BAYESIAN FORECSATING OF VECTOR MOVING AVERAGE PROCESSES

SHERIF S. ALI
ssali1970@yahoo.com
Faculty of Science, Department of Statistics, King Abdulaziz University, KSA.
Permanent Address: Faculty of Economics and Political Science, Department of Statistics, Cairo University, Egypt.

ABSTRACT

Forecasting is the final and one of the most important phases of a multivariate time series analysis. This article develops an approximate Bayesian methodology to forecast the future observations of vector moving average processes. By employing an approximate conditional likelihood and a matrix normal-Wishart, or Jeffreys' vague prior, the proposed Bayesian forecasting methodology is based on deriving an approximate posterior probability density of the future observations in a convenient form. Then one may easily calculate the posterior mean vector and precision of the future vector of observations and hence develops a Highest Predictive Density(HPD) region for the future observations. Four simulation studies, with Jeffreys' vague prior, have been conducted in order to demonstrate the idea of the proposed methodology and test its adequacy in solving the forecasting problems of vector moving average processes. The numerical results show that the proposed methodology can efficiently forecast the vector moving average processes with high precision for moderate and large time series length.

***Keywords**: Forecasting, vector moving average processes, likelihood function, matrix normal-Wishart distribution, predictive density.*

1. INTRODUCTION

The statistical analysis of time series is very important topic and may be found in many areas of application such as economics, business, marketing and environmental studies. The problems of time series analysis are called univariate time series if the observations are made of single response such as hourly temperature and daily spending on food (in money). The vector time series problems arise if the observations are available for several related variables of interest. The multivariate time series may be found in many fields of application such as economics, business, meteorology, hydrology and utilities. In economics, one may record yearly money supply y(t, 1), real interest rate y(t, 2) and gross national product y(t,3) . In business, one my record single–family housing starts y(t, 1) and houses sold y(t, 2) in the U. S. A. These variables are modeled and investigated simultaneously for two reasons. The first reason is to understand the nature of the dynamic relationships between variables. One variable may lead to the other or there may be feedback relationship. The second reason is to increase the precisions of the estimates and forecasts. Better estimates and forecasts can be achieved when the series are modeled jointly if there is information on one series contained in the others. See Tiao and Box (1981).

Usually the Bayesian and non–Bayesian approaches of univariate and multivariate time series are based on a class of parametric models such as autoregressive moving average, denoted by ARMA for short, models. After the model is identified, the parameters are estimated and the future values are forecasted. Model estimation and forecasting are two main phases in time series analysis. Although these two phases are closely connected in time series analysis, they are usually treated as two separate steps. Standard references of the univariate non-Bayesian approach are Box and Jenkins (1970), Granger and Newbold (1973), Harvey (1993), Priestely (1981), Brockwell and Davis (1991), Wei (2005), Box et al. (2008) and Liu (2009). It is a fact that the methodology of Box and Jenkins for identification, estimation, diagnostic checking and forecasting of autoregressive moving average processes is the most favorable non-Bayesian technique to model and forecast time series data arise in most areas of application. Their methodology has grown up in popularity and is today the prevailing technique of modeling and forecasting time series data. The Box-Jenkins methodology has been explained by many others such as Harvey (1993), Priestely

(1981), Bowerman and O'Connell (1987), Chatfield (2001), Wei (2005) and Liu (2009).

Regarding the multivariate version, the estimation and forecasting, from non-Bayesian viewpoint, of vector autoregressive moving average, denoted by VARMA for short, processes have been extensively studied and investigated by a large number of applied and theoretical researchers. However, it seems that the non–Bayesian literature on the estimation problems of multivariate processes traditionally focused on maximum likelihood methods because of their desirable properties. However, it is well- known that the maximum likelihood methods, which estimate simultaneously the coefficients and covariance matrix, require heavy computational efforts in non-linear optimization procedures. Therefore, there have been extensive investigations in order to ease the maximum likelihood routine and make it faster. For instance, Luceno (1994) developed an efficient numerically expression for the likelihood function of stationary and partially non–stationary autoregressive moving average processes. Another development was done by Mauricio (1995) who proposed a new procedure for the exact maximum likelihood estimation of mixed VARMA models.

