Application of the Multi-Model Partitioning Theory for Simultaneous Order and Parameter Estimation of Multivariate ARMA Models

Stylianos. Sp. Pappas*
Department of Information and Communication Systems Engineering,
University of the Aegean, 83200, Karlovassi, Samos, Greece
E-mail: spappas@aegean.gr
*Corresponding author

Vassilios C. Moussas
School of Technological Applications (S.T.E.F.), Technological Educational Institution (T.E.I.) of Athens. Egaleo. Greece
E-mail: vmouss@teiath.gr

Sokratis. K. Katsikas
Dept. of Technology Education & Digital Systems,
University of Piraeus, 150 Androutsou St., Piraeus GR-18532
E-mail: ska@unipi.gr

Abstract: In this paper a study on how to perform simultaneous order and parameter estimation of Multivariate (MV) ARMA (AutoRegressive Moving Average) models under the presence of noise is addressed. The proposed method, which is computationally efficient, is an extension of a previously presented method for MV AR models and is based on the well established and widely applied multi-model partitioning theory. A series of computer simulations indicate that the method is infallible in selecting the correct model order in very few steps. The simultaneous estimation of the Multivariate ARMA parameters is also another benefit of the proposed method. The results are compared with two other established order selection criteria namely Akaike's Information Crieterion (AIC) and Schwarz's Bayesian Information Criterion (BIC). Finally, it is shown that the method is also successful in tracking model order changes, in real time.

Keywords: Order determination, parameter estimation, multivariate, autoregressive, ARMA, multimodel, partitioning, Kalman.

Reference to this paper should be made as follows:

Biographical notes: S.S. Pappas received BEng (Hons) in Electrical and Electronic Engineering (1997) and MSc in Advanced Control (1998) from University of Manchester Institute of Science and Technology (U.M.I.S.T). He is currently a PhD student at the Department of Information and Communication Systems Engineering, University of the Aegean, Karlovassi, Samos, Greece. His research interest is in the area of Partitioning Theory & their applications. He is an I.E.E.E member.

V.C Moussas received his PhD in Computer Engineering & Informatics from University of Patras, Greece in 2002. He is currently Assistant Professor at the Department of Civil Works Technology School of Technological Applications, Technological Educatioanal Institution of Athens, Greece. His current research interests include material testing, fatigue and lifetime prediction, sensor arrays, multi-model algorithms and parallel networked applications.

S.K. Katsikas received the Diploma in Electrical Engineering from the University of Patras, Greece in 1982, the Master of Science in Electrical & Computer Engineering degree from the University of Massachusetts at Amherst, Amherst, USA, in 1984 and the Ph.D. in Computer Engineering & Informatics from the University of Patras, Greece in 1987. In 1990 he joined the University of the Aegean, Greece, where he served as Professor of the Department of Information & Communication Systems Engineering and as the Rector. In 2007 he joined the Dept. of Technology Education and Digital Systems of the University of Piraeus, as a Professor. His research interests lie in the areas of information and communication systems security and of estimation theory and its applications where he has authored or co-authored more than 150 journal publications, book chapters and conference proceedings publications. He is serving on the editorial board of several scientific journals, he has authored/edited 20 books and has served on/chaired the technical programme committee of numerous international conferences.

1 INTRODUCTION

The problem of fitting a Multivariate (MV) ARMA model to a given time series arises in a large variety of applications, such as speech analysis (Chen *et al.* 2007), biomedical applications (Sheng Lu *et al*. 2001), hydrology (Kourosh *et al.* 2006), electric power systems (Derk *et al.* 2007), simulating earthquake ground motions (Mobarakeh *et al.* 2002), effective multi-channel identification of structures under unobservable excitation (Papakos and Fassois, 2003) and many more.

The aim of this paper is not to add yet another ARMA model selection criterion to the rich literature in this area. Rather we focus on an extension to the model order selection criterion proposed for MV AR models by Pappas, Leros and Katsikas (2006). The method is based on the well known adaptive multimodel partitioning theory (Lainiotis 1976a, 1976b, 1971), it is not restricted to the Gaussian case, it is applicable to on line/adaptive operation and it is computationally efficient. Furthermore, it identifies the correct model order and parameters very fast.

An m-variate ARMA model of order (p, q) [ARMA (p, q)] for a stationary time series of vectors $\mathbf{y}$ observed at equally spaced instants $k = 1, 2, \ldots, n$ is defined as:

$$\mathbf{y}_k = \sum_{i=1}^{p} \mathbf{A}_i \mathbf{y}_{k-i} + \sum_{j=1}^{q} \mathbf{B}_j \mathbf{v}_{k-j} + \mathbf{v}_{k,} \quad \mathrm{E}[\mathbf{v}_k \mathbf{v}_k^T] = \mathbf{R} \tag{1.1}$$

where the m-dimensional vector $\mathbf{v_k}$ is uncorrelated random noise, not necessarily Gaussian, with zero mean and covariance matrix $\mathbf{R}$, $\theta = (p, q)$ is the order of the predictor and $\mathbf{A}_1,\ldots,\mathbf{A}_p$, $\mathbf{B}_1,\ldots,\mathbf{B}_q$ are the m x m coefficient matrices of the MV ARMA model.

