AN EFFECTIVENESS STUDY OF BAYESIAN IDENTIFICATION TECHNIQUES FOR ARMA MODELS

Sherif S. Ali
Faculty of Economics and Political Science, Department of Statistics
Cairo University, Egypt

ABSTRACT

Model identification is the first and most important stage when analyzing a time series. As a result of analytical complexity, very little has been done from a Bayesian viewpoint in order to identify the orders of ARMA models. Some analytical and numerical identification techniques have been suggested to handle the Bayesian identification problem. These techniques have been introduced by Monahan (1983), Broemeling and Shaarawy (1987) and Ali (2003). However, the performance of Monahan (1983) and Ali (2003) techniques have not been studied yet through an effectiveness study. This article has three different objectives. The first one is to carry out a simulation study to assess the performance and efficiency of Monahan's technique. The second objective is to carry out a simulation study to test the adequacy of Ali's technique in handling the identification problem of ARMA processes. The third objective is to compare among the three Bayesian identification techniques through a comprehensive simulation study. In addition, the results of the three Bayesian Identification techniques are compared with the well-known non-Bayesian automatic technique, *AIC*. The numerical results support the adequacy of the Bayesian techniques in solving the identification problems of autoregressive moving average processes.

Keywords: *Time series, Identification, ARMA models, AIC, normal gamma density, Jeffreys' prior, posterior probability mass function* .

1. INTRODUCTION

The autoregressive moving average models are very useful in modeling time series data that arise in many real life situations. Time series analysis is usually based on ARMA(p,q) parameterization in which the variable of interest is linearly regressed on a finite number (p) of the previous values of the process and a finite number (q) of the previous random shocks. If the moving average order q is zero, one will have pure autoregressive models. If the autoregressive order p is zero, one will have pure moving average models. The orders p and q of an ARMA model are usually unknown and have to be identified or estimated. In literature, Bayesian and non-Bayesian suggestions are presented to solve this identification problem.

The most popular non-Bayesian technique to identify the orders of ARMA(p,q) models is developed by Box and Jenkins (1970). Their methodology is based on matching the sample autocorrelation and partial autocorrelation functions with their theoretical counterparts. Their technique is explained in many references such as Chatfield (1980), Priestley (1981), and Tong (1990). Another non- Bayesian technique, known as the automatic one, is based on fitting all possible models then computing a certain criterion for each model, such as *AIC*, *FPE* and *SIC*, and choosing the model which minimizes the proposed criterion. For more details about the automatic technique, the reader is referred to Akaike (1973, 1974), Hannan and Quinn (1979), Mills and Prasad (1992), and Beveridge and Oickle (1994).

On the other hand, the Bayesian literature which is devoted to handle the identification problem of ARMA models is sparse. Diaz and Farah (1981) have developed a direct Bayesian method to identify the pure autoregressive models. This technique depends on deriving the posterior probability mass function of the order p in a closed form. Monahan (1983) has made an important contribution to the analysis of low order ARMA models by developing a numerical technique which implements the identification, estimation, and forecasting phases of an ARMA process. Broemeling and Shaarawy (1987) have developed an approximate analytical procedure to identify the orders of ARMA processes depending on approximating the posterior distribution of the coefficients by a multivariate t distribution, then the significance of coefficients is tested by a series of t-tests in a similar fashion to the backward elimination procedure used in linear regression analysis. George and McCulloch (1993) have introduced the stochastic search variable selection (SSVS) which can be used for ARMA model identification via Gibbs Sampling. Most recently, Daif et al. (2003) have studied the efficiency of Diaz and Farah technique and compared it with Broemeling and Shaarawy technique for autoregressive models through a large scale simulation study. Moreover, Shaarawy and Ali (2003) have developed a direct Bayesian technique to identify the orders of seasonal autoregressive models. Ali (2003) has derived the posterior mass function of the orders p and q for ARMA models although he hasn't checked the efficiency of the proposed technique. Shaarawy et al. (2007) have extended the direct Bayesian technique to handle the moving average model identification problem and checked its performance through an effectiveness simulation study. For more details about the identification techniques, the reader can is referred to Newbold (1984,1988).

