ESTIMATING THE ORDERS OF BIVARIATE MIXED ARMA(p,q) PROCESSES USING BAYESIAN APPROACH

Emad E.A. Soliman[1]

Department of Statistics
King Abul-Aziz University

easalam@hotmail.com

ABSTRACT

Estimating the orders of bivariate mixed autoregressive moving average processes, denoted by $ARMA_2(p,q)$, is the first and one of the most important phases in time series analysis. This article has three different objectives. The first one is to develop an indirect Bayesian methodology to estimate the orders of bivariate mixed ARMA processes. Assuming the maximum orders are known, the indirect methodology is based on approximating the posterior distribution of the model's coefficients by a matrix t distribution. Then one may test the significance of the coefficients marginally or conditionally and eliminate insignificance coefficients. The second objective is to develop a pure Bayesian methodology to estimate the orders of $ARMA_2(p,q)$ processes. The pure methodology is based on deriving an approximate joint posterior probability mass function of the orders in a convenient form. Then one may inspect the posterior probabilities and select the orders with maximum probability to be the estimated orders. The third objective is to carry out a simulation study to assess the performance and numerical efficiency of the two proposed methodologies and compare the results with the well-known automatic technique AIC or Akaike's information criterion. The numerical results show that the proposed indirect methodology is the best and can efficiently estimate the orders of bivariate mixed ARMA processes with high precision for moderate and large time series lengths.

Keywords: Bivariate ARMA processes – Pure Bayesian identification – indirect Bayesian identification – Posterior probability density function – Posterior probability mass function – Matrix normal-Wishart prior– Jeffreys' prior.

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

1. INTRODUCTION

Time series identification is a very important topic and may be found in many areas of applications such as economics, business, engineering, physics, chemistry, meteorology, ecology and environmental studies. The problem of identification is called bivariate problem if the observations are available for a set of two related time series. In economic, one may record spending on food y(t,1) and spending on health services y(t,2). In meteorology, one may record temperature y(t,1) and air pressure y(t,2). In hydrology, one may record rain fall y(t,1) and river flow y(t,2). These variables are modeled and examined simultaneously for two reasons. The first is to understand and investigate the nature of the dynamic relation between the two variables and the second is to increase the precision of the estimates and forecasts.

With respect to bivariate autoregressive moving average $ARMA_2(p,q)$ processes, identification means selecting the orders p and q using the observed data. It is a fact that the solution of the identification problem of $ARMA_2(p,q)$ processes 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 completely agreed upon.

Regarding univariate ARMA processes, one may trace two different non-Bayesian approaches to identify the model orders p and q. The first and most favorable one is developed by Box and Jenkins (1970). Their methodology is based on matching the sample autocorrelation and partial autocorrelation functions with their theoretical counterparts. The second non-Bayesian approach to identify the orders of univariate ARMA processes 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 minimizes the criterion. However, one may note that there is no agreement on the criterion to be minimized. The most popular automatic criterion AIC or Akaike's information criterion was introduced by Akaike (1974). For more details, the reader is referred to Schwarz (1978), Mills and Prasad (1992) and Beveridge and Oickle (1994).

On the other hand, the Bayesian methods of identification of univariate ARMA models are being developed. However, for well-understood reasons, most of the Bayesian publications focus on the analysis of pure autoregressive models and pay little attention to mixed ARMA models. This void in the Bayesian literature for the mixed processes is due to the complexity of the likelihood function of such processes.

Monahan (1983) has given a numerical algorithm to handle the estimation problem of the orders of ARMA processes. Broemeling and Shaarawy (1988) have developed an approximate Bayesian technique to estimate the orders of mixed ARMA processes. Shaarawy and Ali (2003) have developed a Bayesian technique to estimate the orders of seasonal autoregressive processes. More recently, Shaarawy et al. (2007) have developed an approximate pure Bayesian methodology to estimate the order of pure moving average processes.

With respect to the bivariate and multivariate version, the identification problem, from non-Bayesian view point, has been studied by Tiao and Box (1981) and Tiao and Tsay (1983) by matching the cross-correlation functions computed from the data with theoretical counterparts. On the other hand, the Bayesian methods of identification of bivariate and multivariate time series are not well-known. Shaarawy and Ali (2008) have introduced a pure Bayesian technique to identify order of multivariate (vector) autoregressive processes. Moreover, Shaarawy and Ali (2012) have developed a pure Bayesian methodology to identify the order of vector moving average processes. Most recently, Shaarawy and Ali (2015) have developed a pure Bayesian methodology to identify the orders of vector seasonal autoregressive processes. However, one may say that the Bayesian approach to estimate the bivariate mixed ARMA processes have not been explored yet.