It is obvious that the problem is twofold. The first task is the successful determination of the predictor's order $\theta = (p, q)$. Once the model order selection task is completed, one proceeds with the second task, i.e. the computation of the predictor's matrix coefficients $\{\mathbf{A}_i, \mathbf{B}_j\}$

Determining the order of the ARMA process is usually the most delicate and crucial part of the problem. Over the past years substantial literature has been produced for this problem and various different information theoretic criteria, such as Akaike's (1969, 1973, 1974), Rissanen's (1978, 1986), Schwarz's (1978), Wax's (1988) have been proposed to implement the order selection process.

Akaike's and Schwarz's classical order selection procedures are based on the minimization of an objective function of the form

$$\textit{Akaike}\ (\textit{AIC})\text{: } AIC(m) = n\log\left(\hat{\sigma}^2_{e,m}\right) + 2m \tag{2.12}$$

$$\textit{Schwarz}\ (\textit{BIC})\text{: } BIC(m) = n\log\left(\hat{\sigma}^2_{e,m}\right) + m\ln(n) \tag{2.13}$$

where n is the sample size, m is the number of parameters and $\hat{\sigma}^2_{e,m}$ is the *estimated residual variance.*

The criteria may be minimized over choices of m to form a tradeoff between the fit of the model (which lowers the sum of squared residuals) and the model's complexity, which is measured by m. Increasing the number of free parameters to be estimated improves the goodness of fit, regardless of the number of free parameters in the data generating process. Hence *AIC* and *BIC* not only reward goodness of fit, but also include a penalty that is an increasing function of the number of estimated parameters. This penalty discourages overfitting. The preferred model is the one with the lowest *AIC* or *BIC* value. The *AIC* and *BIC* methodology attempts to find the model that best explains the data with a minimum of free parameters.

The above mentioned criteria are not optimal and are also known to suffer from deficiencies; for example, Akaike's information criterion (1969) suffers from overfit (Lutkepohl 1985). Also their performance depends on the assumption that the data are Gaussian and upon asymptotic results. In addition to this, their applicability is justified only for large samples; furthermore, they are two pass methods, so they cannot be used in an on line or adaptive fashion.

In addition to the previous criteria one can mention graphic methods for ARMA order identification based on the autocorrelation and partial autocorrelation functions such as ACF and PACF (Box *et al*., 1994).

The paper is organized as follows. In Section 2 the MV ARMA model order selection problem is reformulated so that it can be fitted into the state space under uncertainty estimation problem framework. In the same section the multi-model partitioning filter (MMPF) is briefly described and its application to the specific problem is discussed. In Section 3, simulation examples are presented which demonstrate the performance of our method in

comparison to previously reported ones. Finally, Section 4 summarizes the conclusions.

2 PROBLEM REFORMULATION

In (Katsikas, Likothanassis and Lainiotis 1990) the problem of simultaneously identifying the order and estimating the parameters of a univariate AR model was reformulated in a state-space form. This work was then extended to univariate ARMA models in (Likothanassis, Demiris, Karelis, 1997), multivariate ARX models in (Demiris, Likothanassis, Katsikas 1998) and to multivariate AR models in (Pappas, Leros and Katsikas 2006). Herein, the approach is extended to cover multivariate ARMA models.

If we assume that the model order fitting the data is known and is equal to $\theta = (p, q)$, we can rewrite equation (1.1) in standard state-space form as:

$$\mathbf{x}(k+1) = \mathbf{x}(k) \tag{2.1}$$

$$\mathbf{y}(k) = \mathbf{H}(k)\mathbf{x}(k) + \mathbf{v}(k) \tag{2.2}$$

where $\mathbf{x}(k)$ is an $m^2 (p + q) \times 1$ vector made up from the coefficients of the matrices $\{\mathbf{A}_1, ..., \mathbf{A}_p, \mathbf{B}_1, ..., \mathbf{B}_q\}$, and $\mathbf{H}(k)$ is an $m \times m^2 (p + q)$ observation history matrix of the process $\{\mathbf{y}(k)\}$ up to time $k - (p + q)$.

Assuming that the general form of the matrix $\mathbf{A}_p$ is:

$$\begin{bmatrix} a_{11}^p & \dots & a_{1m}^p \\ \vdots & \ddots & \vdots \\ a_{m1}^p & \cdots & a_{mm}^p \end{bmatrix}$$ and $\mathbf{B}_q$ is: $$\begin{bmatrix} b_{11}^q & \dots & b_{1m}^q \\ \vdots & \ddots & \vdots \\ b_{m1}^q & \cdots & b_{mm}^q \end{bmatrix}$$ then

$$\mathbf{x}(k) \triangleq [\alpha_{11}^1 \alpha_{21}^1 \cdots \alpha_{m1}^1 \vdots \alpha_{12}^1 \alpha_{22}^1 \cdots \alpha_{m2}^1 \vdots \cdots \alpha_{mm}^1 \vdots \cdots \alpha_{mm}^p \vdots b_{11}^1 b_{21}^1 \cdots b_{m1}^1 \vdots b_{12}^1 b_{22}^1 \cdots b_{m2}^1 \vdots \cdots b_{mm}^1 \vdots \cdots b_{mm}^q]^T$$

$$\mathbf{H}(k) \triangleq [y_1(k-1) I \cdots y_m(k-1) I \vdots \cdots \vdots y_1(k-p) I \cdots y_m(k-p) I \vdots v_1(k-1) I \cdots v_m(k-1) I \vdots \cdots \vdots v_1(k-q) I \cdots v_m(k-q) I]$$

where $\mathbf{I}$ is the $m \times m$ identity matrix and $\theta = (p, q)$, is the model order.

