THE INFLUENCE OF NUMERICAL AND OBSERVATIONAL ERRORS ON THE LIKELIHOOD OF AN ARMA SERIES

By S. Rao Jammalamadaka, Chengou Wu and Weiqung Wang

University of California, Santa Barbara, and Nanjing Institute of Meteorology, China

First version received March 1992

Abstract. Formulae are established describing how the round-off and observational errors influence the reduced likelihood function, as a step towards providing guidelines for such likelihood computations in a time series context. Error bounds are set up for Ali's method.

Keywords. ARMA series; likelihood function; error analysis; measurement error; round-off error.

1. INTRODUCTION

When p and q are given, the maximum likelihood (ML) method provides a precise way for estimating the parameters of an autoregressive moving-average (ARMA) series. In the general case, there are three important methods for calculating the likelihood function: the Box–Jenkins method, Ali's method and the innovations algorithm. Box and Jenkins (1970, Ch. 7) developed an iterative method to calculate the approximate likelihood. Galbraith and Galbraith (1974) and Ali (1977) use Woodbury's formula while Chen and Gu (1983) use a partition matrix technique to calculate the inverse of the correlation matrix. All of these provide fast recurrence methods. Harvey and Phillips (1979), Rissanen and Barbosa (1969), Kailath (1968, 1970), Ansley (1979) and others developed the innovations algorithm, while Pan (1981) uses a matrix method, getting the same recurrence procedure. In this paper, we focus on Ali's method.

These are basically three kinds of errors that influence the accuracy of calculation in a statistical procedure: the measurement error, the method error, and the round-off error. We discuss here how these different errors influence the computation of the likelihood in a time series context. In general, most statisticians do not pay much attention to the round-off errors and that can be castastrophic as Faye and Vignes (1985) pointed out. Sometimes the entire calculation can be meaningless because of round-off errors, as we show in our Example 1. Koreisha and Pukkila (1990) also point out some problems arising out of round-off errors in ML estimation. Moreover, the observations data are not always measured exactly, often being in error by 1%–10% of the exact value, as for example when we measure temperature. In this paper, we find that the influence of these errors on the computations cannot be ignored. We develop 'error analysis' formulae based on Wilkinson's (1963) work, to describe how these errors propagate through the large number of computational steps that are involved.

Notations and assumptions about round-off error as well as some basic formulae are introduced in Section 2. In Section 3, we establish the error formulae for Ali's method and consider some asymptotic properties of the propagation of round-off error, while Section 4 provides brief conclusions.

2. NOTATION AND FORMULAE

Consider $\{y_t\}$, an invertible and stationary ARMA series,

$$y_t - \phi_1 y_{t-1} - \cdots - \phi_p y_{t-p} = a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} \tag{2.1}$$

where a_t are independent random variables from $N(0, \sigma^2)$ while σ, $\phi_1, \ldots, \phi_p$ and $\theta_1, \ldots, \theta_q$ are parameters. To simplify notation, we will replace p and q by $r = \max(p, q)$ with the understanding that some coefficients ϕ_i or θ_j are zero. Suppose now $X = (x_1, \ldots, x_N)^T$ denotes the observation vector with error vector $\eta = (\eta_1, \ldots, \eta_N)^T$; i.e. $X = Y + \eta$, where we assume that these observational errors $\{\eta_t\}$ are independent variables from the uniform distribution $U(-0.5d, 0.5d)$ for some $d > 0$. Write the above equations for $t = 1, 2, \ldots, N$ in the following partitioned form:

$$\left(\begin{array}{cccc|cccccccc}
-\phi_r & \cdots & \cdots & -\phi_1 & 1 & 0 & \cdots & & & & & 0 \\
0 & -\phi_r & \cdots & -\phi_2 & -\phi_1 & 1 & & & & & & \vdots \\
0 & 0 & \ddots & \vdots & \vdots & \ddots & \ddots & & & & & \\
0 & 0 & 0 & -\phi_r & -\phi_{r-1} & & & & & & & \\
0 & 0 & 0 & 0 & -\phi_r & & & -\phi_1 & 1 & & & \\
\vdots & \vdots & \vdots & \vdots & 0 & \ddots & & & \ddots & \ddots & & 0 \\
0 & 0 & 0 & 0 & \cdots & 0 & -\phi_r & \cdots & & -\phi_1 & & 1
\end{array}\right)
\times \left(\begin{array}{c} y_{1-r} \\ \vdots \\ y_0 \\ \hline y_1 \\ y_2 \\ \vdots \\ y_N \end{array}\right)$$

