Ann. Inst. Statist. Math.
29 (1977), Part B, 319-332

ON A SEARCH PROCEDURE FOR THE OPTIMAL AR-MA ORDER*

GENSHIRO KITAGAWA

(Received Dec. 1, 1976; revised Apr. 13, 1977)

1. Introduction

In identifying an autoregressive moving average (AR-MA) model

$$x_n + \sum_{m=1}^{M} a_m x_{n-m} = \varepsilon_n + \sum_{m=1}^{L} b_m \varepsilon_{n-m} , \tag{1}$$

we can obtain the AR-coefficients a_m $(m=1,\cdots,M)$ and the MA-coefficients b_m $(m=1,\cdots,L)$ by the maximum likelihood method, provided that the AR-order M and the MA-order L are given. Since the algorithm and the computer programs for the method have already been prepared (Akaike [1] and [2]), the only problem is to determine the order (M, L) of the model. According to the minimum AIC method, the order (M, L) is selected as the one which gives the minimum of

$$\text{AIC}(M, L) = N \log (d(M, L)) + 2(M+L) , \tag{2}$$

where N is the data length, $d(M, L)$ is the maximum likelihood estimate of the innovation variance.

In the case of fitting autoregressive models $(L=0)$, the minimum AIC method can be performed quite easily, since the maximum likelihood estimates are obtained by solving a linear equation. By the Levinson's procedure the coefficients of the autoregressive model for each order are obtained in the process of obtaining the coefficients of the model with the highest order. However, in the case of fitting an AR-MA model, the complete realization of the minimum AIC method is not necessarily easy. There are many numbers of AR-MA models which must be fitted, and worse, we must solve the nonlinear optimization problem to obtain the maximum likelihood estimates of the coefficients. There is no efficient procedure which corresponds to the Levinson's procedure for the fitting of autoregressive models.

The only way to make the fitting of AR-MA models more efficient is to introduce some other information about the order of the model.

* This research was supported in part by the Ministry of Education, Science and Culture, Grant.

Akaike [2], [3] proposed a method for the selection of the order which initially estimates the order by the canonical correlation analysis of the past and future observations and then searches for a locally optimal order.

In this paper, another heuristic procedure for the search of the optimal order is presented. This procedure is composed of a search rule and a computational method to obtain the initial estimates for the iterative computation of likelihood maximization. The search rule is based on the consideration of the behavior of the value of AIC and the innovation variance when AR-MA models with various orders are fitted. In searching for the best order, the procedure fully utilizes the hitherto fitted models. Initial estimates of the AR-MA parameters are given by maximizing the entropy of the model. Numerical examples are given to illustrate how the proposed procedure works. The results show that the optimal orders can often be detected without fitting the whole possible models with various combinations of the AR and MA orders.

2. Entropy of an AR-MA model and its maximization

Suppose that a Gaussian stochastic process $\{x_n\}$ is generated by an autoregressive moving average model with order M and L

$$x_n+\sum_{m=1}^{M} a_m x_{n-m}=\varepsilon_n+\sum_{m=1}^{L} b_m \varepsilon_{n-m}\ , \tag{3}$$

where ε_n is a white noise with variance $\mathrm{E}\,\varepsilon_n^2=\sigma^2$ and $\mathrm{E}\,\varepsilon_n x_{n-k}=0$ $(k=1, 2,\cdots)$. Hereafter, we call it an AR-MA(M, L) model and sometimes denote by

$$\left(1+\sum_{m=1}^{M} a_m B^m\right)x_n=\left(1+\sum_{m=1}^{L} b_m B^m\right)\varepsilon_n\ , \tag{4}$$

using the backshift operator B defined by $Bx_n=x_{n-1}$.

From (3), we have

$$\mathrm{E}\,x_n\varepsilon_{n-k}+\sum_{m=1}^{M} a_m\,\mathrm{E}\,x_{n-m}\varepsilon_{n-k}=\mathrm{E}\,\varepsilon_n\varepsilon_{n-k}+\sum_{m=1}^{L} b_m\,\mathrm{E}\,\varepsilon_{n-m}\varepsilon_{n-k}\ ,$$

$$\mathrm{E}\,x_n x_{n-k}+\sum_{m=1}^{M} a_m\,\mathrm{E}\,x_{n-m}x_{n-k}=\mathrm{E}\,\varepsilon_n x_{n-k}+\sum_{m=1}^{L} b_m\,\mathrm{E}\,\varepsilon_{n-m}x_{n-k}\ ,$$

where E denotes expectation. Thus, the autocovariance function $R_k=\mathrm{E}\,x_n x_{n-k}$ and the cross-covariance function $S_n=\mathrm{E}\,x_n\varepsilon_{n-k}$ $(k=0, 1,\cdots)$ are given by

$$S_k+\sum_{m=1}^{M} a_m S_{k-m}=\sigma^2\left(\delta_{0,k}+\sum_{m=1}^{L} b_m\delta_{m,k}\right)\ , \qquad (k=0, 1,\cdots) \tag{5}$$

$$R_k+\sum_{m=1}^{M} a_m R_{k-m}=S_{-k}+\sum_{m=1}^{L} b_m S_{m-k}, \qquad (k=0, 1, \cdots)$$

where

$$S_k=0 \quad \text{for } k<0, \qquad R_{-k}=R_k,$$

and

$$\delta_{m,k}=\begin{cases} 1 & \text{for } m=k, \\ 0 & \text{for } m\neq k. \end{cases}$$