The Bayesian analysis of time series processes is being developed and most of the Bayesian contributions have occurred since the last few decades. It was not until Zellner (1971) that the systematic analysis of Bayesian analysis on time series began. With respect to Bayesian literature, one may trace three different approaches to analyze univariate time series processes. The first approach is to use the numerical integration. Monahan (1983) used this approach to implement the identification, estimation and forecasting phases of autoregressive moving average models with low orders. This was the first Bayesian attempt to perform a comprehensive time series analysis and was a very valuable contribution. However, the use of numerical integration is difficult and time consuming especially in the cases of multi-parameters and multivariate models

The second approach is the use of analytical approximation in order to have standard posterior distributions. Several approximations have been developed to solve the estimation and forecasting problems of ARMA processes. However, Zellner (1971) derived the posterior and predictive distribution for the first and second order autoregressive processes using Jeffreys' improper prior density. Newbold (1973) developed an approximate Bayesian analysis for transfer function models of which an ARMA process is a special case using Taylor's expansion for the errors as linear

function of the coefficients. Newbold's results were based on a t-approximation for the posterior distribution. Another important development in the area of autoregressive processes was done by Chow (1975), who found the moments of the joint Predictive distribution of future observations. Harison and Stevens (1976) had given a general review of the time series that can be analyzed by Bayesian approach. Shaarawy and Broemeling (1984) and Broemeling and Shaarawy (1988) have developed Bayesian techniques of identification, estimation, and forecasting phases based on a t–approximation to the posterior distribution of the coefficients. Their approximation has been extended to the case of seasonal models by Shaarawy and El–Shawadfy (1987) and Shaarawy and Ismail (1987). The first study has been extended later to the case of bilinear models by Chen (1992).

The third approach to ease Bayesian time series analysis is to use sampling based methods which include the Gibbs sampler technique (Geman and Geman (1984), data augmentation algorithm (Tanner and Wong (1987)) and the importance sampling algorithm (Rubbin (1988)). Recently, Ismail (2003a, 2003b) used Gibbs sampling algorithm to estimate the coefficients and forecast future observations of multiplicative seasonal autoregressive and moving average processes denoted by SAR(p, P) and SMA(q, Q) respectively. Most recently, Philipp (2006) proposed a Bayesian technique to identify the orders of ARMA processes and estimate their coefficients using MCMC algorithm assuming a prior distribution of the initial values. His algorithm is computationally efficient and not too sensitive to additive outliers. Young and Petti (2006) used Gibbs sampling algorithm to calculate Bayes factors for choosing the order of an autoregressive model and measure the effect of a set of observations on these factors. Finally, Abd-Ellah (2009) extended Ismail's work to case of multiplicative seasonal autoregressive moving average processes.

Regarding the Bayesian analysis multivariate time series, a considerably large part of multivariate time series literature tackles the analytical phases that precede forecasting; namely; identification and estimation. On the other hand, the forecasting phase still lacks analysis. With this in mind, Shaarawy (1989) introduced approximate techniques to estimate the parameters of vector autoregressive moving average, denoted by VARMA for short, processes and predict their future values. Shaarawy (1993) has shown that the same approximation can be used to do a complete analysis for multivariate mixed ARMA processes. However, the numerical properties and the

efficiency of this approximation have never been investigated to forecast future observations of pure bivariate or multivariate moving average processes. The current article has three main objectives. The first one is to is to develop an approximate Bayesian methodology to forecast the future observations of pure vector moving average processes. The second objective is to investigate the numerical efficiency of the proposed Bayesian technique in solving the forecasting problems of bivariate moving average processes by conducting a wide simulation study. The last objective is to study and asses the sensitivity of the numerical efficiency with respect to the parameters values and sample size.