$$
= \left(\begin{array}{cccc|ccccccc}
-\theta_r & \cdots & \cdots & -\theta_1 & 1 & 0 & \cdots & & & & 0 \\
0 & -\theta_r & \cdots & -\theta_2 & -\theta_1 & 1 & & & & & \vdots \\
0 & 0 & \ddots & \vdots & \vdots & \ddots & \ddots & & & & \\
0 & 0 & 0 & -\theta_r & -\theta_{r-1} & & & & & & \\
0 & 0 & 0 & 0 & -\theta_r & & & -\theta_1 & 1 & & \\
\vdots & \vdots & \vdots & \vdots & 0 & \ddots & & & \ddots & \ddots & 0 \\
0 & 0 & 0 & 0 & \cdots & 0 & -\theta_r & \cdots & & -\theta_1 & 1
\end{array}\right)
\times \left(\begin{array}{c} a_{1-r} \\ \vdots \\ a_0 \\ \hline a_1 \\ a_2 \\ \vdots \\ a_N \end{array}\right)
$$

which is used to define the matrices in the corresponding equation:

$$
\left(\frac{\Phi^*}{0}\middle|\Phi\right)\left(\frac{Y^*}{Y}\right) = \left(\frac{\Theta^*}{0}\middle|\Theta\right)\left(\frac{a^*}{a}\right).
$$

Then let

$$
Z = \Phi Y = \Theta a + \begin{pmatrix} \Theta^* a^* - \Phi^* Y^* \\ 0 \end{pmatrix} = \Theta a + \begin{pmatrix} d \\ 0 \end{pmatrix} = \Theta a + D
$$

where d is of length r and is uncorrelated with a. Considering the covariance matrix of Z then gives the expression involving the covariance matrix Γ of Y, namely

$$
\Phi \Gamma \Phi^{\mathrm{T}} = \Theta \Theta^{\mathrm{T}} \sigma^2 + \sigma^2 \begin{pmatrix} P & 0 \\ 0 & 0 \end{pmatrix}
$$

where $\sigma^2 P$ is the $r \times r$ covariance matrix of d which may be calculated directly from the ARMA model parameters.

We denote the reduced likelihood by $S = \sigma^2 Y^{\mathrm{T}} \Gamma^{-1} Y$ and its actual computed value, which is influenced by both round-off errors and observational errors, by S_r. Writing $\Theta^{-\mathrm{T}} = (\Theta^{\mathrm{T}})^{-1}$, by Ali's method we have

$$
\begin{aligned}
S &= Y^{\mathrm{T}}\Phi^{\mathrm{T}}\left\{\Theta\Theta^{\mathrm{T}} + \begin{bmatrix} P & 0 \\ 0 & 0 \end{bmatrix}\right\}^{-1}\Phi Y \\
&= Y^{\mathrm{T}}\Phi^{\mathrm{T}}\left\{\Theta^{-\mathrm{T}}\Theta^{-1} - \Theta^{-\mathrm{T}}\Theta^{-1}\begin{bmatrix} M & 0 \\ 0 & 0 \end{bmatrix}\Theta^{-\mathrm{T}}\Theta^{-1}\right\}\Phi Y \qquad (2.2)
\end{aligned}
$$

where an expression for M is

$$M = ([\Theta^{-\mathrm{T}}\Theta^{-1}] + P^{-1})^{-1}$$

and $[\Theta^{-\mathrm{T}}\Theta^{-1}]$ is the first $r \times r$ block of $\Theta^{-\mathrm{T}}\Theta^{-1}$. This expression is derived using a standard matrix inverse identity. It may also be derived from the likelihood expressions which naturally provide an alternative form which is very similar but much less prone to numerical error.

If both T and d were observed and hence both Z and d, the reduced likelihood would be

$$S = a^{\mathrm{T}}a + d^{\mathrm{T}}P^{-1}d = [\Theta^{-1}(Z - D)]^{\mathrm{T}}\Theta^{-1}(Z - D) + d^{\mathrm{T}}P^{-1}d.$$

The required reduced likelihood is achieved by replacing d in S by its minimizing value

$$\hat{d} = ([\Theta^{-\mathrm{T}}\Theta^{-1}] + P^{-1})^{-1}[\Theta^{-\mathrm{T}}]\Theta^{-1}Z$$

where $[\Theta^{-\mathrm{T}}]$ is the matrix formed by the first r rows of $\Theta^{-\mathrm{T}}$.