The main objective of this article is to develop two convenient methodologies to estimate the orders of bivariate mixed ARMA processes. The first methodology, denoted by the indirect one, is based on approximating the posterior distribution of the model coefficients by a matrix t distribution. Then one may test the significance of the coefficients marginally or conditionally by a series of F-tests in a similar fashion to the backward elimination procedure used in regression analysis. The second methodology, denoted by the pure Bayesian one, is based on deriving an approximate joint posterior probability mass function of the model orders in a convenient form. Then one may inspect the posterior probabilities and select the orders at which the joint posterior mass function attains its maximum. A wide simulation study is conducted, using the modern specialized SCA package, in order to examine the numerical efficiency of the proposed Bayesian methodologies and compare results with the well-known automatic technique AIC or Akaike's information criterion.

The rest of this article is structured as follows: Section 2 gives the definition of the bivariate mixed ARMA processes in scalar and matrix notations. Section 3 constructs the conditional likelihood function of the parameters of the mixed $ARMA_2(p,q)$ processes. Section 4 develops the proposed indirect Bayesian technique to estimate the orders of $ARMA_2(p,q)$ processes. Section 5 explains the idea of the proposed pure Bayesian procedure to estimate the orders of $ARMA_2(p,q)$ processes. Section 6 is devoted to examine the numerical effectiveness of the proposed indirect and pure Bayesian procedures in estimating the orders of bivariate mixed processes. The performance of the proposed Bayesian procedures is compared with the well-known AIC technique.

2. BIVARIATE AUTORGRESSIVE MOVING AVERAGE PROCESSES

Let $\{t\}$ be a sequence of integers, $p \in \{1,2,\ldots\}$, $q \in \{1,2,\ldots\}$, Φ_i $(i=1,2,\ldots,p)$ are 2×2 unknown matrices of real constants, Θ_i $(i=1,2\ldots,q)$ are 2×2 unknown matrices of real constants, $\{y_t\}$ is a sequence of 2×1 real observable random vectors and $\{\varepsilon_t\}$ is a sequence of 2×1 independent and normally distributed unobservable random vectors with zero mean and a 2×2 unknown precision matrix. Then the bivariate autoregressive moving average process of orders p and q is defined for n vectors of observations as

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

Where

$$\Phi_p(\mathrm{B}) = \mathrm{I}_2 - \phi_1\mathrm{B} - \phi_2\mathrm{B}^2 - \cdots - \phi_p\mathrm{B}^p \quad , \quad y(t) = \left[y(t,1) \quad y(t,1)\right]',$$

$$\Theta_q(\mathrm{B}) = \mathrm{I}_2 - \theta_1\mathrm{B} - \theta_2\mathrm{B}^2 - \cdots - \theta_q\mathrm{B}^q \quad \text{and} \quad \varepsilon(t) = \left[\varepsilon(t,1) \quad \varepsilon(t,2)\right]'.$$

I_2 is the identity matrix of order 2, and B is the backward shift operator defined by B^r $y(t) = y(t\text{-}r)$. The 2×2 matrix polynomial $\Phi_p(\mathrm{B})$, of degree p in the backshift operator B, is known as the autoregressive operator of order p, while the 2×2 matrix polynomial $\Theta_q(\mathrm{B})$, of degree q in the backshift 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 $|\Phi_p(\mathrm{B})|=0$ lie outside the unit circle, while the process is invertible if all the roots determinantal equation $|\Theta_q(\mathrm{B})|=0$ lie outside the unite circle.

Consider a special case, the $ARMA_2(1,1)$ with coefficients

$$\phi = \phi_1 = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \quad \text{and} \quad \theta = \theta_1 = \begin{bmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{bmatrix}$$

Then the model (2.1) can be written as

$$(\mathrm{I} - \phi\, \mathrm{B})\, y(t) = (\mathrm{I} - \theta\, \mathrm{B})\varepsilon(t)$$

Where y_t and ε_t are defined above,

$$\mathrm{I} - \phi\, \mathrm{B} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} \phi_{11}\mathrm{B} & \phi_{12}\mathrm{B} \\ \phi_{21}\mathrm{B} & \phi_{22}\mathrm{B} \end{bmatrix} = \begin{bmatrix} 1 - \phi_{11}\mathrm{B} & -\phi_{12}\mathrm{B} \\ -\phi_{21}\mathrm{B} & 1 - \phi_{22}\mathrm{B} \end{bmatrix}$$

And

$$\mathrm{I} - \theta\, \mathrm{B} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} \theta_{11}\mathrm{B} & \theta_{12}\mathrm{B} \\ \theta_{21}\mathrm{B} & \theta_{22}\mathrm{B} \end{bmatrix} = \begin{bmatrix} 1 - \theta_{11}\mathrm{B} & -\theta_{12}\mathrm{B} \\ -\theta_{21}\mathrm{B} & 1 - \theta_{22}\mathrm{B} \end{bmatrix}$$

