Indirect and Direct Bayesian Techniques to Identify The orders of Vector ARMA Processes

Samir M. Shaarawy
Department of Quantitative Methods and Information Systems
Kuwait University
samir@cba.edu.kw

Emad E.A. Soliman[1]
Department of Statistics
Faculty of Science
King Abdulaziz University
easalam@hotmail.com

Sherif S. Ali[1]
Department of Statistics
Faculty of Science
King Abdulaziz University
ssali1970@yahoo.com

Abstract

This article develops two Bayesian techniques to identify the orders of vector mixed autoregressive moving average processes, namely the indirect and direct techniques. The proposed indirect technique approximates the joint posterior probability density function of the coefficients of the largest possible model by a matrix t distribution. Then, by employing a series of tests of significance, the insignificant coefficients are eliminated, and the model orders are determined. On the other hand, the proposed direct technique derives an approximate joint posterior probability mass function of the model orders. Then one may select the orders with maximum posterior probability. A wide simulation study is conducted to examine the effectiveness of the proposed procedures and compare their performance with the well-known AIC technique. The numerical results show that the proposed techniques can efficiently identify the orders of vector autoregressive moving average processes for moderate and large time series lengths. Moreover, the indirect technique dominates the direct and AIC ones.

Key words: Vector ARMA processes – indirect Bayesian identification – direct Bayesian identification– posterior probability mass function – Matrix normal-Wishart prior– Jeffreys' prior.

1. Introduction

The class of vector autoregressive moving average models, denoted by VARMA$_k(p,q)$ for short, is one of the most successful, parsimonious and practical class of models which model and forecast vector time series arise in wide variety of applications in many fields such as economics, business, engineering, meteorology and environmental

[1] The permanent position of the second and third authors is Department of Statistics, Faculty of Economics and Political Science, Cairo University.

studies. See Box and Tiao (1973), Tiao et al. (1979), Tiao and Box (1981), Liu (2009) and Tsay (2013,2014). In practice, the orders p and q are usually unknown and should be identified or estimated. Identifying or determining the orders of $VARMA_k(p, q)$ processes is the first and most important phase of vector time series analysis in which we choose the most appropriate model from a general class of linear vector time series models. The identification phase plays an important and difficult part in time series analysis because the other phases depend on its accuracy. It is a fact that the solution of the identification problem depends on subjective opinions as well as mathematical and statistical arguments. However, one may say that there is no panacea for the identification problem since there is no optimum method yet fully agreed upon.

Regarding the univariate autoregressive moving average (ARMA) processes, one may trace three different non-Bayesian approaches to identify the model orders p and q. Box and Jenkins (1970) developed the first and most favorable one (See Box et al. (2016)). Their methodology is based on matching the sample autocorrelation and partial autocorrelation functions with their theoretical counterparts. The second approach is the so-called automatic or exploratory approach. The foundation of this approach is to fit all possible ARMA models and compute a certain criterion for each model; then one may choose that model which optimizes the criterion. The most popular automatic criterion AIC or Akaike's information criterion was introduced by Akaike (1974). The third non-Bayesian approach is called a goodness of fit approach. This approach has been proposed by Box and Pierce (1970) and is based on a Chi-square test to identify the orders of the process. The Chi-square test has been modified and improved by Ljung and Box (1978). For more recent references on univariate ARMA modeling see Walter (2014), Brockwell and Davis (2016) and Shumway and Stoffer (2017).

On the other hand, the Bayesian methods of identification of univariate time series models have been recently developed. However, for well understood reasons, most of the Bayesian publications focus on pure autoregressive processes. This void in the Bayesian literature for the mixed processes is due to the complexity of the likelihood function of such processes. Diaz and Farah (1981) have developed a direct Bayesian identification procedure for pure autoregressive processes. Monahan (1983) has given a numerical algorithm to handle the identification problem of low order ARMA processes. Broemeling and Shaarawy (1988) have developed an approximate indirect Bayesian technique to identify the orders of univariate ARMA processes. Shaarawy and Ali (2003) have

developed a direct Bayesian technique to identify the orders of seasonal autoregressive processes. Moreover, Shaarawy et al. (2007) have developed an approximate direct Bayesian methodology to identify the order of pure univariate moving average processes.

With respect to the vector version, the identification problem, from non-Bayesian view point, has been studied by Granger and Newbold (1977), Tiao and Box (1981) and Tiao and Tsay (1983) by matching the cross correlation functions computed from the data with theoretical counterparts. However, this method is somewhat subjective, difficult and requires a high standard of experience in the case of mixed VARMA models. On the other hand, the Bayesian methods of identification of vector time series have been recently studied. Shaarawy et al. (2006) have introduced a direct Bayesian technique to identify order of bivariate autoregressive processes. Shaarawy and Ali (2008) extended the technique to the case of general pure vector autoregressive processes. More recently, Shaarawy and Ali (2012) have developed a direct Bayesian methodology to identify the order of pure vector moving average processes. Most recently, Shaarawy and Ali (2015) have extended the technique to the case of vector seasonal autoregressive processes. Soliman (2016) has extended the technique to mixed bivariate ARMA(p,q) Processes. However, one may say that both the direct and the indirect Bayesian approaches to identify mixed vector ARMA processes have not been explored yet.

The main objective of this article is to develop two approximate convenient methodologies to identify the orders of mixed VARMA processes. Using n vectors each of k observations and a matrix normal-Wishart prior density (or Jeffreys' vague prior), the indirect approach approximates the joint posterior distribution of the coefficients of the largest possible model by a matrix t distribution. Then the significance of the coefficients are tested, marginally or conditionally, by a series of F or χ^2 statistics in a similar fashion to the backward elimination procedure used in regression analysis. On the other hand, the direct approach uses the above two mentioned priors to develop the joint posterior probability mass function of the model orders in an approximate convenient form. Then, one may inspect the approximate joint posterior probability mass function of the model orders over its grid and select the orders with maximum posterior probability. In order to examine the numerical efficiency of the proposed two Bayesian procedures and compare their performance with the well-known AIC technique, a wide simulation study is conducted using the modern specialized SCA package.

The rest of this article is structured as follows: Section 2 introduces the definition of VARMA$_k$ (p, q) processes in scalar and matrix notations; and presents the main assumptions required to develop the proposed identification procedures for these processes. Section 3 develops the proposed indirect Bayesian technique to identify the orders of VARMA$_k$ (p, q) processes. Section 4 develops the proposed direct Bayesian technique to identify the orders of VARMA$_k$ (p, q) processes. Section 5 is devoted to examine the numerical effectiveness of the proposed two Bayesian procedures in solving the identification problem of VARMA$_k$ (p, q) processes.

