Communications in Statistics - Theory and Methods
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lsta20

Residual variance estimation in moving average models

Raul P. Mentz [a], Pedro A. Morettin [b] & Clélia M.C. Toloi [b]
[a] University of Tucumán and CONICET, Tucumán, 4.000, Argentina
[b] University of São Paulo, São Paulo, 05315-970, Brazil
Published online: 27 Jun 2007.

COMMUN. STATIST.—THEORY METH., 26(8), 1905–1923 (1997)

RESIDUAL VARIANCE ESTIMATION IN MOVING AVERAGE MODELS

Raul P. Mentz[1], Pedro A. Morettin[2] and Clélia M.C. Toloi[2]

[1] University of Tucumán and CONICET, 4.000 Tucumán, Argentina
[2] University of São Paulo, 05315-970 - São Paulo, Brazil

Key words: bias; maximum likelihood; method of moments; time series.

ABSTRACT

We consider time series models of the MA (moving average) family, and deal with the estimation of the residual variance. Results are known for maximum likelihood estimates under normality, both for known or unknown mean, in which case the asymptotic biases depend on the number of parameters (including the mean), and do not depend on the values of the parameters. For moment estimates the situation is different, because we find that the asymptotic biases depend on the values of the parameters, and become large as they approach the boundary of the region of invertibility. Our approach is to use Taylor series expansions, and the objective is to obtain asymptotic biases with error of $o(1/T)$, where T is the sample size. Simulation results are presented, and corrections for bias suggested.

1. INTRODUCTION

The moving average time series models of order q, denoted $MA(q)$, is defined by

$$X_t - \mu = \sum_{k=0}^{q} \alpha_k a_{t-k}, \quad t = \dots, -1, 0, 1, \dots, \tag{1.1}$$

where $\alpha_0 = 1$, X_t is the observable time series, a_t is a white noise residual with zero mean and constant variance σ^2, and the α_j's, μ and σ^2 are parameters, $0 < \sigma^2 < \infty$. We call σ^2 the residual variance of the process.

The process is stationary for any choice of parameters. If the roots of the associated polynomial equation $\sum_{k=0}^{q} \alpha_k w^{q-w} = 0$ satisfy $|w_k| < 1$, $k = 1, \dots, q$, then (1.1) can be inverted into an infinite autoregression, $a_t = \sum_{s=0}^{\infty} \delta_s (X_{t-s} - \mu)$, $t = \dots, -1, 0, 1, \dots$, with $\delta_0 = 1$, where the δ_s are determined by the α_k.

The covariance sequence of the process is

$$\gamma_s = E(X_t - \mu)(X_{t+s} - \mu) = \sigma^2 \sum_{k=0}^{q-s} \alpha_k \alpha_{k+s} = \gamma_{-s}, \quad s = 0, 1, \dots, q, \tag{1.2}$$

and equal to 0 if $|s| > q$, and the correlation sequence has

$$\rho_0 = 1, \quad \rho_s = \frac{\gamma_s}{\gamma_0} = \frac{\alpha_s + \alpha_1 \alpha_{s+1} + \cdots + \alpha_{q-s}\alpha_q}{1 + \alpha_1^2 + \cdots + \alpha_q^2} = \rho_{-s}, \quad s = 1, 2, \dots, q. \tag{1.3}$$

The covariance and correlation sequences are two of the tools of the time domain analysis. The covariance sequence satisfies the inversion formula $\gamma_s = \int_{-\pi}^{\pi} e^{i\lambda s} f(\lambda) d\lambda$ where $f(\lambda) = (\sigma^2/2\pi) \left| \sum_{k=0}^{q} \alpha_k e^{i\lambda k} \right|^2$, $-\pi < \lambda < \pi$, is the spectral density of the process, defined in the frequency domain.

We must specify the theoretical nature of the innovations a_t. One case is when they are i.i.d. normal with parameters 0 and σ^2. This assumption is used to define the likelihood function, and hence the maximum likelihood estimators. For purposes of comparison, we assume this property when defining and investigating other types of estimators; they, however, are sometimes studied under more general conditions, but this question will not be addressed here.

In this paper we consider the statistical problem of estimating σ^2. This is important because estimates of σ^2 (or of σ) enter, for example, in confidence sets for the other parameters, in the estimation of the spectrum, and in expressions for the estimated prediction error. They are also used in criteria for order determination, like AIC or BIC.

For purposes of inference we consider a sample $X_1, X_2, \ldots, X_T$ from (1.1). Estimates of σ^2 come from the method of moments (MM), and maximum likelihood under normality (ML). Some authors have also considered least squares (LS) procedures, or deriving estimators from frequency domain arguments. In spite of their inferential role, not many papers have been written about determination of large-sample biases of residual variance estimators in MA models. Our purposes in this paper are to review the literature, and to present new material. There are more known results about AR models; some of these, and new results, are presented in Mentz, Morettin and Toloi (1995).

2. REVIEW OF THE LITERATURE

The object of the inference may be taken to be the covariances or correlations introduced in (1.2) and (1.3), respectively. For these functions, large-sample expectations, variances, covariances and distributions are available, for several standard definitions of the sample quantities. This point will be briefly considered in Section 3.

When the object of the inference is taken to be the α_j and μ appearing in (1.1), early results available in the literature are surveyed in several sources. For example, the inefficiency of the moment estimator of α_j has been known for many years.

