.198
0.99	exact	0.011	0.027	0.054	0.104
	limit	0.011	0.027	0.052	0.102

Figures 1 and 2 also show that for the purposes of testing the null hypothesis that $\theta = 1$ it is better to use the GLR test for all alternatives. For this reason we will now turn to a comparison of the GLR test and the S_T test of Tanaka.

4. Comparison of the GLR test and the S_T test of Tanaka. The score type test of Tanaka [10] is demonstrated to be locally best invariant unbiased. Hence it will provide a benchmark against which the performance of the generalized likelihood ratio test can be judged. The next two tables provide additional details to that found in Figures 1 and 2. Table 4.1 compares the finite sample performance of the GLR test and the score type test while Table 4.2 gives an asymptotic comparison.

A few observations on Table 4.2 are:

1. The asymptotic values are very good approximations to the finite sample values (compare Tables 4.1 and 4.2).

2. The above results reported for Tanaka's S_T test are computed exactly using Imhof's procedure [5] since the distribution of S_T is equal to the probability that the sum of a linear combination of independent chi-squares is positive.

3. Clearly, Tanaka's S_T, based on Tables 4.1 and 4.2 (see also Figures 1 and 2), has a very small edge on the GLR test up to $\gamma = 5$ or so. Thereafter,

TABLE 4.1. Comparison of the finite sample ($T = 50$) power of Tanaka's score type test S_T with the limiting power of the GLR test for $\alpha = 0.05$

γ	Power of S_T	$P(\hat{Z}_T > b_{\mathrm{GLR}}(\alpha)\|\gamma)$
0.50	0.053	0.053
1.25	0.069	0.070
2.50	0.129	0.123
3.75	0.221	0.215
5.00	0.320	0.317
6.25	0.412	0.427
7.50	0.492	0.524
8.75	0.560	0.625
10.0	0.618	0.695
15.0	0.773	0.872

TABLE 4.2. Comparison of the limiting power of Tanaka's score type test S_T with the limiting power of the GLR test for $\alpha = 0.05$

γ	Power of S_T	$P(\tilde{Z}_\gamma > b_{\mathrm{GLR}}(\alpha))$
0.50	0.053	0.0522
1.00	0.062	0.0599
1.25	0.069	0.0669
2.00	0.099	0.0923
2.50	0.127	0.1218
5.00	0.311	0.3146
6.25	0.402	0.4218
10.0	0.611	0.6664
12.5	0.705	0.7791
15.0	0.772	0.8473
20.0	0.866	0.9329
25.0	0.919	0.9700
30.0	0.949	0.9855
40.0	0.979	0.9970
50.0	0.991	0.9993
60.0	0.996	0.9999

the GLR test increasingly outperforms S_T by a wide margin. The case $\gamma = 5$ corresponds to the alternative that $\theta = 0.9$ when $T = 50$. As seen from Figure 1, the power functions for the tests based on $\hat{\beta}_{\mathrm{LM}}$ and $\hat{\beta}_{\mathrm{MLE}}$ dominate the power function of S_T test for values of $\gamma > 8$ and $\gamma > 8.5$, respectively.

4. Overall we recommend the use of the GLR test.

Appendix

PROPOSITION. (a) *The sequence* $\{\hat{\beta}_T\}$ *is tight.*

(b) $\tilde{\beta}_{\mathrm{MLE},\gamma} < \infty$ *a.s.*

Proof. (a) First note that since $\hat{\beta}_T = T(1-\hat{\theta}_{\mathrm{MLE}})$ and $\hat{\theta}_{\mathrm{MLE}}$ is strongly consistent (see [8]), we have $\hat{\beta}_T/T \to 0$ a.s. Thus to prove (a), it suffices to show that for all $\varepsilon > 0$

(A1) $$\lim_{M\to\infty} \limsup_{T\to\infty} P(M < \hat{\beta}_T \leqslant \varepsilon T) = 0.$$

Now for $\varepsilon > 0$ fixed and small, if $M < \hat{\beta}_T \leqslant \varepsilon T$, then $\sup_{M \leqslant \beta \leqslant \varepsilon T} L'_T(\beta) > 0$, so that (A1) is implied by

(A2) $$\lim_{M\to\infty} \limsup_{T\to\infty} P\Big(\sup_{M \leqslant \beta \leqslant \varepsilon T} L'_T(\beta) > 0\Big) = 0.$$

We only give the proof for the case where $\gamma = 0$; the case $\gamma > 0$ follows in much the same way.