This article studies and compares the performance and efficiency of three Bayesian Identification procedures for ARMA models. The three procedures are Broemeling and Shaarawy (1987), Monahan (1983) and Ali (2003) which will be denoted by ***B-S, M*** and ***D*** techniques respectively. The simulation study has been chosen in such a way to include different ARMA models, different parameter values and different prior distributions of the model orders. In addition, the adopted Bayesian identification procedures are compared with the well known non-Bayesian automatic technique, ***AIC***.

This paper is structured as follows. Section 2 presents ARMA models. Section 3 sheds lights on ***B-S*** technique while section 4 is devoted to review ***M*** and ***D*** techniques as pure Bayesian solutions of the identification problem. The steps and results of the effectiveness study are explained in section 5.

2. AUTOREGRESSIVE MOVING AVERAGE PROCESSES

The mixed autoregressive moving average (ARMA) class of models (2.1) is quite important in modeling time series data, see for instance Box and Jenkins (1970). Let $\{t\}$ be a sequence of integers, $p,q \in \{1,2,\dots\}$, $\{y_t\}$ is a sequence of real observable random variables, and $\{\varepsilon_t\}$ is a sequence of independent and normally distributed unobservable random variables with zero mean and unknown precision $\tau > 0$. The autoregressive moving average model of orders p and q, denoted by ARMA(p,q), is defined for n observations as

$$\Phi(B)y_t = \Theta(B)\varepsilon_t \tag{2.1}$$

where $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ and $\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$, and the backshift operator B is such that $B^r y_t = y_{t-r}$; $r = 1,2,\ldots$

The time series is stationary if the roots of $\Phi(B)$ lie outside the unit circle. On the other hand, the process is invertible if the roots of $\Theta(B)$ lie outside the unit circle. However, ARMA(p,q) model (2.1) can be rewritten explicitly as

$$y_t = \varepsilon_t + \sum_{i=1}^{p} \phi_i y_{t-i} - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \qquad , t=1,2,\ldots n \tag{2.2}$$

Where $\phi_1, \phi_2, \cdots, \phi_p, \theta_1, \theta_2, \cdots, \theta_q$ are the model coefficients.

In practice, the orders p and q are unknowns and one has to determine values for them using n observations $S_n = [y_1, y_2, \ldots, y_n]'$. Thus the statistical question is: "given n observations generated from an autoregressive moving average process, what are the values of p and q?". Broemeling and Shaarawy (1987), Monahan (1983) and Ali (2003) have introduced different Bayesian identification techniques to answer this question. The following two sections introduce these three techniques.

3. INDIRECT BAYESIAN IDENTIFICATION TECHNIQUE

The first step of ***B-S*** technique, which has been introduced by Broemeling and Shaarawy (1987), is to derive the posterior distribution of model coefficients. To derive the posterior distribution of the model coefficients, suppose that there is a time series with n observations $\underline{S}_n = [y_1, y_2, \ldots, y_n]'$. Let $\underline{\Phi}^{'(m_1)} = (\phi_1, \phi_2, \ldots \phi_{m_1})$ and $\underline{\Theta}^{'(m_2)} = (\theta_1, \theta_2, \ldots \theta_{m_2})$. Where m_1 is the known maximum value of p and m_2 is the known maximum value of q, where m_1 and m_2 are non-negative integers. By conditioning on the first m_1 observations and letting $\varepsilon_{m_1} = \cdots = \varepsilon_0 = \cdots = \varepsilon_{m_1-m_2+1} = 0$, where $m_2 > m_1 + 1$ (see Priestley,1981), The conditional likelihood function has the form

$$L_1(\underline{\Phi}^{(m_1)}, \underline{\Theta}^{(m_2)}, \tau \mid \underline{S}_{n-m_1}) = \left(\frac{\tau}{2\pi}\right)^{\frac{n-m_1}{2}} \exp\left\{-\frac{\tau}{2}\sum_{t=m_1+1}^{n}\left(y_t - \sum_{i=1}^{m_1}\phi_i y_{t-i} + \sum_{j=1}^{m_2}\theta_j \varepsilon_{t-j}\right)^2\right\} \tag{3.1}$$