2. VECTOR MOVING AVERAGE PROCESSES

Let { t } be a sequence of integers, $q \in \{1,2,...\}$, $k \in \{2,3,...\}$, 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 independent and normally distributed unobservable random vectors with zero mean and an $k\times k$ unknown precision matrix. Then the vector (multivariate) moving average process of order q, denoted by $MA_k(q)$ for short, is defined for n vectors as

$$y(t)=\theta_q(B)\varepsilon(t) \qquad (2.1)$$

Where

$$y\ (t)=\begin{bmatrix} y(t,1) & y(t,2) & \cdots y(t,k) \end{bmatrix}\ ,$$

$$\theta_q\ (B)=\mathrm{I}_k-\theta_1\mathrm{B}-\theta_2\mathrm{B}^2-\cdots-\theta_q\mathrm{B}^q$$

And

$$\varepsilon\ (t)=\begin{bmatrix} \varepsilon(t,1) & \varepsilon(t,2) & \cdots \varepsilon(t,k) \end{bmatrix}'$$

I_k is the identity matrix of order k, and B is the usual backward shift operator. The $k\times k$ matrix polynomial $\theta_q(B)$, of degree q in the backshift operator, is known as the moving average operator of order q. The process is invertible if all the roots of determinantal equation $|\theta_q(B)|=0$ lie outside the unite circle .

Consider this very important case with moving average coefficients

$$\theta = \theta_1 = \begin{bmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{bmatrix}$$

Then, the model (2.1) can be written as

$$y(t) = (I - \theta B)\varepsilon(t)$$

Where

$$y(t) = [y(t,1) \quad y(t,2)],\ \varepsilon(t) = [\varepsilon(t,1) \quad \varepsilon(t,2)]$$

And

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

Thus, one may write the observations of the $MA_2(1)$ processes in scalar notations as

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

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

$$y(t) = \varepsilon(t) - \theta\varepsilon(t-1), \qquad t=1, 2 \ldots n \qquad (2.3)$$

Where

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

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

Consider another important special case, the $MA_2(2)$ process with moving average coefficients

$$\theta_1 = \begin{bmatrix} \theta_{1.11} & \theta_{1.12} \\ \theta_{1.21} & \theta_{1.22} \end{bmatrix} \text{ and } \theta_2 = \begin{bmatrix} \theta_{2.11} & \theta_{2.12} \\ \theta_{2.21} & \theta_{2.22} \end{bmatrix}$$

Then the model (2.1) can be written as

$$y(t) = (\mathrm{I} - \theta_1 \mathrm{B} - \theta_2 B^2)\varepsilon(t)$$

Where

$$\mathrm{I} - \theta_1 \mathrm{B} - \theta_2 \mathrm{B}^2 = \begin{bmatrix} 1 - \theta_{1.11} B - \theta_{2.11} B^2 & -\theta_{1.12} B - \theta_{2.12} B^2 \\ -\theta_{1.21} B - \theta_{2.21} B^2 & 1 - \theta_{1.22} B - \theta_{2.22} B^2 \end{bmatrix}$$

Thus, the observation of $MA_2(2)$ model can be written in scalar notations as

$$y(t,1) = \varepsilon(t,1) - \theta_{1.11}\varepsilon(t-1,1) - \theta_{1.12}\varepsilon(t-1,2) - \theta_{2.11}\varepsilon(t-2,1) - \theta_{2.12}y(t-2,2)$$

$$y(t,2) = \varepsilon(t,2) - \theta_{1.21}\varepsilon(t-1,1) - \theta_{1.22}\varepsilon(t-1,2) - \theta_{2.21}\varepsilon(t-2,1) - \theta_{2.22}\varepsilon(t-2,2) \quad (2.4)$$

Similarly, the model $MA_2(2)$ can be written in a compact form, using matrix notations, for n observation as

$$y(t) = \varepsilon(t) - \theta_1\varepsilon(t-1) - \theta_2\varepsilon(t-2), \quad t=1, 2 \ldots n \quad (2.5)$$