Substituting this value of $\hat{d}$ in S, expanding and simplifying also leads to (2.2). It may, however, be employed directly in the alternative form $S = \hat{a}^{\mathrm{T}}\hat{a} + \hat{d}^{\mathrm{T}}P^{-1}\hat{d}$ where $\hat{a}\Theta^{-1}(Z - D)$. Expression (2.2) is widely used and we investigate the numerical errors in its computation. The alternative form is a sum of positive quantities. We have seen in particular examples that the latter form eliminates most of the numerical error obtained when using (2.2). Note that $\Theta^{-1}u$ and $\Theta^{-\mathrm{T}}u$ may be calculated by recurrence as follows. Let

$$\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{bmatrix} = \Theta^{-1}\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{bmatrix} = \Theta^{-\mathrm{T}}\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix}.$$

Then, we have

$$
\begin{aligned}
b_1 &= u_1 \\
b_2 &= u_2 + \theta_1 b_1 \\
&\vdots \\
b_{p+1} &= u_{p+1} + \theta_1 b_p + \cdots + \theta_p b_1 \\
b_{p+2} &= u_{p+2} + \theta_1 b_{p+1} + \cdots + \theta_p b_2 \\
&\vdots \\
b_N &= u_N + \theta_1 b_{N-1} + \cdots + \theta_p b_{N-p}
\end{aligned}
\tag{2.3}
$$

and

$$
\begin{aligned}
c_N &= u_N \\
c_{N-1} &= u_{N-1} + \theta_1 c_N \\
&\vdots \\
c_1 &= u_1 + \theta_1 c_2 + \cdots + \theta_p c_{p+1}.
\end{aligned}
\tag{2.4}
$$

Recall that

$$
\Theta^* = \begin{bmatrix} -\theta_q & \cdot & \cdots & -\theta_1 \\ \cdot & -\theta_q & \cdots & -\theta_2 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & -\theta_q \end{bmatrix} \qquad \Phi^* = \begin{bmatrix} -\phi_p & \cdot & \cdots & -\phi_1 \\ \cdot & -\phi_p & \cdots & -\phi_2 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & -\phi_p \end{bmatrix}
$$

$$
Y^* = \begin{bmatrix} y_{1-p} \\ \vdots \\ y_0 \end{bmatrix} \qquad a^* = \begin{bmatrix} a_{1-q} \\ \vdots \\ a_0 \end{bmatrix} \qquad a = \begin{bmatrix} a_1 \\ \vdots \\ a_N \end{bmatrix}
$$

$$
Y^0 = \begin{bmatrix} y_1 \\ y_2 - \phi_1 y_1 \\ \vdots \\ y_r - \phi_1 y_{r-1} - \cdots - \phi_r y_0 \end{bmatrix} \qquad a^0 = \begin{bmatrix} a_{r+1}\theta_1 a_r - \cdots - \theta_q a_{r-q-1} \\ \vdots \\ a_N - \theta_1 a_{N-1} - \cdots - \theta_q a_{N-q} \end{bmatrix}.
$$

Then

$$
\Phi Y = \begin{bmatrix} y^0 \\ a^0 \end{bmatrix} \tag{2.5}
$$

and

$$E(Y^* a^{0^{\mathrm{T}}}) = 0. \tag{2.6}$$

Let

$$\tilde{Z} = \begin{bmatrix} -\phi_p & \cdots & -\phi_1 & 1 & \cdots & \cdots & 0 \\ 0 & -\phi_p & \cdots & -\phi_1 & 1 & \cdots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & 0 & -\phi_p & \cdots & -\phi_1 & 1 \end{bmatrix} \begin{pmatrix} Y^* \\ Y \end{pmatrix}$$

$$= \begin{pmatrix} -\Phi^* Y^* \\ 0 \end{pmatrix} + \Phi Y. \tag{2.7}$$

From (2.5), (2.6) and (2.7), we get

$$E(\tilde{Z}\tilde{Z}^{\mathrm{T}}) = (\Phi\Gamma\Phi^{\mathrm{T}} + \begin{bmatrix} \Phi^* E(Y^* Y^{0^{\mathrm{T}}}) & 0 \\ 0 & 0 \end{bmatrix}$$

$$+ \begin{bmatrix} E(Y^* Y^{0^{\mathrm{T}}})\Phi^{*^{\mathrm{T}}} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} \Phi^*\Gamma\Phi^{*^{\mathrm{T}}} & 0 \\ 0 & 0 \end{bmatrix} \tag{2.8}$$

with

$$\Gamma_p = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{p-1} \\ \gamma_1 & \gamma_2 & \cdots & \cdots \\ \cdots & \cdots & \cdots & \gamma_1 \\ \gamma_{p-1} & \cdots & \gamma_1 & \gamma_0 \end{bmatrix}.$$