Tanaka (1984) suggests a technique for obtaining the Edgeworth-type asymptotic expansion associated with ML estimators in ARMA models. He obtained biases up to order $1/T$ for AR(1), AR(2), MA(1), MA(2) and ARMA(1,1) models, with and without constant terms. Biases for the residual variance estimators are also derived.

Cordeiro and Klein (1994) present a general procedure to obtain the biases of ML estimators in ARMA models. It turns out that their formula is difficult to obtain for models other than the lower order ones, but it can be easily computed numerically.

Other related papers are Davis (1977) and Porat and Friedlander (1986). Good references for techniques and results for asymptotic analysis in time series are Anderson (1971), and Fuller (1996).

3. ESTIMATION OF COVARIANCES AND CORRELATIONS

We consider estimating γ_j by

$$c_j = \frac{1}{T}\sum_{t=1}^{T-j}(X_t - \overline{X})(X_{t+j} - \overline{X}) = c_{-j}, \quad j = 0, 1, \ldots, T-1. \tag{3.1}$$

Other estimators are considered in the literature, for example, by changing in (3.1) the denominator, the value to be subtracted from the X's, or the range of the sums; see, for example, Anderson (1971, Chapter 8), or Fuller (1996, Chapter 6). We use (3.1) because for $T > p$, a covariance matrix with elements $c_{|i-j|}$ is positive definite, a fact that we shall use below.

With these estimators of the covariances, we form estimators of the correlations, $r_j = c_j/c_0$, $j = 1, 2, \ldots, T-1$.

From the above indicated sources, we deduce large-sample moments of these estimators. They are derived in the general case of a linear model $Y_t = \sum_{-\infty}^{\infty} \omega_j e_{t-j}$, under some conditions on the random variables e_t and coefficients ω_j. Our model (1.1) is a special case of such linear models. For example, when $q = 1$, for further reference we note that, with $r = r_1$,

$$\begin{aligned}
&E(c_0 - \gamma_0) \approx -\frac{\sigma^2}{T}(1+\alpha)^2, \quad E(c_1 - \gamma_1) \approx -\frac{\sigma^2}{T}(1 + 3\alpha + \alpha^2), \\
&E(c_0 - \gamma_0)^2 \approx \frac{2\sigma^4}{T}(1 + 4\alpha^2 + \alpha^4), \quad E(c_1 - \gamma_1)^2 \approx \frac{\sigma^4}{T}(1 + 5\alpha^2 + \alpha^4), \\
&E(c_0 - \gamma_0)(c_1 - \gamma_1) \approx \frac{4\sigma^4}{T}\alpha(1+\alpha^2), \\
&E(r - \rho) \approx -\frac{1}{T}\frac{1 + 4\alpha + \alpha^2 + 4\alpha^3 + \alpha^4 + 4\alpha^5 + \alpha^6}{(1+\alpha^2)^3}, \\
&E(r - \rho)^2 \approx \frac{1}{T}\frac{1 + \alpha^2 + 4\alpha^4 + \alpha^6 + \alpha^8}{(1+\alpha^2)^4},
\end{aligned} \tag{3.2}$$

where the expressions have errors $o(1/T)$. These results follow from general expressions in, for example, Anderson (1971) or Fuller (1996), and will be studied further in Section 4.2.3.

4. PARAMETER ESTIMATION BY THE METHOD OF MOMENTS

4.1. Estimation of the Coefficients

The parametric relation (1.3) between α_j's and ρ_j's can be inverted to express the former in terms of the latter, leading to a form of moment estimators of the coefficients. In the simple case of $q = 1$, $\rho_1 = \rho = \gamma_1/\gamma_0 = \alpha/(1+\alpha^2)$, and hence

$$\alpha = \frac{1-\sqrt{1-4\rho^2}}{2\rho}, \quad 0 < |\rho| < 0.5. \tag{4.1}$$

The relation for general q is more easily seen by factoring the spectral density function:

$$\frac{\sigma^2}{2\pi}\left|\sum_{k=0}^{q}\alpha_k e^{i\lambda k}\right|^2 = \frac{\sigma^2}{2\pi}\left(\sum_{k=0}^{q}\alpha_k e^{i\lambda k}\right)\left(\sum_{j=0}^{q}\alpha_j e^{-i\lambda j}\right) = \frac{\gamma_0}{2\pi}\sum_{s=-q}^{q}\rho_s e^{-i\lambda s}, \tag{4.2}$$

so that the moment estimators of the α_j are those satisfying

$$\frac{\left(\sum_{k=0}^{q}\widehat{\alpha}_k z^k\right)\left(\sum_{j=0}^{q}\widehat{\alpha}_j z^{-j}\right)}{1+\widehat{\alpha}_1^2+\cdots\widehat{\alpha}_q^2} = \sum_{s=-q}^{q} r_s z^{-s}\ . \tag{4.3}$$

Asymptotic biases of these estimators can be obtained as follows. Let $\dot{\alpha}_{jk} = \partial\alpha_j/\partial\rho_k$, $\ddot{\alpha}_{jks} = \partial^2\alpha_j/\partial\rho_k\partial\rho_s$, etc. Then

$$\begin{aligned} E(\widehat{\alpha}_j - \alpha_j) \approx & \sum_{k=1}^{q}\dot{\alpha}_{jk}E(r_k-\rho_k) + \frac{1}{2}\sum_{k=1}^{q}\ddot{\alpha}_{jkk}E(r_k-\rho_k)^2 \\ & + \sum_{\substack{k=1 \\ k>s}}^{q}\sum_{s=1}^{q}\ddot{\alpha}_{jks}E(r_k-\rho_k)(r_s-\rho_s), \end{aligned} \tag{4.4}$$

where this is valid with an error $o(1/T)$.