Thus, one may write the observations of the $ARMA_2(1,1)$ process as

$$\begin{aligned} y(t,1) &= \phi_{11}y(t-1,1) + \phi_{12}y(t-1,2) - \theta_{11}\varepsilon(t-1,1) - \theta_{12}\varepsilon(t-1,2) + \varepsilon(t,1) \\ y(t,2) &= \phi_{21}y(t-1,1) + \phi_{22}y(t-1,2) - \theta_{21}\varepsilon(t-1,1) - \theta_{22}\varepsilon(t-1,2) + \varepsilon(t,2) \end{aligned} \tag{2.2}$$

However, the model (2.2) can be written compactly for n observations as

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

Where

$$y(t-1) = \begin{bmatrix} y(t-1,1) \\ y(t-1,2) \end{bmatrix} \quad \text{and} \quad \varepsilon(t-1) = \begin{bmatrix} \varepsilon(t-1,1) \\ \varepsilon(t-1,2) \end{bmatrix}$$

Here we consider $y(t,1)$ and $y(t,2)$ as dependent variables, while $y(t-1,1)$, $y(t-1,2)$, $\varepsilon(t-1,1)$ and $\varepsilon(t-1,2)$ are considered as input or independent variables.

In general, one can write the $ARMA_2(p,q)$ process as

$$\begin{aligned} y(t) = \varepsilon(t) &+ \phi_1 y(t-1) + \phi_2 y(t-2) + \ldots + \phi_p y(t-p) \\ &- \theta_1\varepsilon(t-1) - \theta_2\varepsilon(t-2) - \ldots - \theta_q\varepsilon(t-q) \end{aligned} \tag{2.4}$$

Conditioning on the first p vectors of observations, the model (2.4) can be rewritten in a more compact expression as

$$Y = X\Gamma + U \tag{2.5}$$

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

$$Y = Y_{(n-p)\times 2} = \left[y(p+1) \quad y(p+2) \quad \ldots \quad y(n)\right]'$$

The matrix X is of order $(n\text{-}p)\times 2h$, $h= p + q$, defined by

$$X = X_{(n-p)\times 2h} = \left[X_1 \quad X_2\right] \text{ where}$$

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

Furthermore, Γ is the $2h\times 2$ matrix of coefficients defined as follows

$$\Gamma = \begin{bmatrix} \gamma_1 \\ \cdots \\ \gamma_2 \end{bmatrix} \text{ where, } \gamma_1 = \begin{bmatrix} \phi_1' \\ \cdots \\ \phi_2' \\ \cdots \\ \vdots \\ \cdots \\ \phi_p' \end{bmatrix}_{2p\times 2} \quad \text{and} \quad \gamma_2 = \begin{bmatrix} \theta_1' \\ \cdots \\ \theta_2' \\ \cdots \\ \vdots \\ \cdots \\ \theta_q' \end{bmatrix}_{2q\times 2}$$

$$\phi_i = \begin{bmatrix} \phi_{i.11} & \phi_{i.12} \\ \phi_{i.21} & \phi_{i.22} \end{bmatrix}, \ i = 1,2,\ldots,p \quad \text{and} \quad \theta_i = \begin{bmatrix} \theta_{i.11} & \theta_{i.12} \\ \theta_{i.21} & \theta_{i.22} \end{bmatrix}, \ i = 1,2,\ldots,q$$

The class of bivariate ARMA models (2.5) is an extremely useful, flexible and practical to model and forecast two correlated time series arise in many areas such as business, economics, industry, chemistry, ecology and meteorology. The class $ARMA_2(p,q)$ allows for general dynamic relationships between the two series in the system. In economics, for instance, one knows that a certain rise in prices will generally lead to some rise in wages, which in turn lead to a further rise in prices. Thus, modeling a univariate model for each series is not appropriate. Instead, a bivariate model will be more satisfactory to model and forecast the prices and wages simultaneously.

A specific diagonal element of the lag matrix φ_i, say $\varphi_{i.rr}$, shows how an individual series r is affected by its own past y(t-i, r). For example the element $\varphi_{2.11}$ reflects how y(t,1) is affected by y(t-2,1). An off diagonal element of the lag matrix φ_i, say $\varphi_{i.rj}$, reflects the influence of the series j on the series r. For example, $\varphi_{2.12}$ reflects

how $y(t,1)$ is affected by $y(t-2,2)$. The elements of the lag matrix θ_i are interpreted in a similar way.

3. THE CONDITIONAL LIKELIHOOD FUNCTION

In order to achieve our main goal, let $S_n=[y(1) \quad y(2) \ \dots \ y(n)]'$ be a matrix of $n\times 2$ observations generated from the bivariate ARMA process of orders p and q of the form (2.5) 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

$$L\left(\Gamma(p,q),p,q,T|S_n\right)\propto(2\pi)^{\frac{-2(n-p)}{2}}|T|^{\frac{(n-p)}{2}}\exp\left(-\tfrac{1}{2}tr\sum_{t=p+1}^{n}\varepsilon(t)\varepsilon'(t)T\right) \tag{3.1}$$

Where $\Gamma(p, q) \in R^{2h\times 2}$; $T>0$, $p = 1,2,\dots, P$; $q = 1,2,\dots, Q$ where P and Q are the largest possible orders of the process.