On the other hand, we can also write

$$\tilde{Z} = \begin{bmatrix} -\theta_q & \cdots & -\theta_1 & 1 & 0 & \cdots & 0 \\ 0 & -\theta_q & \cdots & -\theta_1 & 1 & \cdots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & 0 & -\theta_q & \cdots & -\theta_1 & 1 \end{bmatrix} \begin{pmatrix} a^* \\ a \end{pmatrix}$$

$$= \begin{pmatrix} \Theta^* a^* \\ 0 \end{pmatrix} + \Theta a$$

so that

$$E(\tilde{Z}\tilde{Z}^{\mathrm{T}}) = \sigma^2 \Theta\Theta^{\mathrm{T}} \begin{bmatrix} \Theta^*\Theta^{*^{\mathrm{T}}} & 0 \\ 0 & 0 \end{bmatrix}. \tag{2.9}$$

Comparing (2.8) with (2.9), we get

$$\Phi\Gamma\Phi^{\mathrm{T}} = \sigma^2\left(\Theta\Theta^{\mathrm{T}} + \begin{bmatrix} P & 0 \\ 0 & 0 \end{bmatrix}\right) \tag{2.10}$$

and

$$\Gamma = \sigma^2\left(\Phi^{-1}\Theta\Theta^{\mathrm{T}} + \begin{bmatrix} M & 0 \\ 0 & 0 \end{bmatrix}\Phi^{-\mathrm{T}}\right)$$

so that

$$\sigma^2\Gamma^{-1} = \Phi^{\mathrm{T}}\left(\Theta^{-\mathrm{T}}\Theta^{-1} - \Theta^{-\mathrm{T}}\Theta^{-1}\begin{bmatrix} M & 0 \\ 0 & 0 \end{bmatrix}\Theta^{-\mathrm{T}}\Theta^{-1}\right)\Phi^{-\mathrm{T}}.$$

We now make the following standard assumptions regarding the round-off error, during the numerical computations. Because the calculational and observational errors are small, their products are neglected.

ASSUMPTION 1. The error in a product ab is $ab\Psi$ where $|\Psi| < 2^{-g}$ and g is a characteristic of the computer itself. Hence, the error in $\prod_{i=1}^{k} a_i$ is $\sum_{j=1}^{k-1}\Psi_j\prod_{i=1}^{k} a_i$.

ASSUMPTION 2. The error in a sum $a+b$ is $a\Psi_1 + b\Psi_2$. Hence, the error in $\sum_{i=1}^{k} a_i$ is $\sum_{i=1}^{k} a_i\Psi_i + \sum_{i=1}^{k-1} a_i \sum_{j=1}^{k-1}\xi_j$, with $\xi_1 = 0$ and $|\xi_j| < 2^{-g}$.

From these assumptions, we have the following proposition.

PROPOSITION 2.1. If the order of calculating $b + \sum a_i c_i$ is 'first compute $\sum a_i c_i$ and then add b', the accumulated calculated error in Φu is given by $\{(\Phi - I)@(\Psi_1 + \Psi_2 + \xi_1 L_1)\}u + \Psi_3@u$. By the recurrence method, the error in $\Phi^{-1}V$ is $\Theta^{-1}[\{(I-\Theta)@(\Psi_4 + \xi_2 L_2 + \Psi_5)\}\Theta^{-1}v + \Psi_6@v]$, where @ stands for the Hadamard product and

$$\xi = \begin{bmatrix} 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & \cdots & \cdots & \cdots & 0 \\ \xi_{3,1} & 0 & 0 & \cdots & \cdots & 0 \\ \xi_{4,2} & \xi_{4,1} & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \xi_{N,p-1} & \cdots & \xi_{N,1} & 0 & 0 \end{bmatrix} \qquad L_1 = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 0 & 1 & \cdots & 1 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & 1 \end{bmatrix}$$

with ξ_2 and L_2 having similar forms.

The proof of the following proposition is straightforward.