From equation (2.4) in [4] we have

$$L'_T(\beta) = \frac{C_T(\beta)}{S_T(\beta)}[B_T(\beta) - A_T(\beta)S_T(\beta)],$$

where

$$A_T(\beta) = \frac{1}{(T+1)^2}\sum_{t=1}^{T}\frac{2d_t}{1-d_t+(\beta^2/(2T^2))\,a(\beta,\,T)d_t},$$

$$B_T(\beta) = \frac{1}{(T+1)^2}\sum_{t=1}^{T}\frac{2d_t(1-d_t)U_{t,T}^2}{\left(1-d_t+(\beta^2/(2T^2))\,a(\beta,\,T)d_t\right)^2},$$

$$S_T(\beta) = \frac{1}{T}\sum_{t=1}^{T}\frac{(1-d_t)U_{t,T}^2}{1-d_t+(\beta^2/(2T^2))\,a(\beta,\,T)d_t},$$

$$C_T(\beta) = \frac{\beta}{2}\left(\frac{T+1}{T}\right)^2\left(1-\frac{\beta}{2T}\right)a^2(\beta,\,T),$$

$$a(\beta,\,T) = \left(1-\frac{\beta}{T}+\frac{\beta^2}{2T^2}\right)^{-1}, \qquad d_t = \cos\left(\frac{\pi t}{T+1}\right),$$

and

$$U_{t,T} = \sqrt{\frac{2}{T+1}}\,(1-d_t)^{-1/2}\sum_{s=1}^{T}\frac{Y_s}{\sqrt{2\sigma^2}}\sin\left(\frac{\pi st}{T+1}\right).$$

A straightforward calculation shows that

(A3)

$$\mathrm{E}U_{t,T}^2 = 1, \qquad \mathrm{Var}(U_{t,T}^2) \leqslant C_1, \qquad \text{and} \qquad \mathrm{Cov}(U_{t,T}^2,\,U_{s,T}^2) \leqslant C_2/(T+1)$$

for all $s \neq t$, where C_1 and C_2 are constants independent of s and t. Now since $C_T(\beta)/S_T(\beta) > 0$ for all $\beta \in [M,\,\varepsilon T]$, it suffices to show that

(A4)
$$\lim_{M\to\infty}\limsup_{T\to\infty} P[\sup_{M\leqslant\beta\leqslant\varepsilon T} B_T(\beta) - A_T(\beta)S_T(\beta) > 0] = 0.$$

For $\beta_0 < \beta_1$ large we have

(A5)

$$P[\sup_{\beta_0\leqslant\beta\leqslant\beta_1} B_T(\beta) - A_T(\beta)S_T(\beta) > 0] \leqslant P[\sup_{\beta_0\leqslant\beta\leqslant\beta_1} B_T(\beta) - A_T(\beta) > -\varepsilon_T(\beta_1)]$$

$$+P[\sup_{\beta_0\leqslant\beta\leqslant\beta_1} A_T(\beta)[1-S_T(\beta)] > \varepsilon_T(\beta_1)],$$

where $\varepsilon_T(\cdot)$ is a sequence of positive constants to be specified. Now since $A_T(\beta)$ and $B_T(\beta)$ are nonincreasing in β, the first term on the right in (A4) is bounded by

$$P[B_T(\beta_0)-A_T(\beta_1) > -\varepsilon_T(\beta_1)] \leqslant \frac{\operatorname{Var}[B_T(\beta_0)]}{[-\varepsilon_T(\beta_1)+A_T(\beta_1)-\mathrm{E}B_T(\beta_0)]^2}.$$