If n is large relative to p, as it usually is, the conditional likelihood function serves as a good approximation to the exact likelihood function, see Priestley (1981). In general the conditional likelihood function (3.1) is very complicated because the disturbances $\varepsilon(t)$ are non-linear functions of the coefficients φ_i and θ_i. To see this, one may write the disturbances of (2.5) as

$$\varepsilon'(t)=y'(t)-X'_{p,q}(t-1)\Gamma(p,q), \quad t=1, 2, \dots, n \tag{3.2}$$

Where

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

Thus, one may write the m^{th} component of the residual vector $\varepsilon(t)$ as

$$\varepsilon(t,m)=y(t,m)-\sum_{k=1}^{p}\sum_{j=1}^{2}\phi_{k,mj}y(t-k,j)+\sum_{k=1}^{q}\sum_{j=1}^{2}\theta_{k,mj}\varepsilon(t-k,j), \qquad m=1, 2 \tag{3.3}$$

The expression (3.3) is a recurrence relation for the residuals. This recurrence causes the main problem in developing the exact Bayesian solution for the identification problem of the bivariate ARMA processes. However, (3.3) can be used to evaluate the residuals recursively if one knows φ_i's, θ_i's and the initial values of the residuals. Using (3.3), the conditional likelihood function (3.1) can be written as

$$L(\Gamma(p,q),p,q,T|S_n) \propto (2\pi)^{\frac{-(n-p)}{2}} |T|^{\frac{n-p}{2}} \exp\left(-\tfrac{1}{2} tr \sum_{t=p+1}^{n} H(\Gamma(p,q),p,q,t)T\right) \quad (3.4)$$

Where $H(\Gamma(p, q),p,q,t) = (h_{rs})$ is 2×2 matrix and

$$h_{rs} = \left[y(t-r) - \sum_{k=1}^{p}\sum_{j=1}^{2} y(t-k,j)\phi_{k.rj} + \sum_{k=1}^{q}\sum_{j=1}^{2} \varepsilon(t-k,j)\theta_{k.rj} \right] \times \left[y(t-s) - \sum_{k=1}^{p}\sum_{j=1}^{2} y(t-k,j)\phi_{k.sj} + \sum_{k=1}^{q}\sum_{j=1}^{2} \varepsilon(t-k,j)\theta_{k.sj} \right] \quad (3.5)$$

The form (3.5) is not quadratic in the parameters φ's and θ's because ε(t-k, j) is a function of φ's and θ's through the recurrence formula (3.3). If ε(t-k, j) are known, $H(\Gamma(p,q),p,q,t)$ would be a quadratic from in the parameters. The proposed approximation is based on replacing the exact residuals ε(t) by their least squares estimates. The least squares estimates, say $\hat{\varepsilon}(t)$, are obtained by searching over the parameter space for the values of φ and θ, say φ_0 and θ_0, which minimize the residual sum of squares $\sum_t \hat{\varepsilon}^2(t,i)$, i=1,2. Before doing this process, one should have initial adequate values for the orders p and q. It is proposed to obtain such values, say p_0 and q_0 by the indirect Bayesian technique presented in the next section. The least squares estimates φ_0 and θ_0 and the assumed initial values, namely zero, are then substituted in (3.3) to obtain the least squares estimates of the residuals recursively. Substituting these estimates in $X'_{p,q}(t-1)$, one can write (3.2) as

$$\hat{\varepsilon}'(t) = y'(t) - \hat{X}'_{p_0,q_0}(t-1)\Gamma(p,q) \quad (3.6)$$

Where $t = 1,2,\ldots,n$ and $\hat{X}'_{p_0,q_0}(t-1)$ is the same as $X'_{p_0,q_0}(t-1)$ but using the residuals estimates instead of the exact ones. Using the estimates of the residuals, one may rewrite the conditional likelihood function (3.4) approximately as

$$L^*(\Gamma(p,q),p,q,T|S_n) \propto (2\pi)^{\frac{-2(n-p)}{2}} |T|^{\frac{n-p}{2}} \exp\left(-\tfrac{1}{2} tr \sum_{t=p+1}^{n} \left\{ \left[y(t) - \Gamma'(p,q)X_{p_0,q_0}(t-1)\right]\left[y(t) - \Gamma'(p,q)X_{p_0,q_0}(t-1)\right]' T \right\}\right) \quad (3.7)$$