2. General Vector Autoregressive Moving Average Processes

Let $\{t\}$ be a sequence of integers, $p \in \{1,2,\ldots\}$, $q \in \{1,2,\ldots\}$, $k \in \{2,3,\ldots\}$, Φ_i (i=1, 2,..., p) are $k\times k$ unknown matrices of real constants, θ_i (i=1, 2,..., q) are $k\times k$ unknown matrices of real constants, $\{y(t)\}$ is a sequence of $k\times 1$ real observable random vectors and $\{\varepsilon(t)\}$ is a sequence of $k\times 1$ independent and normally distributed unobservable random vectors with zero mean and a $k\times k$ unknown precision matrix. Then the Multivariate (vector) autoregressive moving average process of orders p and q, denoted by VARMA$_k$(p, q) for short, is defined for n vectors as

$$\Phi_p(B)\,y(t) = \Theta_q(B)\,\varepsilon(t) \tag{2.1}$$

Where, $\Phi_p(B) = I_k - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$, $y(t) = \left[y(t,1) \quad y(t,2) \quad \ldots \quad y(t,k)\right]'$, $\Theta_q(B) = I_k - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$ and $\varepsilon(t) = \left[\varepsilon(t,1) \quad \varepsilon(t,2) \quad \ldots \quad \varepsilon(t,k)\right]'$. I_k is the identity matrix of order k and B is the usual backward shift operator. The $k\times k$ matrix polynomial $\Phi_p(B)$, of degree p in the backward shift operator B, is known as the autoregressive operator of order p, while the $k\times k$ matrix polynomial $\Theta_q(B)$, of degree q in the backward shift operator, is known as the moving average operator of order q. The process $y(t)$ is stationary if all the roots of the determinantal equation $|\Phi_p(B)| = 0$ lie outside the unit circle, while the process is invertible if all the roots of determinantal equation $|\Theta_q(B)| = 0$ lie outside the unite circle.

Consider the special case VARMA$_3$(1,1) with autoregressive and moving average coefficients

$$\Phi=\Phi_1=\begin{bmatrix}\phi_{11} & \phi_{12} & \phi_{13}\\ \phi_{21} & \phi_{22} & \phi_{23}\\ \phi_{31} & \phi_{32} & \phi_{33}\end{bmatrix} \quad \text{and} \quad \Theta=\Theta_1=\begin{bmatrix}\theta_{11} & \theta_{12} & \theta_{13}\\ \theta_{21} & \theta_{22} & \theta_{23}\\ \theta_{31} & \theta_{32} & \theta_{33}\end{bmatrix}$$

Then the model (2.1) can be written as

$$\left(I_3-\Phi B\right)y(t)=\left(I_3-\Theta B\right)\varepsilon(t) \tag{2.2}$$

Where $y(t)=\left[y(t,1) \quad y(t,2) \quad y(t,3)\right]'$, $\varepsilon(t)=\left[\varepsilon(t,1) \quad \varepsilon(t,2) \quad \varepsilon(t,3)\right]'$ and

$$I_3-\Phi B=\begin{bmatrix}1-\phi_{11}B & -\phi_{12}B & -\phi_{13}B\\ -\phi_{21}B & 1-\phi_{22}B & -\phi_{23}B\\ -\phi_{31}B & -\phi_{32}B & 1-\phi_{33}B\end{bmatrix} \text{ Similarly, } I_3-\Theta B=\begin{bmatrix}1-\theta_{11}B & -\theta_{12}B & -\theta_{13}B\\ -\theta_{21}B & 1-\theta_{22}B & -\theta_{23}B\\ -\theta_{31}B & -\theta_{32}B & 1-\theta_{33}B\end{bmatrix}$$

Thus, one may write the observation $y(t)$ of the VARMA$_3$(1,1) process as

$$\begin{aligned}y(t,1)=&\phi_{11}y(t-1,1)+\phi_{12}y(t-1,2)+\phi_{13}y(t-1,3)-\theta_{11}\varepsilon(t-1,1)\\&-\theta_{12}\varepsilon(t-1,2)-\theta_{13}\varepsilon(t-1,3)+\varepsilon(t,1)\end{aligned}$$

$$\begin{aligned}y(t,2)=&\phi_{21}y(t-1,1)+\phi_{22}y(t-1,2)+\phi_{23}y(t-1,3)-\theta_{21}\varepsilon(t-1,1)\\&-\theta_{22}\varepsilon(t-1,2)-\theta_{23}\varepsilon(t-1,3)+\varepsilon(t,2)\end{aligned}$$

$$\begin{aligned}y(t,3)=&\phi_{31}y(t-1,1)+\phi_{32}y(t-1,2)+\phi_{33}y(t-1,3)-\theta_{31}\varepsilon(t-1,1)\\&-\theta_{32}\varepsilon(t-1,2)-\theta_{33}\varepsilon(t-1,3)+\varepsilon(t,3)\end{aligned}$$

The model (2.2) can be written compactly for n observations as

$$y(t)=\varepsilon(t)+\phi\, y(t-1)-\theta\,\varepsilon(t-1),\ t=1,2,\ldots,n \tag{2.3}$$

Where

$y(t-1)=\left[y(t-1,1) \quad y(t-1,2) \quad y(t-1,3)\right]'$ and $\varepsilon(t-1)=\left[\varepsilon(t-1,1) \quad \varepsilon(t-1,2) \quad \varepsilon(t-1,3)\right]'$.

Consider another important special case, the VARMA$_3$(2,1) process with autoregressive and moving average coefficients

$$\Phi_1=\begin{bmatrix}\phi_{1.11} & \phi_{1.12} & \phi_{1.13}\\ \phi_{1.21} & \phi_{1.22} & \phi_{1.23}\\ \phi_{1.31} & \phi_{1.32} & \phi_{1.33}\end{bmatrix},\quad \Phi_2=\begin{bmatrix}\phi_{2.11} & \phi_{2.12} & \phi_{2.13}\\ \phi_{2.21} & \phi_{2.22} & \phi_{2.23}\\ \phi_{2.31} & \phi_{2.32} & \phi_{2.33}\end{bmatrix} \text{ and } \Theta_1=\begin{bmatrix}\theta_{1.11} & \theta_{1.12} & \theta_{1.13}\\ \theta_{1.21} & \theta_{1.22} & \theta_{1.23}\\ \theta_{1.31} & \theta_{1.32} & \theta_{1.33}\end{bmatrix}$$

Then the model (2.1) can be written as

$$\left(I_3-\Phi_1B-\Phi_2B^2\right)y(t)=\left(I_3-\Theta_1B\right)\varepsilon(t)$$

where

$$I_3 - \Phi_1 B - \Phi_2 B^2 = \begin{bmatrix} 1-\phi_{1.11}B-\phi_{2.11}B^2 & -\phi_{1.12}B-\phi_{2.12}B^2 & -\phi_{1.13}B-\phi_{2.13}B^2 \\ -\phi_{1.21}B-\phi_{2.21}B^2 & 1-\phi_{1.22}B-\phi_{2.22}B^2 & -\phi_{1.23}B-\phi_{2.23}B^2 \\ -\phi_{1.31}B-\phi_{2.31}B^2 & -\phi_{1.32}B-\phi_{2.32}B^2 & 1-\phi_{1.33}B-\phi_{2.33}B^2 \end{bmatrix}$$

and

$$I_3 - \Theta_1 B = \begin{bmatrix} 1-\theta_{1.11}B & -\theta_{1.12}B & -\theta_{1.13}B \\ -\theta_{1.21}B & 1-\theta_{1.22}B & -\theta_{1.23}B \\ -\theta_{1.31}B & -\theta_{1.32}B & 1-\theta_{1.33}B \end{bmatrix}$$

