Chapter 5

Adaptive MV ARMA identification under the presence of noise

Stylianos Sp. Pappas[1] Vassilios C. Moussas[2] Sokratis K. Katsikas[3]

[1] University of the Aegean, Department of Information and Communication Systems Engineering 83200 Karlovassi Samos, Greece
e-mail: spappas@aegean.gr
[2] School of Technological Applications (S.T.E.F.), Technological Educational Institution (T.E.I.) of Athens. Egaleo. Greece
e-mail: vmouss@teiath.gr
[3] University of Piraeus, Department. of Technology Education & Digital Systems, 150 Androutsou St., Piraeus,18532, Greece.
e-mail: ska@unipi.gr

Abstract. An adaptive method for simultaneous order estimation and parameter identification of Multivariate (MV) ARMA models under the presence of noise is addressed. The proposed method is based on the well known multi-model partitioning (MMP) theory. Computer simulations indicate that the method is 100% successful in selecting the correct model order in very few steps. The results are compared with two other established order selection criteria namely Akaike's Information Criterion (AIC) and Schwarz's Bayesian Information Criterion (BIC).

Keywords. Multivariate, ARMA, Kalman, Partitioning

5.1 Introduction

The problem of fitting a multivariate ARMA model to a given time series is an essential one in speech analysis, biomedical applications, hydrology, electric power systems and many more [1-4].

In this paper, a new method for multivariate ARMA model order selection and parameter identification is presented, as an extension to the one proposed in [5] for MV AR models. The method is based on the well known adaptive multimodel partitioning theory [6,7], 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 very fast.

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

$$\mathbf{y}_k = \sum_{i=1}^{p} \mathbf{A}_i \mathbf{y}_{k-i} + \sum_{j=1}^{q} \mathbf{B}_j \mathbf{v}_{k-j} + \mathbf{v}_{k,} \qquad \mathrm{E}[\mathbf{v}_k \mathbf{v}_k^T] = \mathbf{R} \qquad (5.1)$$

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

It is obvious that the problem is twofold. The first task, which is the most important for the problem under consideration, 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 criteria, such as Akaike's [8], Rissanen's [9,10], Schwarz's [11], Wax's [12] have been proposed to implement the order selection process.

The above mentioned criteria are not optimal and are also known to suffer from deficiencies; for example, Akaike's information criterion suffers from overfit [13]. 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.

The paper is organized as follows. In Section 5.2 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 sec-

tion the multi-model partitioning filter (MMPF) is briefly described and its application to the specific problem is discussed. In Section 5.3, simulation examples are presented which demonstrate the performance of our method in comparison to previously reported ones. Finally, Section 5.4 summarizes the conclusions.

5.2 Problem reformulation

If we assume that the model order fitting the data is known and is equal to $\theta = (p, q)$, we can rewrite equation (5.1) in standard state-space form as:

$$\mathbf{x}(k+1) = \mathbf{x}(k) \qquad (5.2)$$

$$\mathbf{y}(k) = \mathbf{H}(k)\mathbf{x}(k) + \mathbf{v}(k) \qquad (5.3)$$

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

$$\mathbf{A}_p \; is \begin{bmatrix} a_{11}^p & ... & a_{1m}^p \\ \vdots & \ddots & \vdots \\ a_{m1}^p & \cdots & a_{mm}^p \end{bmatrix} \; and$$

$$\mathbf{B}_q \; is \begin{bmatrix} b_{11}^q & ... & b_{1m}^q \\ \vdots & \ddots & \vdots \\ b_{m1}^q & \cdots & b_{mm}^q \end{bmatrix} \; then$$

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

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

where 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.

In the case where the prediction coefficients are subject to random perturbations (5.2), becomes

$$\mathbf{x}(k+1) = \mathbf{x}(k) + \mathbf{w}(k) \tag{5.4}$$

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

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

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 $\mathbf{v}(k)$ 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 in [14]. In the case of coefficients constant in time, or slowly varying, Q is assumed to be zero (just like in equation (5.4)).