This likelihood function is intractable since the exponent is not a quadratic function in the model's coefficients. This leads to a non-standard posterior distribution of the model coefficients. Therefore, making inferences about model coefficients becomes too difficult. To overcome this problem, the above likelihood function can be approximated using B-S approximation (see Broemeling and Shaarwy (1987)) as

$$L_1^*(\underline{\Phi}^{(m_1)},\underline{\Theta}^{(m_2)},\tau \mid \underline{S}_{n-m_1}) = \left(\frac{\tau}{2\pi}\right)^{\frac{n-m_1}{2}} \exp\left\{-\frac{\tau}{2}\sum_{t=m_1+1}^{n}\left(y_t - \sum_{i=1}^{m_1}\phi_i y_{t-i} + \sum_{j=1}^{m_2}\theta_j \hat{\varepsilon}_{t-j}\right)^2\right\} \quad (3.2)$$

where $\hat{\varepsilon}_{t-j}$ is a nonlinear least square estimate of ε_{t-j}. They combined the approximate conditional likelihood function (3.2) with a Jeffreys' prior as a non-informative prior for the model's coefficients which has the form

$$g_1(\underline{\Phi}^{(m_1)},\underline{\Theta}^{(m_2)},\tau) \propto \tau^{-1} \quad , \quad \underline{\Phi}^{(m_1)} \in R^{m_1}, \underline{\Theta}^{(m_2)} \in R^{m_2}, \tau > 0 \quad (3.3)$$

They proved that the marginal posterior distribution of $\underline{\Phi}^{(m_1)},\underline{\Theta}^{(m_2)}$ is a multivariate t-distribution with $n-2m_1-m_2$ degrees of freedom ((see Broemeling and Shaarawy (1987)).

The second step of the ***B-S*** identification technique is to employ a sequence of univariate t-tests about the significance of the model's coefficients, similar to backward elimination procedure. The ***B-S*** identification technique has been extended to include a normal-gamma prior by Broemeling and Shaarawy (1988). An advantage of this technique is that it is simple and easy to program. However, the efficiency of ***B-S*** identification technique has been studied for mixed ARMA models through a large scale simulation study by El-Souda (2000).

4. PURE BAYESIAN IDENTIFICATION TECHNIQUES

This section is devoted to present ***M*** and ***D*** Bayesian identification techniques for mixed ARMA models. Unlike the indirect technique, these Bayesian identification techniques assume the orders p and q to be random variables and the problem is how to find their posterior probability mass function. The conditional likelihood function of ARMA(p,q) model has the form

$$L_2(\underline{\Phi}^{(p)},\underline{\Theta}^{(q)},p,q,\tau \mid \underline{S}_{n-p}) = \left(\frac{\tau}{2\pi}\right)^{\frac{n-p}{2}} \exp\left\{-\frac{\tau}{2}\sum_{t=p+1}^{n}\left(y_t - \sum_{j=1}^{p}\phi_{pj} y_{t-j} + \sum_{i=1}^{q}\theta_{qi}\varepsilon_{t-i}\right)^2\right\}$$

(4.1)

where ϕ_{pj} is the coefficient of the j^{th} lagged value of y_t in the p^{th}-order autoregressive part, $\underline{\Phi}^{'(p)} = \left(\phi_{p1}, \phi_{p2}, \ldots \phi_{pp}\right)$, θ_{qi} is the coefficient of the i^{th} lagged value of ε_t in the q^{th}-order moving average part, $\underline{\Theta}^{'(q)} = \left(\theta_{q1}, \theta_{q2}, \ldots \theta_{qq}\right)$, $\underline{\Phi}^{(p)} \in R^p$, $\underline{\Theta}^{(q)} \in R^q, \tau > 0$ and $p = 1,2,\ldots,P$, $q = 1,2,\ldots,Q$ where P, Q are the largest possible orders of p, q respectively. The likelihood function (4.1) is analytically intractable since the error terms ε_{t-i}'s are nonlinear functions in the model coefficients.