Thus, the observation $y(t)$ of VARMA$_3$(2,1) model can be written as

$$\begin{aligned} y(t,1) &= \varepsilon(t,1) + \phi_{1.11}y(t-1,1) + \phi_{1.12}y(t-1,2) + \phi_{1.13}y(t-1,3) + \phi_{2.11}y(t-2,1) \\ &\quad + \phi_{2.12}y(t-2,2) + \phi_{2.13}y(t-2,3) - \theta_{1.11}\varepsilon(t-1,1) + \theta_{1.12}\varepsilon(t-1,2) - \theta_{1.13}\varepsilon(t-1,3) \\ y(t,2) &= \varepsilon(t,2) + \phi_{1.21}y(t-1,1) + \phi_{1.22}y(t-1,2) + \phi_{1.23}y(t-1,3) + \phi_{2.21}y(t-2,1) \\ &\quad + \phi_{2.22}y(t-2,2) + \phi_{2.23}y(t-2,3) - \theta_{1.21}\varepsilon(t-1,1) + \theta_{1.22}\varepsilon(t-1,2) - \theta_{1.23}\varepsilon(t-1,3) \\ y(t,3) &= \varepsilon(t,3) + \phi_{1.31}y(t-1,1) + \phi_{1.32}y(t-1,2) + \phi_{1.33}y(t-1,3) + \phi_{2.31}y(t-2,1) \\ &\quad + \phi_{2.32}y(t-2,2) + \phi_{2.33}y(t-2,3) - \theta_{1.31}\varepsilon(t-1,1) + \theta_{1.32}\varepsilon(t-1,2) - \theta_{1.33}\varepsilon(t-1,3) \end{aligned}$$

Similarly, the model VARMA$_3$(2,1) can be written in a compact form for n observations as

$$y(t) = \varepsilon(t) + \Phi_1 y(t-1) + \Phi_2 y(t-2) - \Theta_1 \varepsilon(t-1), \quad t = 1, 2, \ldots, n \tag{2.4}$$

Where $y(t)$, $\varepsilon(t)$, $y(t\text{-}1)$, $\varepsilon(t\text{-}1)$ are as defined before, and

$$y(t-2) = \begin{bmatrix} y(t-2,1) & y(t-2,2) & y(t-2,3) \end{bmatrix}'$$

In general, one can write the VARMA$_k$ (p, q) process as

$$\begin{aligned} y(t) &= \varepsilon(t) + \Phi_1 y(t-1) + \Phi_2 y(t-2) + \cdots + \Phi_p y(t-p) \\ &\quad - \Theta_1 \varepsilon(t-1) - \Theta_2 \varepsilon(t-2) - \cdots - \Theta_q \varepsilon(t-q), \; t = 1, 2, \ldots, n \end{aligned} \tag{2.5}$$

Conditioning on the first p vectors of observations, the model (2.5) can be written in more compact expression as

$$Y = X\,\Gamma + U \tag{2.6}$$

where Y is a matrix of order $(n\text{-}p)\times k$ with ij-th element equals $y(p+i, j)$, $i = 1, 2\ldots, n\text{-}p$; $j = 1, 2, \ldots, k$. That is

$$Y = Y_{(n-p)\times k} = \begin{bmatrix} y(p+1) & y(p+2) & \ldots & y(n) \end{bmatrix}'$$

The matrix X is of order $(n\text{-}p)\times kh$, $h = p + q$, defined by

$$X = X_{(n-p)\times kh} = [X_1 \quad X_2]$$

Where

$$X_1 = \begin{bmatrix} y'(p) & y'(p-1) & \dots & y'(1) \\ y'(p+1) & y'(p) & \dots & y'(2) \\ \vdots & \vdots & \vdots & \vdots \\ y'(n-1) & y'(n-2) & \cdots & y'(n-p) \end{bmatrix} \text{ and } X_2 = \begin{bmatrix} -\varepsilon'(p) & -\varepsilon'(p-1) & \dots & -\varepsilon'(p-q+1) \\ -\varepsilon'(p+1) & -\varepsilon'(p) & \dots & -\varepsilon'(p-q+2) \\ \vdots & \vdots & \vdots & \vdots \\ -\varepsilon'(n-1) & -\varepsilon'(n-2) & \cdots & -\varepsilon'(n-q) \end{bmatrix}$$

Furthermore, Γ is the $kh\times k$ matrix of coefficients defined by

$$\Gamma = \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix}, \quad \gamma_1 = \begin{bmatrix} \Phi_1' \\ \Phi_2' \\ \vdots \\ \Phi_p' \end{bmatrix} \quad \text{and} \quad \gamma_2 = \begin{bmatrix} \Theta_1' \\ \Theta_2' \\ \vdots \\ \Theta_q' \end{bmatrix}$$

Where

$$\Phi_i = \begin{bmatrix} \phi_{i.11} & \phi_{i.12} & \cdots & \phi_{i.1k} \\ \phi_{i.21} & \phi_{i.22} & \cdots & \phi_{i.2k} \\ \vdots & \vdots & & \vdots \\ \phi_{i.k1} & \phi_{i.k2} & & \phi_{i.kk} \end{bmatrix} \quad \text{and} \quad \Theta_j = \begin{bmatrix} \theta_{j.11} & \theta_{j.12} & \cdots & \theta_{j.1k} \\ \theta_{j.21} & \theta_{j.22} & \cdots & \theta_{j.2k} \\ \vdots & \vdots & & \vdots \\ \theta_{j.k1} & \theta_{j.k2} & & \theta_{j.kk} \end{bmatrix},$$

$$i = 1,2,\dots,p \; ; \; j = 1,2,\dots,q$$

The class of models (2.5) represents the general class of vector autoregressive moving average models of orders p and q and is usually denoted by VARMA$_k$ (p, q). In practice the orders p and q are usually unknown and it is necessary to identify them using n vectors of observations $S_n = [y(1)\ y(2)\ \dots\ y(n)]'$. Thus the statistical question is the following: given n vectors of observations S_n generated from a vector autoregressive moving average process, what are the most adequate values of p and q?. The indirect Bayesian answer of this question is to find the posterior distribution of the coefficients of the largest possible model and use a series of tests of significance to eliminate insignificant coefficients from the model to determine its orders. Whereas, the direct Bayesian answer is to find the joint posterior mass function of the orders p and q. These two processes have not been done yet because of the complexity of the likelihood function. It is worth noting that the orders p and q obtained by the indirect technique may be used as initial reasonable values to employ the direct technique.

3. Indirect Bayesian Identification of VARMA$_k(p,q)$

The VARMA$_k(p,q)$ class of models is quite useful in modeling vector time series data and frequently $p \leq 2$ and $q \leq 2$. In practice the values of the orders p and q are unknown and should be identified. The indirect Bayesian approach proposed to focus on the posterior distribution of the coefficients

$$\Gamma = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_P & \theta_1 & \theta_2 & \cdots & \theta_Q \end{bmatrix}' = \begin{bmatrix} \overbrace{\Phi}^{kP\times k} \\ \cdots \\ \overbrace{\Theta}^{kQ\times k} \end{bmatrix} \qquad (3.1)$$

Where ϕ_i is a square matrix of order k, $i = 1, 2,\ldots,P$, while θ_j is a square matrix of order k, $j = 1, 2, \ldots Q$. The maximum orders P and Q are assumed to be known. The conditional likelihood function of the parameters Φ, Θ and T is

$$L(\Phi,\Theta,T \mid S_n) \propto |T|^{-\frac{n-P}{2}} \exp\left(-\tfrac{1}{2} tr\{\sum_{t=P+1}^{n} \varepsilon(t)\,\varepsilon'(t)\}\right) \qquad (3.2)$$

where $\Phi \in R^{kP\times k}$, $\Theta \in R^{kQ\times k}$, $T > 0$ and

$$\varepsilon'(t) = y'(t) - X'(t-1)\Gamma \quad , \quad t = 1,2,\ldots,n \qquad (3.3)$$

Where

$$X'(t-1) = \begin{bmatrix} y'(t-1) & y'(t-2) & \ldots & y'(t-P) & -\varepsilon'(t-1) & -\varepsilon'(t-2) & \ldots & -\varepsilon'(t-Q) \end{bmatrix}$$

