THE EGYPTAIN STATISTICAL JOURNAL
ISSR, CAIRO UNIV., VOL.,41, NO.2, 1997

BAYESIAN CLASSIFICATION OF MULTIVARIATE AUTOREGRESSIVE SOURCES WITH UNKNOWN ORDER

SAMIR M. SHAARAWY AHMED L. DAIF MONA A. EL SHAFY

FACULTY OF ECONOMICS AND POLITICAL SCIENCE
CAIRO UNIVERSITY

ABSTRACT

The objective of this paper is to present a Bayesian classification technique that can be used to classify a multivariate time series realization into one of several multivariate autoregressive sources. The main assumption here is that the sources share a common unknown order. This is to be closer to the real situations, whereas the order of a process is usually unknown or at least, has to be estimated. Hence, the order of the processes will be regarded as a nusiance parameter that has a maximum known value. The proposed technique is based on deriving the marginal posterior mass function of a classification vector, then one can assign the multivariate realization to the r-th multivariate source whenever the posterior mass function of the classification vector has its largest value at the r-th mass point. A simulation technique is carried out to study the numerical efficiency of the proposed classification technique. The simulation studies the behavior of the technique with respect to time series length, the maximum value assumed to the order of the process, and the parameter values, on the performance of the proposed technique.

1-INTRODUCTION

Exact specifications of population are the bases of classical statistical methods for estimation and hypotheses testing. In many practical situations, an observation may be assigned to one of several populations. For example, a psychologist classifies candidates for admission to a school, to accept them or not, on the basis of scores they get in a previously decided examination.

The classification problem of time series is to assign a time series realization into one of several time series sources. This problem may be faced in many fields such as economics, business, demography, and environmental data. For example, if an economist aims to classify the development stage of a certain country to one of development stages of several countries, he will take some selected economic variables as a basis. These variables may be national income, investment rate, and consumption expenditure. If these variables are available in time series data, a classification problem of multivariate time series may arise.

Walker (1967) began the work of non-Bayesian classification of time series. He has established an approximate discrimination method for ARMA sources using the idea of Cox (1962). Some other non-Bayesian contributions of the classification problems can be found in Gersch and Brotherton (1979), Gersch (1981), and Kedem and Slud (1982). The first Bayesian contribution to the classification problem of time series was presented by Broemeling and Son (1987). They presented a Bayesian technique to assign a univariate time series realization into one of several autoregressive sources that have common known order and a common unknown

precision. Shaarawy and Haroun (1992) presented an approximate Bayesian classification technique that can be used with univariate moving average sources that share common known order and unknown precision. Shaarawy (1992) is the first to introduce a Bayesian classification technique of multivariate time series. He developed a Bayesian classification technique to classify a multivariate time series into one of multivariate autoregressive sources, with unknown coefficients, that share a common known order and unknown precision matrix. Finally, trying to include more generalized cases, Shaarawy and Al-Mahmeed (1995) developed a Bayesian classification technique that can be used with multivariate autoregressive sources with different orders and common unknown precision.

2- THE CLASSIFICATION PROBLEM

Consider the problem of assigning the multivariate time series realization

$S = [y(1)\ y(2)\ \ldots\ y(n)]'$ where,

y (t) is a sequence of $k \times 1$ real observable random vectors, into one of m multivariate autoregressive, $AR_k(p)$, sources.

$$\underset{(n_i \times k)}{Y_i} = \underset{(n_i \times kp)}{Z_i}\ \underset{(kp \times k)}{\Phi_i} + \underset{(n_i \times k)}{U_i} \qquad i = 1, 2, \ldots m$$

where

$$\underset{(n_i \times k)}{Y_i} = [y_i(1)\ y_i(2) \ldots y_i(n_i)]',$$

$$\underset{(k \times 1)}{y_i(t)} = [y_i(t,1)\ y_i(t,2) \ldots y_i(t,k)]' \qquad t = 1, 2, \ldots n_i$$

$$\underset{(n_i \times kp)}{Z_i} = \begin{bmatrix} z_i'(0) \\ z_i'(1) \\ \vdots \\ z_i'(n_i - 1) \end{bmatrix} = \begin{bmatrix} y_i'(0) & y_i'(-1) & \dots & y_i'(1-p) \\ y_i'(1) & y_i'(0) & \dots & y_i'(2-p) \\ \vdots & & & \\ y_i'(n_i - 1) & y_i'(n_i - 2) & \dots & y_i'(n_i - p) \end{bmatrix}$$