PROPOSITION 2.2. Denote $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$, $\theta(z) = 1 - \theta_1 z - \cdots - \theta_q z^q$. Let u_i, v_i be the coefficients of z^{i-1} in the power series of $\{\theta(z)\}^{-1}$ or $\{\phi(z)\}^{-1}$. Then

$$\Theta^{-1} = \begin{bmatrix} u_1 & 0 & \cdots & 0 \\ u_2 & u_1 & 0 & \cdots \\ \cdots & \cdots & \cdots & 0 \\ u_N & \cdots & u_2 & u_1 \end{bmatrix} \qquad \Phi^{-1} = \begin{bmatrix} v_1 & 0 & \cdots & 0 \\ v_2 & v_1 & 0 & \cdots \\ \cdots & \cdots & \cdots & 0 \\ v_N & \cdots & v_2 & v_1 \end{bmatrix}$$

and

$$\Theta^{-\mathrm{T}}\Theta^{-1} = \begin{bmatrix} \sum_1^N u_j^2 & \sum_1^{N-1} u_j u_{j+1} & \cdots & u_N \\ & \sum_1^{N-1} u_j^2 & \cdots & u_{N-1} \\ & & \cdots & \cdots \\ & & & u_1 \end{bmatrix}$$

where $u_k = \mathrm{O}(\lambda^{-k})$, $v_k = \mathrm{O}(\mu^{-k})$, $0 = u_0 = u_{-1} = \cdots$, $0 = v_0 = v_{-1} = \cdots$, $\lambda = \min\{|\lambda_1|, |\lambda_2|, \ldots, |\lambda_q|\}$, $\mu = \min\{|\mu_1|, |\mu_2|, \ldots, |\mu_p|\}$ and $\{\lambda_i\}$ are the roots of $\theta(z)$, $\{\mu_i\}$ are the roots of $\phi(z)$.

3. ALI'S METHOD

For Ali's method, it is easy to see that the observational errors do not influence M in (2.2) at all, because p, q are small while N is relatively large. We suppose that M is independent of N and the calculational errors in M can be neglected. From Proposition 2.2 we obtain the following theorem.

THEOREM 3.1

$$S_r = V^{\mathrm{T}}V - Z^{\mathrm{T}}\begin{bmatrix} M & 0 \\ 0 & 0 \end{bmatrix} Z + \varepsilon_4$$

where

$$U = \Phi X = \varepsilon_1 \qquad \varepsilon_1 = \{(\Phi - I)@(\Psi_1 + \Psi_2 + \xi_1 L_1)\}Y + \Psi_3@Y$$

$$V = \Theta^{-1}(U + \varepsilon_2) \qquad \varepsilon_2 = \{(I - \Theta)@(\Psi_4 + \Psi_5 + \xi_2 L_2)\}\Theta^{-1}U + \Psi_6@U$$

$$Z = \theta^{-\mathrm{T}}(V + \varepsilon_3) \qquad \varepsilon_3 = T\{(\Theta - I)@(\Psi_7 + \Psi_8 + \xi_3 L_3)\}\Theta^{-1}V + \Psi_9@V$$

$$\varepsilon_4 = \sum_1^N v_i^2(\Psi_{ai} + \Psi_{bi}) + \sum_1^{N-1} v_i^2 \sum_1^{N-1} \xi_{aj} + \sum_{i,j=1}^{r} m_{ij}z_i z_j(\Psi_{aij} + \Psi_{bij} + \Psi_{cij})$$

$$+ \sum_{i=1}^{r}\sum_{j=1}^{r-1} m_{ij}z_i z_j \sum_{j}^{r-1} \xi_{bik} + \sum_1^r \xi_i \sum_1^r m_{ij}z_i z_j + \sum_{i=1}^{r-1}\sum_{j=1}^{r} m_{ij}z_i z_j \sum_{i}^{r-1} \xi_{ck}$$

$$+ \Psi_d \sum_1^N v_i^2 + \Psi_e \sum_{i,j=1}^{r} m_{ij}z_i z_j.$$

We now consider an asymptotic property of these accumulated errors, in the following.

LEMMA 3.1

$$\lim \frac{1}{N}\sum_{1}^{N}\left|a_k \sum_{1}^{N-k+1} u_j a_{k+j-1}\right| = E\left|a_1 \sum_{1}^{\infty} u_j a_j\right| = c_0 \tag{3.1}$$

$$\lim \sum_{1}^{N} \frac{|a_{k-i}\sum_{1}^{N-k+1} u_j a_{k+j-1}|}{N} = E|a_1|E\left|a_1 \sum_{1}^{\infty} u_j a_j\right| = c = 2\frac{\sigma^2}{\pi}\left(\sum_{1}^{\infty} u_j^2\right)^{1/2} \tag{3.2}$$