In the case of $q = 1$, $\partial\rho/\partial\alpha = (1-\alpha^2)/(1+\alpha^2)^2$, $\partial^2\rho/\partial\alpha^2 = -2\alpha(3-\alpha^2)/(1+\alpha^2)^3$, and hence

$$\dot{\alpha} = \frac{\partial\alpha}{\partial\rho} = \frac{(1+\alpha^2)^2}{1-\alpha^2}, \quad \ddot{\alpha} = \frac{\partial^2\alpha}{\partial\rho^2} = \frac{2\alpha(1+\alpha^2)^3(3-\alpha^2)}{(1-\alpha^2)^3}, \tag{4.5}$$

and using (3.2),

$$E(\widehat{\alpha}-\alpha) = -\frac{1}{T}\frac{1+\alpha-2\alpha^2-7\alpha^3+2\alpha^4-4\alpha^5-2\alpha^6+\alpha^7+\alpha^8+\alpha^9}{(1-\alpha^2)^3} + o(1/T). \tag{4.6}$$

Similarly,

$$E(\widehat{\alpha}-\alpha)^2 \approx \mathrm{Var}(\widehat{\alpha}) \approx \frac{(1+\alpha^2)^4}{(1-\alpha^2)^2}E(r-\rho)^2 \approx \frac{1}{T}\frac{1+\alpha^2+4\alpha^4+\alpha^6+\alpha^8}{(1-\alpha^2)^2}, \quad (4.7)$$

a known result, see, for example, Fuller (1996, (8.3.3)). These approximations have errors $o(1/T)$. See Section 4.2.3 below.

4.2. Estimation of the Residual Variance

A moment estimator of σ^2 can be deduced by making $s = 0$ in (1.3), that is,

$$\widehat{\sigma}^2_{MM} = \frac{c_0}{1+\widehat{\alpha}_1^2+\cdots+\widehat{\alpha}_q^2}, \quad (4.8)$$

where the $\widehat{\alpha}_j$ are those defined in Section 4.1.

To study the asymptotic bias, we derive an expansion valid for arbitrary $q \geq 1$, and then evaluate this explicity for the special case of $q = 1$.

4.2.1. An Expansion for the MA(q) Model

Applying to (4.8) a Taylor expansion that retains terms contributing up to second order, and expanding again terms involving $\widehat{\alpha}_j - \alpha_j$ as functions of $r_k - \rho_k$, as we did to obtain (4.4), we prove that

$$\begin{aligned}E(\hat{\sigma}^2_{MM}-\sigma^2) &= \frac{\sigma^2}{\gamma_0}E(c_0-\gamma_0)+\frac{2\sigma^4}{\gamma_0^4}A_1E(c_0-\gamma_0)^2\\ &-\frac{2\sigma^4}{\gamma_0^3}\sum_{k=1}^{q}A_{2k}E(c_0-\gamma_0)(c_k-\gamma_k)-\frac{2\sigma^4}{\gamma_0}\sum_{k=1}^{q}A_{2k}E(r_k-\rho_k)\\ &+\sum_{k=1}^{q}A_{3k}E(r_k-\rho_k)^2+\sum_{\substack{k=1\\k>s}}^{q}\sum_{s=1}^{q}A_{4ks}E(r_k-\rho_k)(r_s-\rho_s)+o(1/T), \quad (4.9)\end{aligned}$$

where

$$A_1 = \sum_{j=1}^{q}\sum_{k=1}^{q}\alpha_j\gamma_k\dot{\alpha}_{jk}, \qquad A_{2k} = \sum_{j=1}^{q}\alpha_j\dot{\alpha}_{jk}, \quad (4.10)$$

$$A_{3k} = \sum_{j=1}^{q}\left[-\frac{\sigma^4}{\gamma_0}\alpha_j\ddot{\alpha}_{jkk}-\frac{\sigma^4}{\gamma_0}\dot{\alpha}_{jk}^2+\frac{4\sigma^6}{\gamma_0^2}\alpha_j^2\dot{\alpha}_{jk}^2\right]+\frac{8\sigma^6}{\gamma_0^2}\sum_{\substack{j=1\\j>i}}^{q}\sum_{i=1}^{q}\alpha_j\alpha_i\dot{\alpha}_{jk}\dot{\alpha}_{ik},$$

$$A_{4ks} = \sum_{j=1}^{q} \left[-\frac{2\sigma^4}{\gamma_0} \alpha_j \ddot{\alpha}_{jks} - \frac{2\sigma^4}{\gamma_0} \dot{\alpha}_{jk} \dot{\alpha}_{js} + \frac{8\sigma^6}{\gamma_0^2} \alpha_j^2 \dot{\alpha}_{jk} \dot{\alpha}_{js} \right] + \frac{16\sigma^6}{\gamma_0^2} \sum_{\substack{j=1 \\ j>i}}^{q} \sum_{i=1}^{q} \alpha_j \alpha_i \dot{\alpha}_{jk} \dot{\alpha}_{is}.$$

