Discussion Paper
8401
ON FAST ALGORITHMS FOR SEVERAL
PROBLEMS IN TIME SERIES MODELS.
by
Guy Melard

Abstract
In a recent paper the author has given a Fortran program for the
computation of the likelihood function of an ARMA process. The
algorithm is extended here to handle exactly and efficiently
four related problems:
(a) the computation of forecasts
(b) the generation of artificial time series according to a given
ARMA model
(c) the computation of the likelihood function of a transfer function
model or a regression model with ARMA disturbances
(d) the computation of the first-order derivatives of the log-likelihood of an ARMA \((p,q)\) process with respect to the coefficients.

1. INTRODUCTION

In a recent paper, the author (Mélard, 1984) has given a Fortran program for the computation of the likelihood function of a stationary ARMA(p,q) process, defined by the equation

$$w_t = \phi_1 w_{t-1} + \ldots + \phi_p w_{t-p} + a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q}$$  \hspace{1cm} (1)

where the $a_t$ are n.i.d. random variables with mean zero and variance $\sigma^2$. Using the AR and MA polynomials in the backshift operator $\theta$, (1) is written as

$$\phi(\theta) w_t = \theta(\theta) a_t.$$  \hspace{1cm} (1)

Let $\phi = (\phi_1, ..., \phi_p)'$ and $\theta = (\theta_1, ..., \theta_q)'$, the parameters of the model. Given a time series $w = (w_1, ..., w_n)$, the likelihood function is obtained by

$$L(\phi, \theta, \sigma^2; w) = (2\pi)^{-n/2} \left( \frac{1}{\sigma} \right)^{-1} \exp \left\{ -\frac{1}{2} \sum_{t=1}^{n} \left( \frac{\hat{a}_t}{\hat{\sigma}_t} \right)^2 \right\}$$

where $\hat{a}_t = w_t - \hat{w}_t$, $\hat{w}_t$ is the orthogonal projection of $w_t$ into the subspace spanned by $(w_1, ..., w_{t-1})$ and $\hat{\sigma}_t$ is the norm of $\hat{a}_t$. For maximizing the likelihood function with respect to $\phi$ and $\theta$ all what is required is the sum of squares

$$S(\phi, \theta) = \left( \frac{1}{n} \sum_{t=1}^{n} h_t^2 \right)^{1/n} \sum_{t=1}^{n} \left( \frac{\hat{a}_t}{\hat{\sigma}_t} \right)^2,$$  \hspace{1cm} (2)

where $h_t = \hat{\sigma}_t / \sigma$.

Let $r = \max\{p, q+1\}$ and define the $r \times 1$ vectors $\varphi = (\phi_1, ..., \phi_r)'$, $\vartheta = (-1, \theta_1, ..., \theta_{r-1})'$ and $H' = (1, 0, ..., 0)'$ where $\phi_i = 0$ for $i > p$ and $\theta_j = 0$ for $j > q$. We consider the state space representation of (1)
\[ w_t = H \hat{w}_t \]
\[ \hat{w}_t = F \hat{w}_{t-1} + G a_t \]

with \( F = \begin{pmatrix} \Phi & I_{r-1} \\ 0 & 0 \end{pmatrix} \)

where \( \hat{w}_t \) is the \( rl \) state vector and \( I_{r-1} \) is the identity matrix of order \( r-1 \).

According to Pearlman (1980), the \( a_t \) can be computed by the following recursions

\[ a_t = w_t - H \hat{w}_t \] (3)

\[ \hat{w}_{t+1} = F \hat{w}_t + K_t (a_t/h_t^2) \] (4)

\[ K_{t+1} = K_t - \alpha_t F L_t \] with \( \alpha_t = H L_t/h_t^2 \) (5)

\[ L_{t+1} = F L_t - \alpha_t K_t \] (6)

\[ h_{t+1}^2 = h_t^2 (1 - \alpha_t^2) \] (7)