The expression (3.3) is a recurrence relation for the residuals and the *m-th* component of the residual $\varepsilon(t)$ can be written as

$$\varepsilon(t,m) = y(t,m) - \sum_{i=1}^{P}\sum_{j=1}^{k} \phi_{i,mj}\, y(t-i,j) + \sum_{i=1}^{Q}\sum_{j=1}^{k} \theta_{i,mj}\, \varepsilon(t-i,j)$$
$$t=1,2,\ldots,n;\ m=1,2,\ldots,k \qquad (3.4)$$

The recurrence relation (3.4) causes the main problem in developing the exact analysis of VARMA$_K(p,q)$ processes. However, this recurrence may be used to evaluate the residuals recursively if ϕ_i, θ_i and the initial values of the residuals are known. The proposed Bayesian approximation is based on replacing the exact residuals by their least squares estimates and assuming that the initial residuals equal their means, namely zero. Thus, we estimate the residuals recursively by

$$\hat{\varepsilon}(t,m) = y(t,m) - \sum_{i=1}^{P}\sum_{j=1}^{k}\hat{\phi}_{i.mj}\, y(t-i,j) + \sum_{i=1}^{Q}\sum_{j=1}^{k}\hat{\theta}_{i.mj}\,\hat{\varepsilon}(t-i,j)$$
$$t=1,2,\ldots,n;\ m=1,2,\ldots,k \qquad (3.5)$$

Where $\hat{\phi}_{i.mj}$ and $\hat{\theta}_{i.mj}$ are the nonlinear least squares estimates of parameters $\phi_{i.mj}$ and $\theta_{i.mj}$. Using the estimates of the residuals, it is possible to write the conditional likelihood function approximately as

$$L^{*}\left(\Phi,\Theta,\Gamma|S_n\right) \propto |T|^{\frac{n-P}{2}} \exp\left(-\tfrac{1}{2}tr\{\sum_{t=P+1}^{n}[y(t)-\Gamma'\hat{X}(t-1)][y(t)-\Gamma'\hat{X}(t-1)]'T\}\right) \qquad (3.6)$$

Where $\hat{X}(t-1)$ is the same as $X(t-1)$ but using the estimated residuals instead of the exact ones. A convenient choice of the joint prior density of the parameters Γ and T is the following matrix normal-Wishart distribution

$$\xi(\Gamma,T) = \xi_1(\Gamma|T)\xi_2(T) \qquad (3.7)$$

Where

$$\xi_1(\Gamma|T) \propto |T|^{\frac{k(P+Q)}{2}} \exp\left(-\tfrac{1}{2}tr[\Gamma-D]'W[\Gamma-D]T\right) \text{ and}$$

$$\xi_2(T) \propto |T|^{\frac{a-(k+1)}{2}} \exp\left(-\tfrac{1}{2}tr\,\Psi T\right)$$

The hyper-parameters $D \in R^{k(P+Q)\times k}$, W is a $k(P+Q)\times k(P+Q)$ positive definite matrix, a is a positive scalar and Ψ is a $k\times k$ positive definite matrix. If there is little information about the parameters, a priori, it is possible to use Jeffreys' vague prior

$$\xi(\Gamma,T) \propto |T|^{\frac{-(k+1)}{2}}, \quad \Gamma \in R^{k(P+Q)\times k}, \ T>0 \qquad (3.8)$$

Theorem 3.1

If the approximate conditional likelihood function (3.6) is combined with the joint prior density (3.7), the marginal posterior distribution of Γ is a matrix t with parameters: $(A^{-1}B,\ A^{-1},\ C\text{-}B'A^{-1}B,\ \upsilon)$ where

$$A = W + \sum_{t=P+1}^{n}\hat{X}(t-1)\hat{X}'(t-1),$$

$$B = WD + \sum_{t=P+1}^{n}\hat{X}(t-1)y'(t),$$

$$C = D'WD + \Psi + \sum_{t=P+1}^{n}y(t)y'(t) \text{ and}$$

$$\upsilon = n - P + a - k + 1$$

Corollary 3.1

If the approximate conditional likelihood function (3.6) is combined with the Jeffreys' vague prior (3.8), the marginal posterior distribution of Γ is a matrix t with parameters: $(A^{-1}B, A^{-1}, C\text{-}B'A^{-1}B, \upsilon)$. However, the quantities A, B, C and υ will be modified by letting $W \rightarrow 0$ $(k(P+Q)\times k(P+Q))$, $a \rightarrow -k(P+Q)$ and $\Psi \rightarrow 0$ $(k\times k)$.

Since Γ has a matrix t distribution, one can test any subset of rows to be zero, marginally or conditionally, using F statistic if k=2 and χ^2 statistic if $k \geq 3$. The forms of F and χ^2 statistics can be found in Box and Tiao (1973). Theorem 3.1 and corollary 3.1 give the machinery necessary to do an indirect procedure to identify the orders p and q of the general vector mixed ARMA$_K(p,q)$ processes as follows:

(1) Test $\mathbf{H_0}$: θ_Q=0 versus $\mathbf{H_1}$: $\theta_Q \neq 0$ using the marginal posterior distribution of θ_Q, which is a matrix t.

(2) If $\mathbf{H_0}$ is not rejected test $\mathbf{H_0}$: $\theta_{Q-1} = 0$ versus $\mathbf{H_1}$: $\theta_{Q-1} \neq 0$ using the conditional distribution of θ_{Q-1} given θ_Q= 0, which is also a matrix t.

(3) If the above $\mathbf{H_0}$ is not rejected, test $\mathbf{H_0}$: $\theta_{Q-2} = 0$ versus $\mathbf{H_1}$: $\theta_{Q-2} \neq 0$ using the conditional distribution of θ_{Q-2} given $\theta_Q = \theta_{Q-1} = 0$, which is also a matrix t, etc.

(4) On the other hand, if $\mathbf{H_0}$ of (1) is rejected, test $\mathbf{H_0}$: $\phi_P = 0$ versus $\mathbf{H_1}$: $\phi_P \neq 0$ with the marginal distribution of ϕ_P, which is also a matrix t.

(5) And so on.

Continuing in this fashion, one will arrive at a stage where he rejects the hypotheses $\phi_{p_0} = 0$ and $\theta_{q_0} = 0$. One concludes that the data was generated by a VARMA$_K(p_0,q_0)$ process, where $0 < p_0 \leq P$ and $0 < q_0 \leq Q$.