4.2.2. The Special Case of MA(1)

When $q = 1$, (4.10) becomes

$$A_1 = \sigma^2 \alpha^2 \frac{(1+\alpha^2)^2}{1-\alpha^2}, \quad A_{2k} = \alpha \frac{(1+\alpha^2)^2}{1-\alpha^2}, \quad A_{3k} = -\sigma^2 \frac{(1+\alpha^2)^4}{(1-\alpha^2)^3} \tag{4.11}$$

while the necessary asymptotic expectations are given in (3.2). Substitution leads to

$$E(\hat{\sigma}^2_{MM} - \sigma^2) = -\frac{\sigma^2}{T} \frac{2 - 6\alpha^2 - 2\alpha^3 + 15\alpha^4 + 4\alpha^5 - 4\alpha^6 - 2\alpha^7 + \alpha^8}{(1-\alpha^2)^3} + o(1/T). \tag{4.12}$$

In this special case we can check (4.12) by using directly an expansion of its left-hand side stemming from (4.8), plus $E(c_0 - \gamma_0)$, $E(\hat{\alpha} - \alpha)$, $E(\hat{\alpha} - \alpha)^2$ given, respectively, in (3.2), (4.6), (4.7), and $E(c_0 - \gamma_0)(\hat{\alpha} - \alpha) \approx (\sigma^2/T)[2\alpha(1+\alpha^4)]/(1-\alpha^2)$.

In figure 1 we graph T/σ^2 times (4.12). The graph is flat, at the level of -2, for α small in absolute value; -2 is the value corresponding to $\alpha = 0$, and in Section 5 we will find its connection with ML estimation. As α increases in absolute value(i.e. tends to leave the region of invertibility) the values increase substantially their negative values, since in fact, for fixed α and T, (4.12) approaches $-\infty$ as $|\alpha|$ approaches 1.

The depicted values, times σ^2/T, are the negative biases in the estimation of the residual variance by the MM, as a function of the parameter α. Hence, the proposed estimator tends to lead to values that, in the average and approximately, are smaller than desired. A consequence will be that confidence intervals with scale parameter estimated by MM will tend to be unduly optimistic, that is, short. Further, this defect tends to become more important for values of α approaching 1 in absolute value.

Note, however, that these biases in the estimation of the residual variance must be divided by the sample size T, so that, as expected, larger sample sizes will lead to better results.

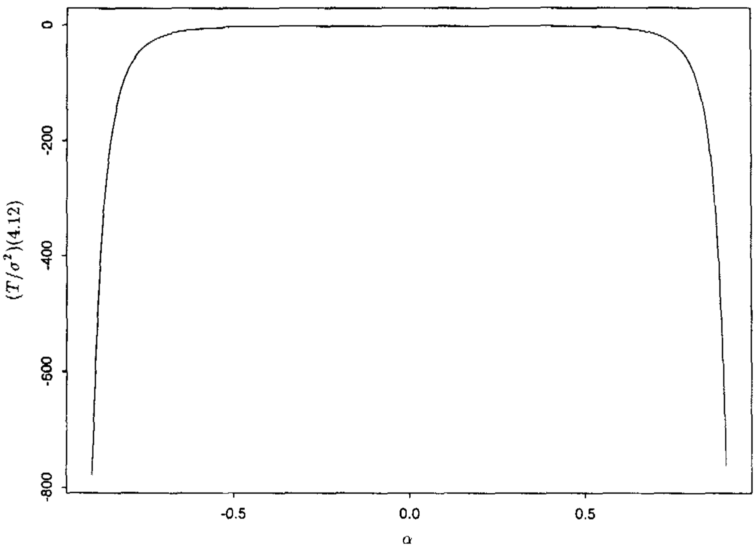

FIGURE 1: Representation of T/σ^2 times (4.12).

4.2.3. A Simulation Study for the MA(1) Model

Expression (4.12) is to be added to comparable and previously known asymptotic biases for the MA(1) model: for moment estimators, the next to last line of (3.2) corresponds to r and (4.6) to α; the first line of (3.2) has the asymptotic biases of the usual estimators of γ_0 and γ_1. A graphical analysis of these functions shows, in general, dependence on the values of α: the α-effect illustrated in Figure 1 for (4.12), tends to hold for several of the other expressions, and anticipates difficulties near the boundary of the region of invertibility.

To study the behaviour of these estimation procedures, a simulation study was performed. Two of the main purposes of such a study were to investigate the sample sizes needed for reasonable fits of the asymptotic theory to the empirical results, and to relate the analysis to values of α.

In the application of these procedures we found that the sample correlation r not always satisfies $|r| < 0.50$ as suggested by the parametric relation (4.1), so that the corresponding estimator of α is not real. For small values of α, no value of r was found to violate this restriction, but as we considered larger values of α, replications with $|r| > 0.5$ were found. The following definition was introduced: use the sample analog of (4.1) only for $0 < |r| \leq 0.50$, and then set $\hat{\alpha} = -1$ if $r < -0.5$, $\hat{\alpha} = 1$ if $r > 0.5$, and $\hat{\alpha} = 0$ if $r = 0$ (Fuller, 1996, Section 8.3). The problem of noninvertibility of sample estimators in MA models has been studied in other contexts, for example, in ML estimation under normality: see, for example, Cryer and Ledolter (1981), Pesaran (1983), Anderson and Takemura (1986).

The results of the simulations are reported as follows. Besides the estimation of the residual variance σ^2 by (4.8), we also consider the estimation of $\gamma_0 = \sigma^2(1+\sigma^2)$, $\gamma_1 = \sigma^2\alpha$, $\rho = \gamma_1/\gamma_0$, and α, with estimators introduced in Section 3, and (4.3) (that for $q = 1$ leads to the sample analog of (4.1)), respectively. The objective of this design is to facilitate the comparison of (4.12) with existing similar expressions. We simulated T values corresponding to (1.1), with the a_t pseudorandom, independent normal (0,1).