for \( t = 1, \ldots, n \). Using improvements proposed by Mølard (1984) no more than \( p + 3q + \min \{ p, q \} \) multiplications or divisions are needed for each \( t \) and the total storage size is limited to three vectors of length \( r+1 \). The starting values for these recursions are \( \hat{w}_1 = 0, K_1 = L_1 = F \mu \) and \( h_1^2 = H \mu \) where the \( i \)-th component of \( \mu \) is given by

\[ u_i = \sum_{j=1}^{r} (\phi_j \gamma_{j-i+1} - \theta_{j-1} \lambda_{j-i}) \] (8)

where the \( \gamma_j = \text{cov} (w_t, w_{t-j})/\sigma^2 \) and \( \lambda_j = \text{cov} (w_t, a_{t-j})/\sigma^2 \) can be obtained by the algorithm of Wilson (1979) with a number of operations bounded by a quadratic function of \( p \) and \( q \). Memory space is also limited to three vectors of length \( r+1 \).
Despite its apparent complexity, this new algorithm for computing the likelihood function of the ARMA(p,q) process is presently the most efficient in terms of computer time and storage. In practice the method is even faster than the traditional unconditional least squares method which can give only approximate (and sometimes bad) results.

In this paper, we extend the algorithm in order to handle four related problems.

2. COMPUTATION OF THE FORECASTS

If the \( a_t \) were known for \( t = 1, \ldots, n \), (1) could be used to compute forecasts at times \( n+1, \ldots, n+h \) (Box and Jenkins, 1976). Since we can only compute exactly the \( \hat{a}_t \), the forecasts should be derived from them (Ansley and Newbold, 1980). After the algorithm of Section 1 has been used the following relations are used for \( t = n+1, \ldots, n+h \):

\[
\hat{w}_t = H \hat{w}_t
\]

\[
\hat{w}_{t+1} = F \hat{w}_t.
\]

(9)

The \( \hat{w}_t \) are the forecasts.

3. GENERATION OF ARTIFICIAL TIME SERIES USING ARMA MODELS

Generally, (1) is used for \( t > p \) and \( a_{p-1} = a_{p-2} = \ldots = 0 \). To avoid the transient caused by these starting values, a certain number of data points are neglected at the beginning of the series. Ansley and Newbold (1980) and
Wilson (1979) have recommended the use of a more exact method. It appears that a very efficient method is given by the algorithm of Section 1, with the following modification: (3) is replaced by

$$w_t = \hat{a}_t + \frac{H}{t} \hat{W}_t$$

(10)

where the $(\hat{a}_t/h_t)$, $t = 1, \ldots, n$, are generated by pseudo-independent random variables with a given arbitrary distribution.

4. THE LIKELIHOOD OF A TRANSFER FUNCTION MODEL

The reader is referred to Miu (1983) for methods of estimation of transfer function models. Let the model be

$$w_t = \sum_{i=1}^{k} \frac{\omega_i(B)}{\delta_i(B)} X_{i,t} + \frac{\theta(B)}{\phi(B)} a_t$$

(11)

where $\omega_i(B)$, $\delta_i(B)$, $\theta(B)$, $\phi(B)$ are polynomials of respective degrees $s_i$, $r_i$, $q$, $p$, such that $\delta_i(0) = \theta(0) = \phi(0) = 1$ for $i = 1, \ldots, k$. It is assumed that data for $w_t$ are available for $t > 0$ and that $X_{i,t}$ are available for $t > t_i$. For each $i = 1, \ldots, k$, two cases are possible:

1° $X_{i,t}$ is non-stochastic and $t_i = -\infty$. This occurs e.g. if $X_i$ is an intervention variable, with the convention that $X_{i,t} = 0$ before the beginning of the series. Then, let $X_{i,t} = \delta_i^{-1}(B) \omega_i(B) X_{i,t}$, $\omega_i(B) = \delta_i^1(B) = 1$.

2° otherwise we let $X_{i,t} = X_{i,t}$, $\omega_i(B) = \omega_i(B)$ and $\delta_i(B) = \delta_i(B)$.