An appropriate choice of the conditional prior of Γ(p,q) given p, q and T is

$$g\left(\Gamma(p,q)|p,q,T\right)=(2\pi)^{-\frac{4h}{2}}|R(p,q)|^{\frac{2}{2}}|T|^{\frac{2h}{2}} \exp\left(-\tfrac{1}{2}tr\left\{\left[\Gamma(p,q)-D(p,q)\right]' R(p,q)\left[\Gamma(p,q)-D(p,q)\right]T\right\}\right) \quad (3.8)$$

Where the hyper-parameters $D(p,q) \in R^{2h\times 2}$ and $R(p,q)$ is a $2h\times 2h$ positive definite matrix. The precision matrix is assigned, a priori the Wishart distribution

$$g_2(T)\propto |T|^{\frac{a-3}{2}}\exp\left(-\frac{1}{2}tr[\psi T]\right)$$

Where, ψ is a 2×2 positive definite matrix. Thus, the joint prior distribution of $\Gamma(p,q)$ and T given p and q is assumed to be

$$g_3\left(\Gamma(p,q),T|p,q\right)\propto g_1\left(\Gamma(p,q)|p,q,T\right)g_2(T) \quad (3.9)$$

The class of prior distributions (3.9) is called matrix Normal–Wishart class of distributions. Let β_{ij} be the prior probability mass function of the orders p and q, i.e.

$$\beta_{ij}=\Pr[p=i,q=j]\ ;\ \ i=1,2,\ldots,P\ ;\ j=1,2,\ldots,Q \quad (3.10)$$

From (3.9) and (3.10), the joint prior distribution of the parameters $\Gamma(p,q)$, T, p and q is

$$g\left(\Gamma(p,q),p,q,T\right)=\beta_{ij}(2\pi)^{-\frac{4h}{2}}|R(p,q)|^{\frac{2}{2}}|T|^{\frac{[2h+a-3]}{2}} \exp\left(-\tfrac{1}{2}tr\left\{\left[\Gamma(p,q)-D(p,q)\right]' R(p,q)\left[\Gamma(p,q)-D(p,q)\right]+\psi\right\}T\right) \quad (3.11)$$

If one can't or unwilling to specify the hyper-parameters $D(p,q)$, $R(p,q)$, a, ψ and β_{ij}, one might use jeffreys' vague prior

$$g\left(\Gamma(p,q),p,q,T\right)\propto |T|^{\frac{-3}{2}} \quad (3.12)$$

4. INDIRECT BAYESIAN METHODOLOGY

The $ARMA_2(p,q)$ class of models is quite useful in modeling and forecasting the two dimensional time series data and frequently p and q are not excess of 2. In practice the values of the orders p and q are unknown and one has to estimate them using the observed n vectors of 2×1 observations. The pure Bayesian approach to estimate the orders p and q is to find the joint posterior probability mass function of p and q and choose the orders at which the joint posterior mass function attains its maximum.

The approach taken here is somewhat different from the pure Bayesian approach. Instead of working directly with joint posterior distribution of p and q, it is proposed to focus the marginal posterior distribution of the coefficients

$$\Gamma = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_P \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_Q \end{bmatrix} = \begin{bmatrix} \overbrace{\Phi}^{2P\times 2} \\ \cdots \\ \overbrace{\Theta}^{2Q\times 2} \end{bmatrix} \tag{4.1}$$

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

$$L(\Phi,\Theta,T|S_n) \propto |T|^{\frac{n-p}{2}} \exp\left(-\tfrac{1}{2} tr\left\{ \sum_{t=p+1}^{n} \varepsilon(t)\,\varepsilon'(t)\,\mathrm{T} \right\} \right) \tag{4.2}$$

Where, $\Phi \in R^{2P\times 2}$, $\Theta \in R^{2Q\times 2}$, T is a positive definite matrix, and

$$\varepsilon'(t) = y'(t) - X'(t-1)\Gamma \quad , \quad \text{t= 1,2,...,}n \tag{4.3}$$

Where

$$X'(t-1)=\left[y'(t-1) \;\; y'(t-2) \;\; \ldots \;\; y'(t-P) \;\; -\varepsilon'(t-1) \;\; -\varepsilon'(t-2) \;\; \ldots \;\; -\varepsilon'(t-Q)\right]$$

The expression (4.3) is a recurrence relation for the residuals and the m^{th} component of the residual ε(t) can be written as

$$\varepsilon(t,m) = y(t,m) - \sum_{k=1}^{p}\sum_{j=1}^{2} \phi_{k.mj}\, y(t-k,j) + \sum_{k=1}^{Q}\sum_{j=1}^{2} \theta_{k.mj}\, \varepsilon(t-k,j) \quad , m=1,2 \tag{4.4}$$