Several values of α ranging from 0.2 to 0.9 were considered and various sample sizes starting at $T = 50$. To simplify, and to stress the nature of our findings, we restrict our presentation to $\alpha = 0.4$, 0.6 and 0.8, and $T = 400$, 600, 800 and 1,000. For each choice of α and T, 1,000 replications were done.

Table I summarizes the results of the simulation study, by presenting the averages over the 1,000 replications, of the estimates obtained by using the indicated procedures. The signs of the biases correspond to those of the theoretical expressions, except for the small variability of $\hat{\alpha}$ around 0.40 for $\alpha = 0.40$. The last line contains the numbers of cases where $r > 0.50$, and we see that occurences are frequent for $\alpha = 0.80$. This tends to raise the values of $\hat{\alpha}$, and hence to lower those of $\hat{\sigma}^2$.

Table II presents an inferential analysis of the results of the simulations done for $\alpha = 0.40$ and $T = 400$. For this combination of values (as well as for smaller $|\alpha|$ or for larger T) no case of $r > 0.5$ was observed. Column 2

TABLE I: Average estimates over 1,000 replications, usual estimators of γ_0, γ_1, MM estimators of ρ, α, σ^2.

Parameter to be estimated	Parameter value	$\alpha = 0.40$ Sample sizes 400	600	800	1,000
γ_0	1.160	1.154	1.156	1.160	1.160
γ_1	0.400	0.395	0.397	0.401	0.398
ρ	0.345	0.341	0.343	0.345	0.343
α	0.400	0.400	0.401	0.402	0.399
σ^2	1.000	0.992	0.994	0.997	0.999
$\lvert r\rvert > 0.5$		(0)	(0)	(0)	(0)

Parameter to be estimated	Parameter value	$\alpha = 0.60$ Sample sizes 400	600	800	1,000
γ_0	1.360	1.354	1.357	1.357	1.359
γ_1	0.600	0.595	0.598	0.598	0.598
ρ	0.441	0.437	0.440	0.440	0.440
α	0.600	0.618	0.614	0.612	0.606
σ^2	1.000	0.978	0.984	0.986	0.993
$\lvert r\rvert > 0.5$		(52)	(19)	(9)	(7)

Parameter to be estimated	Parameter value	$\alpha = 0.80$ Sample sizes 400	600	800	1,000
γ_0	1.640	1.631	1.637	1.637	1.638
γ_1	0.800	0.792	0.797	0.796	0.799
ρ	0.488	0.476	0.481	0.481	0.483
α	0.800	0.800	0.816	0.815	0.823
σ^2	1.000	0.998	0.986	0.987	0.980
$\lvert r\rvert > 0.5$		(340)	(315)	(287)	(283)

TABLE II: Inferential analysis of simulations, usual estimators of γ_0, γ_1, MM estimators of ρ, α, σ^2. $\alpha = 0.40$, $T = 400$

1. Parameter to be estimated	2. Average estimate over 1,000 replications	3. Empirical bias	4. Asymptotic bias	5. Empirical standard error	6. Studentized bias
γ_0	1.1545	-0.0055	-0.0049	0.0045	-0.1438
γ_1	0.3953	-0.0047	-0.0059	0.0034	0.3437
ρ	0.3410	-0.0038	-0.0077	0.0022	1.7847
α	0.3998	-0.0002	-0.0027	0.0036	0.6962
σ^2	0.9919	-0.0081	-0.0056	0.0036	-0.6954

contains the averages over replications (also reported in Table I), and column 3 the corresponding estimated biases; these should be compared with theoretical values in column 4. To judge the significance of the differences between columns 3 and 4, column 5 contains the empirical standard errors, from which the studentized differences are formed in column 6.

We observe the following: (1) Columns 3 and 4 visually show good concordance; (2) The studentized differences are small for all rows. We conclude that for $\alpha = 0.40$, $T = 400$ is large enough to lead to a good fit of the asymptotic theory to the simulated results.

Tables similar to Table II were constructed for each combination of α and T in the indicated ranges. The studentized differences of the biases in these tables are presented in Table III. We observe the following: (1) When the estimation of γ_0 and γ_1 is considered, all values are small, with a maximum of 2.07 for $\alpha = 0.40$ and $T = 800$; (2) Most values corresponding to the estimation of ρ are large, and increase as α gets larger; (3) Studentized biases in the estimation of α are smaller than 3 for $\alpha = 0.40$ and 0.60, and very large negative for $\alpha = 0.80$; (4) Studentized biases in the estimation of σ^2 are smaller than 3 (all but one are smaller than 2) for $\alpha = 0.40$ and 0.60, and very large and positive for $\alpha = 0.80$.

TABLE III: Studentized biases, usual estimators of γ_0, γ_1, MM estimators of ρ, α, σ^2.