4.1 Monahan's Identification Technique (*M*-Technique)

The ***M***-technique, which has been introduced by Monahan (1983), merged the above likelihood function (4.1) with a prior function, assume for illustration the Jeffreys' one of the form

$$g_2\left(\underline{\Phi}^{(p)}, \underline{\Theta}^{(q)}, p, q, \tau\right) \propto \tau^{-1} \qquad , \ \tau > 0 \tag{4.2}$$

This results in a posterior probability function of the form

$$\zeta(\underline{\Phi}^{(p)}, \underline{\Theta}^{(q)}, p, q, \tau \mid \underline{S}_{n-p}) = \frac{\tau^{\frac{n-p}{2}-1}}{(2\pi)^{\frac{n-p}{2}}} \exp\left\{-\frac{\tau}{2}\sum_{t=p+1}^{n}\left(y_t - \sum_{j=1}^{p}\phi_{pj}y_{t-j} + \sum_{i=1}^{q}\theta_{qi}\varepsilon_{t-i}\right)^2\right\} \tag{4.3}$$

This function is not a known standard form of a density function since ε_{t-i} is unknown; therefore mathematical integrations cannot carried out to get the marginal posterior mass function of p and q. The ***M***-technique solves this problem numerically by doing numerical integrations (see Monahan (1983)). Disadvantages of using numerical integrations are complexity and time consuming especially for high dimensions cases. However, the efficiency of this technique has not been studied yet through a large scale simulation study.

4.2 Direct Identification Technique (*D*-Technique)

Instead of using numerical integration to get the marginal posterior distribution of the orders, One can use an analytical approximation to simplify the form of the likelihood function in (4.1). However, one can't approximate the unknowns ε_{t-i}'s without determining the values of p and q. Thus, we need another identification technique to determine initial values of p and q, say $p_\circ$ and $q_\circ$

respectively. To identify theses initial orders, Ali (2003) has selected ***B-S*** identification technique (explained in section 3) which gives high percentage of correct identification (see El-Souda, 2000). Consequently, the likelihood function in (4.1) is approximated by the following form

$$L_2^*(\underline{\Phi}^{(p)},\underline{\Theta}^{(q)},p,q,\tau\mid \underline{S}_{n-p})=\left(\frac{\tau}{2\pi}\right)^{\frac{n-p}{2}}\exp\left\{-\frac{\tau}{2}\sum_{t=p+1}^{n}\left(y_t-\sum_{j=1}^{p}\phi_{pj}y_{t-j}+\sum_{i=1}^{q}\theta_{qi}\hat{\varepsilon}_{t-i}\right)^2\right\}$$

(4.4)

where $\hat{\varepsilon}_t=y_t-\sum_{j=1}^{p_0}\hat{\phi}_j y_{t-j}+\sum_{i=1}^{q_0}\hat{\theta}_i\hat{\varepsilon}_{t-i}$; $\hat{\phi}_j, j=1,2,\ldots,p_0$ and $\hat{\theta}_j, j=1,2,\ldots,q_0$ are the nonlinear least square estimate of the coefficients obtained by fitting the model ARMA(p_0, q_0).

Combining the approximate likelihood function (4.4) with the Jeffreys' prior (4.2) and integrating with respect to $\underline{\Phi}^{(p)},\underline{\Theta}^{(q)}$ and τ, results in a closed form of the marginal posterior mass function of p and q, see Ali (2003).

Also, it is important to mention that the posterior mass function of p and q has been derived using a normal gamma prior, a conjugate one (see Ali (2003)). However, the efficiency of the direct Bayesian identification technique for mixed ARMA models has not been checked yet.

5. AN EFEECTIVENESS STUDY

This section aims to assess and compare the performance and efficiency of the three considered Bayesian identification techniques in selecting the orders of mixed autoregressive moving average processes (ARMA) through simulation studies. Moreover, this comparison will be carried out with the well known non-Bayesian identification technique, ***AIC***. Three different prior distributions of the model orders are employed to identify the orders of ARMA(1,1), ARMA(1,2) and ARMA(2,1) sources with various assumed parameter values. The parameters of some cases are chosen to be well inside the stationarity-invertibility domains while in other cases they are chosen to be near the boundaries. All computations are performed using MATLAB 7.