$$\underset{(kp \times k)}{\Phi_i} = \begin{bmatrix} \phi_i(1,1) & \phi_i(1,2) & \dots & \phi_i(1,k) \\ \phi_i(2,1) & \phi_i(2,2) & \dots & \phi_i(2,k) \\ \vdots & \vdots & & \vdots \\ \phi_i(kp,1) & \phi_i(kp,2) & \dots & \phi_i(kp,k) \end{bmatrix}$$

and

$$\underset{(n_i \times k)}{U_i} = \left[\varepsilon_i(1)\ \varepsilon_i(2)\ \dots\ \varepsilon_i(n_i) \right]', \text{ where}$$

$$\underset{(k \times 1)}{\varepsilon_i(t)} = \left[\varepsilon_i(t,1)\ \varepsilon_i(t,2)\ \dots\ \varepsilon_i(t,k) \right]' \qquad t = 1, 2, \dots n_i$$

It is assumed that $\{\varepsilon_i(t), i = 1, 2, \dots, m\}$ is a sequence of independent and normally distributed k×1 unobservable random vectors with zero mean and k × k positive definite unknown precision matrix T. It is also assumed that the order p is unknown. The parameters Φ_1, Φ_2, . . . , and Φ_m, are unknown matrices, thus it is assumed that there is a realization matrix available from each of the m sources, say S_i where,

$$\underset{(n_i \times k)}{S_i} = \left[y_i(1)\ y_i(2)\ \dots\ y_i(n_i) \right]', \qquad i = 1, 2, \dots, m$$

The main problem is to use the Bayesian approach to assign a multivariate time series realization S to one of the $AR_k(p)$ sources S_i. The proposed Bayesian technique is based on introducing the classification vector:

$\lambda = [\lambda_1\ \lambda_2\ \ldots\ \lambda_m]'$, where $\lambda_i = 0, 1$ and $\sum_{i=1}^{m} \lambda_i = 1$

This means that the classification vector λ has m mass points $[1, 0, \ldots, 0]'$, $[0, 1, 0, \ldots, 0]'$, $\ldots$, $[0, 0, \ldots 0, 1]'$. Our proposed technique is to assign the multivariate time series realization S to the rth $AR_k(p)$ source whenever the marginal posterior mass function of λ has its maximum value at the r-th mass point $[0, 0, \ldots, 1, 0, \ldots, 0]'$ i.e., a vector with 1 in position r and 0 elsewhere. In order to obtain the conditional likelihood function, one should combine the contribution of the multivariate realization S with the contribution of the m multivariate realizations $S_1, S_2, \ldots, S_m$.

The contribution of the multivariate realization to be classified, S to the likelihood function is

$$L(S) \propto |T|^{\frac{n-p}{2}} (2\Pi)^{\frac{(n-p)}{2}} \exp -\frac{1}{2}\sum_{i=1}^{m}\lambda_i \operatorname{tr}\{\sum_{t=p+1}^{n}(y(t) - \Phi_i'(p)\, Z(t-1))(y(t) - \Phi_i'(p) Z(t-1))'\, T\} \tag{2.1}$$

Where,

$$\underset{(n \times k\, p)}{Z(t-1)} = [y(t-1)\ \ y(t-2)\ \ldots\ \ y(t-p)]$$

The contribution of the m multivariate realizations $S_1, S_2, \ldots, S_m$ is

$$L(S_1, S_2, \ldots, S_m) \propto |T|^{\frac{\sum_{i=1}^{m} n_i - mp}{2}} (2\Pi)^{-\frac{\sum_{i=1}^{m} n_i - mp}{2}} \exp -\frac{1}{2}\sum_{i=1}^{m} \operatorname{tr}[\sum_{t=p+1}^{n_i}(y_i(t) - \Phi_i'(p)\, Z_i(t-1)) \times (y_i(t) - \Phi_i'(p)\ Z_i(t-1))'\, T] \tag{2.2}$$

From (2.1) and (2.2), one may write the conditional likelihood function of the parameters $\Phi = [\Phi_1, \Phi_2, \ldots, \Phi_m]$, T, p, and λ as

$$L(\Phi, T, p, \lambda|D) \propto (2\Pi)^{\frac{p+mp}{2}} |T|^{a} \exp -\frac{1}{2}[G_1(\Phi,T,p,t) + G_2(\Phi,T,p,t,\lambda)] \qquad (2.3)$$