If the system model and its statistics were completely known, the Kalman filter (KF) in its various forms would be the optimal estimator in the minimum variance sense.

Remarks

1) In the case where the prediction coefficients are subject to random perturbations (2.1) becomes

$$\mathbf{x}(k+1) = \mathbf{x}(k) + \mathbf{w}(k) \tag{2.3}$$

$$\mathbf{w}(k) \triangleq [w_{11}^1 w_{21}^1 \cdots w_{m1}^1 \vdots w_{12}^1 w_{22}^1 \cdots w_{m2}^1 \vdots \cdots w_{mm}^1 \vdots \cdots w_{mm}^p \vdots w_{11}^1 w_{21}^1 \cdots w_{m1}^1 \vdots w_{12}^1 w_{22}^1 \cdots w_{m2}^1 \vdots \cdots w_{mm}^1 \vdots \cdots w_{mm}^q]^T \tag{2.4}$$

$\mathbf{v}(k)$, $\mathbf{w}(k)$ are independent, zero–mean, white processes, not necessarily Gaussian.

2) A complete system description requires the value assignments of the variances of the random processes $\mathbf{w}(k)$ and $\mathbf{v}(k)$. We adopt the usual assumption that $\mathbf{w}(k)$ and $\mathbf{v}(k)$ are at least wide sense stationary processes, hence their variances, $\mathbf{Q}$ and $\mathbf{R}$ respectively are time invariant. Obtaining these values is not always trivial. If $\mathbf{Q}$ and $\mathbf{R}$ are not known they can be estimated by using a method such as the one described by Sage and Husa (1969).

In the case of coefficients constant in time, or slowly varying, $\mathbf{Q}$ is assumed to be zero (just like in equation (2.1)).

3) It is also necessary to assume an a priori mean and variance for each $\{\mathbf{A}_i, \mathbf{B}_i\}$. The a priori mean of the $\mathbf{A}_i(0)$'s and $\mathbf{B}_i(0)$'s can be set to zero if no knowledge about their values is available before any measurements are taken (the most likely case). On the other hand the usual choice of the initial variance of the $\mathbf{A}_i$'s and $\mathbf{B}_i$'s, denoted by $\mathbf{P}_0$ is $\mathbf{P}_0 = n\mathbf{I}$, where n is a large integer.

Let us now consider the case where the system model is not completely known. The adaptive multimodel partitioning filter (MMPF) is one of the most widely used approaches for similar problems. This approach was introduced by Lainiotis (1971, 1976a, 1976b) and summarizes the parametric model uncertainty into an unknown, finite dimensional parameter vector whose values are assumed to lie within a known set of finite cardinality. In our problem assume that the model uncertainty is the lack of knowledge of the model order θ. Let us further assume that the model order θ lies within a known sample space of finite cardinality, i.e. that $1 \leq \theta \leq M$, $\theta \in \mathfrak{I}$, where $\mathfrak{I}$ denotes the set of integers.

The MMPF operates on the following discrete-time model:

$$\mathbf{x}(k+1) = \mathbf{F}(k+1, k/\theta)\,\mathbf{x}(k) + \mathbf{w}(k) \tag{2.5}$$

$$\mathbf{y}(k) = \mathbf{H}(k/\theta)\,\mathbf{x}(k) + \mathbf{v}(k) \tag{2.6}$$

where θ is the unknown parameter - the model order in this case- $\mathbf{F}$ is the state transition matrix and $\mathbf{w}(k)$ is independent, zero mean, white noise not necessarily Gaussian with covariance $\mathbf{Q}$ which is usually set to a small positive non zero constant. The optimal MMSE (Minimum Mean Square Error) estimate of $\mathbf{x}(k)$ is given by

$$\hat{\mathbf{x}}(k/k) = \sum_{j=1}^{M} \hat{\mathbf{x}}(k/k; \theta_j)\, p(\theta_j / k) \tag{2.7}$$

A finite set of models is designed, each matching one value of the parameter vector. If the prior probabilities $p(\theta_j / k)$ for each model are already known, these are assigned to each model. In the absence of any prior knowledge, these are set to $p(\theta_j / k) = 1/M$, where M is the cardinality of the model set.

A bank of conventional elemental filters (non adaptive, e.g. Kalman) is then applied, one for each model, which can be run in parallel. At each iteration the MMPF selects the model which corresponds to the maximum posterior probability as the correct one. This probability tends to one, while the others tend to zero. The overall optimal estimate can be taken either to be the individual estimate of the elemental filter exhibiting the maximum posterior probability (MAP) (Lainiotis *et al* 1988), which is the case used in this paper, or the weighted average of the estimates produced by the elemental filters, as described in Eq (2.7). The weights are determined by the posterior probability that each model in the model set is in fact the true model.