The used effectiveness criterion of the considered techniques is the percentage of correct identification. Such effectiveness criterion is computed with respect to different time series lengths (sample size) and various parameters sets. For all models, the precision of the noise term is fixed at 2.

For Illustration, simulation 1, begins with generating 500 data sets of normal variates, each of size 500, to represent ε_t. These data sets are then used to generate 500 realizations, each of size 300, from ARMA(1,1) process with coefficients $\phi = -0.5$, $\theta = 0.8$. Note that, the first 200 observations are ignored to remove the initialization effect. For a specific prior, the second step of simulation 1 is computing the posterior distributions of the three identification techniques. All computations are done, assuming maximums (2,2) for orders (*p*,*q*), to identify each of the 500 realizations and finding the percentage of correct identification. The sample size n is taken to be 50, 100, 150, 200 and 300. With respect to the prior probability mass function of orders p and q, which is combined with the vague prior of $\underline{\Phi}^{(p)}$, $\underline{\Theta}^{(q)}$ and τ, the following three priors are used.

Prior1: $g(p,q) = \dfrac{1}{(P+1)(Q+1)-1}$ $\quad p = 0,1,2,\cdots,P\,, q = 0,1,2,\cdots,Q$; where $(p,q) \neq (0,0)$

Prior2 : $g(p,q) \propto (0.5)^{p+q}$ $\qquad p = 0,1,2,\cdots,P\,, q = 0,1,2,\cdots,Q$; where $(p,q) \neq (0,0)$

Prior3: for $P=2$ and $Q=2$

$$g(0,1) = g(1,0) \propto 0.4, g(1,1) = g(0,2) = g(2,0) \propto 0.3, g(1,2) = g(2,1) \propto 0.2 \text{ and } g(2,2) \propto 0.1$$

The first prior assigns equal probabilities to all possible values of the orders p and q. The second prior is chosen in such a way to give probabilities that decline exponentially with the order, while the third prior is chosen in such a way to give probabilities that decrease with an absolute amount 0.1 as the order increases.

All other simulations are done in similar steps but using different parameter sets and different models as shown in the following table.

Table (5.1): Parameter Sets for Simulated ARMA models

PROCESS	*CASE I*	*CASE II*	*CASEIII*	*CASE VI*
ARMA(1,1)	(-0.5,0.8)	(-0.9,0.5)	(-0.6,0.6)	(-0.9,0.9)
ARMA(1,2)	(-0.5,0.1,0.8)	(0.9,0.3,-0.6)	(0.5,0.3,-0.6)	(0.9,0.1,0.8)
ARMA(2,1)	(0.3,-0.6,0.9)	(0.1,0.8,-0.5)	(0.3,-0.6,0.5)	(0.1,0.8,0.9)

The results of the simulations studies for ARMA(1,1), ARMA(1,2) and ARMA(2,1) models are displayed in tables (5.2), (5.3) and (5.4) respectively.

Table (5.2): Percentages of Correct Identification of the Bayesian Techniques for ARMA(*1,1*) Processes Assuming *(m1,m2)=(2,2)*