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 $\mathbf{P}_0 = nI$, where n is a large integer.

Let us now consider the case where the system model is not completely known The adaptive multi-model partitioning filter (MMPF) is one of the most widely used approaches for similar problems. This approach was introduced by Lainiotis in [6, 7] 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. A non - exhaustive list of the reformulation, extension and application of the MMPF approach as well as its application to a variety of problems by many authors can be found in [15] and [16-19]. 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 set of finite cardinality: $1 \le \boldsymbol{\theta} \le \mathrm{M}$, where $\theta = (p, q)$, is the model order.

The MMPF operates on the following discrete model:

$$\mathbf{x}(k+1) = \mathbf{F}(k+1, k/\theta)\,\mathbf{x}(k) + \mathbf{w}(k) \tag{5.5}$$

$$\mathbf{y}(k) = \mathbf{H}(k/\theta)\,\mathbf{x}(k) + \mathbf{v}(k) \tag{5.6}$$

where $\theta = (p, q)$ is the unknown parameter, the model order in this case. A block diagram of the MMPF is presented in Figure 5.1.
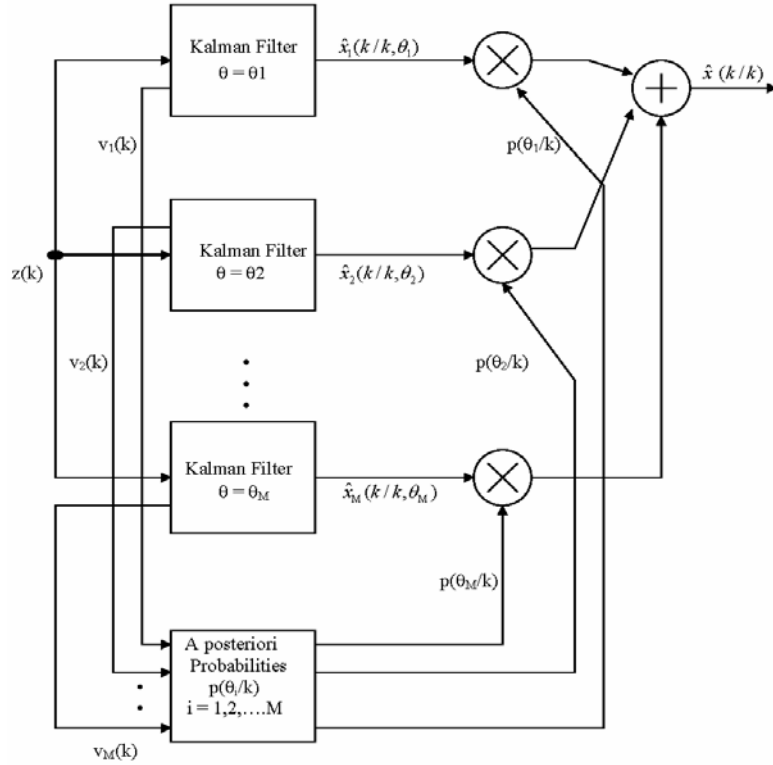
Fig. 5.1. MMPF Block Diagram

In the Gaussian case the optimal MMSE estimate of $\mathbf{x}(k)$ is given by

$$\hat{\mathbf{x}}(k/k) = \sum_{j=1}^{M} \hat{\mathbf{x}}(k/k; \theta_j) \, p(\theta_j/k) \tag{5.7}$$

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 $p(\theta_j/k) = \dfrac{1}{M}$ 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 posteriori 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

highest posterior probability, called the maximum a posteriory (MAP) estimate , in [20], which is the case used in this paper, or the weighted average of the estimates produced by the elemental filters, as described in equation (5.7). The weights are determined by the posterior probability that each model in the model set is in fact the true model.