Parameter to be estimated	$\alpha = 0.40$ Sample sizes				$\alpha = 0.60$ Sample sizes				$\alpha = 0.80$ Sample sizes			
	400	600	800	1000	400	600	800	1000	400	600	800	1000
γ_0	-0.14	-0.26	1.15	1.07	0.10	0.31	-0.08	0.53	-0.06	0.43	0.17	0.46
γ_1	0.34	0.59	2.07	0.59	0.55	1.07	1.04	0.89	0.31	1.12	0.23	1.53
ρ	1.78	2.21	3.45	1.39	4.53	5.70	5.77	5.19	7.23	8.78	6.76	8.35
α	0.70	1.08	2.16	0.24	1.62	2.18	2.55	1.38	-22.64	-18.30	-16.07	-12.60
σ^2	-0.70	-1.08	-0.14	0.93	-1.45	-1.79	-2.62	-0.87	23.36	19.21	17.46	13.59

Our conclusions are: (1) Moment estimators of ρ, α and σ^2 have biases, both asymptotic and for finite samples, that depend on the true value of α that generated the MA(1) series; (2) For a sample size of $T = 400$ or larger, the finite-sample biases are well predicted by the asymptotic theory if $\alpha = 0.40$ or smaller; (3) For $\alpha = 0.80$, even for T as large as 1,000, the finite-sample biases are not well predicted by the asymptotic theory: in some cases the approximations tend to improve as sample size increases, but clearly larger values of T will be needed to observe good fits; (4) For values of α intermediate between 0.40 and 0.80, and values of T intermediate between 400 and 1,000, results of the simulations tend to fill in the trends suggested by the extreme values chosen for the simulations; (5) These situations contrast with the analysis of the estimation of the covariances by the standard sample quantities, since for them the simulated biases are in agreement with those predicted by the asymptotic theory, for all chosen values of α and T.

In terms of the method of moments estimation of the residual variance in MA(1) models, if α is close to the boundary of the region of invertibility, very large sample sizes will be required to make the asymptotic theory for its bias useful in practice. This contrasts with the analysis for the method of ML, to be considered below. Users of the "preliminary estimators" of the residual variance given by the method of moments, should be aware of the difficulties in predicting the biases associated with such a choice, and the differences that may exist with

ML estimation. The situation is similar when MM estimation of ρ and α is considered, a fact that has not been sufficiently stressed in the literature, even when asymptotic results are known. On the other side, the situation differs markedly from that of estimating the covariances of the process, for which the asymptotic results are useful for all α and sample sizes. An important role in these findings is played by the fact that the autocorrelation estimator r tends to give values $r > 0.50$ with high frequency, when $\alpha > 0$ is large.

5. PARAMETER ESTIMATION BY MAXIMUM LIKELIHOOD

When the residuals in (1.1) are independent $N(0, \sigma^2)$, the likelihood function of the observations $X_1, X_2, \ldots, X_T$ is the function of μ, σ and the α_j's given by

$$L = (2\pi)^{-\frac{1}{2}T}|\mathbf{\Sigma}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{X}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu})\right\}, \qquad (5.1)$$

where $\mathbf{X} = (X_1, \ldots, X_T)'$, $\boldsymbol{\mu} = (\mu, \ldots, \mu)'$, and $\mathbf{\Sigma}$ is the covariance matrix with components $\gamma_{|i-j|}$. This can also be written as

$$L = (2\pi\sigma^2)^{-\frac{1}{2}T}|\mathbf{P}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{X}-\boldsymbol{\mu})'\mathbf{P}^{-1}(\mathbf{X}-\boldsymbol{\mu})\right\}, \qquad (5.2)$$

where $\mathbf{\Sigma} = \sigma^2\mathbf{P}$. Maximum likelihood estimators of $\boldsymbol{\mu}$, $\alpha_1, \ldots, \alpha_q$ and σ^2 are obtained by maximizing (5.2) over its parameter space. No explicit formulas for these estimators are known, not even in the case of $q = 1$. However, asymptotic expressions for some of the biases are known: Tanaka (1984) and Cordeiro and Klein (1994) give, for the MA(1) model with $\alpha_1 = \alpha$,

$$E(\hat{\alpha} - \alpha) = \frac{1+2\alpha}{T} + o(1/T). \qquad (5.3)$$

Maximizing (5.2) with respect to σ^2 only, shows that

$$\hat{\sigma}^2_{ML} = \frac{1}{T}(\mathbf{X}-\hat{\boldsymbol{\mu}})'\hat{\mathbf{P}}^{-1}(\mathbf{X}-\hat{\boldsymbol{\mu}}). \qquad (5.4)$$

For this estimator, Tanaka (1984) and Cordeiro and Klein (1994) give, for the MA(1) model

$$E(\hat{\sigma}^2_{ML} - \sigma^2) = -\frac{2\sigma^2}{T} + o(1/T). \qquad (5.5)$$

5.1. An Approximated Closed-form Expression for $\hat{\sigma}^2_{ML}$ in MA(1)

We operate with (5.4). If $\mathbf{G}$ denotes the $T \times T$ matrix that has components equal to 1 in its two diagonals adjacent to the main diagonal, and 0 elsewhere, then $\mathbf{P} = (1+\alpha^2)\mathbf{I} + \alpha\mathbf{G} = (1+\alpha^2)[\mathbf{I}+\rho\mathbf{G}]$, and $\mathbf{P}^{-1} = (1+\alpha^2)^{-1}[\mathbf{I}+\rho\mathbf{G}]^{-1}$, so that it suffices to invert the $\mathbf{I}+\rho\mathbf{G}$ matrix. This matrix and its inverse are symmetric Toplitz matrices, and the inverse was studied in detail in Mentz (1976) and Shaman (1969). In the former the components of $[\mathbf{I}+\rho\mathbf{G}]^{-1}$ are given as

$$\frac{\left(1 - x_1^{2T-2j+2}\right)\left(x_1^{j+i+1} - x_1^{j-i+1}\right)}{\rho\left(1-x_1^2\right)\left(1-x_1^{2T+2}\right)} \tag{5.6}$$

where x_1 is a root of the associated polynomial equation $\rho x^2 + x + \rho = 0$, and hence $x_1 = -\alpha$.

