ON THE IDENTIFICATION OF ARIMA PROCESSES

G.A. El-Kobrosy, M.A. Shalaby
Dept. of Eng. Math. and Phys., Faculty of Eng., Alexandria University, Alexandria, Egypt.

and E.E. AFIFY
Eng. Math. Department, Faculty of Engineering, Menoufia Univ., Shebin Elkom, Egypt.

ABSTRACT

The identification of the autoregressive integrated moving average processes is considered. A method, based on the autocorrelation function of the given data, is used for the order determination of the process. The method has been applied to three examples and compared with the other methods. An algorithm for the parameter estimation of the model is given and, in contrary to earlier investigators, can be extended to higher order processes. As a practical application, the above mentioned results have been applied to the monthly electric power of the "Yarn and Textile Company", Shebin Elkom, Egypt, During the years 1982-1987.

1. INTRODUCTION

The identification problem of ARIMA process is one of the fundamental problems for time series analysis and signal processing. The problem has been considered by many authors [4-8], and the application of ARIMA process to system identification and adaptive control can be found, for example [3,4]. The identification problem of ARIMA process can be established in two steps. The first is to determine the order of the process and usually the "Residual variance plots" are used [4], and this method depends on the parameters of the process that are practically unknown. The second step is to estimate the parameters of the process and usually the time series data are fitted by a high order AR (autoregressive) model and a reduced order ARIMA model from the high order AR model is derived by applying various model reduction methods [2,4].

In this paper, a method based on the auto-correlation function of the given data is used for the order determination of the process and there is no need to calculate the "Residual variance plots", and to compare the results, the AIC (Akaike's information criterion) [1,4] expression has been used. An algorithm for the parameter estimation of the model is given and, in contrary to earlier investigators, can be extended to higher order processes.

2. PROBLEM FORMULATION

Consider the autoregressive integrated moving average process, ARIMA (p,d,q), given by

$$G(B)\, u(t) = C(B)\, e(t)$$

$$t = 1, 2, \ldots, N, \; u(t) = 0, \text{ for } t \leq 0 \tag{1}$$

where N is the number of observations, $\{e(t)\}$ is a sequence of i.i.d. Gaussian variables having variance σ_e^2, and B is the backward shift operator defined by $Bu(t) = u(t-1)$. The polynomial G(B) is a nonstationary autoregressive operator such that d of the roots of $G(B) = 0$ are unity and the remainder lie outside the unit circle. The polynomial C(B) is the moving average operator, it is assumed to be invertible, that is, the roots of $C(B) = 0$ lie outside the unit circle. Equation (1) can be written in the form

$$A(B)\,(1-B)^d\, u(t) = C(B)\, e(t) \tag{2}$$

where A(B) is the autoregressive operator and it is assumed to be stationary, that is, the roots of $A(B) = 0$ lie outside the unit circle. The polynomials A(B) and C(B) have the form

$$\begin{aligned} A(B) &= 1 - a_1B - a_2B^2 - \ldots - a_pB^p \\ C(B) &= 1 - c_1B - c_2B^2 - \ldots - c_pB^p \end{aligned} \tag{3}$$

Equation (2) can be written in the form

$$A(B)\, Z(t) = C(B)\, e(t) \tag{4}$$

where Z(t) is the dth difference of the series and is given by

$$Z(t) = (1-B)^d\, u(t) \tag{5}$$

Thus, the model corresponding to the dth difference of the series can be represented by a stationary, invertible ARMA process. For a given time series it is required to

Ldetermine the orders p, d and q of the model (2) and to estimate the parameters of the polynomials A(B) and C(B) of the model (4).

3. MODEL IDENTIFICATION

For different values of the orders p, d and q, the autocorrelation function has been calculated and saved in the memory of the computer. For a given time series the autocorrelation function will be compared with the saved plots. To ensure the accuracy of the results, the AIC is used, where

$$AIC(r) = n \log \hat{\sigma}_{\varepsilon}^{2} + 2r \quad (6)$$

where r is the number of independently adjusted parameters, n is the number of effective observations, and $\hat{\sigma}_{\varepsilon}^{2}$ is the maximum likelihood estimate of the residual variance. For the estimation of the parameters the model (4) is used, where the residual is given by

$$\varepsilon(t) = Z(t) - a_1 Z(t-1) - \ldots - a_p Z(t-p)$$

$$+ c_1 e(t-1) + \ldots + c_q e(t-q) \quad (7)$$

The sum of squares of the residuals is given by

$$S(a,c) = \sum_{t=1}^{N} [\varepsilon(t)]^2$$

By differentiating S(a,c) with respect to the unknown parameters, we get the normal equations which can be written in vector form as