PARAMETERS	*N*	*M PRIOR1*	*M PRIOR2*	*M PRIOR3*	*B-S*	*D PRIOR1*	*D PRIOR2*	*D PRIOR3*	*AIC*
$\phi=-0.5$ $\theta=0.8$	50	36.2	59.0	66.4	72.6	3.0	9.4	12.2	61.0
	100	48.2	64.0	74.0	84.6	16.0	24.4	26.2	70.8
	150	48.2	66.0	76.6	83.0	19.8	24.4	27.6	74.2
	200	54.4	65.0	74.2	83.4	22.0	26.6	28.0	77.0
	300	62.2	69.0	77.4	84.2	29.6	33.0	34.4	76.0
$\phi=-0.9$ $\theta=0.5$	50	57.6	72.8	75.8	32.2	0.6	7.0	13.4	61.8
	100	75.4	82.4	85.4	55.2	6.4	24.2	31.2	71.6
	150	88.0	91.2	92.4	71.0	14.0	29.0	37.4	73.8
	200	93.4	94.6	94.6	82.4	22.8	36.0	41.4	75.4
	300	97.0	97.6	98.0	86.2	27.0	37.6	41.4	75.8
$\phi=-0.6$ $\theta=0.6$	50	37.2	65.0	68.8	41.8	1.0	7.8	12.0	56.0
	100	68.2	79.8	85.6	69.4	6.6	17.8	21.6	67.2
	150	83.0	88.6	92.4	79.8	16.2	24.2	27.0	72.0
	200	90.0	93.0	94.6	84.2	19.6	27.6	29.2	72.6
	300	98.4	98.8	98.8	85.6	23.8	29.4	31.0	75.8
$\phi=-0.9$ $\theta=0.9$	50	78.4	87.2	89.6	80.0	19.4	35.4	41.8	77.2
	100	94.2	96.4	97.4	77.2	33.2	43.6	47.4	77.4
	150	97.6	98.8	99.0	80.0	42.6	49.4	51.8	75.2
	200	99.0	99.0	99.6	80.6	45.4	50.2	53.8	76.4
	300	99.8	100.0	100.0	78.0	42.6	46.8	49.6	74.6

Source : Simulated Data

Table (5.3): Percentages of Correct Identification of the Bayesian Techniques for ARMA(*1,2*) Processes Assuming *(m1,m2)=(2,2)*

PARAMETERS	N	M PRIOR1	M PRIOR2	M PRIOR3	B-S	D PRIOR1	D PRIOR2	D PRIOR3	AIC
$\phi=-0.5$	50	50.4	34.6	21.6	45.4	10.8	32.0	30.6	45.2
$\theta_1=0.1$	100	62.4	58.6	51.8	70.2	33.0	53.4	52.6	70.2
$\theta_2=0.8$	150	65.6	65.0	60.2	78.6	44.8	61.6	61.2	75.8
	200	67.0	67.0	64.4	82.8	52.6	65.4	65.4	82.6
	300	69.4	69.0	66.8	89.0	60.0	74.0	74.0	84.6
$\phi=0.9$	50	74.4	75.8	72.2	83.8	4.8	39.6	39.0	81.0
$\theta_1=0.3$	100	87.0	86.2	84.6	87.2	20.6	56.8	56.8	83.8
$\theta_2=-0.6$	150	90.2	89.6	88.2	86.8	36.2	62.0	62.0	81.4
	200	93.0	92.2	92.0	90.2	47.0	69.0	69.0	83.6
	300	96.0	95.8	95.6	90.6	56.8	72.2	72.2	84.0
$\phi=0.5$	50	50.4	50.8	42.0	45.2	3.8	27.2	25.4	49.4
$\theta_1=0.3$	100	58.6	62.0	57.8	73.4	20.0	51.0	48.6	73.8
$\theta_2=-0.6$	150	63.4	66.0	64.0	86.8	33.0	62.2	61.8	83.4
	200	65.4	68.8	67.6	88.0	44.4	67.8	67.6	82.4
	300	64.6	68.8	68.4	88.2	53.4	70.8	70.8	84.6
$\phi=0.9$	50	23.8	5.6	1.8	34.8	7.6	25.0	24.4	7.4
$\theta_1=0.1$	100	32.8	11.4	2.6	41.4	20.2	34.6	33.2	10.8
$\theta_2=0.8$	150	43.6	20.2	9.2	48.6	30.0	41.4	39.8	15.8
	200	53.8	27.4	13.6	57.6	36.2	48.6	46.8	24.0
	300	60.8	38.4	24.6	66.2	48.4	59.6	58.2	34.0

Source : Simulated Data

Table (5.4): Percentages of Correct Identification of the Bayesian Techniques for ARMA(*2,1*) Processes Assuming *(m1,m2)=(2,2)*