The probabilities are calculated on – line in a recursive manner as it is shown by equations (2.8) and (2.9).

$$p(\theta_j/k)=\frac{L(k/k;\theta_j)}{\sum_{j=1}^{M} L(k/k;\theta_j)\, p(\theta_j/k\text{-}1)}\, p(\theta_j/k\text{-}1) \tag{2.8}$$

$$L(k/k;\theta_j)=\left|\mathbf{P}_{\tilde{\mathbf{y}}}(k/k\text{-}1;\theta_j)\right|^{-1/2}\cdot \exp[-\frac{1}{2}\tilde{\mathbf{y}}^{T}(k/k\text{-}1;\theta_j)\,\mathbf{P}_{\tilde{\mathbf{y}}}^{-1}(k/k\text{-}1;\theta_j)\,\tilde{\mathbf{y}}(k/k\text{-}1;\theta_j)] \tag{2.9}$$

where the innovation process

$$\tilde{\mathbf{y}}(k/k\text{-}1;\theta_j)=\mathbf{y}(k)-\mathbf{H}(k;\theta_j)\hat{\mathbf{x}}(k/k\text{-}1;\theta_j) \tag{2.10}$$

is a zero mean white process with covariance matrix

$$\mathbf{P}_{\tilde{y}}(k/k\text{-}1;\theta_j)=\mathbf{H}(k;\theta_j)\mathbf{P}(k/k;\theta_j)\mathbf{H}^{\mathbf{T}}(k;\theta_j)+\mathbf{R} \tag{2.11}$$

For equations (2.8) – (2.11) $j = 1,2, \ldots, M$.

3 EXAMPLES

In order to assess the performance of our method, several simulation experiments were conducted. All of these experiments were conducted for 100 Monte Carlo Runs. For details for the application of stochastic Monte Carlo techniques see Shiryaev (1996) or Christakis (1998). The models used were those of (2.1) and (2.2), with cardinality $M= 10$.

Example 1. ARMA (*1, 1*). $\theta = (1, 1) = 2$.
and cardinality $M = 10$.

$$\mathbf{A}_1=\begin{bmatrix}-0.85 & 0.75\\ 0.65 & -0.55\end{bmatrix},\ \mathbf{B}_1=\begin{bmatrix}-1.9833 & 1.889\\ 1.7 & 1.9833\end{bmatrix},$$

$$\mathbf{R}=\begin{bmatrix}1.5625 & 1.5\\ 1.5 & 1.5625\end{bmatrix}$$

Figure 2 depicts the posterior probabilities associated with each value of θ plotted against normalised time intervals. *Figure 3* shows the criteria comparison for two data sets, one relatively small (50 samples) and one larger (100 samples); and *Table 1* shows the estimated ARMA parameter coefficients.

From *Figure 2*, is obvious that the MMPF identifies the correct probability $\theta = (1, 1) = 2$ very fast, in just *17* steps. Convergence is taken to occur when the posterior probability of the model exceeds *0.9*.

From *Figure 3* we deduce that MMPF is 100% successful in selecting the correct model order for both data sets, while only BIC matches its performance for the larger data set.

Also *Table 1*, shows that the parameter coefficient estimation is very accurate to the considered amount of noise.(*RMSE* – Root Mean Square Error is very small).

Example 2 ARMA (*1, 1*). $\theta = (1, 1) = 2$. This is a more complex MV ARMA since $m = 3$.

$$\mathbf{A}_1=\begin{bmatrix}1 & 0.2 & 0.23\\ 0.15 & 0.18 & 0.16\\ 0.17 & 0.24 & 0.21\end{bmatrix},\ \mathbf{B}_1=\begin{bmatrix}1 & 0.15 & 0.09\\ 0.1 & -0.1 & 0.05\\ 0.05 & 0.13 & 0.075\end{bmatrix}$$

$\mathbf{R}$= **diag** [(*0.42*, *0.01*, *0.16*)].

Figure 4 depicts the posterior probabilities associated with each value of θ plotted against normalised time intervals. *Figure 5* shows the criteria comparison for two data sets, one relatively small (50 samples) and one larger (100 samples); and *Table 2* shows the estimated ARMA parameter coefficients.

MMPF identifies the correct probability $\theta = (1, 1) = 2$ very fast, in just *18* steps (*Figure 4*). Convergence is taken to occur when the posterior probability of the model exceeds *0.9*. Moreover only MMPF is 100% successful in selecting the correct model order for both data sets (*Figure 5*).

Example 3 ARMA (*2, 2*). $\theta = (2, 2) = 4$.

$$\mathbf{A}_1=\begin{bmatrix}-0.17 & 0.14\\ -0.19 & -0.1\end{bmatrix},\ \mathbf{A}_2=\begin{bmatrix}-0.2 & 0.12\\ 0.22 & -0.25\end{bmatrix},$$

$$\mathbf{B}_1=\begin{bmatrix}-0.45 & 0.52\\ -0.32 & -0.7\end{bmatrix},\ \mathbf{B}_2=\begin{bmatrix}-0.85 & 0.75\\ -0.65 & -0.55\end{bmatrix},$$

$$\mathbf{R}=\begin{bmatrix}1 & -0.08\\ -0.08 & 1\end{bmatrix}.$$

Figure 6 depicts the posterior probabilities associated with each value of θ plotted against normalised time intervals. *Figure 7* shows the criteria comparison for two data sets, one relatively small (50 samples) and one larger (100 samples) and *Table 3* shows the estimated ARMA parameter coefficients.

From *Figure 6*, is obvious that the MMPF identifies the correct probability $\theta = (2, 2) = 4$ very fast, in just 24 steps. Convergence is taken to occur when the posterior probability of the model exceeds *0.9*.

From *Figure 7* we deduce that MMPF is 100% successful in selecting the correct model order for both data sets, only BIC matches its performance for the larger data set.

As *Table 3*, clearly shows the parameter estimation is again accurate since the Root Mean Square Error (RMSE) is very small.