Where, $2a = \sum_{i=1}^{m} n_i - mp + n - p$

$$G_1(\Phi, T, p, t) = \sum_{i=1}^{m} tr\left[\sum_{t=p+1}^{n_i} (y_i(t) - \Phi_i'(p) Z_i(t-1))(y_i(t) - \Phi_i'(p) Z_i(t-1))' T\right]$$

$$G_2(\Phi, T, p, t, \lambda) = \sum_{i=1}^{m} \lambda_i \, tr\left[\sum_{t=1}^{n} (y(t) - \Phi_i'(p) z(t-1))(y(t) - \Phi_i'(p) Z(t-1))' T\right]$$

and

$$D = S_1 U S_2 U \ldots U S_m U S.$$

3- THE PROPOSED CLASSIFICATION TECHNIQUE

In order to develop the marginal posterior mass function of the classification vector λ, one should combine the conditional likelihood function (2.3) with the prior density of the parameters via Bayes theorem. An appropriate choice of the prior mass function of the classification vector λ is

$$\zeta_1(\lambda) \propto \prod_{i=1}^{m} \alpha_i^{\lambda_i} \quad \text{Where, } \alpha_i > 0 \text{ and } \sum_{i=1}^{m} \alpha_i = 1 \qquad (3.1)$$

and α_i is the prior probability that the realization S is generated from the ith process. If one has "little" information about the classification vector, one may use the prior mass function.

$$\zeta_2(\lambda) = 1/m \qquad (3.2)$$

The parameters Φ and T are assigned a non- informative Jeffreys' prior .

$$\zeta_3(T, \Phi) \propto |T|^{-\frac{(k+1)}{2}}, \quad \Phi_i \in R^{kp \times k} \qquad i = 1, 2, \ldots, m \qquad (3.3)$$

The marginal prior mass function of p is assumed to be

$$\zeta_4(p) = 1 / w \qquad , \; p = 1, 2, \ldots , w \tag{3.4}$$

Where w, is the maximum possible value of p, that is assumed to be known.

Combining (3.1), (3.3) and (3.4), the joint prior density of the parameters Φ, T, p and λ is

$$\zeta(\Phi, T, p, \lambda) \propto |T|^{-\frac{(k+1)}{2}} \prod_{i=1}^{m} \alpha_i^{\lambda_i} \; w^{-1} \tag{3.5}$$

Combining the prior density (3 .5) and the conditional likelihood function (2.3), the joint posterior density of the parameters is

$$\zeta(\Phi, T, p, \lambda|D) \propto (2\Pi)^{\frac{(p+mp)}{2}} |T|^{\nu} \prod_{i=1}^{m} \alpha_i^{\lambda_i} \exp -\frac{1}{2} \left[G_1(\Phi, T, p, t) + G_2(\Phi, T, p, t, \lambda) \right] \tag{3.6}$$

Where, $\nu = \sum_{i=1}^{m} n_i + n - mp - p - k - 1$

The marginal posterior mass function of the classification vector λ can be obtained by integrating (3.6) over Φ and T, and then taking summation over p.

Theorem 1

Using the conditional likelihood function (2.3) and the prior density (3.5), the marginal posterior mass function of the classification vector λ will be

$$g(\lambda | D) \propto \sum_{p=1}^{W} g(p, \lambda | D) \tag{3.7}$$

Where

$$g(p,\lambda|D) \propto \frac{(2\Pi)^{\frac{mp+p+mpk}{2}} 2^{\frac{k}{2}\left(\sum_{i=1}^{m} n_i - mp + n - p - mpk\right)} \prod_{i=1}^{m} \alpha_i^{\lambda_i} \prod_{\alpha=1}^{k} \left(\frac{\sum_{i=1}^{m} n_i - mp + n - p - mpk + k - \alpha}{2}\right)}{|A(\lambda,p)|^{\frac{k}{2}} \left|C - B'(\lambda,p)A^{-1}(\lambda,p)B(\lambda,p)\right|^{\nu}} \quad (3.8)$$

where

A (λ, p) is a square matrix of order m k p

B (λ, p) is a matrix of order m k p × k

C (λ, p) is a matrix of order k × k.