PROOF. We prove only (3.1) since (3.2) is similar. Let

$$f_{N,k} = \sum_{1}^{N-k+1} u_j a_{k+j-1} \qquad f_N = \sum_{1}^{\infty} u_j a_{k+j-1}.$$

Then

$$\left(\sum_{1}^{N} |f_{Nk} a_k| - c_0\right)^2 \leqslant 2\left\{\sum_{1}^{N} |a_k|(|f_{Nk}| - |f_k|)\right\}^2 + 2\left(\sum_{1}^{N} |f_k a_k| - c_0\right)^2$$

$$\left\{\sum_{1}^{N} |a_k|(|f_{Nk}| - |f_k|)\right\}^2 \leqslant 2\left\{\sum_{1}^{N-m} |a_k|(|f_{Nk}| - |f_k|)\right\}^2 + 2\left\{\sum_{N-m+1}^{N} |a_k|(|f_{Nk}| - |f_k|)\right\}^2$$

where m is selected so that $E\{|a_k|(|f_{Nk}| - |f_k|)\}^2 < \epsilon$ when $k < N - m$. Since $E\{|a_k|(|f_{Nk}| - |f_k|)\}^2 < c_1$, when N is large enough,

$$\frac{1}{N} E\left\{\sum_{1}^{N} |a_k|(|f_{Nk}| - |f_k|)\right\}^2 \leqslant 3\epsilon.$$

On the other hand,

$$E\left\{\sum_{1}^{N}(|a_k f_k| - c_0)\right\}^2$$

$$= NE(|a_1 f_1| - c_0)^2 + 2(N-1)E\{(|a_1 f_1| - c_0)(|a_2 f_2| - c_0)\} + \cdots$$

$$+ 2E\{(|a_1 f_1| - c_0)(|a_N f_N| - c_0)\}.$$

However,

$$\begin{aligned} E\{(|a_1 f_1| - c_0)(|a_k f_k| - c_0) &= E(|a_1 f_1 a_k f_k| - c_0^2) \\ &= E\left\{\left(|a_1|\left|\sum_1^{k-1} + \sum_k^{\infty} u_j a_j\right|\right)|a_k f_k|\right\} - c_0^2 \\ &= E\left\{\left(\left|a_1 \sum_1^{k-1} u_j a_j\right|\right\} c_0 - c_0^2 + \mathrm{o}(1) \\ &= \mathrm{o}(1). \end{aligned}$$

Hence, (3.1) is proved.

Analogously, we can obtain the next lemma.

LEMMA 3.2

$$\lim \frac{1}{N}\sum_1^N \left| y_{k-i} \sum_1^{N-k+1} u_j a_{k+j-1}\right| = E y_{k-i} \sum_1^{\infty} u_j a_j \Big| = d_i \tag{3.3}$$

where

$$d_i = d = \sigma\left(\gamma_0 \sum_1^{\infty} u_j^2\right)^{1/2} \qquad i = 1, 2, \ldots .$$

Since

$$\begin{aligned} S_r - S &= 2(\Phi\eta + \varepsilon_1 + \varepsilon_2)^{\mathrm{T}} \Theta^{-\mathrm{T}} \Theta^{-1} \Phi Y \\ &\quad - 2\{\Theta^{-1}(\Phi\eta + \varepsilon_1 + \varepsilon_2) + \varepsilon_3\}^{\mathrm{T}} \Theta^{-1} \begin{bmatrix} M & 0 \\ 0 & 0 \end{bmatrix} \Theta^{-\mathrm{T}} \Theta^{-1} \Phi Y \end{aligned}$$

we can estimate the influence of ε_1, ε_2 and ε_3. Let $a = (a_1, \ldots, a_N)^{\mathrm{T}}$.

THEOREM 3.2. *If ϕ_i, θ_j are true parameters*

$$\frac{\eta^{\mathrm{T}} \Phi^{\mathrm{T}} \Theta^{-\mathrm{T}} \Theta^{-1} \Phi Y}{N} = \mathrm{o}(1) \tag{3.4}$$

$$\frac{1}{N} |\varepsilon_1 \theta^{-\mathrm{T}} \Theta^{-1} \Phi Y| \leqslant \rho \left\{ d_0 + \sum_1^p (p - t + 3)|\phi_t| d \right\} + \mathrm{o}(1) \tag{3.5}$$

$$\frac{1}{N}|\varepsilon_2^{\mathrm{T}}\Theta^{-\mathrm{T}}\Theta^{-1}\Phi y| \leqslant \rho\left\{c_0 + \sum_1^q (q-t+4)|\theta_t|c\right\} + \mathrm{o}(1) \tag{3.6}$$

$$\frac{1}{N}\left[\{\varepsilon_3 + \Theta^{-1}(\varepsilon_1 + \varepsilon_2 + \Phi\eta)\}^{\mathrm{T}}\Theta^{-1}\begin{bmatrix} M & 0 \\ 0 & 0 \end{bmatrix}\Theta^{-\mathrm{T}}\Theta^{-1}\Phi Y\right] = \mathrm{o}(1). \tag{3.7}$$