$$Z\theta = Y$$

where

$$\theta = [a_1 \ a_2 \ \ldots \ a_p \ c_1 \ c_2 \ \ldots \ c_q]_T$$

$$Y = [Y(1)\ Y(2) \ldots Y(p)\ X(1)\ X(2) \ldots X(q)]^T$$

where

$$Y(i) = \sum_{t=1}^{N} z(t) z(t-i), \quad X(j) = \sum_{t=1}^{N} z(t) e(t-j)$$
$$i = 1, 2, \ldots, p, \quad j = 1, 2, \ldots, q$$

and the matrix Z is given by

$$Z = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$$

where the elements of the submatrices are given by

$$M_{11}(i,j) = \sum_{t=1}^{N} z(t-i) z(t-j)$$
$$i = 1, 2, \ldots, p, \quad j = 1, 2, \ldots, p$$

$$M_{12}(i,j) = -\sum_{t=1}^{N} z(t-i) e(t-j)$$
$$i = 1, 2, \ldots, p, \quad j = 1, 2, \ldots, q$$

$$M_{21}(i,j) = \sum_{t=1}^{N} e(t-i) z(t-j)$$
$$i = 1, 2, \ldots, q, \quad j = 1, 2, \ldots, p$$

$$M_{22}(i,j) = -\sum_{t=1}^{N} e(t-i) e(t-j)$$
$$i = 1, 2, \ldots, q, \quad j = 1, 2, \ldots, q$$

4. SIMULATION RESULTS

Example 4.1. The series in Figure (1) represents a 100 viscosity readings of a chemical process. The autocorrelation function R(K) is shown in Figure (2). According to the above mentioned results, it has been shown that the series follows an ARIMA(2,0,0), where $a_1 = 1.8$ and $a_2 = -0.85$, and from equation (7) the model is given by

$$u(t) = 1.8\, u(t-1) - 0.85\, u(t-2) + e(t)$$

Example 4.2. The series in Figure (3) represents the monthly electric power of the "Yarn and Textile Company", Shebin Elkom, Egypt, during the years 1982-1987. The autocorrelation function is shown in Figure (4). The series follows an ARIMA(1,0,0) where $a_1 = 0.95$, and the model is given by

$$u(t) = 0.95\, u(t-1) + e(t)$$

Example 4.3. The series in Figure (5) represents the series L from Box and Jenkins [5]. The autocorrelation function R(K) is shown in Figure (6). The series follows an ARIMA(1,1,1), where $a_1 = -0.85$, $c_1 = -0.9$ and $d = 1$. The model is given by

$$u(t) = 0.15\, u(t-1) + 0.85 u(t-2) + e(t) + 0.9\, e(t-1)$$

5. CONCLUSION

The identification of the autoregressive integrated moving average process has been considered. A method, based on the autocorrelation function of the given data, is used to determine the order of the process and the AIC expression has been used to ensure the results. An algorithm for parameter estimation of the model is given and in contrary to earlier work, can be extended to higher order processes. The application of ARIMA processes to system identification and adaptive control has not been considered here and will be published elsewhere.

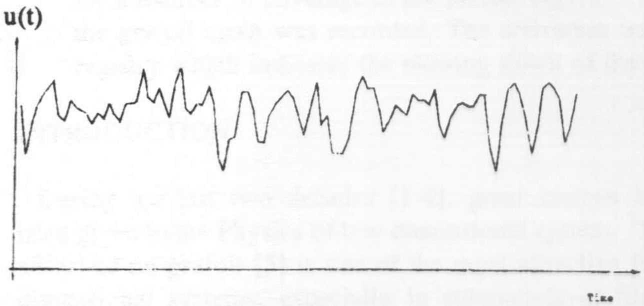

Figure 1. The series given in the example 1.

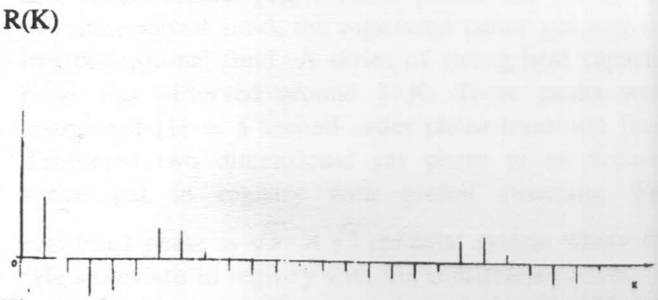

Figure 2. Autocorrelation function for the series in example 1.