The recurrence relation (4.4) causes the main problem in developing the exact Bayesian analysis of the bivariate ARMA processes. However, this recurrence may be used to evaluate the residuals recursively if one knows Φ, Θ and the initial values of the residuals. The proposed 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

$$\hat{\varepsilon}(t,m) = y(t,m) - \sum_{k=1}^{p}\sum_{j=1}^{2}\hat{\phi}_{k.mj}\,y(t-k,j) + \sum_{k=1}^{Q}\sum_{j=1}^{2}\hat{\theta}_{k.mj}\,\varepsilon(t-k,j)\,,\ m=1,2 \tag{4.5}$$

Where t=1,2,…,n; m=1,2 and $\hat{\phi}_{k.mj}$ and $\hat{\theta}_{k.mj}$ are the nonlinear least squares estimates of the parameters $\phi_{k.mj}$ and $\theta_{k.mj}$. Using the estimates of the residuals, one may rewrite the conditional likelihood function approximately as

$$L^{*}(\Phi,\Theta,T|S_n) \propto |T|^{\frac{n-p}{2}} \exp\left(-\tfrac{1}{2}tr\left\{\sum_{t=p+1}^{n}\left[y(t)-\Gamma'\hat{X}(t-1)\right]\left[y(t)-\Gamma'\hat{X}(t-1)\right]' T\right\}\right) \tag{4.6}$$

Where $\hat{X}(t-1)$ is the same as $X(t-1)$ but using the estimated residuals instead of the exact ones.

An appropriate choice of the prior density of the parameters Γ and T is the following matrix normal–Wishart distribution:

$$\xi(\Gamma, T) = \xi_1(\Gamma|T)\,\xi_2(T) \tag{4.7}$$

Where

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

And

$$\xi_2(T) \propto |T|^{\frac{\alpha-3}{2}} \exp\left(-\frac{1}{2}tr\,\psi T\right)$$

Where the hyper-parameters $D \in R^{2(P+Q)\times 2}$, W is a $2(P+Q)\times 2(P+Q)$ positive definite matrix, a > 0 and ψ is a 2×2 positive definite matrix. If one has little information about the parameters, a priori, he may use Jefferys' vague prior.

$$\xi(\Gamma, T) \propto |T|^{\frac{-3}{2}}, \qquad \Gamma \in R^{2(P+Q)\times 2},\ T>0 \tag{4.8}$$

Theorem 4.1

Using the approximate conditional likelihood function (4.6) and the matrix normal–Wishart prior density (4.7), the marginal posterior distribution of Γ is a matrix t distribution with parameters ($A^{-1}B$, A^{-1}, C-$B'A^{-1}B$, υ). Where

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

$$C = D'WD + \sum_{t=P+1}^{n} y(t)y'(t) \text{ , and } \quad \upsilon = n - P + a - 1$$

Corollary 4.1

Using the approximate conditional likelihood function (4.6) and Jeffreys' prior (4.8), the marginal posterior distribution of Γ is a matrix t distribution with parameters (A^{-1} B, A^{-1}, C-$B'A^{-1}$ B, υ). However, the quantities A, B, C and υ will be modified by letting $W \rightarrow 0$ ($2(P+Q)\times 2(P+Q)$), a $\rightarrow$ -$2(P+Q)$, and $\psi \rightarrow 0$ (2×2).

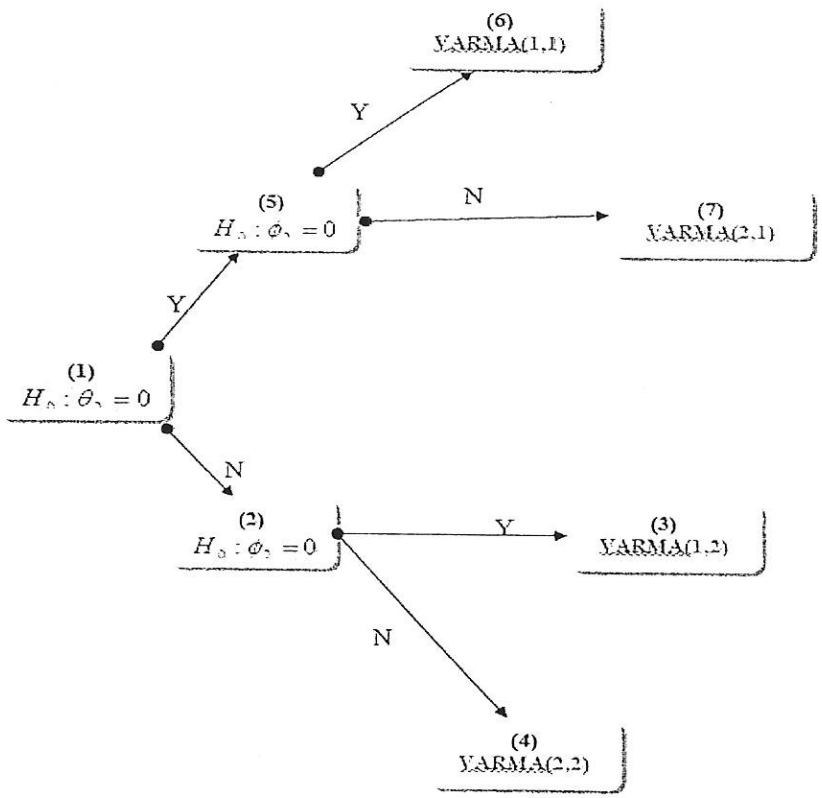
The reader is referred to Box and Tiao (1973) for the form and properties of the matrix t density function. Since Γ has a matrix t distribution, any subset of rows has a matrix t distribution; particularly the marginal posterior distribution of the matrix Φ or Θ is a matrix t distribution. In addition, the conditional distribution of any subset of rows given any other subset of rows is a matrix t distribution. Furthermore, one can test any subset of rows to be zero (marginally or conditionally) using F statistics. The forms of the F statistics can be found in Box and Tiao (1973, pp.451-455).