Furthermore, the matrix

$A(\lambda, p) = \text{Diag}(A_1, A_2, \ldots . A_m)$

where A_i is a square matrix of order kp and

$$A_i = \sum_{t=p+1}^{n_i} Z_i(t-1)Z_i'(t-1) + \lambda_i \sum_{t=p+1}^{n} Z(t-1)Z'(t-1) \quad i = 1, 2, \ldots m$$

The m p k × k matrix $B(\lambda, p) = (B_1' \; B_2' \ldots B_m')$

where the matrix B_i is of order k p × k and

$$B_i = \sum_{t=p+1}^{n_i} Z_i(t_i - 1)y_i'(t) + \lambda_i \sum_{t=p+1}^{n} Z(t-1)y'(t) \quad i = 1, 2, \ldots m$$

The k x k matrix C is $C = \sum_{i=1}^{m} C_i$ where the matrix C_i is of order k × k and

$$C_i = \sum_{t=p+1}^{n_i} y_i(t)y_i'(t) + \lambda_i \sum_{t=p+1}^{n} y(t)y'(t) \quad i = 1, 2, \ldots m$$

The scalar ν is

$$2\nu = \sum_{i=1}^{m} n_i + n - mp - p - m\,p\,k$$

the prior probabilities $\alpha_1, \alpha_2, \ldots, \alpha_m$ are supplied by the user. If one has no information about α_i ' s one may set $\alpha_i = 1/m$. $i = 1, 2, \ldots m$.

Theorem1 gives an analytical form to the marginal posterior mass function of the classification vector that can be easily handled by computers. Denoting the marginal posterior probabilities by α_1^*, α_2^*, . . . , and α_m^* respectively, and let $\alpha_r^* = \max_{j=1}^{m} \alpha_j^*$, then our Bayesian procedure is to assign the multivariate time series realization S to the r th AR_k (p) source.

4-NUMERICAL STUDY

Our objective of this simulation study, is to evaluate the adequacy of the proposed technique in handling classification problems of bivariate autoregressive sources. The behavior of the rule to the parameter values, the time series length, and the maximum value assumed to the order of the process will be studied. The main idea of this simulation is to use the technique to classify a time series realization of length n, whose source is known , into one of two bivariate autoregressive time series sources, that share a common order equal to 1, and one of them is the true source. Then, the percentage of the correct classification is computed to be taken as an indicator to the efficiency of the proposed technique. For all the cases considered here, it is assumed that $\alpha_1 = \alpha_2 = 0.5$, so apriori, the realization has the same chance of being assigned to each source. For all original sources and realizations the covariance matrix of the noise term is fixed at $\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$.To study the behavior of the proposed procedure with respect to the closeness of the coefficients of the two original sources, our study involves four simulations. For every simulation, different two parameter matrices (Φ_1 and Φ_2) are chosen for the two sources. i.e., the first source is bivariate autoregressive of order 1 with parameter matrix Φ_1 and covariance matrix $\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$ and the second source is bivariate autoregressive of order 1 with parameter matrix Φ_2 and the same covariance matrix. To show the effect of the time series length on the efficiency of the technique, we select n (the time series length) to be 20, 30, 50, 100, 150, 200, 250, and

300. Whereas the effect of the value of w (the maximum value of the autoregressive order p) is studied by selecting w to be 1, 2, 3, and 4.

Simulation I, as an illustration, begins with the generation of a pair of 500 data sets of bivariate normal variates, each of size 300, to represent $\varepsilon_1(t)$ and $\varepsilon_2(t)$ respectively. These data sets are then used, two times. to generate a pair of 500 realizations, each of size 300, from the $AR_2(1)$ process, with parameter matrix $\Phi_1 = \begin{bmatrix} 0.5 & 0.4 \\ -0.5 & -0.4 \end{bmatrix}$ in the first time and to generate a pair of 500 realizations, each of size 300, from the $AR_2(1)$ process with parameter matrix $\Phi_2 = \begin{bmatrix} -0.5 & 0.4 \\ 0.5 & -0.4 \end{bmatrix}$ in the second, assuming the starting values are zero. Such realizations are used as the time series to be classified in run 1 and run 2 respectively.

The second step of simulation I involves the generation of two data sets of bivariate normal variates, each of size 300. The first data set is used to generate a sample of 300 observations of bivariate autoregressive process of order 1 using the same above value of Φ_1 as its parameter. Such a sample is considered as the first true known source. The second data is used later for the generation of a sample of 300 observations to the second source using the above value of Φ_2. For a given specific sample size, these two data sets are fixed throughout the classification process.

The third step of simulation I is to carry out the computations required to classify each of the 500 realizations, which have been generated from the first source, to one of the two known sources, and to find the percentage of correct classification. Each of the 500 realizations, which have been generated from the second source, are also classified and the percentage of correct classification is calculated. Such classifications are done for a given time series length n using the first n observations of the two known sources as well as the realizations to be classified.