4. Direct Bayesian Identification of VARMA$_k(p,q)$

Given the initial values p_0 and q_0, the main objective of this section is to develop an approximate joint posterior probability mass function for the orders p and q of the VARMA$_k(p,q)$ process. In order to achieve our main objective, let $S_n = [y(1)\ y(2)\ \ldots\ y(n)]'$ be a matrix of $n\times k$ observations from VARMA$_k(p,q)$ process of the form (2.8) where the

orders p and q are unknown positive integers. Conditioning on the first p observed vectors, the likelihood function of the parameters $\Gamma(p,q)$, p, q and T is

$$L\left(\Gamma(p,q),p,q,T|S_n\right)\propto(2\pi)^{-\frac{k(n-p)}{2}}|T|^{\frac{n-p}{2}}\exp\left(-\tfrac{1}{2}tr\sum_{t=p+1}^{n}\varepsilon(t)\varepsilon'(t)T\right) \qquad (4.1)$$

Where $\Gamma(p,q)\in R^{kh\times k}$, $T>0$; p=1, 2, …, P ; q=1,2,…,Q; where P and Q are the largest possible orders of the process and h=P+Q. In general the conditional likelihood function (4.1) is very complicated because the disturbances $\varepsilon(t)$ are non-linear functions of the coefficients ϕ_i and θ_j. However, one may write the disturbances' vectors, from (2.8), as

$$\varepsilon'(t)=y'(t)-X'_{p,q}(t-1)\Gamma(p,q)\ , \quad t=1,2,\ldots,n \qquad (4.2)$$

Where

$$X'_{p,q}(t-1)=\left[y'(t-1)\ \ y'(t-2)\ \ldots y'(t-p)\ -\varepsilon'(t-1)\ \ -\varepsilon'(t-2)\ \ldots\ -\varepsilon'(t-q)\right]$$

The recurrence formula (4.2) causes the main problem in developing the exact Bayesian solution for the identification problem of VARMA$_K$(p,q) processes. The proposed Bayesian approximation is based on replacing the exact residuals ε(t) by their least squares estimates, say $\hat{\varepsilon}(t)$, which are obtained by searching over the parameter space for the values of Φ and Θ, say $\hat{\Phi}$ and $\hat{\Theta}$, which minimize the residual sum of squares $\sum\hat{\varepsilon}^2(t,i)$, i=1,2,…,k. Before doing this process, one should have initial values for the orders p and q. It is proposed to obtain such values, say p_0 and say q_0, by the indirect Bayesian technique presented in the previous section. The least squares estimates of Φ and Θ and the assumed initial values of the residuals, namely zero, are then substituted in (4.2) to obtain the least squares estimates of the residuals recursively. Substituting these estimates in $X'_{p,q}$(t-1), one can write (4.2) as

$$\hat{\varepsilon}'(t)=y'(t)-\hat{X}'_{p_0,q_0}(t-1)\Gamma(p,q)\ , \quad t=1,2,\ldots,n \qquad (4.3)$$

Where $\hat{X}'_{p_0,q_0}(t-1)$ is the same as $X'_{p,q}$(t-1) but using the residual estimates instead of the exact ones. Using the estimates of the residuals, one may rewrite the conditional likelihood function (4.1) approximately as

$$L^*\left(\Gamma(p,q),p,q,T|S_n\right)\propto(2\pi)^{-\frac{k(n-p)}{2}}|T|^{\frac{n-p}{2}}\exp\Big(-\frac{1}{2} tr\left\{\sum_{t=p+1}^{n}\left[y(t)-\Gamma'(p,q)\hat{X}_{p_0,q_0}(t-1)\right]\left[y(t)-\Gamma'(p,q)\hat{X}_{p_0,q_0}(t-1)\right]'T\right\}\Big) \qquad (4.4)$$

An adequate choice of the conditional prior of $\Gamma(p,q)$ given p, q and T is

$$f_1\left(\Gamma(p,q)\middle|p,q,T\right)\propto(2\pi)^{-\frac{hk^2}{2}}\left|R(p,q)\right|^{\frac{k}{2}}\left|T\right|^{\frac{hk}{2}}\exp(-\frac{1}{2}$$
$$tr\left\{\left[\Gamma(p,q)-D(p,q)\right]'R(p,q)\left[\Gamma(p,q)-D(p,q)\right]T\right\}) \quad (4.5)$$

Where the hyperparameters $D(p,q)\in R^{hk\times k}$ and $R(p,q)$ is an $hk\times hk$ positive definite matrix. The precision matrix is assigned, a priori, the Wishart distribution

$$f_2(T)\propto|T|^{\frac{[a-(k+1)]}{2}}\exp\left(-\frac{1}{2}tr\Psi T\right)$$

Where Ψ is a $k\times k$ positive definite matrix. Thus, the joint prior distribution of the parameters $\Gamma(p,q)$ and T given p and q is assumed to be

$$f_3\left(\Gamma(p,q),T\middle|p,q\right)\propto f_1\left(\Gamma(p,q)\middle|p,q,T\right)f_2(T) \quad (4.6)$$

The class of prior distributions (4.6) is called a matrix normal-Wishart class of distributions. Let β_{ij} be the prior probability mass function of the orders p and q, i.e.

$$\beta_{ij}=\Pr[p=i,q=j];\quad i=1,2,\ldots P\quad;\quad j=1,2,\ldots,Q \quad (4.7)$$

From (4.6) and (4.7), the joint prior distribution of the parameters $\Gamma(p,q)$, p, q and T is

$$f\left(\Gamma(p,q),p,q,T\right)\propto\beta_{ij}(2\pi)^{\frac{-hk^2}{2}}\left|R(p,q)\right|^{\frac{k}{2}}\left|T\right|^{\frac{hk+a-3}{2}}$$
$$\exp\left(-\tfrac{1}{2}tr\left\{\left[\Gamma(p,q)-D(p,q)\right]'R(p,q)\left[\Gamma(p,q)-D(p,q)\right]+\psi\right\}T\right) \quad (4.8)$$

If one can't or unwilling to specify the hyperparameters $D(p,q)$, $R(p,q)$, a, Ψ and β_{ij}, one might use Jeffreys' vague prior

$$f\left(\Gamma(p,q),p,q,T\right)\propto\left|T\right|^{-\frac{k+1}{2}} \quad (4.9)$$

Combining the approximate conditional likelihood function (4.4), via Bayes theorem, with the prior distribution of the parameters (4.8), one may write the joint posterior distribution of the parameters $\Gamma(p,q)$, p, q and T as

$$g\left(\Gamma(p,q),p,q,T\middle|S_n\right)\propto\beta_{ij}(2\pi)^{\frac{-k(n-p)-hk^2}{2}}\left|R(p,q)\right|^{\frac{k}{2}}\left|T\right|^{\frac{\alpha(p,q)}{2}}$$
$$\exp(-\tfrac{1}{2}tr\{\left[\Gamma(p,q)-D(p,q)\right]'R(p,q)\left[\Gamma(p,q)-D(p,q)\right]+\Psi+ \quad (4.10)$$
$$\sum_{t=p+1}^{n}\left[y(t)-\Gamma'(p,q)\hat{X}_{p,q}(t-1)\right]\left[y(t)-\Gamma'(p,q)\hat{X}_{p,q}(t-1)\right]'\}T)$$

Where

$$\alpha(p,q)=n-P+kh+a-k-1 \quad (4.11)$$

Theorem (4.1)

Using the approximate likelihood function (4.4) and the prior distribution (4.8), the joint posterior probability mass function of the orders p and q is

$$h(p,q|S_n) \propto \beta_{ij}(2\pi)^{\frac{pk}{2}} |R(p,q)|^{\frac{k}{2}} |A(p,q)|^{-\frac{k}{2}} |C(p,q)|^{-\frac{n-p+a}{2}} \prod_{j=1}^{k} \Gamma\left(\frac{n-P+a+j-k}{2}\right)$$

$$; n > P\text{-}a\text{+}k\text{-}1$$

Where

$$A(p,q) = R(p,q) + \sum_{t=p+1}^{n} \hat{X}_{p,q}(t-1)\hat{X}'_{p,q}(t-1),$$

$$B(p,q) = R(p,q)D(p,q) + \sum_{t=p+1}^{n} \hat{X}_{p,q}(t-1)y'(t-1) \quad \text{and}$$

$$C(p,q) = D'(p,q)R(p,q)D(p,q) + \Psi + \sum_{t=P+1}^{n} y(t)y'(t) - B'(p,q)A^{-1}(p,q)B(p,q)$$