For the autoregressive parameters a_m $(m=1,\cdots,M)$, the matrix

$$\begin{bmatrix} 1 & a_1 & a_2 & \cdots a_{M-1} & a_M \\ a_1 & 1+a_2 & a_3 & \cdots a_M & 0 \\ a_2 & a_1+a_3 & 1+a_4 & \cdots 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{M-1} & a_M+a_{M-2} & a_{M-3} & \cdots 1 & 0 \\ a_M & a_{M-1} & a_{M-2} & \cdots a_1 & 1 \end{bmatrix}$$

is regular [4]. Thus we can compute the autocovariance function R_k and the cross-covariance function S_k, provided that the parameters a_m $(m=1,\cdots,M)$ and b_m $(m=1,\cdots,L)$ of the model (4) are given.

Let an AR-MA (I, J) model

$$(7) \qquad \left(1+\sum_{m=1}^{I} c_m B^m\right)x_n=\left(1+\sum_{m=1}^{J} d_m B^m\right)e_n$$

with $\mathrm{E}\, e_n^2=\sigma_e^2$ be another one which represents the same process. The entropy [5] of the AR-MA (M, L) with respect to the AR-MA (I, J) model is given by

$$(8) \qquad \begin{aligned} I(f_0; f) &= \mathrm{E} \log \left\{\frac{f(x|c_1,\cdots,c_I, d_1,\cdots,d_J)}{f_0(x|a_1,\cdots,a_M, b_1,\cdots,b_L)}\right\} \\ &= \mathrm{E} \log f(x|c_1,\cdots,c_I, d_1,\cdots,d_J) \\ &\quad -\mathrm{E} \log f_0(x|a_1,\cdots,a_M, b_1,\cdots,b_L) \end{aligned}$$

where $f_0(x|a_1,\cdots,a_M, b_1,\cdots,b_L)$ is the probability density function of the AR-MA (M, L) model given in (3), $f(x|c_1,\cdots,c_I, d_1,\cdots,d_J)$ is that of the AR-MA (I, J) model given in (7) and E denotes the expectation operator with respect to the AR-MA (M, L) process. The second term of the right-hand side of (8) is a constant and by neglecting the effect of the initial condition the first term is given by

$$(9) \qquad \mathrm{E}\left\{-\frac{1}{2}\log(2\pi\sigma_e^2)-\frac{1}{2\sigma_e^2}\left(x_n+\sum_{m=1}^{I} c_m x_{n-m}-\sum_{m=1}^{J} d_m e_{n-m}\right)^2\right\}$$

$$=-\frac{1}{2}\log 2\pi-\frac{1}{2}\log\sigma_e^2-\frac{1}{2\sigma_e^2}\mathrm{E}\{e_n^2\}\ ,$$

where $\mathrm{E}\{e_n^2\}$ is given as the residual variance of the AR-MA $(M+J, L+I)$ process

$$e_n=\frac{\left(1+\sum_{m=1}^{I}c_mB^m\right)\left(1+\sum_{m=1}^{L}b_mB^m\right)}{\left(1+\sum_{m=1}^{J}d_mB^m\right)\left(1+\sum_{m=1}^{M}a_mB^m\right)}\varepsilon_n\ . \tag{10}$$

A model which maximizes the entropy

$$I(f_0;f)\equiv I(c_1,\cdots,c_I,d_1,\cdots,d_J,\sigma_e^2)$$

is considered to be the best one in the family of the AR-MA (I, J) models. The maximization of the entropy function $I(c_1,\cdots,c_I,d_1,\cdots,d_J,\sigma_e^2)$ can be performed as follows.

Since $(\partial I/\partial\sigma_e^2)_{\sigma_e^2=\hat{\sigma}_e^2}=0$ holds for the parameter $\hat{\sigma}_e^2$ of σ^2 which maximizes the entropy, it follows that

$$\hat{\sigma}_e^2=\mathrm{E}\{e_n^2\}\ .$$

Thus the maximization of the entropy function leads to the problem of minimizing

$$\hat{L}(\theta)=\log\hat{\sigma}_e^2(c_1,\cdots,c_I,d_1,\cdots,d_J)\ ,$$

or equivalently

$$L(\theta)=\hat{\sigma}_e^2\ ,$$

where $\theta=(\theta_1,\cdots,\theta_{I+J})\equiv(c_1,\cdots,c_I,d_1,\cdots,d_J)$. Putting

$$A_m=\sum_{k=1}^{M+J}a_kd_{m-k}$$

and

$$C_m=\sum_{k=1}^{L+I}b_kc_{m-k}\ ,$$

equation (10) is rewritten as

$$\left(1+\sum_{m=1}^{M+J}A_mB^m\right)e_n=\left(1+\sum_{m=1}^{L+I}C_mB^m\right)\varepsilon_n\ . \tag{11}$$

The function $L(\theta)=\mathrm{E}\{e_n^2\}$ is the theoretical residual variance R_0 of the AR-MA $(M+J, L+I)$ process (10). R_0 is obtained as the solution of the equations

$$(12)\quad \begin{aligned} &R_k+\sum_{m=1}^{M+J} A_m R_{k-m}=S_k+\sum_{m=1}^{L+I} C_m S_{k-m}\,, \qquad (k=0,1,\cdots,M+J)\\ &R_{-k}=R_k\,, \end{aligned}$$

where S_k $(k=0,1,\cdots,L+I)$ are given by

$$\begin{aligned} &S_0=\sigma^2\\ &S_k+\sum_{m=1}^{M+J} A_m S_{k-m}=C_k\sigma^2\,, \qquad (k=1,2,\cdots,L+I)\\ &S_k=0\,, \qquad (k<0)\,. \end{aligned}$$