The last classification step is repeated for each chosen time series length n and for each value of w. The results of simulation I are shown in table I.a and I.b. The first table I.a (run 1) presents the percentages of correct classification of the 500 realizations generated from the first known source for all chosen time series lengths and all the values of w, while the second table I.b (run 2) gives such percentages of

the 500 realizations generated from the second source.

The other simulation studies are done in a similar way, but using different values of the matrix parameter. The results are presented in cases II, III, and IV. The parameter matrices in every case are chosen in such a way to satisfy the stationarity conditions, see Harvey (1981). The choice of the parameters is made to show the effect of the closeness of these parameters to that of the true source on the classification technique. In simulation IV, the parameters are chosen to be close to the stationarity border, see tables IV.a and IV.b.

The numerical results of the simulation studies are presented in tables I.a, I.b II.a, II.b, III.a, III.b, IV.a and IV.b. It is noticed that, in general, there is an increasing trend in the efficiency of the proposed technique as the sample size increases. In simulations I, II where the parameter matrices are different , the increasing trend starts from the beginning. Whereas in simulation III where the parameter matrices are close, this trend starts at time series length 150 in run 1 and at time series length 50 in run 2. In simulation IV, the parameter matrices are chosen to be close to each other and also close to the stationarity border, the trend starts from the time series length 100 in run 1 and from the beginning in run 2, see table IV.a and IV.b. For all the cases, the best work for the classification technique is at $w = 1$, which introduces the exact value of p, but the difference in the correct classification percentage as the value of w increasing is low. This means that, whatever the maximum value of p, the performance of the classification technique is slightly affected.

Behavior of the Classification Technique for Bivariate AR (1)

CASE I

$$\Phi_1 = \begin{bmatrix} 0.5 & 0.4 \\ -0.5 & -0.4 \end{bmatrix} \qquad \Phi_2 = \begin{bmatrix} -0.5 & 0.4 \\ 0.5 & -0.4 \end{bmatrix}$$

Table (I.a)
Run (1)

	Time Series Length (n)							
	20	30	50	100	150	200	250	300
w=1	99.6	100	100	100	100	100	100	100
w=2	99	99.8	100	100	100	100	100	100
w=3	98.2	99.8	100	100	100	100	100	100
w=4	96.6	98.8	100	100	100	100	100	100

Table (I.b)
Run (2)

	Time Series Length (n)							
	20	30	50	100	150	200	250	300
w=1	99.4	99.4	100	100	100	100	100	100
w=2	99	99.8	100	100	100	100	100	100
w=3	98.6	99	100	100	100	100	100	100
w=4	97.4	98.8	99.8	100	100	100	100	100

Behavior of the Classification Technique for Bivariate AR (1)

CASE II

$$\Phi_1 = \begin{bmatrix} -0.5 & 0.4 \\ 0.5 & -0.4 \end{bmatrix} \qquad \Phi_2 = \begin{bmatrix} -0.2 & 0.2 \\ 0.2 & -0.2 \end{bmatrix}$$

Table (II.a)
Run (1)

	Time Series Length (n)							
	20	30	50	100	150	200	250	300
w=1	85.4	85.4	94.6	99.2	99.6	99.8	99.8	99.8
w=2	84.4	80.6	94.4	98.8	99	99.2	99.6	99.6
w=3	83.4	77	94	99.2	99.2	99.4	99.8	99.8
w=4	82.4	69.4	93.4	99	99.2	99.4	99.6	99.8

Table (II.b)
Run (2)

	Time Series Length (n)							
	20	30	50	100	150	200	250	300
w=1	86.2	88.6	99.6	99.8	100	100	100	100
w=2	84	91.6	99.2	100	100	100	100	100
w=3	83.2	92.6	99	99.8	100	100	100	100
w=4	79.4	91.1	98.4	99.8	100	100	100	100

Behavior of the Classification Technique for Bivariate AR (1)

CASE III

$$\Phi_1 = \begin{bmatrix} -0.5 & 0.4 \\ 0.5 & -0.4 \end{bmatrix} \qquad \Phi_2 = \begin{bmatrix} -0.4 & 0.3 \\ 0.4 & -0.4 \end{bmatrix}$$

Table (III.a)
Run (1)

	Time Series Length (n)							
	20	30	50	100	150	200	250	300
w=1	70.6	60	63.2	88.4	73.2	84.8	91	94.8
w=2	68.2	47.8	56.2	82.2	59	76	84.6	91.8
w=3	64.6	39	50.4	73.8	48	76.4	80.6	90.4
w=4	62	32	50.8	65	42	73.6	76.6	88.6