PARAMETERS	N	M PRIOR1	M PRIOR2	M PRIOR3	B-S	D PRIOR1	D PRIOR2	D PRIOR3	AIC
$\phi_1=0.3$	50	80.2	81.2	77.2	86.0	60.6	81.4	81.2	77.8
	100	96.2	96.4	96.0	88.4	79.6	86.4	86.4	82.6
$\phi_2=-0.6$	150	99.0	99.4	99.2	87.4	81.8	86.2	86.2	83.8
	200	99.8	99.8	99.8	87.2	83.6	86.2	86.2	86.6
$\theta=0.9$	300	100.0	100.0	100.0	83.8	82.0	85.6	85.6	85.4
$\phi_1=0.1$	50	46.8	38.6	30.0	25.4	42.8	52.8	46.8	40.0
	100	54.0	48.4	45.0	57.0	71.0	71.8	66.8	67.8
$\phi_2=0.8$	150	57.6	54.6	53.2	78.2	83.2	83.2	81.0	79.2
	200	61.2	58.6	55.8	85.2	89.6	89.8	88.2	84.0
$\theta=-0.5$	300	62.8	59.8	58.0	92.0	92.2	95.4	95.2	86.2
$\phi_1=0.3$	50	46.8	55.6	48.2	50.8	55.6	64.6	59.6	56.0
	100	59.0	65.0	62.4	81.4	79.6	83.4	82.0	77.6
$\phi_2=-0.6$	150	58.0	64.4	62.4	89.4	87.4	90.2	89.8	82.4
	200	58.0	62.6	61.4	90.2	87.4	90.4	90.4	82.6
$\theta=0.5$	300	56.6	61.8	60.8	91.4	90.6	93.0	93.0	82.0
$\phi_1=0.1$	50	26.6	8.6	3.8	13.2	43.0	47.2	45.8	8.2
	100	44.8	22.6	13.4	23.4	48.2	51.0	49.0	19.2
$\phi_2=0.8$	150	57.0	34.6	22.4	30.2	49.0	51.4	49.2	24.0
	200	63.0	41.8	29.2	37.2	55.2	57.8	55.4	31.6
$\theta=0.9$	300	69.4	53.6	42.2	46.2	61.2	64.0	61.2	39.8

Source : Simulated Data

From the above tables, one can summarize the results as follows:

1. The percentages of correct identification of the four considered technique increase as the time series length *n* increases.

2. For ARMA(1,1) process (table (5.2)), it looks that the ***M***-technique, which is the exact one, dominates all other techniques except the case in which the parameter of the moving average part is near the invertibility boundary and the parameter of the autoregressive part is well inside the stationarity region. The ***B-S*** technique gives high percentages of correct identification for all cases and for all n especially with n>50. Also, it is clear that the ***D***-technique is the worst one for all cases and for all *n*. The ***AIC*** technique gives good results being greater than 55% for all cases and for all *n*.

3. For ARMA(1,2) process (table (5.3)), the ***B-S*** technique dominates all other techniques almost everywhere. The results of the last case in which the parameter values are near the invertibility and stationarity boundaries are the worst for the all techniques as it is expected. In general, the results of ***M***-technique are less than those of the case of table (5.2) because the prior probabilities give more weight to the model with small order. Unsimilar to ARMA(1,1) cases, the percentages of correct identification of ***M***-Prior1 are better than those of ***M***-Prior2 and ***M***-Prior3. In addition, one may say that the results of the D-technique are better than its results in table (5.1).

4. For ARMA(2,1) process (table (5.4)), all results are fairly high except the case in which the parameters are near invertibility and stationarity boundaries. The results of the ***D***-technique for Prior3 are the best almost everywhere. Similar to ARMA(1,2) cases, the ***M***-Prior1 is still superior to the other two priors. The ***B-S*** technique is still stable.

5. For the matter of comparison, One may say that the ***B-S*** and ***AIC*** techniques are more stable than the other techniques.

CONCLUSION

The article has studied the efficiency of three Bayesian identification techniques, ***M***, ***D*** and ***B-S*** techniques, and used to identify different mixed ARMA sources through large scales simulation studies. The percentages of correct identification have been calculated using the three proposed identification techniques and the automatic ***AIC*** technique. The efficiency of each technique depends on the assumed parameter set of the sources, time series length and the order of the assumed source. For sufficient large *n*, all techniques are fairly adequate in identifying mixed ARMA models. Generally, for the purpose of comparison, the study shows that ***B-S*** technique is more stable than the other techniques. However, the analysis of the numerical results shows that the adopted three Bayesian identification techniques can efficiently identify the order of autoregressive moving average processes.