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

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

In general, one can write the $MA_2(q)$ process as

$$y(t) = \varepsilon(t) - \theta_1\varepsilon(t-1) - \theta_2\varepsilon(t-2) - \ldots - \theta_q\varepsilon(t-q) \quad (2.6)$$

The model (2.6) can be rewritten in a more compact expression as

$$Y = X\Theta + U \quad (2.7)$$

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

$$Y = Y_{n\times 2} = [y(1) \quad y(2) \quad \ldots \quad y(n)]'$$

The matrix X is of order $n \times 2q$ defined by

$$X=\begin{bmatrix} -\varepsilon'(0) & -\varepsilon'(-1) & \dots & -\varepsilon'(1-q) \\ -\varepsilon'(1) & -\varepsilon'(0) & \dots & -\varepsilon'(2-q) \\ \vdots & \vdots & \dots & \vdots \\ -\varepsilon'(n-1) & -\varepsilon'(n-2) & \dots & -\varepsilon'(n-q) \end{bmatrix}$$

Furthermore, Θ is the $2q\times 2$ matrix of coefficients defined as follows:

$$\Theta=\begin{bmatrix} \theta_1' \\ \cdots \\ \theta_2' \\ \cdots \\ \vdots \\ \cdots \\ \theta_q' \end{bmatrix}_{2q\times 2} \quad \text{and} \quad \theta_i=\begin{bmatrix} \theta_{i.11} & \theta_{i.12} \\ \theta_{i.21} & \theta_{i.22} \end{bmatrix}, \ i=1, 2 \ldots, q$$

The class of two dimensional pure moving average models (2.7) is an extremely useful, flexible and practical to model and forecast two correlated time series arise in many areas of application such as business, economics, industry, chemistry, ecology and meteorology.

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

$$y(t)=\varepsilon(t)-\theta_1\varepsilon(t-1)-\theta_2\varepsilon(t-2)-\phi_2\cdots-\theta_q\varepsilon(t-q) \quad ,t=1,2\ldots,n \tag{2.8}$$

The model (2.8) can be written in more compact expression as

$$Y=X\Theta+U \tag{2.9}$$

where Y is a matrix of order n×k with ij-th element equals y(i,j), i=1,2…,n; j=1,2,…,k. . That is

$$Y=Y_{n\times k}=[y(1) \quad y(2) \quad \cdots y(n)]'$$

The matrix X is order n×kq defined by

$$X=\begin{bmatrix} -\varepsilon'(0) & -\varepsilon'(-1) & \cdots & -\varepsilon'(1-q) \\ -\varepsilon'(1) & -\varepsilon'(0) & \cdots & -\varepsilon'(2-q) \\ \vdots & \vdots & & \vdots \\ -\varepsilon'(n-1) & -\varepsilon'(n-2) & \cdots & -\varepsilon'(n-q) \end{bmatrix}$$

Furthermore, Θ is the kq×k matrix of coefficients defined by

$$\Theta=\begin{bmatrix} \overbrace{\theta_1'}^{k\times k} \\ \cdots\cdots \\ \overbrace{\theta_2'}^{k\times k} \\ \cdots\cdots \\ \vdots \\ \cdots\cdots \\ \overbrace{\theta_q'}^{k\times k} \end{bmatrix} \quad where \quad \theta_i=\begin{bmatrix} \theta_{i.11} & \theta_{i.12} & \cdots & \theta_{i.1k} \\ \theta_{i21} & \theta_{i.22} & \cdots & \theta_{i.2k} \\ \vdots & \vdots & & \vdots \\ \theta_{i.k1} & \theta_{i.k2} & & \theta_{i.kk} \end{bmatrix}, \; i=1,2,\ldots,$$

The class of models (2.9) represents the general class of vector pure moving average models of order q and is usually denoted by $MA_K(q)$. This class of models allows the feedback to happen among the variables.