The gradient $\partial L(\theta)/\partial\theta=\partial R_0/\partial\theta$ is given by solving the equations

$$(13)\quad \begin{aligned} &\frac{\partial S_0}{\partial\theta_i}=0\,,\\ &\frac{\partial S_k}{\partial\theta_i}+\sum_{m=1}^{M+J} A_m\frac{\partial S_{k-m}}{\partial\theta_i}=\sigma^2\frac{\partial C_k}{\partial\theta_i}-\sum_{m=1}^{M+J}\frac{\partial A_m}{\partial\theta_i}S_{k-m}\,,\\ &\frac{\partial R_k}{\partial\theta_i}+\sum_{m=1}^{M+J} A_m\frac{\partial R_{k-m}}{\partial\theta_i}=\frac{\partial S_k}{\partial\theta_i}+\sum_{m=1}^{L+I} C_m\frac{\partial S_{k-m}}{\partial\theta_i}+\sum_{m=1}^{L+I}\frac{\partial C_m}{\partial\theta_i}S_{k-m}\\ &\qquad\qquad -\sum_{m=1}^{M+J}\frac{\partial A_m}{\partial\theta_i}R_{k-m}\,, \end{aligned}$$

where

$$\frac{\partial A_m}{\partial\theta_i}=\begin{cases} 0 & 1\leqq i\leqq I+m-1\\ 1 & i=I+m\\ a_{m-i+I} & i>I+m\,, \end{cases}$$

$$\frac{\partial C_m}{\partial\theta_i}=\begin{cases} 0 & i>m \text{ or } i>I\\ 1 & i=m,\ i\leqq I\\ b_{m-i} & i<m,\ i\leqq I\,. \end{cases}$$

Thus, the value $L(\theta)$ and the gradient $\partial L(\theta)/\partial\theta$ can be obtained, provided that a_i $(i=1,\cdots,M)$, b_i $(i=1,\cdots,L)$, σ^2 and θ_i $(i=1,\cdots,I+J)$ are given. Using these, the minimum of $L(\theta)$ can be found by the gradient method, e.g., by the Davidon method. An efficient method of obtaining an initial estimate of the parameter is given in the appendix.

3. Computational experiments

To give some feeling of the behavior of the residual variance, we consider here the AR-MA process which is characterized by the param-

Table 1. Parameters of an AR-MA (M, L) model $x_n + \sum_{m=1}^{M} a_m x_{n-m} = \varepsilon_n + \sum_{m=1}^{L} b_m \varepsilon_{n-m}$, with $\mathrm{E}\,\varepsilon_n^2 = \sigma^2$

m	$M=8,\ L=2,\ \sigma^2=1.0$	
	a_m	b_m
1	−2.30880	−1.36690
2	2.01490	0.48766
3	−0.68625	
4	0.23839	
5	−0.28178	
6	−0.22088	
7	0.54607	
8	−0.28580	

eters given in Table 1.

Table 2 shows the value of the residual variances of the AR-MA (i, j) models obtained by maximizing the entropy with respect to the model. Fig. 1 illustrates the behavior of the residual variance. The value of residual variance is shown as the height of the three dimensional figure and it ranges from 1.00 to 1.20. Eleven contour lines with the values $1.01+0.02n$ $(n=0, 1, \cdots, 10)$ are also plotted.

The AR-MA (i, j) models are fitted by the maximum likelihood method for one realization of the process with $N=1000$. The residual variances of the fitted models produce Fig. 2 which is quite similar to Fig. 1. In the figure, the value of AIC is represented as the height and it ranges from 0.0 to 200.0. Nine contour lines with the value $3.0+4.0n$ $(n=0, 1, \cdots, 8)$ are also plotted in the figure.

These figures indicate that the contour lines of the residual vari-

Table 2. Residual variances of AR-MA (I, J) models which maximize the entropy of the AR-MA (8, 2) model characterized by the parameters given in Table 1

I \ J	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	4.039	1.950	1.498	1.363	1.216	1.213	1.181	1.163	1.150	1.149	1.097	1.058	1.038	1.031	1.030
1	1.797	1.461	1.399	1.304	1.216	1.179	1.179	1.114	1.075	1.043	1.032	1.032	1.017	1.010	1.009
2	1.269	1.230	1.189	1.133	1.130	1.128	1.082	1.068	1.055	1.036	1.032	1.013	1.011	1.009	
3	1.223	1.222	1.158	1.131	1.092	1.041	1.026	1.015	1.013	1.012	1.005	1.004	1.003		
4	1.219	1.132	1.132	1.122	1.083	1.034	1.016	1.014	1.013	1.012	1.005	1.004			
5	1.154	1.079	1.077	1.041	1.040	1.030	1.012	1.011	1.008	1.007	1.004				
6	1.106	1.075	1.074	1.039	1.033	1.030	1.010	1.010	1.008	1.007					
7	1.096	1.074	1.048	1.036	1.032	1.022	1.008	1.003	1.001						
8	1.090	1.016	1.000	1.000	1.000	1.000	1.000	1.000							
9	1.070	1.005	1.000	1.000	1.000	1.000	1.000								
10	1.047	1.001	1.000	1.000	1.000	1.000									
11	1.029	1.001	1.000	1.000	1.000										
12	1.016	1.000	1.000	1.000											
13	1.008	1.000	1.000												
14	1.004	1.000													