Example 4: The aim of this example is to show that the proposed method is able to track changes in the model structure in real time, and thus, the method should handle successfully the problem of the time varying model order. At each iteration, the proposed algorithm selects the model that corresponds to the maximum a posteriori probability as the correct one. This probability tends (asymptotically) to one, while the remaining probabilities tend to zero (*Figure 8*). Thus the algorithm is adaptive in the sense of being able to track model order changes in real time. A noisy set of data is used, which is generated as shown in *Table 4*

The MV ARMA models used in order to create the data set are the ones used in Examples 1 and 3 plus a new MV ARMA (*2,1*), (i.e. $\theta = 3$) with coefficients:

$$\mathbf{A}_1=\begin{bmatrix}0.6 & 0\\ 0.5 & -0.5\end{bmatrix},\ \mathbf{A}_2=\begin{bmatrix}0.8 & -0.2\\ 0.3 & -0.1\end{bmatrix},$$

$$\mathbf{B}_1=\begin{bmatrix}-0.85 & -0.8\\ 0.1 & 0.23\end{bmatrix}\ \text{and}$$

The noise covariance matrix R and the cardinality M for all three models are $\mathbf{R}=\begin{bmatrix} 1 & -0.08 \\ -0.08 & 1 \end{bmatrix}$ and M= *10* respectively.

100 Monte Carlo Runs were used; as *Figure 8* shows, the MMPF is capable of identifying the correct model order (the corresponding posterior probability exceeds *0.9*) for every time interval. When a model order change occurs, the MMPF algorithm needs *M* steps (in our case *M = 10*) to perform variable initialization. This is the reason for having a *10*-step gap before the MMPF starts to increase the posterior probability corresponding to the true model. This initialization is performed automatically, without user intervention.

4 CONCLUSIONS

A new method for simultaneously selecting the order θ (p, q) and for estimating the parameters of a MV ARMA model has been developed, as an extension of the method proposed for the MV AR case (Pappas *et al.* 2006). The proposed method successfully selects the correct model order in very few steps and identifies very accurately the ARMA parameters. Comparison with other established order selection criteria (AIC and BIC) show that the method needs the shortest data set for successful order identification and accurate parameter estimation for all the simulated models, whereas the other criteria require longer data sets as the model order increases. The method performs equally well when the complexity of the MV ARMA model is increased. Finally, the method is capable of tracking, in real time, any model order changes. As a further step to this research is left the task to optimise the algorithm in order to be able to identify the order of the AR component (p) and MA component (q) separately and to apply the algorithm on financial time series analysis (Liatsis and Hussain, 2001).

REFERENCES

Akaike, H., (1969), 'Fitting autoregressive models for prediction', *Ann. Inst. Stat. Math*., vol. 21, pp. 243-247.

Akaike, H., (1973), 'Information theory and an extension of the maximum likelihood principle'. In Petrov, B. N., Csàki F (Eds), *Proceedings of the second international symposium on information theory,* Akadémia Kiadó, Budapest, pp. 267-281.

Akaike, H., (1974), 'A new look at the statistical model identification', *IEEE Trans. on Automatic Control* AC-19, pp. 716-723.

Anderson, B. D. O. , Brinsmead T. S., Bruyne F. De, Hespanha J., Liberzon D., and Morse A. S. (2000), 'Multiple model adaptive control, Part 1: Finite controller coverings', *Int. J. Robust Nonlinear Contr*., vol 10, pp.:909-929.

Box, G. E. P, Jenkins G. M, Reinsel, G. C., (1994), 'Time Series Analysis: Forecasting and Control', Third Edition, Prentice Hall, New Jersey, ISBN: 0130607746.

Chen, C.P.; Bilmes, J. A., (2007), 'MVA Processing of Speech Features, Audio, Speech and Language Processing', *IEEE Transactions on*, vol 15, Issue 1, pp. 257-270.

Christakis, N., (1998), 'Modelling of the microphysical processes that lead to warm rainformation', PhD Thesis, UMIST Manchester.

Demiris, E.N; Likothanassis S.D. and Katsikas, S.K., (1998), "Multivariate ARX Model Identification and Control with Unknown Process Order", in *Proceedings, 2nd IMACS International Conference on Circuits, Systems and Computers,* Vol. II, Piraeus, Greece, pp. 798-803.

Katsikas, S.K.; Likothanassis, S.D. and Lainiotis, D.G., (1990). "AR Model Identification With Unknown Process Order", *IEEE Transactions on Acoustics, Speech, and Signal Processing,* Vol. ASSP-38, no. 5, pp. 872-876.

Kourosh, M.; Eslami H.R. and Kahawita R., (2006) 'Parameter estimation of an ARMA model for river flow forecasting using goal programming', *Journal of Hydrology*, vol 331, Issues 1-2, pp. 293-299.

Lainiotis, D. G., (1971), 'Optimal adaptive estimation: Structure and parameter adaptation'*, IEEE Trans. Automat. Contr*., vol. AC-16, pp. 160-170.

Lainiotis, D. G., (1976a), 'Partitioning: A unifying framework for adaptive systems I: Estimation', *Proc. IEEE*, vol. 64, pp. 1126-1143.

Lainiotis, D. G., (1976b), 'Partitioning: A unifying framework for adaptive systems II: Control', *Proc. IEEE*, vol. 64, pp. 1182-1198.

Liatsis, P. and Hussain, A.J., (2001), 'Financial time series prediction using second-order pipeline recurrent neural networks', *Proc. IWSSIP'01, Bucharest, Romania*, pp. 51-54, June 7-9.