In Shaman (1969) it is suggested that components of the inverse matrix can be approximated by $x_1^k(1-4\rho^2)^{-\frac{1}{2}}$ that in turn, substituting $\rho = \alpha(1+\alpha^2)^{-1}$, can be expressed as $(-\alpha)^k(1+\alpha^2)(1-\alpha^2)^{-1}$, $k = 0, 1, \ldots, T-1$. The effect of this approximation is as follows: $\mathbf{P}$ times a matrix with these components, produces a matrix that differs from the identity matrix in the components of the first row and column, and in the component in row T, column T.

Denoting by $p^{ij} = p^{|i-j|}$ the components of $\mathbf{P}^{-1}$, $p^k \approx (-\alpha)^k(1-\alpha^2)^{-1}$, and hence (5.4) becomes

$$\hat{\sigma}^2_{ML} = \sum_{i=1}^{T}\sum_{j=1}^{T} \hat{p}^{ij}\frac{1}{T}(X_i - \hat{\mu})(X_j - \hat{\mu}) \approx \frac{c_0 + 2\sum_{j=1}^{T-1}(-\hat{\alpha})^j c_j}{1-\hat{\alpha}^2}. \tag{5.7}$$

Hence, the maximum likelihood estimator of σ^2 has an approximate representation as a weighted sum of c_j's, weights involving powers of $-\hat{\alpha}$, where $\hat{\alpha}$ is the ML estimator of α.

It can be shown that we can arrive at this expression (with the corresponding least squares estimator of α) by starting from a definition of σ^2 in terms of a residual sum of squares (see, for example, Brockwell and Davis, 1991a), and using a "long autoregression" approximation to be able to write the defining expression in a regression type of format (see the technical report by the authors, 1995). The fact that the LS and ML estimators, under some approximations, coincide, has been noted elsewhere. Brockwell and Davis (1991a, Section 8.7) note that if the determinant of the likelihood function is asymptotically negligible compared with the sum of squares in the exponent, as in the case when the parameter is constrained to be invertible, then minimization of the sum of squares will be equivalent to minimization of the likelihood and the least squares and maximum likelihood estimators will have similar asymptotic properties.

For comparison, the MM estimator can be expressed in terms of powers of $-\widehat{\alpha}^2$ as follows,

$$\widehat{\sigma}^2_{MM} = \frac{c_0}{1+\widehat{\alpha}^2} \approx c_0 + \sum_{j=1}^{T-1} c_0(-\widehat{\alpha}^2)^j, \tag{5.8}$$

where now $\widehat{\alpha}$ is the MM estimator of α.

5.2. Simulations for ML Estimation of the Residual Variance

To illustrate the biases in the ML estimation of the residual variance of a MA(1) model, a small simulation study was performed. The values $\alpha = -0.40$, 0.20 and 0.80 were used to generate the simulated observations, and sample sizes $T = 50$, 100, 200 and 400 were considered. Only 100 replications were done for each pair of α and T. The ML procedure in Brockwell and Davis (1991b) was used, in the understanding that it represents accurately the theory presented in this section, and further that the procedure is one that is used frequently in empirical studies. The main results are collected in Table IV.

In column 3 there are some sign changes in relation to what is expected. However, column 6 of studentized biases contains only one value as large as 3, which occurs for the smallest sample size $T = 50$. The assertion that the biases do not depend upon the values of α is well supported by these figures.

In spite of the small number of replications, the differences between this study for ML estimation and that in Section 4.2.3 for the MM are clear. The asymptotic theory is seen to be a valid approximation for a wider range of values of the basic parameter σ, and for smaller sample sizes.

TABLE IV: Inferential analysis of simulations, ML estimators of σ^2.

1. Value of α	2. Sample size	3. Empirical bias	4. Asymptotic bias	5. Empirical standard error	6. Studentized bias
-0.4	50	-0.022	-0.040	0.019	0.947
-0.4	100	-0.002	-0.020	0.013	1.385
-0.4	200	-0.012	-0.010	0.009	-0.222
-0.4	400	-0.004	-0.005	0.006	0.167
0.2	50	0.014	-0.040	0.018	3.000
0.2	100	0.007	-0.020	0.013	2.077
0.2	200	-0.007	-0.010	0.010	0.300
0.2	400	-0.009	-0.005	0.006	-0.667
0.8	50	-0.004	-0.040	0.022	1.636
0.8	100	0.013	-0.020	0.016	2.063
0.8	200	0.007	-0.010	0.011	1.545
0.8	400	0.007	-0.005	0.007	1.714

6. CONCLUDING REMARKS

We considered the estimation of the variance of the white noise component of the MA(q) model defined by (1.1). This is a nuisance parameter, and is important because estimates enter into prediction and confidence intervals, tests of hypotheses, spectral estimates, and other inferential procedures.

In spite of the indicated usefulness, not many results are available about properties of estimators of the residual variance in MA models, except for maximum likelihood estimators under normality. The situation is better for AR models, as indicated in Mentz, Morettin and Toloi (1995).