Setting $\beta_1 = r\beta_0$, where $r\in(1, \sqrt{2})$ is a fixed constant and choosing

$$\varepsilon < 1-\left[r^2\left(1-\frac{M}{T}+\frac{M^2}{2T^2}\right)-1\right]^{1/2}$$

for T large, we have

$$\begin{aligned}
&A_T(\beta_1)-\mathrm{E}B_T(\beta_0)\\
&=\sum_{t=1}^{T}\frac{(1/T^2)d_t^2(1-d_t)[2\beta_0^2a(\beta_0, T)-\beta_1^2a(\beta_1, T)]+(d_t^3\beta_0^4/(2T^4))a^2(\beta_0, T)}{(T+1)^2\left(1-d_t+(\beta_1^2/(2T^2))a(\beta_1, T)d_t\right)\left(1-d_t+(\beta_0^2/(2T^2))a(\beta_0, T)d_t\right)^2}\\
&\geqslant\sum_{t=1}^{T/4}\frac{(1/T^2)d_t^2(1-d_t)[2\beta_0^2a(\beta_0, T)-\beta_1^2a(\beta_1, T)]}{(T+1)^2\left(1-d_t+(\beta_1^2/(2T^2))a(\beta_1, T)d_t\right)\left(1-d_t+(\beta_0^2/(2T^2))a(\beta_0, T)d_t\right)^2}\\
&\geqslant\sum_{t=1}^{T/4}\frac{C_3d_t^2\beta_1^2}{(T+1)^4\left(1-d_t+(\beta_1^2/(2T^2))a(\beta_1, T)d_t\right)^2}\,\frac{1-d_t}{\left(1-d_t+(\beta_1^2/(2T^2))a(\beta_1, T)d_t\right)},
\end{aligned}$$

where

$$\begin{aligned}
&2\beta_0^2a(\beta_0, T)-\beta_1^2a(\beta_1, T)\\
&\qquad\geqslant\beta_1^2(2r^{-2}a(M, T)-a(\varepsilon T, T))\geqslant C_3\beta_1^2 \quad \text{for some } C_3>0,
\end{aligned}$$

by the choice of ε. Since

(A6)

$$\frac{x^2}{2\sqrt{2}}\leqslant 1-\cos x\leqslant\frac{x^2}{4} \text{ for } 0<x<\frac{\pi}{4} \quad\text{and}\quad \frac{\pi^2t^2}{2\sqrt{2}}<(T+1)^2(1-d_t)<\frac{\pi^2t^2}{2},$$

the sum above is bounded below by

$$C_4\beta_1^2\sum_{t=1}^{T/4}\frac{\pi^2t^2}{(\pi^2t^2+\beta_1^2)^3},\quad \text{where } C_4\geqslant 4C_3\cos^2\frac{\pi T}{4(T+1)}.$$

Evaluation of a suitable integral approximation to this sum gives a lower bound of $C_5\beta_1^{-3}$ for $\beta_1\geqslant 16$, and hence

(A7) $$A_T(\beta_1)-\mathrm{E}B_T(\beta_0)\geqslant C_6/\beta_1 \quad \text{for } \beta_1\geqslant 16.$$