Likothanassis, S.D., Demiris, E.N., Karelis, D.G, (1997), 'A real-time ARMA model structure identification method', *Proceedings of 13th International Conference on Digital Signal Processing* Volume 2, pp. 915-918.

Lutkepohl, H., (1985), 'Comparison of criteria for estimating the order of a vector AR process', *J. Time Ser. Anal. 6*, vol. 6, pp. 35-52.

Mobarakeh, A. A, Rofooei F. R, Ahmadi G, (2002), 'Simulation of earthquake record using time – varying Arma (2,1) model', *Probabilistic Engineering Mechanics,* vol 17, pp. 15-34.

Papakos, V, Fassois, D. S, (2003), 'Multichannel Identification of aircraft skeleton structures under unobservable excitation: a vector AR / ARMA framework', *Mechanical Systems and Signal Processing*, vol 17, Issue 6, pp. 1271-1290

Pappas, Sp. S; Leros A. K. ; Katsikas S. K, (2006), 'Joint order and parameter estimation of multivariate autoregressive models using multi-model partitioning theory', *Digital Signal Processing*, vol 16, Issue 6, pp.782-795.

Rissanen, J., (1978), 'Modeling by shortest data description', *Automatica*, vol. 14, pp. 465-471.

Rissanen, J., (1986), 'A predictive least squares principle', *IMA J. Math. Contr. Inform*., vol. 3, pp. 211-222.

Sage, A., and Husa, G., (1969), 'Algorithms for sequential adaptive estimation of prior statistics', *Proc. Joint. Automatic Control Conf*,. Boulder, Colorado,pp. 760-769..

Schwarz, G., (1978), 'Estimation of the dimension of the model', *Ann. Statist*., vol. 6, pp.461-464.

Sengbush, R. L. and Lainiotis D. G., (1969), 'Simplified parameter quantization procedure for adaptive estimation', *IEEE Trans. Automat. Contr*., vol AC-14, pp.:424-425.

Sheng Lu; Ki Hwan Ju; Chon, K.H, (2001), 'A new algorithm for linear and nonlinear ARMA model parameter estimation using affine geometry [and application to blood flow/pressure data]', *Biomedical Engineering*, IEEE Transactions on, vol 48, Issue 10, pp. 1116-1124.

Shiryaev A. N.,(1996) 'Porbability', Springer-Verlag, New York Inc.

Swider, D., J., Weber C., (2007), 'Extended ARMA models for estimating price developments on day-ahead electricity

markets', *Electric Power Systems Research*, vol 77, Issues 5-6, p.p. 583-593.

Wax, M., (1988), 'Order selection for AR models by predictive least squares', *IEEE Trans. Acoust., Speech, Signal Proc*, vol. 36, pp. 581-588.

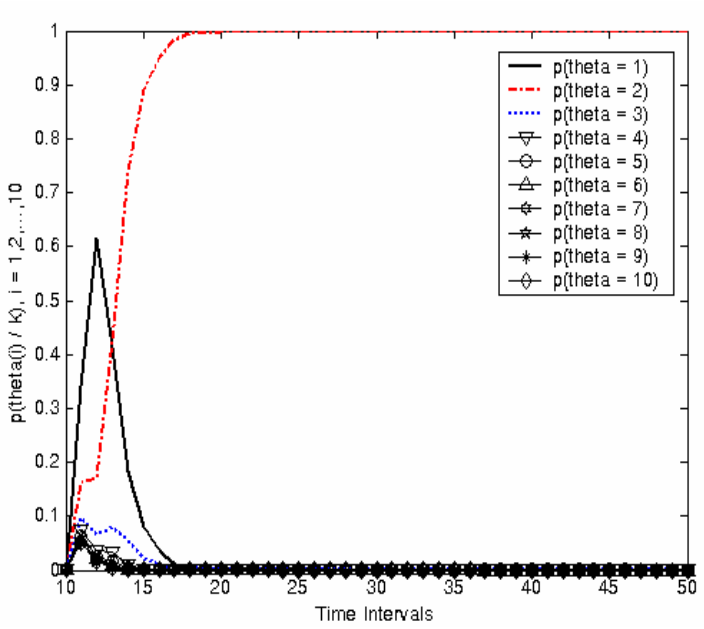

Figure 2 *Example 1, Posterior Probabilities Sequence*

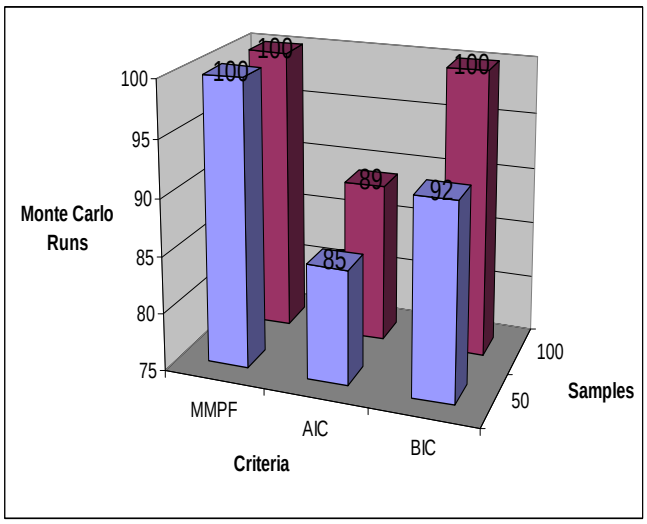

Figure 3 *Example 1, Criteria Comparison– Number of Correct model order identifications out of 100 Monte Carlo Runs*