Table (III.b)
Run (2)

	Time Series Length (n)							
	20	30	50	100	150	200	250	300
w=1	55.2	53.6	69.8	90	100	99.8	100	99.6
w=2	59.2	63.2	74.6	93.4	100	100	100	100
w=3	62.4	69.6	71	93.4	100	100	100	100
w=4	59.6	73.6	65.4	92	100	100	99.8	100

ISSR, CAIRO UNIV., VOL.,41, NO.2, 1997 PP.205-221

Behavior of the Classification Technique for Bivariate AR (1)

CASE IV

$$\Phi_1 = \begin{bmatrix} 1 & -0.2 \\ 1.1 & -1 \end{bmatrix} \qquad \Phi_2 = \begin{bmatrix} 0.9 & -0.2 \\ 1.1 & -0.9 \end{bmatrix}$$

Table (IV.a)
Run (1)

	Time Series Length (n)							
	20	30	50	100	150	200	250	300
w=1	78.2	32.2	90.4	77.2	78.6	82.4	94	97.4
w=2	81.2	40	87.2	65.8	71.8	81.6	95	97.8
w=3	77	32.8	84.2	66	72	85.4	94.8	98
w=4	72.8	28.4	80.6	62.8	67.4	85.2	92.6	97.2

Table (IV.b)
Run (2)

	Time Series Length (n)							
	20	30	50	100	150	200	250	300
w=1	33.8	64	72.4	98.4	99.8	100	100	100
w=2	26.4	56.4	68	96	96.6	96	95.4	95
w=3	29.6	62.4	68	97.4	99.8	99.8	99.8	99.6
w=4	32.8	66	68	97.6	99.8	99.8	99.6	99.6

CONCLUSION

To sum up, we can conclude that the suggested technique has many advantages. One of these advantages is that the posterior distribution of the classification vector can be computed easily and without parameters estimation. Another point is that it is derived for the case of m sources, which means that, there is no conditions on the number of the sources that can be used. A third point is that the simulation study shows that the proposed technique can, in most cases, provide a high precision level of correct classification with short time series. It also shows that the performance of the technique is slightly affected by the maximum value assumed to the order of the process. Besides, even if the case is close to the stationarity border, it can give good results with moderate time series length. Finally, there is no alternative Bayesian technique can be used to solve such classification problem.

REFERENCES

1) Broemeling, L.D.and Son, M. S. (1987). " The Classification Problem with Autoregressive Processes". *Communications in Statistics. Theory and Methods*, Vol 16, part 4, pp. 927-936 .

2) Cox, D.R. (1962). "Further Results on Tests of Separate Families of Hypotheses". *Journal of the Royal Statistics Society*, Series B 24, pp. 406-424.

3) Gersch, W. (1981). "Nearest Neighbor Rule Classification of Stationary and Non-Stationary Time Series". *Applied Time Series Analysis II*, Edited by David Findley. Academic Press, New York.

4) Gersch, W. and Brotherton, T. (1979). "Discrimination in Locally Stationary Time Series". *Proceeding D.D.C*, San Diego.

5) Harvey, A. C. (1981). *Time Series Models*. John Wiley and Sons Inc. New York.

6) Kedem, B. and Slud, E. (1982). "Time Series Discrimination by Higher Order Crossings". *Annals of the Institute of Statistics*, Vol.10, pp.786-791.

7) Shaarawy, S. M. (1992). "Bayesian Classification with Multivariate Autoregressive Process".*The Egyptian Statistical Journal, ISSR*, Cairo University Vol.36, No. 2, pp. 346-356.

8) Shaarawy, S. M. and Al- Mahmeed, M. (1995). "Bayesian Classification with Multivariate Autoregressive Sources that might have Different Orders". *Communications in Statistics: Simulation and Computation*, Vol. 24, part 3,567-581.

9) Shaarawy, S. M . and Haroun, A. (1992). "Bayesian Classification with General Moving Average Sources". Presented at the Third Islamic Countries Conference on Statistical Sciences (ISSOS-3) pp. 25-34. Rabat, Morocco.

10)Tiao, G. C. and Box, G. E. P. (1981). "Modeling Multiple Time Series with Application". *Journal of the American Statistical Association*, Vol. 76, No, 376, pp. 802-816.

11)Walker, A. M. (1967). "Some Tests of Separate Families of Hypotheses in Time Series Analysis". *Biometrica*, Vol. 54, pp 39-67.