Next, we write

$$\mathrm{Var}(B_T(\beta_0)) \leqslant 9(S(1, T/4)+S(T/4+1, T/2)+S(T/2+1, T)),$$

where, using (A3),

$$S(T_1, T_2)$$

$$= \sum_{s=T_1}^{T_2} \sum_{t=T_1}^{T_2} \frac{4d_s d_t(1-d_s)(1-d_t)\mathrm{Cov}(U_{t,T}^2, U_{s,T}^2)}{(T+1)^4(1-d_s+(\beta_0^2/(2T^2))a(\beta_0, T)d_s)^2(1-d_t+(\beta_0^2/(2T^2))a(\beta_0, T)d_t)^2}$$

$$\leqslant \sum_{t=T_1}^{T_2} \frac{4C_1 d_t^2}{(T+1)^4(1-d_t+(\beta_0^2/(2T^2))a(\beta_0, T)d_t)^2}$$

$$+\frac{4C_2}{T+1}\left(\sum_{t=T_1}^{T_2} \frac{d_t}{(T+1)^2(1-d_t+(\beta_0^2/(2T^2))a(\beta_0, T)d_t)}\right)^2.$$

Using (A6) and an integral approximation, we obtain the bounds

$$S(1, T/4) \leqslant \sum_{t=1}^{T/4} \frac{C_7}{(\pi^2 t^2+\beta_0^2)^2}+\frac{C_8}{T+1}\left(\sum_{t=1}^{T/4} \frac{1}{\pi^2 t^2+\beta_0^2}\right)^2$$

$$\leqslant C_9\beta_0^{-3}+C_{10}(T+1)^{-1}\beta_0^{-2} \leqslant \kappa_1\beta_0^{-3},$$

$$S(T/4+1, T/2)$$

$$\leqslant \sum_{t=T/4+1}^{T/2} \frac{C_{11}}{(T+1)^4(1-\sqrt{2}/2)^2}+\frac{C_{12}}{T+1}\left(\sum_{t=T/4+1}^{T/2} \frac{1}{(T+1)^2(1-\sqrt{2}/2)}\right)^2$$

$$\leqslant \kappa_2 T^{-3}$$

and

$$S(T/2+1, T) \leqslant \sum_{t=T/2+1}^{T} \frac{C_{13}}{(T+1)^4}+\frac{C_{14}}{T+1}\Big(\sum_{t=T/2+1}^{T} (T+1)^{-2}\Big)^2 \leqslant \kappa_3 T^{-3},$$

so that

(A8) $$\mathrm{Var}(B_T(\beta_0)) \leqslant \kappa_4/\beta_0^3.$$

Combining the bounds in (A7) and (A8) and choosing $\varepsilon_T(\beta_1) = C_6/(2\beta_1)$, we obtain

(A9) $$P[\sup_{\beta_0 \leqslant \beta \leqslant \beta_1} B_T(\beta)-A_T(\beta) > -\varepsilon_T(\beta_1)] \leqslant \frac{4\kappa_4\beta_1^2}{C_6^2\beta_0^3} = \frac{\kappa_5\beta_1^2}{\beta_0^3}.$$

To handle the second term in (A5), we have

$$P[\sup_{\beta_0 \leqslant \beta \leqslant \beta_1} A_T(\beta)[1-S_T(\beta)] > \varepsilon_T(\beta_1)]$$

$$\leqslant P\left[\frac{1}{2}-\frac{1}{T}\sum_{t=1}^{T/2}\frac{(1-d_t)U_{t,T}^2}{1-d_t+(\beta_1^2/(2T^2))\,a(\beta_1,\,T)d_t} > \frac{\varepsilon_T(\beta_1)}{2A_T(\beta_0)}\right]$$

$$+P\left[\frac{1}{2}-\frac{1}{T}\sum_{t=T/2+1}^{T} U_{t,T}^2 > \frac{\varepsilon_T(\beta_1)}{2A_T(\beta_0)}\right]$$

$$= P_1+P_2,$$

say. To obtain an upper bound for P_1 observe that

$$\mathrm{Var}\left(\frac{1}{T}\sum_{t=1}^{T/2}\frac{(1-d_t)U_{t,T}^2}{1-d_t+(\beta_1^2/(2T^2))\,a(\beta_1,\,T)d_t}\right)$$

$$\leqslant \frac{C_1}{T^2}\sum_{t=1}^{T/2}\frac{(1-d_t)^2}{\left(1-d_t+(\beta_1^2/(2T^2))\,a(\beta_1,\,T)d_t\right)^2}$$