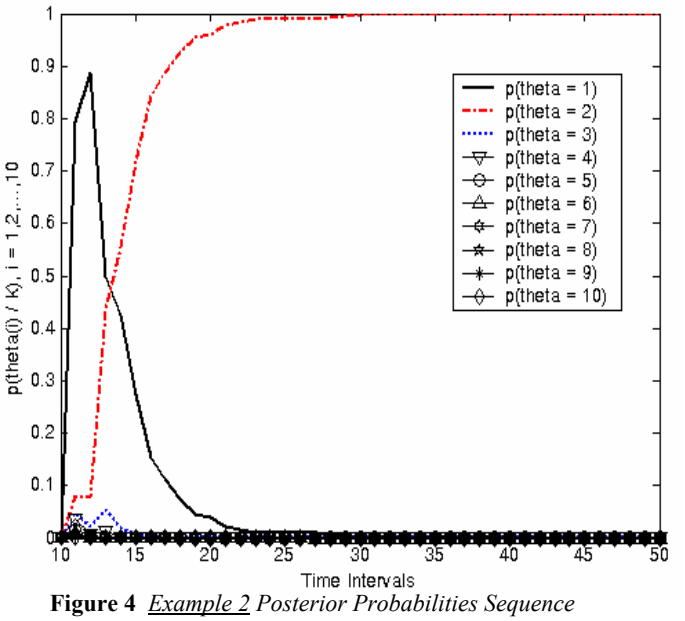

Figure 4 *Example 2 Posterior Probabilities Sequence*

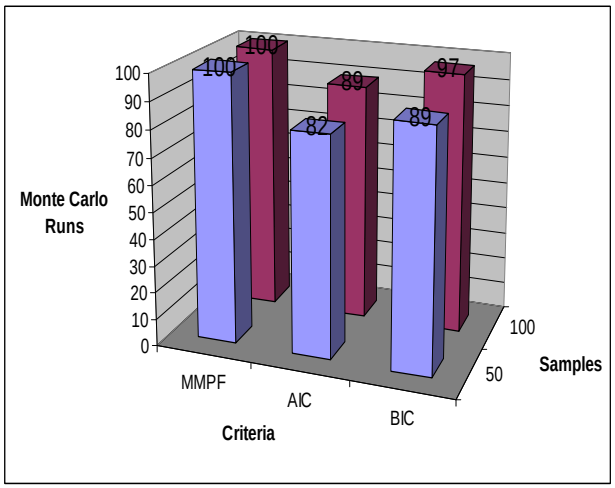

Figure 5 *Example 2, Criteria Comparison– Number of Correct model order identifications out of 100 Monte Carlo Runs*

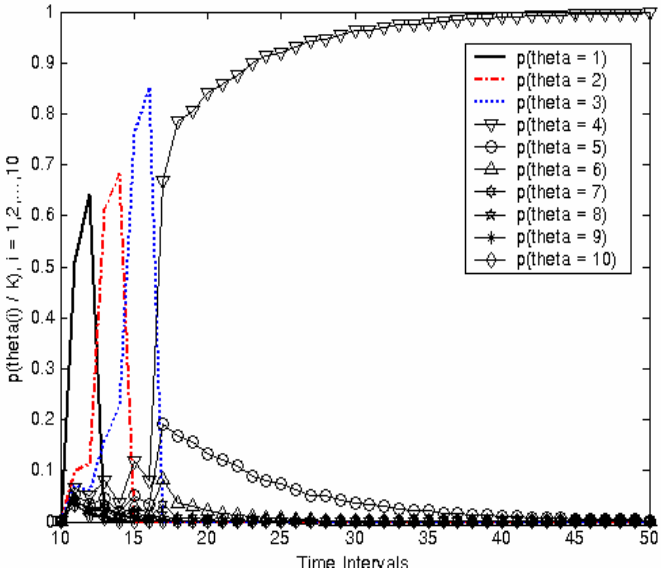

Figure 6 *Example 3 Posterior Probabilities Sequence*

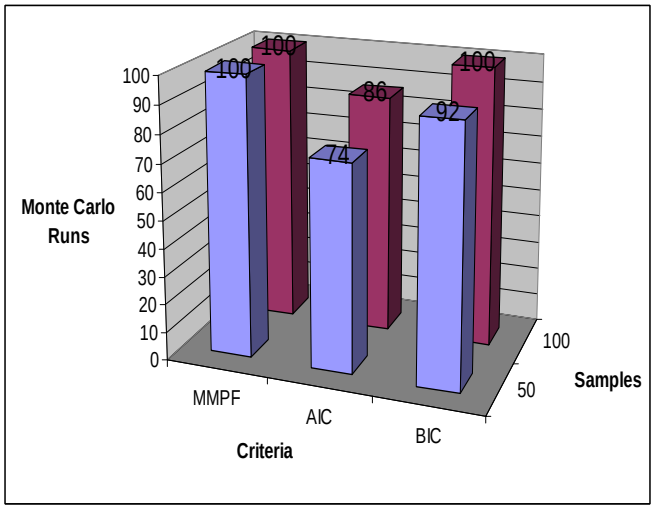

Figure 7 *Example 3, Criteria Comparison– Number of Correct model order identifications out of 100 Monte Carlo Runs*

Estimated Parameters	RMS Error
-0.8499	0.0033
0.6508	0.0036
0.7501	0.0040
-0.5501	0.0032
-1.9823	0.0092
1.7011	0.0064
1.8891	0.0074
1.9831	0.0057