The following diagram, figure 1, gives a binary decision tree which depicts paths by which a particular $ARMA_2(p,q)$ is selected assuming $P = Q = 2$. Thus for example, one first tests $\theta_2 = 0$ and if rejected tests $\varphi_2 = 0$, and if not rejected tests $\varphi_1 = 0$ given $\varphi_2 = 0$, which if rejected then concludes the model is an $ARMA_2(1,2)$. The hypothesis $\theta_2 = 0$ is tested using the marginal distribution of θ_2, which is a matrix t, and the hypothesis $\varphi_2 = 0$ is tested with the marginal distribution of φ_2, but the hypothesis $\varphi_1 = 0$ is tested using the conditional distribution of φ_1 given $\varphi_2 = 0$, which is also a matrix t. The decision procedure begins with the moving average coefficient θ_2 and if it is decided that $\theta_2 \neq 0$, switches the test to the autoregressive parameter $\varphi_2 = 0$. It is somewhat arbitrary to begin with the moving average coefficient instead of the autoregressive part, but it was felt that the moving average coefficient should be given the first opportunity to be eliminated.

In general the proposed indirect Bayesian procedure to identify the orders p and q is continued in this fashion until the two hypotheses H_0: $\theta_{q_0} = 0$ and H_0: $\phi_{p_0} = 0$ are

rejected for some q_0 and p_0 where $0< q_0 \leq Q$ and $0 < p_0 \leq P$. The values p_0 and q_0 are the proposed indirect Bayesian estimates of the orders p and q of bivariate mixed ARMA processes.

FIGURE 1:
Decision Tree For Identification of $ARMA_2(p,q)$ Process Assuming P = Q = 2



5. A PURE BAYESIAN IDENTIFICATION METHODOLOGY

Given the initial values p_0 and q_0, the main objective of this section is to develop an approximate Bayesian procedure to identify the orders of bivariate ARMA processes. Unlike the indirect technique, outlined above, the orders p and q are assumed to be random variables and the problem becomes how to find their joint posterior probability mass function in an easy and convenient form. Combining the approximate conditional likelihood function (3.7), via Bayes theorem, with the prior distribution of the parameters (3.11), one may write the joint posterior distribution of the parameters Γ(p,q), p, q and T as

$$f\left(\Gamma(p,q),p,q,T|S_n\right)\propto \beta_{ij}(2\pi)^{\frac{-2(n-p)-4h}{2}}|R(p,q)|^{\frac{2}{2}}|T|^{\frac{\alpha(p,q)}{2}}\exp(-\tfrac{1}{2}tr$$
$$\{[\Gamma(p,q)-D(p,q)]'R(p,q)[\Gamma(p,q)-D(p.q)\ ]+\Psi+ \quad (5.1)$$
$$\sum_{t=p+1}^{n}[y(t)-\Gamma'(p,q)\hat{X}_{p,q}(t-1)][y(t)-\Gamma'(p,q)\hat{X}_{p,q}(t-1)]'\}T)$$

Where $\alpha(p, q) = n\text{-}p+2h+a\text{-}3$ (5.2)

Theorem (5.1)

Using the approximate conditional likelihood function (3.7) and prior distribution (3.11), the joint posterior mass function of the orders p and q is

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

Where

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

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

And

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

Theorem 5.2

Using the approximate conditional likelihood function (3.7) and the non–informative prior distribution (3.12), the joint posterior probability mass function of the model orders p and q is

$$h_1(p,q|S_n)\propto(\pi)^{3p+2q}|A^*(p,q)|^{-1}|C^*(p,q)|^{\frac{-(n-3p-2q)}{2}}\prod_{j=1}^{2}\Gamma(\frac{n-3p+2q+j-2}{2});$$
$$n>1+3p+2q$$

Where

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

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

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

The forms of the joint posterior probability mass function given by theorems (5.1) and (5.2) are convenient and easily programmed. Then, one may inspect the posterior probability mass function over the grid of the orders p and q and select the values at

which the joint probability mass function attains its maximum to be the identified orders of the bivariate time series being analyzed.