$$+\frac{C_2}{T^2(T+1)}\left(\sum_{t=1}^{T/2}\frac{1-d_t}{1-d_t+(\beta_1^2/(2T^2))\,a(\beta_1,\,T)d_t}\right)^2 \leqslant \frac{C_{15}}{T},$$

and

$$\frac{1}{2}-\mathrm{E}\left[\frac{1}{T}\sum_{t=1}^{T/2}\frac{(1-d_t)U_{t,T}^2}{1-d_t+(\beta_1^2/(2T^2))\,a(\beta_1,\,T)d_t}\right] = \frac{1}{T}\sum_{t=1}^{T/2}\frac{(\beta_1^2/(2T^2))a(\beta_1,\,T)d_t}{1-d_t+(\beta_1^2/(2T^2))a(\beta_1,\,T)d_t},$$

which, after splitting the sum up into two pieces and bounding the summands, is bounded by

$$\frac{1}{T}\sum_{t=1}^{\delta T} 1+\frac{\beta_1^2}{2T^3}\sum_{t=\delta T+1}^{T/2}\frac{1}{1-\cos\left(\pi(\delta T+1)/(T+1)\right)} \leqslant \delta+\left(\frac{1}{2}-\delta\right)\frac{\varepsilon^2}{2}\frac{1}{1-\cos\pi\delta}.$$

Since $(x^2/2)\cos x < 1-\cos x < x^2/2$ and $d_t < 0$ for $t = T/2+1, \ldots, T$,

$$A_T(\beta_0) \leqslant C_{16}\sum_{t=1}^{T/2}\frac{1}{\pi^2 t^2+\beta_0^2} \leqslant C_{17}/\beta_0.$$

Now choose $\delta > 0$ and $\varepsilon > 0$ sufficiently small so that

$$\delta+\frac{(1/2-\delta)\varepsilon^2}{2(1-\cos\pi\delta)} < \frac{C_6}{8C_{17}r}.$$

Then

$$\frac{\varepsilon_T(\beta_1)}{2A_T(\beta_0)} \geqslant \frac{C_6\beta_0}{4C_{17}\beta_1} = \frac{C_6}{4C_{17}r},$$

and hence

(A10) $$P_1 \leqslant C_{15}T^{-1}\left(\frac{\varepsilon_T(\beta_1)}{2A_T(\beta_0)} - \frac{C_6}{8C_{17}r}\right)^{-2} < \frac{C_{18}}{T}.$$

Turning to P_2,

$$\mathrm{Var}\left(\frac{1}{T}\sum_{t=T/2+1}^{T} U_{t,T}^2\right) \leqslant \frac{C_1}{T^2}\sum_{t=T/2+1}^{T} 1 + \frac{C_2}{T^2(T+1)}\Big(\sum_{t=T/2+1}^{T} 1\Big)^2 \leqslant \frac{C_{19}}{T},$$

and

$$\mathrm{E}\left(\frac{1}{T}\sum_{t=T/2+1}^{T} U_{t,T}^2\right) = \frac{1}{2},$$

so that

(A11) $$P_2 \leqslant \frac{C_{19}}{T}\left(\frac{2A_T(\beta_0)}{\varepsilon_T(\beta_1)}\right)^2 \leqslant \frac{C_{20}}{T}.$$

Putting all the pieces together, we have

$$P[\sup_{M\leqslant\beta\leqslant\varepsilon T} B_T(\beta) - A_T(\beta)S_T(\beta) > 0]$$

$$\leqslant \sum_{n\geqslant \log_r M}^{\log_r \varepsilon T} P[\sup_{r^n\leqslant\beta\leqslant r^{n+1}} B_T(\beta) - A_T(\beta)S_T(\beta) > 0]$$

$$\leqslant \sum_{n\geqslant \log_r M}^{\log_r \varepsilon T} P[\sup_{r^n\leqslant\beta\leqslant r^{n+1}} B_T(\beta) - A_T(\beta) > -\varepsilon_T(r^{n+1})]$$