3. AN APPROXIMATE LIKELIHOOD FUNCTION OF THE GENERAL VECTOR MOVING AVERAGE PROCESSES

In order to achieve our main goal, let $S_n=[y(1)\ y(2)...y(n)]'$ be a matrix of $n\times k$ observations generated from k dimensional moving average process of order q of the form (2.8) where the order q is known positive integer. The likelihood function of the parameters Θ and T is

$$L(\Theta,T|S_n)\propto(2\pi)^{\frac{-kn}{2}}|T|^{\frac{n}{2}}exp\left(-\tfrac{1}{2}tr\sum_{t=1}^{n}\varepsilon(t)\varepsilon'(t)T\right) \qquad (3.1)$$

Where $\Theta\in R^{kq*k}$ and $T>0$.

In general the likelihood function (3.1) is very complicated because the disturbance $\varepsilon(t)$ are nonlinear functions of the coefficients θ_i. To see this, one may write the disturbances of (2.8) as

$$\varepsilon'(t) = y'(t) - x'(t-1)\Theta, \quad t=1, 2...n \tag{3.2}$$

Where

$$x'(t-1) = [-\varepsilon(t-1) \quad -\varepsilon(t-2) \quad \cdots \quad -\varepsilon(t-q)\ ]$$

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

$$\varepsilon(t,m) = y(t,m) - \sum_{i=1}^{q}\sum_{j=1}^{k}\theta_{i.mj}\varepsilon(t-i,j), \quad m=1, 2, ...,k \tag{3.3}$$

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 forecasting problems of the general multivariate moving average processes. However, (3.3) can be used to evaluate the residuals recursively if one knows θ_i and the initial values of the residuals. Using (3.3), the likelihood function (3.1) can be written as

$$L(\Theta,T|S_n) \propto (2\pi)^{\frac{-kn}{2}} |T|^{\frac{n}{2}} exp\left(-\tfrac{1}{2}tr\sum_{t=1}^{n}H(\Theta,t)T\right) \tag{3.4}$$

Where $H(\Theta, t) = (h_{rs})$ is $k \times k$ matrix and

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

The form (3.5) is not quadratic in the parameters θ's because $\varepsilon(t-i, j)$ is a function of θ's through the recurrence formula (3.3). If $\varepsilon(t-i, j)$ are known, $H(\Theta, t)$ would be a quadratic form in the parameters. The proposed approximation is based on replacing the exact residuals $\varepsilon(t)$ 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 θ, say θ_0, which minimize the residual sum of squares $\sum_t \varepsilon^2(t,i)$, $i=1, 2, ..., k$. The least squares estimates 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'(t-1), one can write (3.2) as

$$\hat{\varepsilon}'(t) = y'(t) - \hat{x}'(t-1)\Theta \tag{3.6}$$

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

$$L^*(\Theta, T|S_n) \propto |T|^{\frac{n}{2}} exp\left(-\tfrac{1}{2} tr \sum_{t=1}^{n} \left\{ \left[y(t) - \Theta' \hat{x}(t-1) \right] \left[y(t) - \Theta' \hat{x}(t-1) \right]' T \right\}\right) \quad (3.7)$$

An appropriate choice of the conditional prior of Θ given T is

$$g(\Theta|T) \propto |T|^{\frac{kq}{2}} exp\left(-\tfrac{1}{2} tr\left\{ [\Theta - D]' R [\Theta - D] T \right\}\right) \quad (3.8)$$

Where the hyperparameters $D \in R^{kq \times 2}$ and R is a $kq \times kq$ positive definite matrix. The precision matrix T is assigned, a priori, the Wishart distribution

$$g_2(T) \propto |T|^{\frac{a-(k+1)}{2}} exp\left(-\frac{1}{2} tr[\psi T]\right) \quad (3.9)$$

where Ψ is a $k \times k$ positive definite matrix. The joint prior distribution of Θ and T is assumed to be

$$g_3(\Theta, T) \propto g_1(\Theta)|T) g_2(T) \quad (3.10)$$

The class of prior distributions (3.9) is called matrix Normal–Wishart class of distributions.