** Residual variance is reduced monotonously with the lower bound 1.00, as the order I or J increases.

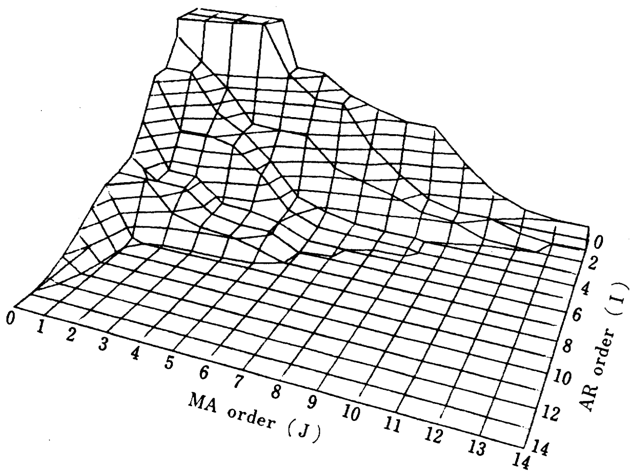

Fig. 1. Behavior of the residual variance

Table 3. Table of AIC $(I, J)+3.2$ for a realization of AR-MA (8, 2) process with $N=1000$

$I \backslash J$	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	1480.7	731.5	437.2	318.0	201.9	202.5	182.5	161.5	142.1	144.1	110.7	55.1	32.1	29.3
1	645.6	413.5	354.4	269.7	203.1	173.3	165.6	155.2	102.3	59.7	40.9	33.7	31.1	30.7
2	224.2	184.1	139.7	106.8	106.8	108.5	64.5	60.0	41.2	26.1	25.7	27.0	29.0	27.5
3	173.1	175.1	121.2	107.4	108.8	100.5	31.4	23.1	20.5	18.2	11.6	13.5	12.6	12.2
4	174.2	108.6	110.5	105.8	89.6	36.2	22.7	22.8	22.4	20.2	13.6	15.0	12.3	13.7
5	121.7	72.5	72.5	47.5	49.5	49.0	20.6	22.1	20.4	11.5	8.6	10.1	11.7	13.0
6	91.3	70.1	71.9	45.4	40.8	42.1	21.6	23.6	21.7	13.5	10.4	12.1	13.6	11.4
7	83.1	66.3	51.4	42.7	42.6	34.0	13.5	9.4	10.4	9.2	11.1	13.5	15.5	13.4
8	76.9	19.1	0	0.9	2.7	4.5	6.2	8.2	10.0	11.1	12.4	9.8	11.5	13.3
9	65.0	12.8	1.1	2.8	4.7	6.3	8.2	9.7	11.3	12.9	14.4	11.1	13.1	15.1
10	47.7	8.1	2.6	4.6	2.0	3.4	5.4	7.4	9.4	11.4	13.4	13.1	15.1	17.1
11	24.5	5.2	4.5	6.3	3.3	5.3	7.3	9.3	11.3	13.3	15.3	15.1	17.1	19.1
12	12.5	6.3	6.2	8.2	5.4	7.3	9.3	11.3	13.3	15.3	17.3	17.1	19.1	21.1
13	7.7	7.8	8.1	8.5	7.9	9.3	11.3	13.3	15.3	17.3	19.3	19.1	21.1	23.1

ances and AIC lie along the line $i+j=$constant. This implies that an order which attains the minimum AIC in the family of autoregressive models gives important information about the minimum of the AIC (i,j). It is also observed that the value AIC $(0, k)$ of the purely moving average model is reduced significantly at a point where the inequality $k \geqq M+L$ holds.

Fig. 3 complements Fig. 2. It shows the behavior of the partial autocorrelation, residual variance and AIC when the autoregressive models with various orders are fitted. The figure indicates that the partial autocorrelations become insignificant after the initial $M+L$ or-

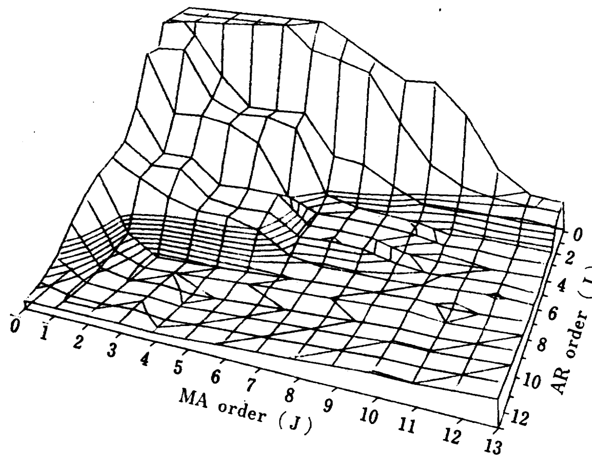

Fig. 2. Behavior of AIC (I, J)