6. THE EFFECIVENESS OF THE STUDY

One of the main objectives of this article is to study the effectiveness of the proposed two Bayesian identification methodologies, namely the pure and the indirect ones, in solving the identification problem of Bivariate mixed ARMA processes. In order to achieve this objective, two simulation studies have been conducted. The proposed two Bayesian methodologies are employed, with three different prior distributions, to identify the orders of $VARMA_2(1,1)$ and $VARMA_2(2,1)$ processes. 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 two Bayesian methodologies by calculating the percentages of correct identification for each one. Such efficiency will be examined with respect to the time series length (n). Note that, 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, each of size 500, from $VARMA_2(1,1)$ process with

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

Assume that the starting values are zeros. 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 using each of the two proposed Bayesian methodologies. It might be important to mention that the computations of the pure Bayesian methodology include the application of indirect Bayesian technique in order to get 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 50,100,200 and 300. With respect to the prior probability mass function of the orders p and q, which is combined with the non-informative prior of $\Gamma(p,q)$ and T defined in (3.12), 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:

$\beta_{11} = 0.35$; $\beta_{12} = \beta_{21} = 0.25$; $\beta_{22} = 0.15$ 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 for $ARMA_2(2,1)$ with

$$\Phi_1 = \begin{bmatrix} 0.5 & -0.4 \\ -0.4 & 0.3 \end{bmatrix}, \quad \Phi_2 = \begin{bmatrix} 0.4 & 0.2 \\ 0.5 & -0.5 \end{bmatrix} \quad \text{and} \quad \Theta_1 = \begin{bmatrix} -0.1 & 0.4 \\ -0.3 & 0.5 \end{bmatrix}$$

The results of simulations I and II are reported in tables 1 and 2 respectively. For each simulation, the percentages of correct identification using the two proposed Bayesian techniques as well as the well-known AIC are reported in the same table.

Table 1: Percentages of Correct Identification of the Proposed Bayesian Methodologies and AIC for Simulation I

PARAMETERS	*n*	INDIRECT	DIRECT PRIOR1	DIRECT PRIOR2	DIRECT PRIOR3	AIC
$\Phi_1 = \begin{bmatrix} 0.1 & 0.1 \\ 0.1 & 0.1 \end{bmatrix}$	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

Source: Simulated Data

Inspection of the results, given in the above table, shows an increasing trend for efficiency of the proposed pure Bayesian methodology 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 pure methodology and the AIC. Moreover, the percentages of correct identification achieved using the third prior in the pure methodology is 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.

The results of simulation II are reported in table 2.The reader can notice that these results are similar to those given in table 1.

Table 2: Percentages of Correct Identification of the of the Proposed Bayesian Methodologies and AIC for Simulation II

PARAMETERS	n	INDIRECT	DIRECT PRIOR1	DIRECT PRIOR2	DIRECT PRIOR3	AIC
$\Phi_1 = \begin{bmatrix} 0.5 & -0.4 \\ -0.4 & 0.3 \end{bmatrix}$	50	32.0	26.8	30.8	30.2	40.2
$\Phi_2 = \begin{bmatrix} 0.4 & 0.2 \\ 0.5 & -0.5 \end{bmatrix}$	100	63.2	32.6	38.4	37.8	48.6
	150	75.8	33.4	36.4	35.8	59.0
$\Theta_1 = \begin{bmatrix} -0.1 & 0.4 \\ -0.3 & 0.5 \end{bmatrix}$	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

Considering the above comments, one may say that the numerical results support the adequacy of using the proposed two Bayesian methodologies in estimating the orders of bivariate mixed ARMA processes when a moderate or a large time series length is used.

References

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

(2) Beveridge, S., Oickle, C. (1994). A comparison of Box–Jenkins and objectives methods for determining the order of a non-seasonal ARMA model. *Journal of Forecasting*13:419–434.

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

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

(5) Broemeling, L., Shaarawy, S. (1988). Time series: A Bayesian analysis in time domain.. *Bayesian Analysis of Time Series and Dynamic Models edited by Spall, J, pp. 1–22.*

(6) Mills, J., Prasad, K. (1992). A comparison of model selection criteria. *Econometric Reviews*11:201–233.

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

(8) Priestley, M. (1981). Spectral Analysis of Time Series. *London: Academic Press.*

(9) Schwarz, G. (1978). Estimating the Dimension of a Model. *Annals of the Institute of Statistical Mathematics, Vol. 6, pp. 461-464*

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

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

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

(13) Shaarawy, S and Ali S. (2012). Bayesian Model Order Selection of Vector Moving Average Processes. *Communications in Statistics- Theory and Methods, Vol.41 ,Issue 4,684-698*

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

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

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