Proof

Consider the terms between { } in the exponent of (4.10). They can be written as

$$\Gamma'(p,q)\left[R(p,q) + \sum_{t=p+1}^{n} \hat{X}_{p,q}(t-1)\hat{X}'_{p,q}(t-1)\right]\Gamma(p,q) - \Gamma'(p,q)\left[R(p,q)D(p,q) + \sum_{t=p+1}^{n} \hat{X}_{p,q}(t-1)y'(t)\right]$$

$$-\left[D'(p,q)R(p,q) + \sum_{t=p+1}^{n} y(t)\hat{X}_{p,q}(t-1)\right]\Gamma(p,q) + D'(p,q)R(p,q)D(p,q) + \Psi + \sum_{t=p+1}^{n} y(t)y'(t)$$

$$= \left[\Gamma'(p,q) - A^{-1}(p,q)B'(p,q)\right]' A(p,q)\left[\Gamma'(p,q) - A^{-1}(p,q)B'(p,q)\right] + C(p,q) \tag{4.12}$$

Substitute from (4.12) into (4.10), one may have

$$g(\Gamma(p,q),p,q,T|S_n) \propto \beta_{ij}(2\pi)^{-\frac{k(n-p)-hk}{2}} |R(p,q)|^{\frac{k}{2}} |T|^{\frac{\alpha(p,q)}{2}} \exp\left(-\tfrac{1}{2} tr\, C(p,q)T\right) \times$$

$$\exp\left(-\tfrac{1}{2} tr\left[\Gamma(p,q) - A^{-1}(p,q)B(p,q)\right]' A(p,q)\left[\Gamma(p,q) - A^{-1}(p,q)B(p,q)\right]T\right) \tag{4.13}$$

Integrating (4.13) with respect to $\Gamma(p,q)$, one may have the joint posterior distribution of p, q and T as

$$g_1(p,q,T|S_n) \propto \beta_{ij}(2\pi)^{-\frac{k(n-p)-hk}{2}} |R(p,q)|^{\frac{k}{2}} |T|^{\frac{\alpha(p,q)}{2}} \exp\left(-\tfrac{1}{2} tr\, C(p,q)T\right) \times \left[(2\pi)^{\frac{hk^2}{2}} |T|^{\frac{-hk}{2}} |A(p,q)|^{\frac{-k}{2}}\right] \tag{4.14}$$

However, (4.14) can be written as

$$g_1(p,q,T|S_n) \propto \beta_{ij}(2\pi)^{\frac{pk}{2}} |R(p,q)|^{\frac{k}{2}} |T|^{\frac{\alpha(p,q)-kh+2}{2}-1} |A(p,q)|^{\frac{-k}{2}} \exp\left(-\tfrac{1}{2} tr\, C(p,q)T\right) \tag{4.15}$$

Integrating (4.15) with respect to T, one may have the joint posterior mass function of p and q as

$$h(p,q|S_n) \propto \beta_{ij}(2\pi)^{\frac{pk}{2}} |A(p,q)|^{-\frac{k}{2}} |R(p,q)|^{\frac{k}{2}} |C(p,q)|^{-\frac{b(p,q)+k-1}{2}} 2^{\frac{k[b(p,q)+k-1]}{2}} \Gamma_k\left(\frac{b(p,q)+k-1}{2}\right) \tag{4.16}$$

Where $b(p,q) = \alpha(p,q)-kh+2$ and $\Gamma_r(s)$ is the generalized gamma function, see Box and Tiao (1973), defined by

$$\Gamma_r(s) = \left[\Gamma\left(\tfrac{1}{2}\right)\right]^{\frac{r(r-1)}{2}} \prod_{j=1}^{r} \Gamma\left(s+\tfrac{j-r}{2}\right) \; ; \quad s > \frac{r-1}{2}$$

Thus (4.16) can be written as

$$h(p,q|S_n) \propto \beta_{ij}(2\pi)^{\frac{pk}{2}} |A(p,q)|^{-\frac{k}{2}} |R(p,q)|^{\frac{k}{2}} |C(p,q)|^{-\frac{b(p,q)+k-1}{2}} .2^{\frac{k.b(p+q)}{2}} \prod_{j=1}^{k} \Gamma\left(\frac{b(p,q)+j-1}{2}\right),$$

$$b(p,q) > 0$$

Finally

$$h(p,q|S_n) \propto \beta_{ij}(2\pi)^{\frac{pk}{2}} |A(p,q)|^{-\frac{k}{2}} |R(p,q)|^{\frac{k}{2}} |C(p,q)|^{-\frac{n-p+a}{2}} \prod_{j=1}^{k} \Gamma\left(\frac{n-P+a+j-k}{2}\right), \quad n > P\text{-}a+k\text{-}1$$

This completes the proof.

Corollary (4.1)

Using the approximate likelihood function (4.4) and vague prior distribution (4.9), the joint posterior probability mass function of the orders p and q is

$$h_1(p,q|S_n) \propto \beta_{ij}(2\pi)^{\frac{k(p+kh)}{2}} |A^*(p,q)|^{\frac{-k}{2}} |C^*(p,q)|^{\frac{-(n-p-kh)}{2}} \prod_{j=1}^{k} \Gamma\left(\frac{n-p-kh+j-k}{2}\right), \quad n > P+kh+k\text{-}1$$

Where

$$A^*(p,q) = \sum_{t=p+1}^{n} \hat{X}_{p,q}(t-1)\hat{X}'_{p,q}(t-1) \; ,$$

$$B^*(p,q) = \sum_{t=p+1}^{n} \hat{X}_{p,q}(t-1)y'(t) \quad \text{and}$$

$$C^*(p,q) = \sum_{t=p+1}^{n} y(t)y'(t) - B^{*'}(p,q)A^{*-1}(p,q)B^*(p,q)$$

Proof

The proof of corollary (4.1) is similar to that of Theorem (4.1) by letting $D(p,q)=0$ $(hk\times k)$, $R(p,q) = 0$ $(hk\times hk)$, $\Psi=0$ $(k\times k)$ and $\alpha = -kh - k + 2$.

5. An Effectiveness Study

The main objective of this section is to study the effectiveness of the two proposed Bayesian techniques in solving the identification problem of multivariate (vector) mixed autoregressive moving average processes. In order to achieve this objective, six simulation

studies have been conducted. The proposed direct and indirect Bayesian techniques are employed to identify the orders of $VARMA_2(1,1)$ and $VARMA_2(2,1)$ processes with various parameters' values. The direct technique is employed with three different priors. All computations were performed on a PC using the most modern package SCA.

Our main concern is to study the numerical efficiency of the proposed indirect and direct Bayesian techniques by calculating the percentages of correct identification. Such efficiency will be examined with respect to the time series length (n) as well as the parameters of the selected model. For all values of the time series length (n) and the parameters, the variance-covariance matrix of the noise term is fixed at

$$T^{-1} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$$

Simulation I, as an illustration, begins with the generation of 500 data sets of bivariate normal variates, each of size 500, to represent $\varepsilon(t,1)$ and $\varepsilon(t,2)$ respectively. These data sets are then used to generate pairs of 500 realizations $y(t,1)$ and $y(t,2)$, each of size 500, from $VARMA_2(1,1)$ process with coefficients

$$\Phi = \begin{bmatrix} 0.1 & 0.1 \\ 0.1 & 0.1 \end{bmatrix} \quad \text{and} \quad \Theta = \begin{bmatrix} 0.2 & 0.2 \\ 0.2 & 0.2 \end{bmatrix}$$

By substituting in the equations

$$y(t,1) = \phi_{11} y(t-1,1) + \phi_{12} y(t-1,2) - \theta_{11}\varepsilon(t-1,1) - \theta_{12}\varepsilon(t-1,2) + \varepsilon(t,1)$$
$$y(t,2) = \phi_{21} y(t-1,1) + \phi_{22} y(t-1,2) - \theta_{21}\varepsilon(t-1,1) - \theta_{22}\varepsilon(t-1,2) + \varepsilon(t,2)$$

assuming that the starting values are zeros, $\varepsilon(0,1) = \varepsilon(0,2) = y(0,1) = y(0,2) = 0$.