Table 1 *Example 1, Estimated ARMA coefficient parameters*

Estimated Parameters		RMS Error	
0.9932	1.0217	0.0152	0.0026
0.2023	0.1516	0.0035	0.0016
0.2310	0.0894	0.0021	0.0059
0.1519	0.1013	0.0143	0.0023
0.1829	-0.1027	0.0091	-0.0127
0.1612	0.0059	0.0044	0.0019
0.1702	-0.0048	0.0030	0.0017
0.2408	0.1351	0.0371	0.0046
0.2143	0.0742	0.0045	0.0028

Table 2 *Example 2, Estimated ARMA coefficient parameters*

Estimated Parameters		RMS Error	
-0.1691	-0.4535	0.0047	0.0094
-0.1896	-0.3260	0.0054	0.0109
0.1458	0.5244	0.0137	0.0090
-0.0899	-0.6931	0.0161	0.0108
-0.1982	-0.8441	0.0064	0.0122
0.2234	-0.6407	0.0073	0.0147
0.1154	0.7508	0.0094	0.0086
-0.2573	-0.5471	0.0114	0.0094

Table 3 *Example 3, Estimated ARMA coefficients*

Steps	**Model order**	**Steps**	**Model order**
1 – 40	3	121 – 160	3
41- 80	4	161 – 200	4
81 – 120	2		

Table 4 *Time Varying MV ARMA Order Sequence*

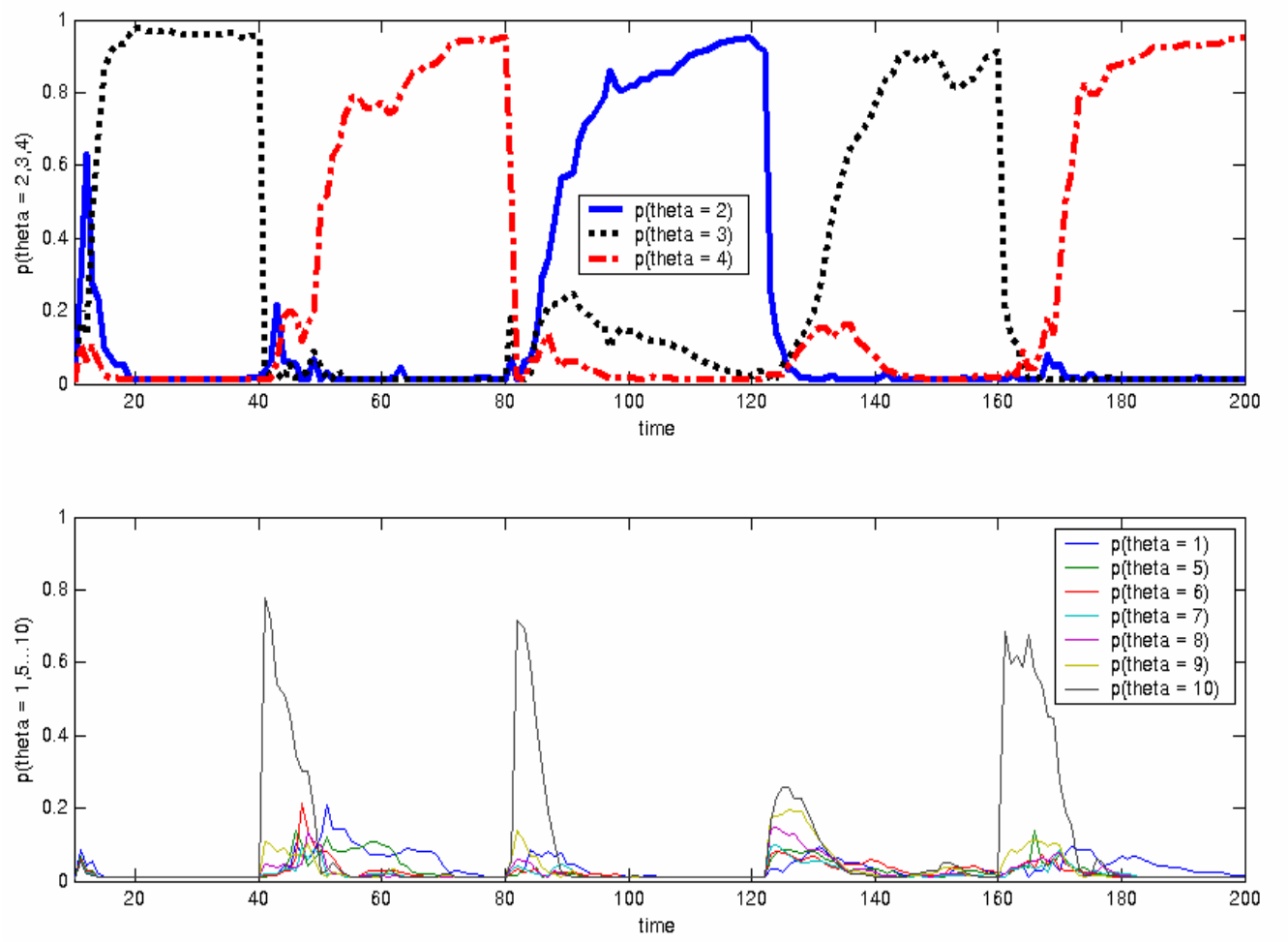

Figure 8 *Example 4 Posterior Probabilities Sequence*