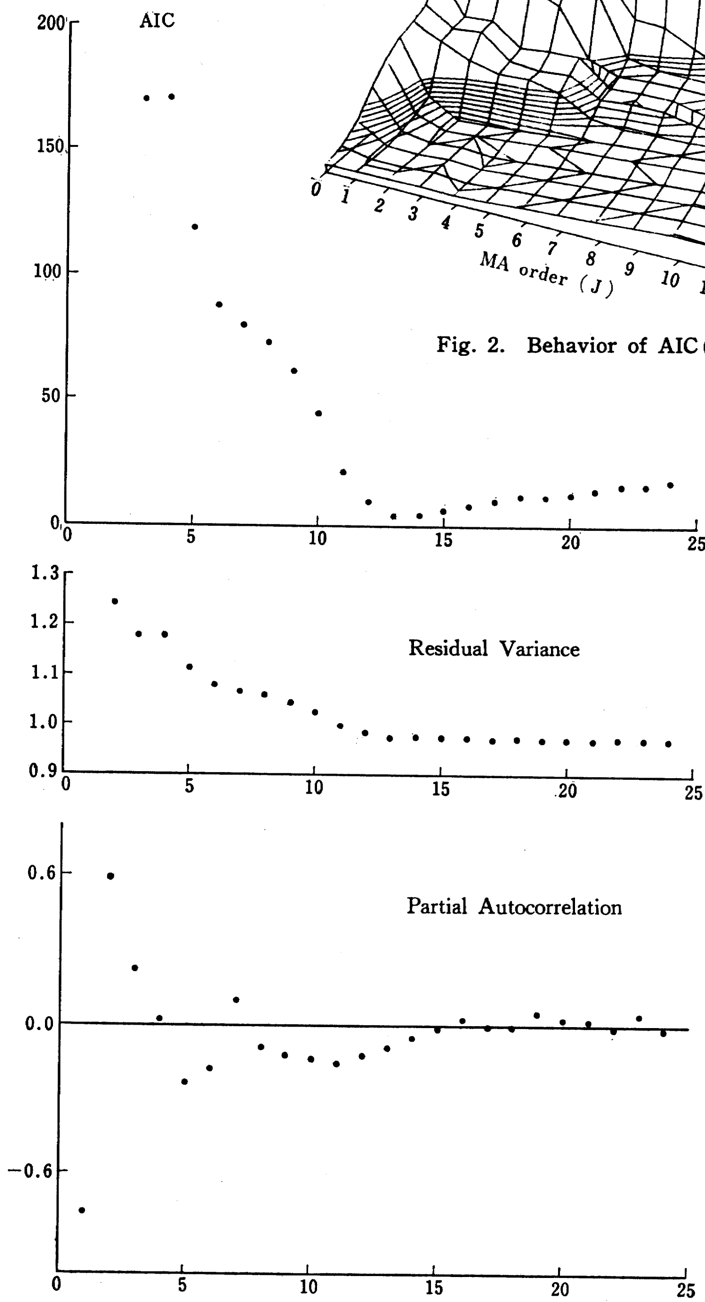

Fig. 3. Behavior of AICs, residual variances and partial autocorrelations of autoregressive models for a realization of AR-MA (8, 2) process with $N=1000$

ders. It also shows that the reduction of the residual variance of the AR-MA $(m, 0)$ model becomes insignificant for $m \gg M+L$. The minimum of AIC $(m, 0)$ is attained at $m=13$.

4. An order search algorithm

Taking into account the preceding observations, we now propose a heuristic order search algorithm as follows.

(I) Determine the upper limit K of the order of autoregressive models to be fitted to the data.

(II) For $k=1, \cdots, K$, fit kth order autoregressive model and compute the AIC $(k, 0)$.

(III) Find minimal points $(k_i, 0)$ $(i=1, \cdots, p)$ of the AIC $(k, 0)$, $(k=0, 1, \cdots, K)$.

(IV) For $i=1, \cdots, p$;

i) for $l=1, \cdots, k_i$ and $m=k_i-l$, fit an AR-MA (m, l) model and compute the AIC (m, l).

ii) find the minimal points (m_j^i, l_j^i), $(j=1, \cdots, q_i)$ of the AIC (m, l), $(l=1, \cdots, q_i;\ m=k_i-l)$.