The posterior probabilities are calculated on-line in a recursive manner as follows

$$p(\theta_j/k) = \frac{L(k/k;\theta_j)}{\sum_{j=1}^{M} L(k/k;\theta_j)\, p(\theta_j/k\text{-}1)}\, p(\theta_j/k\text{-}1) \tag{5.8}$$

$$L(k/k;\theta_j) = \left| \mathbf{P}_{\tilde{\mathbf{y}}}(k/k\text{-}1;\theta_j) \right|^{-\frac{1}{2}} \cdot$$

$$\exp[-\frac{1}{2}\tilde{\mathbf{y}}^T(k/k\text{-}1;\theta_j)\, \mathbf{P}_{\tilde{\mathbf{y}}}^{-1}(k/k\text{-}1;\theta_j)\tilde{\mathbf{y}}(k/k\text{-}1;\theta_j)] \tag{5.9}$$

where the innovation process

$$\tilde{\mathbf{y}}(k/k\text{-}1;\theta_j) = \mathbf{y}(k) - \mathbf{H}(k;\theta_j)\hat{\mathbf{x}}(k/k\text{-}1;\theta_j) \tag{5.10}$$

is a zero mean white process with covariance matrix

$$\mathbf{P}_{\tilde{y}}(k/k\text{-}1;\theta_j) = \mathbf{H}(k;\theta_j)\mathbf{P}(k/k;\theta_j)\mathbf{H}^T(k;\theta_j) + \mathbf{R} \tag{5.11}$$

For equations (5.8) – (5.11) j = 1,2, …, M.

An important feature of the MMPF is that all the Kalman filters needed to implement can be independently realized. This enables us to implement them in parallel, thus saving us enormous computational time [20].

Equations (5.7), (5.8) refer to our case where the sample space is naturally discrete. However in real world applications, θ's probability density function (pdf) is continuous and an infinite number of Kalman filters have to be applied for the exact realization of the optimal estimator. The usual approximation considered to overcome this difficulty is to somehow approximate θ's pdf by a finite sum. Many discretization strategies have been proposed at times and some of them are presented in [21-22].

When the true parameter value lies outside the assumed sample space, the adaptive estimator converges to the value that in the sample space which is closer (i.e. minimizes the Kullback Information Measure) to the true value, [23]. This means that the value of the unknown parameter cannot be exactly defined. The application of a variable structure MMPF is able to overcome this difficulty [17].

5.3 Examples

In order to assess the performance of our method, several simulation experiments were conducted. All of these experiments were conducted 100 times (100 Monte Carlo Runs). The models used was that of (5.2) and (5.3), with cardinality $M = 10$.

5.3.1 Example 1

ARMA $(1, 1)$. $\theta = (1, 1) = 2$.

$$\mathbf{A}=\begin{bmatrix} -0.85 & 0.75 \\ 0.65 & -0.55 \end{bmatrix} \mathbf{B}=\begin{bmatrix} -1.9833 & 1.889 \\ 1.7 & 1.9833 \end{bmatrix} \mathbf{R}=\begin{bmatrix} 1.5625 & 1.5 \\ 1.5 & 1.5625 \end{bmatrix}$$

Figure 5.2 depicts the posterior probabilities associated with each value of θ. Figure 5.3 shows the criteria comparison for two data sets, one relatively small (50 samples) and one larger (100 samples) and Table 5.1 shows the estimated ARMA parameter coefficients.

From Figure 5.2, is obvious that the MMPF identifies the correct $\theta = (1, 1) = 2$ very fast, in just 17 steps. Convergence is taken to occur when the posterior probability of the model exceeds 0.9.

From Figure 5.3 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 5.1 shows that the parameter coefficient estimation is very accurate.(RMSE – *R*oot *M*ean *S*quare *E*rror is very small).