Proof. Consider

$$\begin{aligned}
\frac{1}{N}\eta^{\mathrm{T}}\Phi^{\mathrm{T}}\Theta^{-\mathrm{T}}\Theta^{-1}\Phi Y &= \frac{1}{N}\eta^{\mathrm{T}}\Phi^{\mathrm{T}}\Theta^{-\mathrm{T}}a + \mathrm{o}(1) \\
&= \frac{1}{N}\sum_1^N \phi(B)u_k \sum_1^{N-k-1}\eta_j a_{j+k-1} + \mathrm{o}(1) \\
&= \mathrm{o}(1).
\end{aligned}$$

By Lemma 3.2, it is easy to see that

$$\begin{aligned}
\frac{1}{N}|[\{(\Phi - I)@\Psi_1\}Y]^{\mathrm{T}}\Phi^{-\mathrm{T}}\Phi^{-1}\Phi Y| &\leqslant \frac{1}{N}\sum_1^p \rho|\phi_t|\sum_1^N |y_{k-t}f_{Nk}| + \mathrm{o}(1) \\
&= \rho\sum_1^p |\phi_t|d + \mathrm{o}(1)
\end{aligned}$$

$$\begin{aligned}
\frac{1}{N}|(\Psi_3@Y)^{\mathrm{T}}\Theta^{-\mathrm{T}}\Theta^{-1}\Phi Y| &\leqslant \frac{1}{N}\rho\sum_1^N |y_{k-t}f_{Nk}| + \mathrm{o}(1) \\
&= \rho d_0 + \mathrm{o}(1)
\end{aligned}$$

and

$$\begin{aligned}
&\frac{1}{N}|[\{\Phi - I)@\xi_1 L_1\}Y]^{\mathrm{T}}\Theta^{-\mathrm{T}}\Theta^{-1}\Phi Y| \\
&\leqslant \frac{1}{N}\sum_1^p \rho(p-t+1)|\phi_t|\sum_1^N |y_{k-t}f_{Nk}| + \mathrm{o}(1) \\
&= \rho\sum_1^p |\phi_t|(p-t+1)d_t + \mathrm{o}(1).
\end{aligned}$$

Hence, (3.5) holds. The proof of (3.6) is analogous. Let α_i be the ith column of Θ^{-1}, $b = \varepsilon_3 + \Theta^{-1}(\varepsilon_1 + \varepsilon_2 + \Phi\eta)$, $h = \Theta^{-1}\Phi Y = a + \mathrm{o}(1)$. There exists c_2 such that $\|b_i\| < c_2\|h_i\| < c_2$ for all i. Hence,

$$\frac{1}{N}\left|b^{\mathrm{T}}\Theta^{-1}\begin{bmatrix} M & 0 \\ 0 & 0 \end{bmatrix}\Theta^{-\mathrm{T}}h\right| = \frac{1}{N}\left\|\sum m_{ij}(b^{\mathrm{T}}a_i)(a_j^{\mathrm{T}}h)\right\|$$
$$= \mathrm{o}(1).$$

Our results indicate that, because of the large number of computational steps involved in computing the likelihood, the numerical errors accumulate and can become huge, especially when the u_j do not decrease. This can result in gross errors in parameter estimation. Using double or triple precision (i.e. increasing the g in Assumptions 1 and 2) would alleviate this numerical problem to some extent. Our results also indicate that the observational errors sometimes play a more important role and overwhelm the numerical errors. In practice, the errors will counteract each other, so that the influence of ε_1, ε_2 and ε_3 will become smaller. On the other hand, if $v^{\mathrm{T}}v$, $\sum m_{ij}z_iz_j$ are large, the error in calculation of M may be large. Hence, a large propagation of round-off error may be produced, as the following example indicates.