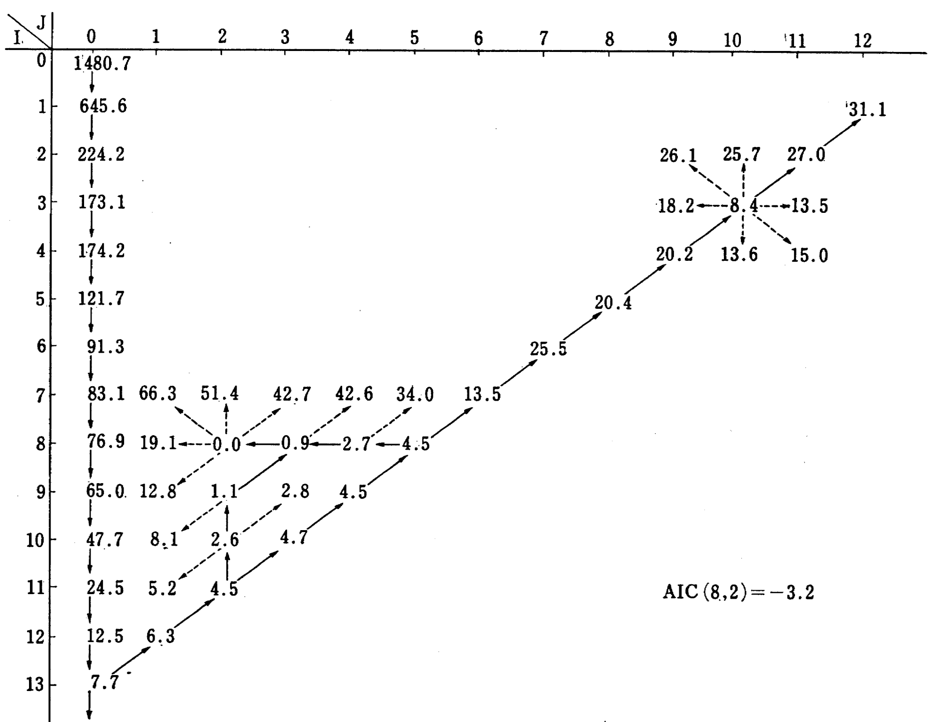

Fig. 4. Illustration of the order search procedure. This figure shows how the procedure worked when it was applied to the same data as that of Table 3 and Fig. 2

(V) For $j=1,\cdots,q_i$; $i=1,\cdots,p$, find the local minimum of AIC(m,l) around the point (m_j^i, l_j^i).

The procedure has an advantage that it always selects an AR-MA model with the AIC less than (or equal to, at worst) the minimum AIC in the family of the autoregressive models.

In the procedure, the models are fitted by the maximum likelihood method. The maximizing procedure of the likelihood function needs an initial value of the parameters. The goodness of the initial estimates affects not only the computing time but also the crucial point whether or not it converges to the global maximum. As stated in Section 2, if we know the true structure of the AR-MA process, we can determine the parameters of the AR-MA(i, j) model which maximize the entropy of the process with respect to the model. We intend to utilize the parameters in our search procedure. We consider the situation where the procedure already fitted an AR-MA(I, J) model. Then, it searches for the AR-MA(M, L) model with the least AIC among the hitherto fitted AR-MA(i, j) models around the AR-MA(I, J). The initial value for the iterative computation of the maximum likelihood method is given by maximizing the entropy of the AR-MA(M, L) model with respect to the AR-MA(I, J) model.

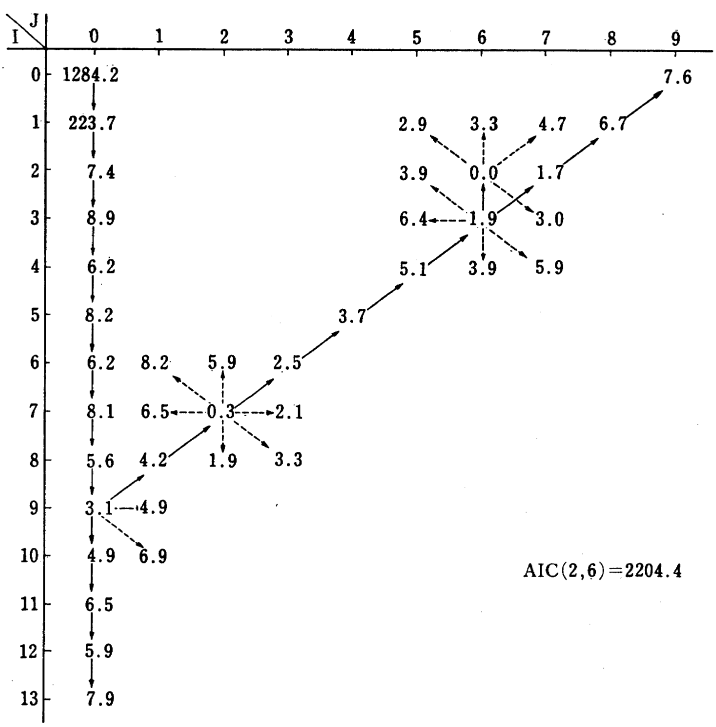

Fig. 5. Illustration of the order search procedure. This figure shows how the procedure worked when it was applied to the same data as that of Table 4 and Fig. 6

5. Numerical examples

Fig. 4 illustrates how the proposed procedure worked when it was applied to the data used in Fig. 2. According to the procedure, the

Table 4. AIC's of the AR-MA (I, J) models fitted to a real record of ship's yaw motion

I \ J	0	1	2	3	4	5	6	7	8	9	10	11
0	1284.2	728.2	377.8	209.5	130.6	67.9	23.5	13.0	5.7	7.6	7.3	9.3
1	223.7	53.8	24.9	22.3	11.5	2.9	3.3	4.7	6.7	7.8	8.5	10.5
2	7.4	8.7	7.2	9.2	5.9	3.9	0.0	2.0	4.0	6.0	6.4	7.3
3	8.9	9.6	9.2	9.5	4.0	5.9	1.9	3.0	5.0	7.0	5.0	6.3
4	6.2	8.2	4.6	4.4	5.4	5.1	3.9	4.8	3.6	5.7	6.7	8.2
5	8.2	5.5	4.4	5.4	3.7	1.7	3.3	4.1	5.3	6.8	8.6	5.4
6	6.2	7.5	5.9	2.5	2.0	3.4	4.7	5.8	4.9	5.8	5.2	7.0
7	8.1	6.5	0.3	2.1	3.5	5.1	4.2	5.1	3.5	5.1	6.6	8.6
8	5.6	4.2	1.9	3.3	5.3	7.1	5.7	7.1	5.5	7.1	8.6	10.5
9	3.1	4.0	3.2	5.2	7.2	9.1	7.0	8.2	6.4	8.4	10.4	11.2
10	4.9	6.9	5.1	2.5	3.9	5.7	7.6	9.6	8.4	9.5	11.5	11.0
11	6.5	3.3	5.3	4.3	5.8	7.6	9.5	11.3	10.4	11.5	13.5	13.0