$$+ \sum_{n\geqslant \log_r M}^{\log_r \varepsilon T} P[\sup_{r^n\leqslant\beta\leqslant r^{n+1}} A_T(\beta)[1 - S_T(\beta)] > \varepsilon_T(r^{n+1})],$$

which, by using the bounds from (A9)–(A11), is less than or equal to

$$\sum_{n\geqslant \log_r M}^{\log_r \varepsilon T}\left(\frac{\kappa_5 r^{2n+2}}{r^{3n}} + \frac{C_{18}+C_{20}}{T}\right) \leqslant (\mathrm{const})\left[M^{-1}\left(1-\frac{M}{rT}\right) + \frac{\log_r \varepsilon T}{T}\right]$$

$$\to (\mathrm{const})M^{-1} \quad (\text{as } T\to\infty)$$

$$\to 0 \quad (\text{as } M\to\infty).$$

This completes the proof of (A4), and hence part (a) of the Proposition.

(b) The weak convergence of L'_T to Z'_γ given in Theorem 2.1 implies that, for M_1 and M_2 fixed positive constants,

$$P(\sup_{M_1 \leqslant \beta \leqslant M_2} Z'_\gamma(\beta) > 0) \leqslant \limsup_{T \to \infty} P(\sup_{M_1 \leqslant \beta \leqslant \varepsilon T} L'_T(\beta) > 0).$$

Now, taking the limits $M_2 \to \infty$, and then $M_1 \to \infty$, we have from (A2) the relation

$$\lim_{M_1 \to \infty} P(\tilde{\beta}_{\mathrm{MLE},\gamma} > M_1) \leqslant \lim_{M_1 \to \infty} P(\sup_{M_1 \leqslant \beta} Z'_\gamma(\beta) > 0) = 0,$$

and hence $\tilde{\beta}_{\mathrm{MLE},\gamma} < \infty$ a.s. ∎

REFERENCES

[1] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, 2-nd edition, Springer, New York 1991.

[2] N. H. Chan and R. S. Tsay, *Asymptotic inference for non-invertible moving average time series*, Technical Report #527, Carnegie Mellon University, Pittsburgh, PA, 1991.

[3] D. C. Chant, *On asymptotic tests of composite hypotheses in non-standard conditions*, Biometrika 61 (1974), pp. 291–298.

[4] R. A. Davis and W. T. M. Dunsmuir, *Maximum likelihood estimation for* MA(1) *processes with a root on or near the unit circle*, Econometric Theory (1996).

[5] J. P. Imhof, *Computing the distribution of quadratic forms in normal variables*, Biometrika 48 (1961), pp. 419–426.

[6] P. A. P. Moran, *Maximum likelihood estimation in non-standard conditions*, Proc. Cambridge Philos. Soc. 70 (1971), pp. 441–450.

[7] J. Neyman, *On optimal tests of composite hypothesis*, in: *Probability and Statistics, The Harold Cramer Volume*, U. Grenander (Ed.), 1959, pp. 213–234; Almquist and Wiksells, Uppsala 1959.

[8] B. M. Pötscher, *Noninvertibility and pseudomaximum likelihood estimation of misspecified* ARMA *models*, Econometric Theory 7 (1991), pp. 435–449.

[9] P. Saikkonen and R. Luukkonen, *Testing for a moving average unit root in autoregressive integrated moving average models*, J. Amer. Statist. Assoc. 88 (1993), pp. 596–601.

[10] K. Tanaka, *Testing for a moving average unit root*, Econometric Theory 6 (1990), pp. 433–444.

Richard A. Davis and Meiching Chen
Department of Statistics
Colorado State University
Fort Collins, CO 80523, U.S.A.

William T. M. Dunsmuir
School of Mathematics
University of New South Wales
Sydney, NSW 2052
Australia

Received on 8.3.1994