EXAMPLE 1. We consider ARMA(0, 15), i.e. $y_t = (1 - 0.5B)^{15}a_t$, and take 380 pseudo-random numbers from N(0, 1) as $\{a_t\}$ to get Y with $N = 365$. Although the exact value of $S = 344.123$, using Ali's method we obtain $\sum m_{ij}z_iz_j = 0.523944 \times 10^{12}$ and $S_r = 0.358259 \times 10^{10}$, whereas $S_r = 0.5768 \times 10^3$ by the innovation algorithm. When using triple precision, these two methods yield 5000 and 350 respectively. Convergence does not occur at all by the Box–Jenkins method. The problem here is that the u_j are large and do not decrease with j.

When $d^2/2^{-g}$ is large, $\eta^{\mathrm{T}}\Phi^{\mathrm{T}}\Theta^{-\mathrm{T}}\Theta^{-1}F\eta$ cannot be neglected. By a lengthy deduction analogous to arguments in Lemma 3.1, we can establish the following theorem.

THEOREM 3.3

$$\lim \frac{1}{N}\eta^{\mathrm{T}}\Phi^{\mathrm{T}}\Theta^{-\mathrm{T}}\Theta^{\mathrm{T}}\Theta\eta = \frac{1}{12}\delta^2\sum_{1}^{\infty}\{\Phi(B)u_j\}^2.$$

When the η_i are not independent, (3.4) does not hold and their influence is large because of the following theorem.

THEOREM 3.4. *When q_i, f_j are true parameters,*

$$\frac{1}{N}|\eta^{\mathrm{T}}\Phi^{\mathrm{T}}\Theta^{-\mathrm{T}}\Theta^{-1}\Theta Y| \leqslant \frac{1}{2}\delta\sigma\left[\sum_{1}^{\infty}\{\Phi(B)u_j\}^2\right]^{1/2} + \mathrm{o}(1).$$

4. CONCLUSIONS

1. When γ_0 or $\sum u_j^2$ is very large, as for example when one of the roots of $\phi(z)\theta(z)$ is near the unit circle, round-off errors play an important role when computing the ML estimator.

2. Ali's method is convenient for theoretical deductions, but from the viewpoint of error analysis it is an ill-conditioned method. Because $\lim \nu_n = 1$, the innovations method is not an ill-conditioned algorithm.

3. When the round-off errors are considered as random variables (see Faye and Vignes, 1985), the variance of the ML estimator can be highly inflated.

4. If $\delta^2/2^{-g}$ is large or if the observational errors are not independent, their influence could become larger than the influence of the round-off error.

ACKNOWLEDGEMENTS

We thank an anonymous referee for a very thorough and careful reading and detailed comments which considerably improved the presentation.

REFERENCES

Ali, M. M. (1977) Analysis of autoregressive-moving average models: estimation and prediction. *Biometrika* 64, 535–45.

Ansley, C. F. (1979) An algorithm for the exact likelihood of mixed autoregressive-moving average process. *Biometrika* 66, 59–65.

Box, G. E. P. and Jenkins, G. M. (1970) *Time Series Analysis: Forecasting and Control*. San Francisco, CA: Holden Day.

Chen, Z. G. and Gu, L. (1983) On the inverse of the covariance matrix and the parameter estimation of ARMA series. *Acta Math. Appl. Sin.* 6, 257–66.

Faye, J. P. and Vignes, J. (1985) Stochastic approach of the permutation–perturbation method for round-off error analysis. *J. Appl. Numer. Math.* 1, 349–62.

Galbraith, R. F. and Galbraith, J. I. (1974) On the inverse of some patterned matrices arising in the theory of stationary time series. *J. Appl. Probab.* 11, 63–71.

Harvey, A. C. and Phillips, G. D. A. (1979) Maximum likelihood estimation of regression model with autogressive-moving average disturbances. *Biometrika* 66, 49–58.

Kailath, T. (1968) An innovations approach to least square estimation—Part 1: Linear filtering in additive white noise. *IEEE Trans. Autom. Control* AC-13, 646–54.

—— (1970) The innovations approach to detection and estimation theory. *Proc. IEEE* 58, 680–95.

Koreisha, S. and Pukkila, T. (1990) A generalized least squares approach for estimation of autoregressive-moving averge models. *J. Time Ser. Anal.* 11, 139–51.

Pan, Y. M. (1981) The recursion algorithm for a kind of statistic. *Acta Math. Appl. Sin.* 4, 190–95.

Rissanen, J. and Barbosa, L. (1969) Properties of infinite covariance matrices and stability of optimum predictors. *Inf. Sci.* 1, 221–36.

Wilkinson, J. (1963) *Rounding Errors in Algebraic Processes*. Englewood Cliffs, NJ: Prentice-Hall.