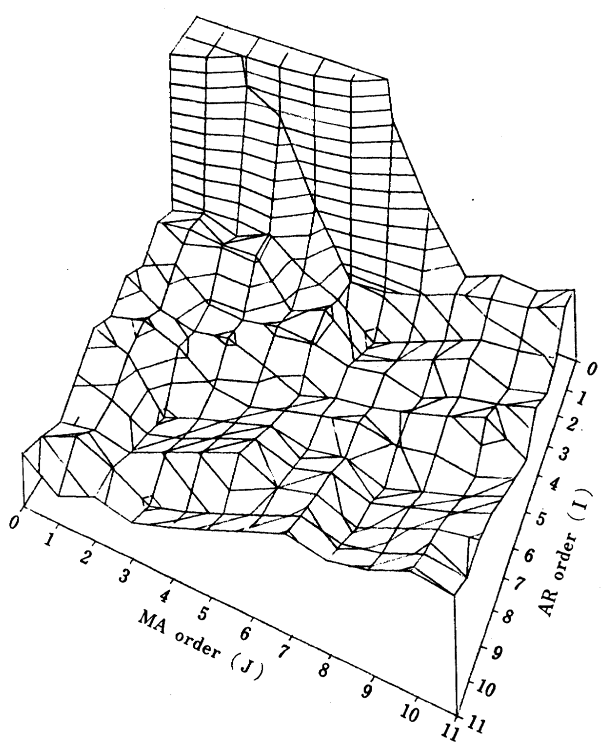

Fig. 6. Behavior of AIC for a real record of ship's yaw motion

autoregressive models ($L=0$, $M=1,\cdots$) were fitted and the minimum AIC was attained at $M=13$. Then the AR-MA(i, j) models $(i+j=13$, $i=1,\cdots,12)$ were fitted and three local minima of AIC were found at AR-MA(11, 2), AR-MA(8, 5) and AR-MA(3, 10). Starting from AR-MA (11, 2), the AIC was reduced successively through AR-MA(10, 2), (9, 2), (8, 3) to AR-MA(8, 2). Starting from AR-MA(8, 5), AR-MA(8, 2) was obtained again. AR-MA(3, 10) produced a local minimum. Fig. 2 indicates that the global minimum was obtained by the proposed procedure.

Fig. 5 shows the case when the procedure was applied to the real record of a ship's yaw motion. Referring to Table 4 and Fig. 6, it is observed that the global minimum was found out quite efficiently.

6. Discussion and conclusion

The behavior of the residual variance and the AIC of the AR-MA model is investigated and a new computational method for the fitting of an AR-MA model is presented. Computational experiments were made to give some feeling of the behavior of the residual variance and AIC. A heuristic procedure for the search of the optimal order is presented which is composed of the search rule given in Section 4 and the computational method of obtaining the initial estimates. By the proposed procedure, we can make use of the information obtained from the hitherto fitted models. The method usually gives good initial estimates of the parameters to get the global maximum of the likelihood function, in addition to the advantage in terms of the computing time. Using the search rule, we can get the optimal order without fitting the whole possible models with various AR and MA orders.

The initial estimates obtained by the procedure given in the appendix sometimes give fairly good estimates of the parameters. It needs much less computing time compared with the method of Section 2. It depends on the data whether we should use the procedure of Section 2 or just use the one of the appendix.

Two numerical examples given in the paper show that the procedure works well with real data.

Appendix

We will present here a method of obtaining a rough estimate of AR-MA(I, J) model. We assume here that we know the parameters a_m $(m=1,\cdots,M)$, b_m $(m=1,\cdots,L)$ and σ^2 of the AR-MA(M, L) model. The parameters c_m $(m=1,\cdots,I)$ and d_m $(m=1,\cdots,J)$ are determined so as to minimize the mean square value

$$\mathrm{E}\,(x_n-\hat{x}_n)^2\,,$$

with $\hat{x}_n=\sum_{j=1}^{J} d_j\varepsilon_{n-j}-\sum_{i=1}^{I} c_i x_{n-i}$. Define e_n by

$$e_n=x_n-\hat{x}_n=x_n-\sum_{j=1}^{J} d_j\varepsilon_{n-j}+\sum_{i=1}^{I} c_i x_{n-i}\ ,$$

then

$$\begin{aligned}\left(1+\sum_{m=1}^{M} a_m B^m\right)e_n&=\left(1+\sum_{m=1}^{M} a_m B^m\right)\left(x_n-\sum_{j=1}^{J} d_j B^m\varepsilon_n+\sum_{i=1}^{I} c_i B^i x_n\right)\\&=\left(1+\sum_{m=1}^{L} b_m B^m\right)\left(1+\sum_{i=1}^{I} c_i B^i\right)\varepsilon_n-\left(1+\sum_{m=1}^{M} a_m B^m\right)\\&\quad\cdot\sum_{j=1}^{J} d_j B^j\varepsilon_n\ .\end{aligned}$$