5.3.2 Example 2

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

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

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

Figure 5.4 depicts the posterior probabilities associated with each value of θ. Figure 5.5 shows the criteria comparison for two data sets, one relatively small (50 samples) and one larger (100 samples) and Table 5.2 shows the estimated ARMA parameter coefficients.

From Figure 5.4, is obvious that the MMPF identifies the correct $\theta = (1, 1) = 2$ very fast, in just 18 steps. Convergence is taken to occur when the posterior probability of the model exceeds 0.9.

From Figure 5.5 we deduce that MMPF is 100% successful in selecting the correct model order for both data sets, while none of the two other criteria achieve a similar performance for either data set.

As Table 5.2, clearly shows the parameter estimation is again accurate since the Root Mean Square Error (RMSE) is very small.

5.3.3 Example 3

ARMA (2, 2). $\theta = (2, 2) = 4$.

$$A_1 = \begin{bmatrix} -0.17 & 0.14 \\ -0.19 & -0.1 \end{bmatrix}, A_2 = \begin{bmatrix} -0.2 & 0.12 \\ 0.22 & -0.25 \end{bmatrix}$$

$$B_1 = \begin{bmatrix} -0.45 & 0.52 \\ -0.32 & -0.7 \end{bmatrix}, B_2 = \begin{bmatrix} -0.85 & 0.75 \\ -0.65 & -0.55 \end{bmatrix},$$

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

Figure 5.6 depicts the posterior probabilities associated with each value of θ. Figure 5.7 shows the criteria comparison for two data sets, one relatively small (50 samples) and one larger (100 samples) and Table 5.3 shows the estimated ARMA parameter coefficients.

From Figure 5.6, is obvious that the MMPF identifies the correct $\theta = (1, 1) = 2$ very fast, in just 24 steps. Convergence is taken to occur when the posterior probability of the model exceeds 0.9.

From Figure 5.7 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 5.3, clearly shows the parameter estimation is again accurate since the Root Mean Square Error (RMSE) is very small.

5.4 Conclusions

A new method for simultaneously selecting the order 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. 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, BIC) show that the proposed method only 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.
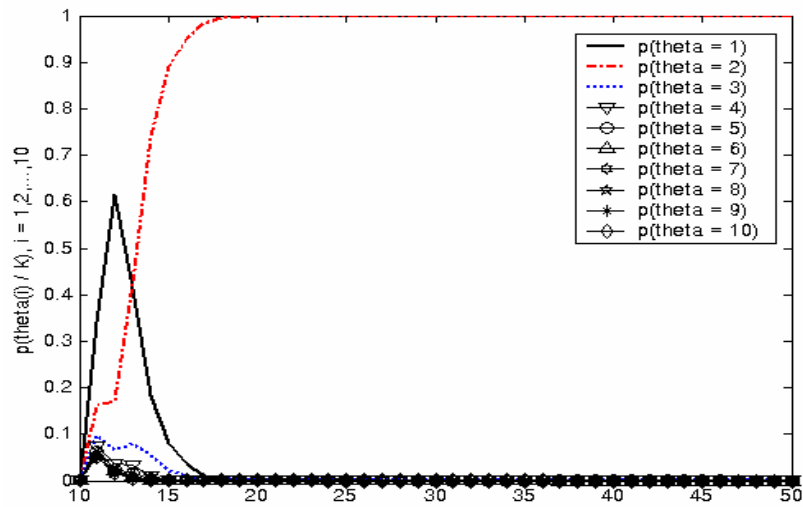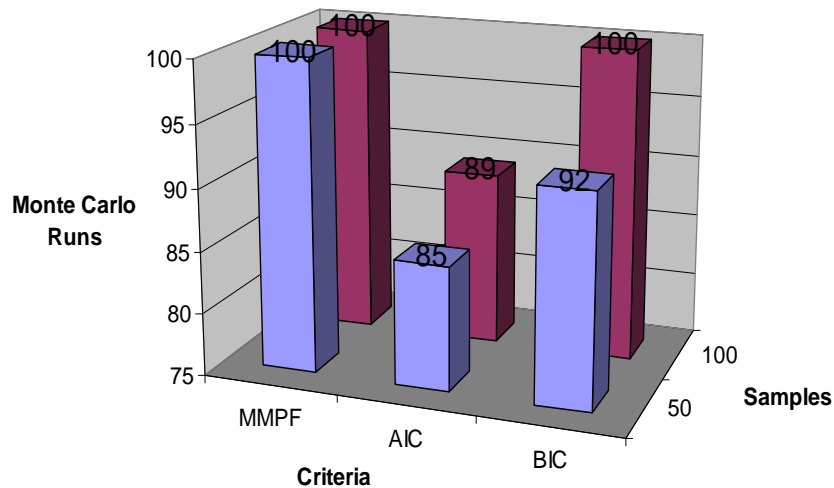