Available results for ML estimates under normality can be interpreted as follows: with $X_1, X_2, \ldots, X_T$ i.i.d. $N(\mu, \sigma^2)$, the ML estimator of σ^2 is $\widehat{\sigma}^2 = \sum_{i=1}^{T}(X_i - \overline{X})^2/T$, and it is biased, it underestimates σ^2, $E(\widehat{\sigma}^2 - \sigma^2) = -\sigma^2/T$, the reason being that the denominator T is too large. Intervals formed with $\widehat{\sigma}$ tend to be too short, too optimistic. Note, however, that the bias is $O(1/T)$. In MA(q) models, the ML estimators under normality of $\sigma^2 = \text{Var}(a_t)$, satisfy $E(\widehat{\sigma}_{ML} - \sigma^2) \approx -k\sigma^2/T$, where k is the number of parameters in the model, including $\mu = E(X_t)$; hence, $k = 2$ for the MA(1) model, $k = 3$ for the MA(2) models, etc., when μ is unknown.

In the analytical part of our study we concentrate in the study of the asymptotic biases of the estimators by the method of moments (MM). We rely on Taylor-type expansions, and search for results to $O(1/T)$, where T is sample size. Incidentally, for simplicity we also tried to rely on first-order Taylor expansions, but found that for MA models they can be misleading, in that for certain ranges of the parameter space, they gave results with the wrong sign.

Section 4.2.1 contains an expansion valid for $q \geq 1$. However, our explicit results, analyses and simulations are for $q = 1$.

For the MA(1) model, Figure 1 shows the behaviour of the asymptotic bias for the MM estimator, and it can be compared with the corresponding value for the ML estimator, namely -2 for all values of α: (1) $\hat{\sigma}^2_{MM}$ underestimates σ^2 for all α; (2) For approximately $|\alpha| \leq 0.5$, $\hat{\sigma}^2_{MM}$ and $\hat{\sigma}^2_{ML}$ have approximately the same (asymptotic) bias; (3) For approximately $|\alpha| > 0.5$, the negative bias of $\hat{\sigma}^2_{MM}$ is larger than that of $\hat{\sigma}^2_{ML}$; (4) The asymptotic bias of $\hat{\sigma}^2_{MM}$ tends to $-\infty$ as $|\alpha|$ tends to 1.

Our simulations tend to confirm some of our expectations, for example, that the ML estimator works better than the MM, or that better fits are obtained for larger sample sizes. The fit of the simulated results to the asymptotic theory is much better for ML than for MM procedures.

Correction for bias when using ML estimators of the residual variance is simple, since it does not depend on the parameter values. A word of caution should be expressed when MM estimators are used for MA(1) models whose parameters tend to be near the region of invertibility, since for the usual sample sizes, we expect to find difficulties in using the available asymptotic reults for the biases.

7. ACKNOWLEDGEMENTS

We express our appreciation to a referee for comments that led to a substantial improvement of the paper, to Professor Carlos E. Harle (São Paulo) for his help with some mathematical problems, and to Eng. Carlos I. Martinez (Tucumán) for his help with the computations. Interchange of visits by the authors were facilitated by a research subsidy from "Fundación Antorchas".

8. BIBLIOGRAPHY

Anderson, T.W. (1971). *The Statistical Analysis of Time Series.* New York: John Wiley and Sons.

Anderson, T.W. and Takemura, A. (1986). Why do noninvertible, estimated moving averages occur? *J. of Time Series Analysis*, **7**, 235–254.

Brockwell, P.J. and Davis, R.A. (1991a). *Time Series: Theory and Methods.* New York: Springer Verlag.

——— (1991b). *ITSM: An Interactive Time Series Modelling Package for the PC.* New York: Springer Verlag.

Cordeiro, G.M. and Klein, R. (1994). Bias correction of maximum likelihood estimates for ARMA models. *Stat. and Prob. Letters*, 169–176.

Cryer, J.D. and Ledolter, J. (1981). Small-sample properties of the maximum likelihood estimator in the first-order moving average model. *Biometrika*, **68**, 691–694.

Davis, W.W. (1977). Robust interval estimation of the innovation variance of an ARMA model. *The Annals of Statistics*, **5**, 700–708.

Fuller, W.A. (1996). *Introduction to Statistical Time Series.* New York: John Wiley and Sons.

Mentz, R.P. (1976). On the inverse of some covariance matrices of Toeplitz type. *SIAM J. Appl. Math.*, **31**, 326–337.

Mentz, R.P., Morettin, P.A. and Toloi, C.M.C. (1995). On residual variance estimation in autoregressive models. Forthcoming.

_____ (1995). Residual variance estimation in moving average models. Cuaderno 68, INIE, U. Nacional de Tucumán.

Pesaran, M.H. (1983). A note on the maximum likelihood estimation of regression models with first order moving average errors with roots on the unit circle. *Australian J. of Stat.*, **25**, 442–448.

Porat, B. and Friedlander, B. (1986). On the estimation of variance for autoregressive and moving average processes. *IEEE Trans. on Inf. Theory*, IT-32, No.1, 120–125.

Priestley, M.B. (1981). *Spectral Analysis and Time Series.* London: Academic Press.

Shaman, P. (1969). On the inverse of the covariance matrix of a first-order moving average. *Biometrika*, **60**, 193–196.

Tanaka, K. (1984). An asymptotic expansion associated with the maximum likelihood estimator in ARMA models. *Journal of the Royal Stat. Soc.*, Series B, **46**, 58–67.

Received August, 1995; Revised April, 1997.