REFERENCES

(1) Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. *Budapest: Akademic Kiado,* pp. 267-281.

(2) Akaike, H. (1974). A New Look at Statistical Model Identification. *IEEE Transaction on Automatic Control*, Vol. 19, pp. 716-723.

(3) Ali, S. S. (2003). *Bayesian Identification of ARMA models.* Unpublished Ph.D. Thesis, Faculty of Economics and Political Science, Cairo University, Egypt.

(4) Beveridge, S. and Oickle, C. (1994). A Comparison of Box-Jenkins and Objectives Methods for Determining the Order of a Non-seasonal ARMA Model. *Journal of Forecasting,* Vol. 13, pp. 419-434.

(5) Box, G. and Jenkins, G. (1970). *Time Series Analysis, Forecasting and Control.* Holden-Day, San Francisco.

(6) Broemeling, L. and Shaarawy, S. (1987). Bayesian Identification of Time Series. *The 22nd Annual Conference in Statistics, Computer Science and Operation Research, Institute of Statistical Studies and Research*, Cairo, Egypt, Vol.2, pp. 146-159.

(7) Broemeling, L. and Shaarawy, S. (1988). Time Series: A Bayesian Analysis in Time Domain. *Bayesian Analysis of Time Series and Dynamic Models*, Elsevier, New York, pp. 1-21.

(8) Chatfield, C. (1980). *The Analysis of Time Series: Theory and Practice.* Chapman and Hill Ltd, London.

(9) Daif, A., Soliman, E. and Ali, S. (2003). On Direct and Indirect Bayesian Identification of Autoregressive Models. *The 14th Annual Conference on Statistics*

and Computer Modeling in Human and Social Sciences, Faculty of Economics and Political Science, Cairo University.

(10) Diaz, J. and Farah, J. L. (1981). Bayesian Identification of Autoregressive Process. *Presented at the 22nd NBER-NSF Seminar on Bayesian Inference in Econometrics.*

(11) El-Souda, R. M. (2000). *Time Series Identification. Unpublished M.Sc. Thesis, Department of Statistics*, Faculty of Economics and Political Science, Cairo University, Egypt.

(12) George, E. I. and McCulloch, R. E. (1993). Variable Selection Via Gibbs Sampling. *Journal of American Statistical Society*, Vol. 88, pp. 881-889.

(13) Hannan, E. J. and Quinn, B. G. (1979). The Determination of an Autoregression. *Journal of the Royal Statistical Society, Series B*, Vol. 42, pp. 190-195.

(14) Mills, J. and Prasad, K. (1992). A Comparison of Model Selection Criteria. *Econometric Reviews,* Vol. 11, p. 201-233.

(15) Monahan, J. F. (1983). Fully Bayesian Analysis of ARIMA Time Series Models. *Journal of Econometrics,* Vol. 21, pp. 307-331.

(16) Newbold P. (1984). Some Recent Developments in Time Series Analysis II, *International Statistical Review*, Vol. 52, pp. 183-192.

(17) Newbold P. (1988). Some Recent Developments in Time Series Analysis III, *International Statistical Review*, Vol. 56, pp. 17-29.

(18) Priestley, M. (1981). *Spectral Analysis and time Series*. Academic Press, London.

(19) Shaarawy, S. and Ali, S. (2003). Bayesian Identification of Seasonal Autoregressive Models. *Communications in Statistics- Theory and Methods*, Vol. 32, Issue 5, pp. 1067-1084.

(20) Shaarawy, S., Soliman, E. and Ali S. (2007). Bayesian Identification of Moving Average Models. *Communications in Statistics- Theory and Methods*, Vol. 36, Issue 12. pp. 2301-2312.

(21) Tong, H. (1990). *Non-Linear Time Series: A Dynamical System Approach.* Oxford University Press, New York.