Therefore, the minimization of $\mathrm{E}(x_n-\hat{x}_n)^2$ is not equivalent to that of the residual variance of the AR-MA$(M+J, L+I)$ process

$$\left(1+\sum_{m=1}^{M} a_m B^m\right)\left(1+\sum_{m=1}^{J} d_m B^m\right)e_n=\left(1+\sum_{m=1}^{L} b_m B^m\right)\left(1+\sum_{m=1}^{I} c_m B^m\right)\varepsilon_n\ .$$

Nevertheless, according to the author's experience, the present procedure gives fairly good estimates, especially when $\mathrm{E}(x_n-\hat{x}_n)^2$ is close to σ^2, that is when the AR-MA(I, J) model is a good alternative of the AR-MA(M, L) model.

Since

$$\begin{aligned}\mathrm{E}(x_n-\hat{x}_n)^2&=\mathrm{E}\left(\varepsilon_n+\sum_{m=1}^{L} b_m\varepsilon_{n-m}-\sum_{m=1}^{M} a_m x_{n-m}-\sum_{j=1}^{J} d_j\varepsilon_{n-j}+\sum_{i=1}^{I} c_i x_{n-i}\right)^2\\&=\sigma^2\left(1+\sum_{m=1}^{L} b_m^2+\sum_{j=1}^{J} d_j^2-2\sum_{j=1}^{\min\{L,J\}} b_j d_j\right)+\sum_{m=1}^{M}\sum_{k=1}^{M} a_m a_k R_{m-k}\\&\quad+\sum_{i=1}^{I}\sum_{j=1}^{I} c_i c_j R_{i-j}-2\sum_{m=1}^{M}\sum_{i=1}^{I} a_m c_i R_{m-i}-2\sum_{j=1}^{L}\sum_{m=1}^{M} a_m b_j S_{j-m}\\&\quad+2\sum_{j=1}^{L}\sum_{i=1}^{I} b_j c_i S_{j-i}+2\sum_{m=1}^{M}\sum_{j=1}^{J} a_m d_j S_{j-m}-2\sum_{j=1}^{J}\sum_{i=1}^{I} c_i d_j S_{j-i}\ ,\end{aligned}$$

the coefficients which attain the minimum of $\mathrm{E}(x_n-\hat{x}_n)^2$ are given by solving

$$(14)\quad\begin{aligned}&\sum_{j=1}^{I} c_j R_{i-j}+\sum_{j=1}^{L} b_j S_{j-i}-\sum_{j=1}^{M} a_j R_{j-i}-\sum_{j=1}^{J} d_j S_{j-i}=0\qquad(i=1,\cdots,I)\ ,\\&d_i\sigma^2-b_i\sigma^2+\sum_{j=1}^{M} a_j S_{i-j}-\sum_{j=1}^{I} c_j S_{i-j}=0\qquad(i=1,\cdots,J)\ .\end{aligned}$$

By the formulae

$$\sum_{m=1}^{M} a_m R_{m-i}-\sum_{m=1}^{L} b_m S_{m-i}=-R_i$$

and

$$\sum_{m=1}^{M} a_m S_{i-m} - b_i \sigma^2 = -S_i ,$$

the equations (14) is reduced to the linear equations

$$(15)\quad \begin{bmatrix} R_0 & R_1 & \cdots R_{I-1} & -S_0 & -S_1 \cdots -S_{J-1} \\ R_1 & R_0 & \cdots R_{I-2} & 0 & -S_0 \cdots -S_{J-2} \\ \vdots & \vdots & \ddots \quad \vdots & \vdots & \ddots \quad \vdots \\ R_{I-1} & R_{I-2} & \cdots R_0 & 0 & \\ -S_0 & & & \sigma^2 & 0 \\ -S_1 & -S_0 & & & \sigma^2 \\ \vdots & \vdots & \ddots & & \ddots \\ -S_{J-1} & -S_{J-2} & \cdots & 0 & \sigma^2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_I \\ d_1 \\ d_2 \\ \vdots \\ d_J \end{bmatrix} = \begin{bmatrix} -R_1 \\ -R_2 \\ \vdots \\ -R_I \\ S_1 \\ S_2 \\ \vdots \\ S_J \end{bmatrix} .$$

It should be noted that in the case of an autoregressive model, (15) gives exactly the same parameters as that given by maximizing the entropy.

Acknowledgements

The author would like to thank Dr. H. Akaike for his helpful suggestions and encouragement in preparing this paper. The author is also grateful to Miss F. Tada for her assistance in programming.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] Akaike, H. (1974). Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes, *Ann. Inst. Statist. Math.*, **26**, 367-387.

[2] Akaike, H., Arahata, M. and Ozaki, T. (1975). TIMSAC-74, A time series analysis and control program package—(1), *Computer Science Monographs*, No. 5, The Institute of Statistical Mathematics, Tokyo.

[3] Akaike, H. (1976). Canonical correlation analysis of time series and the use of an information criterion, System Identification: *Advances and Case studies*, (R. K. Mehra and D. G. Lainiotis, eds.), Academic Press, New York, 27-96.

[4] Akaike, H. (1976). Personal communication.

[5] Akaike, H. (1977). On entropy maximization principle, *Proc. Symposium on Applications of Statistics*, P. R. Krishnaiah, ed., North-Holland, Amsterdam, to appear.