From (3.8) and (3.9), the joint prior distribution of the parameters Θ and T, equation (3.10), can be written as

$$g(\Theta, T) \propto |T|^{\frac{[kq+a-(k+1)]}{2}} exp\left(-\tfrac{1}{2} tr\left\{ [\Theta - D]' R[\Theta - D] + \psi \right\} T\right) \quad (3.11)$$

If one can't or unwilling to specify the hyperparmeters D, R, a, and ψ, one might use Jeffreys' vague prior

$$g(\Theta, T) \propto |T|^{\frac{-(k+1)}{2}} \quad (3.12)$$

The Jeffreys' prior (3.12) is a special case from the matrix normal-Wishart prior (3.11) when $R=0$ $(kq \times kq)$, $a= -kq$ and $\psi = 0$ $(k \times k)$.

4. THE PREDICTIVE DENSITY OF FUTURE OBSERVATIONS

The $MA_k(q)$ class of models is quite useful in modeling and forecasting the general multivariate time series data arise in many areas of applications. In developing the proposed Bayesian forecasting technique, we will assume that the order q is known. The posterior density of the model parameters Θ and T is the Bayesian tool to estimate the unknown parameters and forecast its future observations. The posterior density $\xi(\Theta, T|S_n)$ is the conditional density of the parameters Θ and T given the observations S_n.

Combining the approximate likelihood function (3.7) and the prior density (3.11), according to Bayes' theorem, one can write the joint posterior distribution of the parameters Θ and T as

$$\xi(\Theta,T|S_n)\propto|T|^{\frac{[n+kq+a-1]}{2}-1}\exp(-\tfrac{1}{2}tr\{\sum_{t=1}^{n}[y(t)-\Theta'\hat{x}(t-1)][y(t)-\Theta'\hat{x}(t-1)]' + \Psi+(\Theta-D)'R(\Theta-D)\}T) \quad (4.1)$$

The exponent in (4.1) can be written as

$$-\tfrac{1}{2}tr\{\Theta'[R+\sum_{t=1}^{n}\hat{x}(t-1)\hat{x}'(t-1)]\Theta-2\Theta'[RD+\sum_{t=1}^{n}\hat{x}(t-1)y(t)]+D'RD+\Psi+\sum_{t=1}^{n}y(t)y'(t)\}T$$

$$=-\tfrac{1}{2}tr\{\Theta'A\Theta-2\Theta'B+C\}T \quad (4.2)$$

Where

$$A=R+\sum_{t=1}^{n}\hat{x}(t-1)\hat{x}'(t-1)\ ,\ B=RD+\sum_{t=1}^{n}\hat{x}(t-1)y'(t) \text{ and}$$

$$C=D'RD+\Psi+\sum_{t=1}^{n}y(t)y'(t)$$

Completing the square of the exponent in (4.2), we will have

$$-\tfrac{1}{2}tr\{(\Theta-A^{-1}B)'A(\Theta-A^{-1}B)+C-BA^{-1}B\}T \quad (4.3)$$

Using (4.3), one may write the approximate posterior density (4.1) as

$$\xi(\Theta,T|S_n)\propto|T|^{\frac{[n+kq+a-1]}{2}-1}\exp(-\tfrac{1}{2}tr\{[(\Theta-A^{-1}B)'A(\Theta-A^{-1}B)+C-B'A^{-1}B]T\}) \quad (4.4)$$

Usually with multivariate time series analysis, the final goal is to forecast future values $Y(n+1)$, $Y(n+2)$... of the process $\{y(t)\}$ given the past observation S_n. The posterior predictive distribution of a future value is a convenient Bayesian

tool to solve the forecasting problems. Of special interest is to forecast the next observation vector *Y(n+1)*. The main goal of this subsection is to drive an approximate one step-ahead predictive density in the case of general multivariate pure moving average processes in a standard form, namely a multivariate t distribution. In addition, it aims to explain how to use this predictive density in constructing a point estimate and highest posterior density (HPD) for the first future vector of observation. Moreover, it explains how to develop an approximate conditional predictive for the m^{th} future vector of observations *Y(n+m)*.