Fig. 5.2. Example 1, posterior probabilities

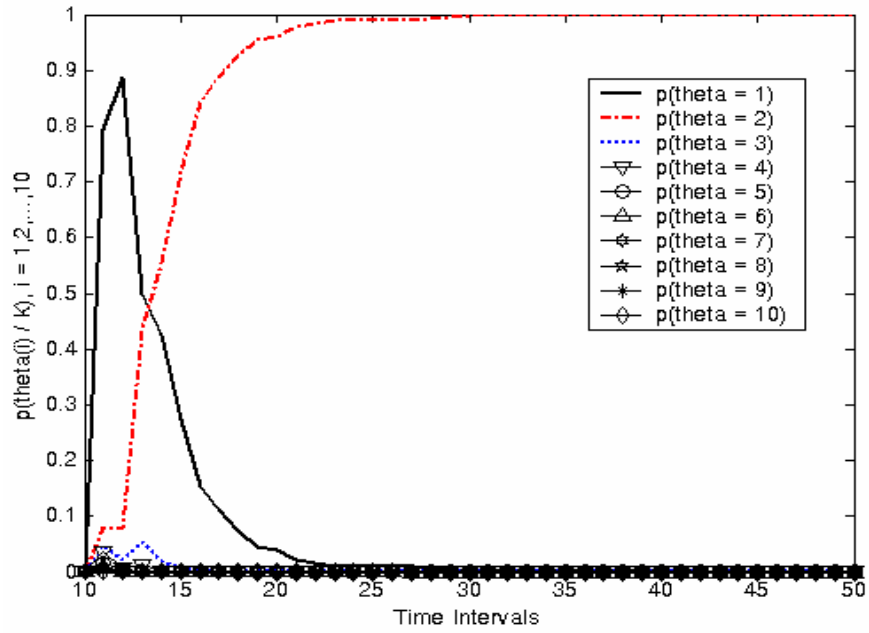**Fig. 5.3.** Example 1, Criteria Comparison– Correct model order identification
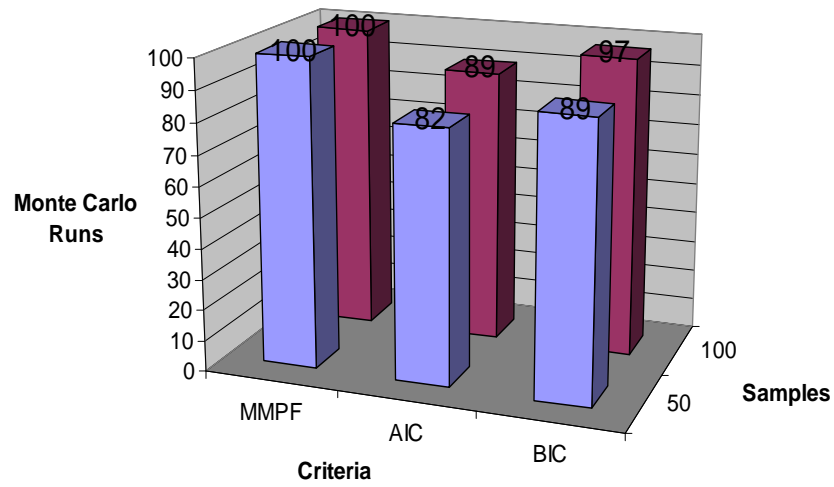
Fig. 5.4. Example 2, posterior probabilities