The first 200 pairs of observations $y(t,1)$ and $y(t,2)$ are ignored to remove the effect of the initial conditions. Thus each generated realization will be of size 300. For a specific prior, the second step is to carry out all computations, assuming the maximum orders P=2 and Q=2, required to identify each of the 500 realizations and finding the percentages of correct identification for the proposed indirect and direct Bayesian methodologies. It might be important to mention that the indirect Bayesian technique is used as an intermediate step in the computations of the direct technique, in order to have adequate initial values p_0 and q_0. Such computations are done for a specific time series length n using the first n observations of each generated data set. This second step is repeated for each chosen time series length and prior combinations. The time series length n is taken to be 30, 50, 100,

150, 200 and 300. With respect to the prior probability mass function of the orders p and q, which is combined with the non-informative prior (4.9), the following priors are used:

Prior 1:

$$\beta_{ij} = \frac{1}{P} \times \frac{1}{Q} \quad , \qquad i = 1,2,\ldots,P \; ; \; j = 1.2,\ldots,Q$$

Prior 2:

$$\beta_{ij} \propto (0.5)^{i+j} \quad , \qquad i = 1,2,\ldots,P \; ; \; j = 1.2,\ldots,Q$$

Prior 3:

$$\beta_{11} = 0.35 \; ; \; \beta_{12} = \beta_{21} = 0.25 \; ; \; \beta_{22} = 0.15 \text{ for } P = Q = 2$$

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 orders p and q, while the third prior is chosen in such a way to give probabilities that decrease with an absolute amount 0.1 as the orders increase. Simulation II is done in a similar way but using

$$\Phi = \begin{bmatrix} -0.2 & -0.2 \\ -0.2 & -0.2 \end{bmatrix} \quad \text{and} \quad \Theta = \begin{bmatrix} -0.2 & -0.2 \\ -0.2 & -0.2 \end{bmatrix}$$

Simulation III is done in a similar way but using different values of the parameters Φ and Θ. Where,

$$\Phi = \begin{bmatrix} 1.1 & 0.3 \\ -0.2 & 0.6 \end{bmatrix} \quad \text{and} \quad \Theta = \begin{bmatrix} 0.4 & 0.4 \\ 0.4 & 0.4 \end{bmatrix}$$

The results of simulations I, II and III are reported in table 1. For each of these simulations, the percentages of correct identification using the indirect Bayesian technique, with non-informative prior (3.8), are also reported. Furthermore, the percentages of correct identification using the well-known AIC are reported as well.

Table (1): Percentages of Correct Identification of the Bayesian Techniques and AIC for Simulations I, II, III

PARAMETERS	n	INDIRECT	DIRECT PRIOR1	DIRECT PRIOR2	DIRECT PRIOR3	AIC
$\Phi_1=\begin{bmatrix}0.1 & 0.1\\ 0.1 & 0.1\end{bmatrix}$	30	92.6	1.2	1.4	1.2	47.6
	50	78.0	3.8	5.8	4.8	40.2
	100	70.6	9.6	14.0	11.4	39.8
$\Theta_1=\begin{bmatrix}0.2 & 0.2\\ 0.2 & 0.2\end{bmatrix}$	150	72.8	19.0	24.0	20.6	39.6
	200	77.0	25.4	33.2	28.8	45.0
	300	78.8	29.4	38.0	32.8	50.4
$\Phi_1=\begin{bmatrix}-0.2 & -0.2\\ -0.2 & -0.2\end{bmatrix}$	30	90.2	0.6	1.0	0.6	47.8
	50	77.4	3.8	5.8	5.0	46.8
	100	71.4	13.8	18.4	16.2	45.2
$\Theta_1=\begin{bmatrix}-0.2 & -0.2\\ -0.2 & -0.2\end{bmatrix}$	150	73.0	22.8	28.4	25.4	42.4
	200	74.6	33.0	38.8	36.0	48.0
	300	76.2	39.0	46.2	42.0	52.4
$\Phi_1=\begin{bmatrix}1.1 & 0.3\\ -0.2 & 0.6\end{bmatrix}$	30	93.0	3.6	6.8	5.4	51.2
	50	67.2	8.6	11.4	10.0	48.4
	100	61.0	18.2	22.8	20.0	50.2
$\Theta_1=\begin{bmatrix}0.4 & 0.4\\ 0.4 & 0.4\end{bmatrix}$	150	62.6	22.8	29.6	25.8	51.4
	200	63.6	24.8	31.2	27.2	56.6
	300	67.0	28.6	33.2	30.6	56.2

Source: Simulated Data

Inspection of the results, given in the above table, shows an increasing trend for efficiency of the proposed Bayesian methodologies as the series length n increases. In addition, the percentages of correct identification achieved using the indirect Bayesian methodology are higher than those obtained by both the direct technique and the AIC. Moreover, the percentages of correct identification achieved using the third prior in the direct technique are higher than the corresponding percentages achieved using the first prior, while the corresponding percentages achieved using the second prior are the highest among the three priors.

Another set of simulation studies were employed in a similar fashion but using $VARMA_2(2,1)$ processes and the results are reported in table 2.The reader can notice that these results are similar to those given above in table 1.

In simulation IV the coefficients are

$$\Phi_1=\begin{bmatrix}0.2 & -0.3\\ -0.3 & 0.4\end{bmatrix} \qquad \Phi_2=\begin{bmatrix}-0.1 & 0.4\\ -0.3 & 0.5\end{bmatrix} \qquad \Theta_1=\begin{bmatrix}0.5 & -0.4\\ -0.4 & 0.3\end{bmatrix}$$

In simulation V the coefficients are

$$\Phi_1=\begin{bmatrix}0.5 & -0.4\\ -0.4 & 0.3\end{bmatrix} \qquad \Phi_2=\begin{bmatrix}0.4 & 0.2\\ 0.5 & -0.5\end{bmatrix} \qquad \Theta_1=\begin{bmatrix}0.2 & -0.3\\ -0.3 & 0.4\end{bmatrix}$$

In simulation VI the coefficients are

$$\Phi_1 = \begin{bmatrix} 0.5 & -0.4 \\ -0.4 & 0.3 \end{bmatrix} \qquad \Phi_2 = \begin{bmatrix} 0.4 & 0.2 \\ 0.5 & -0.5 \end{bmatrix} \qquad \Theta_1 = \begin{bmatrix} -0.1 & 0.4 \\ -0.3 & 0.5 \end{bmatrix}$$

Table (2): Percentages of Correct Identification of the Bayesian Techniques and AIC for Simulation IV, V, VI