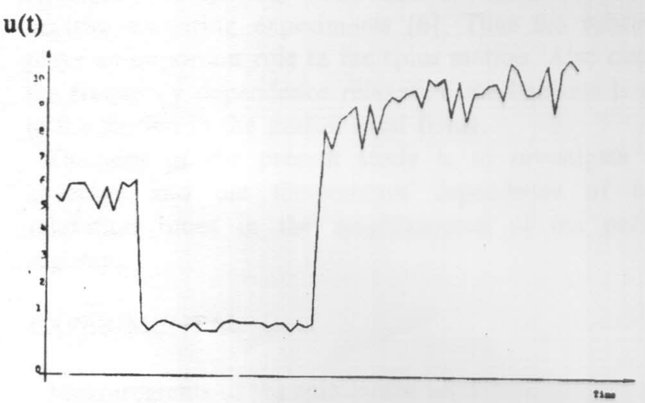

Figure 3. The series given in the example 2.

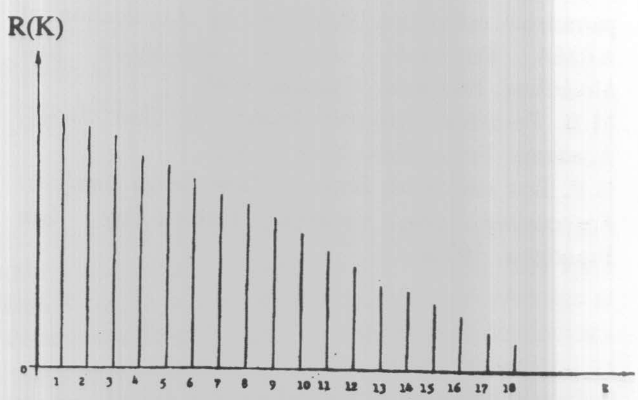

Figure 4. Autocorrelation function for the series in example 2.

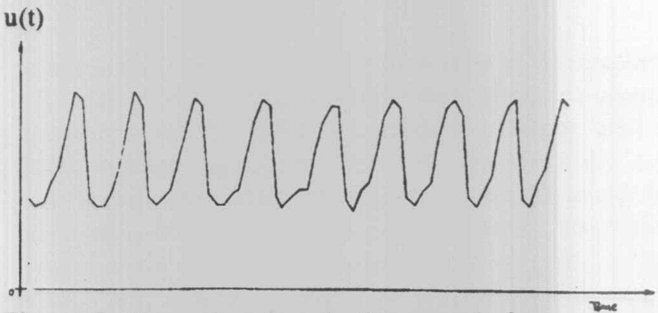

Figure 5. The series given in the example 3.

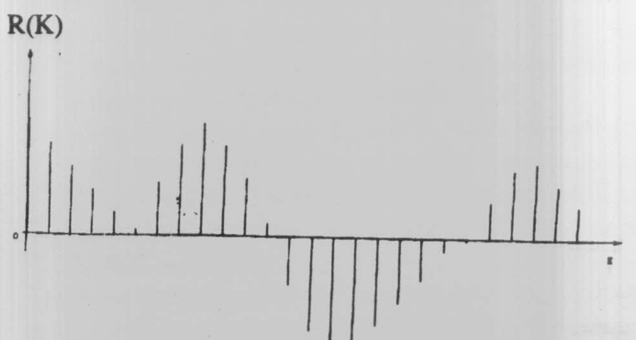

Figure 6. Autocorrelation function for the series in example 3.

REFERENCES

[1] I.J Taneja: On Generalized Information Measures and Their Applications, *adv. Elect. Phys.*, Academic Press, 76, 327/413, 1989.

[2] H. Tanaka and S. Sugimoto: ARMA Modeling via Model Reduction Methods, *The 23rd ISCIE*, Nov. 6-8, Osaka, Japan, 105/108, 1991.

[3] G.A. ElKobrosy: Achievable Accuracy in System parameter estimation Problems by Application of ARMA Processes, *Applied Simulation and Modelling*, Montreal, Canada, 1985.

[4] M.B. Priestley: *Spectral Analysis of Time Series*, Academic Press, New York, 1981.

[5] G.P. Box and G.M. Jenkins: *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, 1976.

[6] T. Ozaki: On the Order Determination of ARIMA Models, *Appl. Stat.*, 26, 290/301, 1977.

[7] V. Solo: The Order of Differencing in ARIMA Models, *JASA*, 79, 388/397, 1984.

[8] R.D. Martin: *Robust Method for ARIMA Models*, Time Series Analysis, Theory and Practice, O.D. Anderson (editor), North Holland, 1983.