Denote by $\delta^*(B)$ the smallest common multiple of $\delta_i(B)$, $i = 1, \ldots, k$, and $r^*$ its degree. Then, the polynomials $\delta_i^*(B) = \delta^*(B)/\delta_i^1(B)$ are computed and
their degrees denoted by \( r_i^* \). With these notations (11) is written under the form

\[
\delta^*(B) w_t = \sum_{i=1}^{k} \omega_i (B) \delta_i^*(B) X_{i,t} + \frac{\delta^*(B) \theta(B)}{\phi(B)} a_t
\]  

(12)

for \( t > t_o = \max \{ r; \max_{i=1,\ldots,k} (t_i + s_i^1 + r_i^*) > 0 \} \), where \( s_i^1 \) is the degree of \( \omega_i(B) \). By cancellation of common factors \( \delta^*(B) \theta(B)/\phi(B) \) is written as \( \varepsilon^*(B)/\phi^*(B) \). An algorithm for computing the likelihood function of a transfer function model is now:

1° for \( t > t_o \) evaluate

\[
w_t^* = \delta^*(B) w_t - \sum_{i=1}^{k} \omega_i (B) \delta_i^*(B) X_{i,t}
\]  

(13)

2° use the algorithm of Section 1 with two modifications:

- \( t \) varies from \( t_o+1 \) to \( n \) instead of from 1 to \( n \);
- with the present notations the model is written

\[
\phi^*(B) w_t^* = \varepsilon^*(B) a_t
\]

The regression model with autoregressive-moving average disturbances of Harvey and Phillips(1979) is a special case of (11) with \( \delta_i(B) = 1, \omega_i(B) = \omega_i \). Consequently the algorithm of Section 1 can be used directly on

\[
w_t^* = w_t - \sum_{i=1}^{k} \omega_i X_{i,t}
\]

with a significant improvement in efficiency.
5. THE EXACT FIRST-ORDER DERIVATIVES OF THE LIKELIHOOD

Minimization of (2) can be made easier if the exact derivatives with respect to the \( \phi_i \) and \( \theta_j \) are known. Box and Jenkins (1976) have given recursions for the derivatives of the conditional log-likelihood, i.e. when starting values of the \( w_t \) and \( a_t \) are provided. In that case \( S(\phi, \theta) = \sum_{t=1}^{n} \hat{a}_t^2 \), where the \( \hat{a}_t \) are linear with respect to the \( \phi_i \) and \( \theta_j \). Khabie-Zeitounne (1980) has described a method for the derivatives of (2) where the first factor is omitted. Furthermore, the algorithm is based on the method of Levinson (1949) which requires a number of operations of order \( n^2 \). Given that the recursions used by Mélard (1984) are simple, it is easy to obtain similar recursions for the first-order derivatives. For example, the \( i \)-th relation of (6) is

\[
(L_t+1)_i = \phi_i(L_t)_1 + (L_t)_i + \alpha_t(K_t)_i,
\]

and differentiation with respect to \( \phi_k \) gives the recursion

\[
\frac{\partial (L_{t+1})_i}{\partial \phi_k} = \delta_{ik}(L_t)_1 + \phi_i \frac{\partial (L_t)_1}{\partial \phi_k} + \frac{\partial (L_t)_i}{\partial \phi_k} + \frac{\partial (L_t)_i + 1}{\partial \phi_k} - \frac{\partial \alpha_t}{\partial \phi_k} (K_t)_i - \alpha_t \frac{\partial (K_t)_i}{\partial \phi_k}
\]

Similarly \( \frac{\partial \mu}{\partial \phi_k} \) requires knowledge of the first-order derivatives of the \( \gamma_i \) and the \( \lambda_i \) which can be obtained by differentiating all the recursions in the algorithm of Wilson (1979). The number of operations is multiplied by a factor less than \( 2(p+q) \) and the storage is multiplied by \( (p+q) \). Thanks to the inherent efficiency of the algorithm for computing (2), the resources needed are still quite reasonable.
ACKNOWLEDGMENT

This paper was written while the author was at the Université de Montréal with a grant from the co-operation between the Province of Quebec and the French Community of Belgium.

REFERENCES