PARAMETERS	N	INDIRECT	DIRECT PRIOR1	DIRECT PRIOR2	DIRECT PRIOR3	AIC
$\Phi_1 = \begin{bmatrix} 0.2 & -0.3 \\ -0.3 & 0.4 \end{bmatrix}$ $\Phi_2 = \begin{bmatrix} -0.1 & 0.4 \\ -0.3 & 0.5 \end{bmatrix}$ $\Theta_1 = \begin{bmatrix} 0.5 & -0.4 \\ -0.4 & 0.3 \end{bmatrix}$	30	9.8	25.6	25.4	25.6	39.0
	50	30.0	21.0	23.8	22.4	33.8
	100	58.8	21.8	26.4	24.6	33.4
	150	64.6	22.8	26.6	25.8	37.4
	200	71.4	23.0	29.4	28.4	39.4
	300	72.6	25.8	30.0	29.0	41.6
$\Phi_1 = \begin{bmatrix} 0.5 & -0.4 \\ -0.4 & 0.3 \end{bmatrix}$ $\Phi_2 = \begin{bmatrix} 0.4 & 0.2 \\ 0.5 & -0.5 \end{bmatrix}$ $\Theta_1 = \begin{bmatrix} 0.2 & -0.3 \\ -0.3 & 0.4 \end{bmatrix}$	30	22.2	24.4	25.4	25.2	48.0
	50	52.8	20.8	27.4	25.6	52.0
	100	74.8	28.0	34.0	33.2	59.6
	150	74.4	32.4	35.0	34.8	61.4
	200	73.8	37.4	38.2	38.0	62.0
	300	70.2	36.8	37.8	37.6	55.2
$\Phi_1 = \begin{bmatrix} 0.5 & -0.4 \\ -0.4 & 0.3 \end{bmatrix}$ $\Phi_2 = \begin{bmatrix} 0.4 & 0.2 \\ 0.5 & -0.5 \end{bmatrix}$ $\Theta_1 = \begin{bmatrix} -0.1 & 0.4 \\ -0.3 & 0.5 \end{bmatrix}$	30	10.4	27.4	27.8	27.8	43.0
	50	32.0	26.8	30.8	30.2	40.2
	100	63.2	32.6	38.4	37.8	48.6
	150	75.8	33.4	36.4	35.8	59.0
	200	75.0	34.2	35.2	35.0	57.4
	300	70.4	34.8	35.2	35.2	62.0

Source: Simulated Data

Taking the above comments into consideration, one may say that the numerical results support the adequacy of using the proposed indirect Bayesian procedure in identifying the orders of vector mixed ARMA processes when a moderate or a large time series length is used.

Conclusion

The current study has developed two Bayesian techniques to identify the orders of vector mixed autoregressive moving average $VARMA_k(p,q)$ processes, namely the indirect and direct Bayesian identification techniques. The proposed indirect technique approximates the joint posterior probability density function of the coefficients of the largest possible model by a matrix *t* distribution. Then, by employing a series of tests of significance, the insignificant coefficients are eliminated, and the model orders are determined. On the other hand, by employing an approximate conditional likelihood function and a matrix Normal–Wishart, or Jeffreys' vague prior, the proposed direct identification technique is based on deriving an approximate joint posterior probability

mass function of the model orders in a convenient form. Then one may easily calculate the posterior probabilities for all possible values of the model orders and select the orders with maximum probability. A wide simulation study is conducted to examine the numerical efficiency of the proposed two Bayesian identification procedures and compare their performance with the well-known AIC technique. The numerical results show that the proposed techniques can efficiently identify the orders of vector autoregressive moving average processes for moderate and large time series lengths. Moreover, the indirect technique dominates the direct and AIC ones.

References

(1) Akaike, H. (1974). A new look at statistical model identification. *IEEE Transaction on Automatic Control, 19:716–723.*

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

(3) Box, G.E.P., Jenkins, G.M, Reinsel, G.C. and Ljung, G.M (2016). *Time Series Analysis: Forecasting and Control.* 5th ed., John Wiley & Sons.

(4) Box, G.E.P. and Pierce, D.A. (1970). Distribution of Residual Autocorrelations in Autoregressive Integrated Moving Average Time Series Models. *Journal of American Statistical Association, 65:1509-1526.*

(5) Box, G. E. P., Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis.* Reading, MA: Addison Wesley.

(6) Broemeling, L., Shaarawy, S. (1988). Time series: A Bayesian analysis in time domain. *Studies in Bayesian Analysis of Time Series and Dynamic Models, Edited by J. spall, Marcel Dekker Inc., New York, pp. 1–22.*

(7) Brockwell, P. J., Davis, R. A. (2016). Introduction to Time Series and Forecasting. *3rd ed. Springer International Publishing.*

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

(9) Granger, C. and Newbold, P. (1977). *Forecasting Economic Time Series.* New York: Academic Press.

(10) Liu, L. M. (2009). *Time Series Analysis and Forecasting.* 2nd ed. Villa Park, IL: Scientific Computing Associates.

(11) Ljung, G.M., Box, C.E.P. (1978). On a Measure of Lack of Fit in Time Series Models. *Biometrica, 65:297-304.*

(12) Monahan, J. F. (1983). Fully Bayesian analysis of ARIMA time series models. *Journal of Econometrics,* 21:307–331.

(13) Shaarawy, S. M. and Ali, S. S. (2003). Bayesian identification of seasonal autoregressive models. *Communications in Statistics- Theory and Methods, 32:1067–1084.*

(14) Shaarawy, S. M., Al Bassam, M. and Ali, S. S. (2006). A direct Bayesian technique to select the order of bivariate autoregressive process. *The Egyptian Statistical Journal, 50(1):1–22.*

(15) Shaarawy, S. M., Soliman, E. A., Ali, S. S. (2007). Bayesian identification of moving average models. *Communications in Statistics- Theory and Methods, 36:2301–2312.*

(16) Shaarawy, S. M. and Ali, S. S. (2008). Bayesian identification of vector autoregressive processes. *Communications in Statistics- Theory and Methods, 37:791–802.*

(17) Shaarawy, S. M. and Ali, S. S. (2012). Bayesian model order selection of vector moving average processes. *Communications in Statistics- Theory and Methods, 41(4):684–698.*

(18) Shaarawy, S M. and Ali, S. S. (2015). Bayesian identification of seasonal vector autoregressive processes. *Communications in Statistics- Theory and Methods, 44: 823–836.*

(19) Shumway, R.H. and Stoffer, D.S. (2017). *Time Series Analysis and its Applications with R Examples.* 4th Ed., Springer International Publishing.

(20) Soliman, E.E.A. (2016). Estimating the Orders of Bivariate Mixed ARMA(p,q) Processes using Bayesian Approach. *The Egyptian Statistical Journal, 60(1):1-19.*

(21) Tiao, G. C., Box, G. E. P., Grupe, M. R., Hudak, G. B., Bell, W. R., Change, I. (1979). *The Wisconsin Multiple Time Series (WMTS-1) Program: A Preliminary Guide.* University of Wisconsin, Madison: Department of Statistics.

(22) Tiao, G. C., Box, G. E. P. (1981). Modeling multiple time series with applications. *Journal of the American Statistical Association, 76:802–816.*

(23) Tiao, G. C., Tsay, R. (1983). Multiple time series modeling and extended sample cross correlation. *Journal of Business and Economic Statistics, 1:43–56.*

(24) Tsay, R. S. (2013). *Multivariate Time Series Analysis: With R and Financial Applications.* John Wiley & Sons.

(25) Tsay, R. S. (2014). *Multivariate Time Series Analysis.* John Wiley, Hoboken, NJ.

(26) Walter, E. (2014). *Applied Econometric Time Series.* 4th ed., John Wiley & Sons.