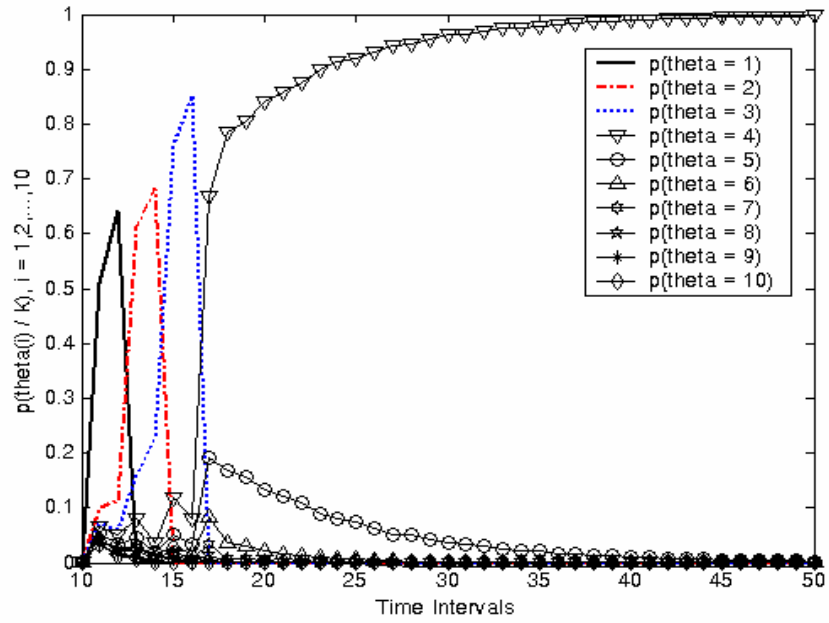**Fig. 5.5.** Example 2, Criteria Comparison– Correct model order identification

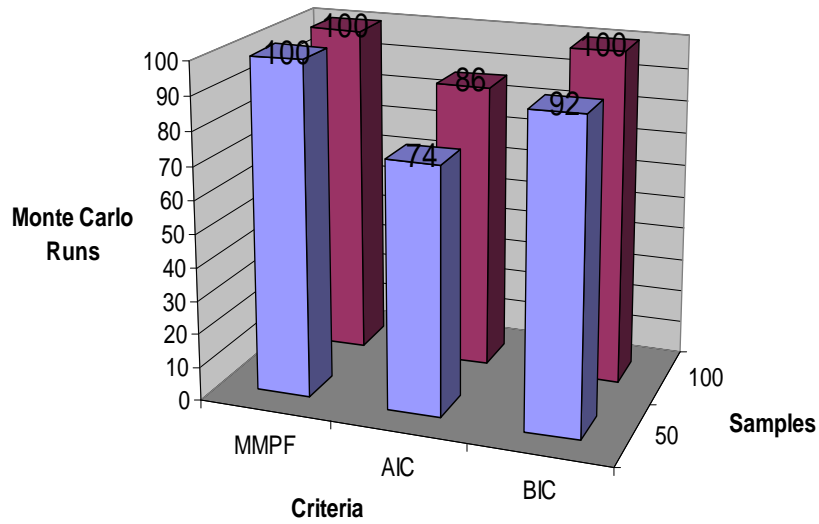**Fig. 5.6.** Example 3, posterior probabilities

Fig. 5.7. Example 2, Criteria Comparison– Correct model order identification

Table 5.1. Example 1, Estimated ARMA coefficient parameters

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 5.2. Example 2, 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 5.3. Example 3, 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

References

1. Sheng Lu, Ki Hwan Ju, Chon KH (2001) A new algorithm for linear and non-linear ARMA model parameter estimation using affine geometry [and application to blood flow/pressure data]. IEEE Transactions on Biomedical Engineering 48(10):1116-1124
2. Derk J, Weber S, Weber C (2007) Extended ARMA models for estimating price developments on day-ahead electricity markets. Electric Power Systems Research 77(5-6):583-593
3. Chen CP, Bilmes JA (2007) MVA Processing of speech features. IEEE Transactions on Audio Speech and Language Processing 15(1):257-270
4. Kourosh M, Eslami HR, Kahawita R (2006) Parameter estimation of an ARMA model for river flow forecasting using goal programming. Journal of Hydrology 331(1-2):293-299
5. Pappas Sp S, Leros AK, Katsikas SK (2006) Joint order and parameter estimation of multivariate autoregressive models using multi-model partitioning theory. Digital Signal Processing 16(6):782-795
6. Lainiotis DG (1976) Partitioning: A unifying framework for adaptive systems I: Estimation. Proc IEEE 64:1126-1143
7. Lainiotis DG (1976) Partitioning: A unifying framework for adaptive systems I: Control. Proc IEEE 64:1182-1198
8. Akaike H (1969) Fitting autoregressive models for prediction. Ann Inst Stat Math 21:243–247
9. Rissanen J (1978) Modelling by shortest data description. Automatica 14:465-471
10. Rissanen J (1986) A predictive least squares principle. IMA J Math Contr Inform 3:211-222
11. Schwarz G (1978) Estimation of the dimension of the model. Ann Statist 6:461-464
12. Wax M (1988) Order selection for AR models by predictive least squares. IEEE Trans Acoust Speech Signal Processing 36:581-588
13. Lutkepohl H (1985) Comparison of criteria for estimating the order of a vector AR process J Time Ser Anal 6:35-52
14. Sage AP, Husa GW (1969) Adaptive filtering with unknown prior statistics. In: Proc Joint Automatic Control Conf, Boulder, Colorado, pp 760-769
15. Watanabe K, (1992) Adaptive Estimation and Control: Partitioning Approach. Englewood Cliffs NJ Prentice Hall
16. Katsikas SK, Leros AK, Lainiotis DG (1994) Passive tracking of a maneuvering target: an adaptive approach. IEEE Transactions on Signal Processing 42(7):1820-1825
17. Katsikas SK, Likothanassis SD, Beligiannis GN, Berkeris KG, Fotakis DA (2001) Genetically determined variable structure multiple model estimation. IEEE Transactions on Signal Processing 49(10):2253-2261

18. Lainiotis DG, Papaparaskeva PA (1998) Partitioned adaptive approach to non-linear channel equalization. IEEE Transactions on Communications 46(10):1325-1336
19. Moussas VC, Likothanassis SD, Katsikas SK, Leros AK (2005) Adaptive On-Line Multiple Source Detection. In: IEEE International Conference on Acoustics Speech and Signal Processing Proceedings (ICASSP '05,), 4:1029-1032
20. Lainiotis DG, Katsikas SK, Likothanassis SD (1988) Adaptive deconvolution of seismic signals: Performance, computational analysis parallelism. IEEE Trans Acoust Speech Signal Processing 36(11):1715-1734
21. Sengbush RL, Lainiotis DG (1969) Simplified parameter quantization procedure for adaptive estimation. IEEE Trans Automat. Contr. AC-14:424-425
22. Anderson BDO, Brinsmead TS, Bruyne FDe, Hespanha J, Liberzon D, Morse AS (2000) Multiple model adaptive control Part 1: Finite controller coverings. Int J Robust Nonlinear Contr 10:909-929
23. Hawks RM, Moore JB (1976) Performance of Bayesian parameter estimators for linear signal models. IEEE Trans Automat. Contr. AC–21:523-527