In general the predictive density of *Y(n+1)* is obtained by multiplying the conditional density of *Y(1), Y(2),..., Y(n), Y(n+1)* given the parameters Θ and *T* by the prior density of Θ and *T*; then eliminating the parameters Θ and *T* by integration . The predictive posterior density of *Y(n+1)* can be denoted by $g(y(n+1)\,|\,S_n)$.

Theorem 4.1:

If the approximate likelihood function (3.7) is combined with the joint prior density (3.11), the approximate conditional predictive distribution of $Y'(n+1)$ is a t distribution in *k* dimensions with $\upsilon=(n+a-k+1)$ degrees of freedom, location parameter μ and precision *P* defined as

$$\begin{cases}\mu = EXP\left(Y(n+1)|S_n\right)=\hat{E}'\hat{F}^{-1}\\ P = PRE\left(Y(n+1)|S_n\right)=\hat{F}\left(\hat{L}-\hat{E}'\hat{F}^{-1}\hat{E}\right)(n+a-k+1)\end{cases}$$

The quantities $\hat{F}$, $\hat{E}$ and $\hat{L}$ are defined as

$\hat{F} = 1-\hat{x}'(n)\left[A+\hat{x}(n)\hat{x}'(n)\right]^{-1}\hat{x}'$, $\hat{E}_{1\times 2}=\hat{x}'(n)\left[A+\hat{x}(n)\hat{x}'(n)\right]^{-1}$ and

$\hat{L}_{2\times 2}=C-B'\left[A+\hat{x}(n)\hat{x}'(n)\right]^{-1}B$.

Where

$$\hat{x}'(n)=[\hat{\varepsilon}'(n)\quad \hat{\varepsilon}'(n-1)\quad \cdots\hat{\varepsilon}(n+1-q)]$$

$$=[\hat{\varepsilon}(n,1)\quad \hat{\varepsilon}(n,2)\quad \hat{\varepsilon}(n-1,1)\quad \hat{\varepsilon}(n-1,2)\quad \cdots\hat{\varepsilon}(n+1-q,1)\quad \hat{\varepsilon}(n+1-q,2]$$

With regard to point estimate of $Y'(n+1)$, one may use the posterior expectation $\hat{E}'\hat{F}^{-1}$. One may also use Theorem (4.1) to construct an $(1-\alpha)$ HPD region as follows:

$$R_\alpha(y'(n+1)) = \{y'(n+1) : [y'(n+1) - F^{-1}E'](I - E'F^{-1}E)^{-1}[y'(n+1) - F^{-1}E']F\upsilon \le kF_{\alpha,k,\upsilon}\}$$

Where $F_{\alpha,k,\upsilon}$ is the upper (100 α %) percentage of an F distribution with k and υ degrees of freedom.

Another use of the credible region is to test

H_0: $Y'(n+1) = Y_0'(n+1)$ versus H_1: $Y'(n+1) \neq Y_0'(n+1)$

Where H_0 is rejected at level α if $Y_0'(n+1)$ is not a member of the HPD region. One may also notice that any subset of the vector $Y'(n+1)$ has a multivariate t distribution. Thus, one may also construct an HPD region for any subset of the vector $Y'(n+1)$. Furthermore, may construct an HPD for any element of $Y'(n+1)$ using student t distribution, see Box and Tiao (1973) for the properties of the multivariate t distribution.

On the other hand if one is not sure about the hyperparameters (to express on· s prior information), one may use Jeffreys' vague prior (3.12). If this is the case, the parameters of the predictive density should by revised by letting $R \to 0$ $(kq \times kq)$, $a \to -k(q)$, and $\psi \to 0$ $(k \times k)$.

5. AN EFFECIVENESS STUDY

One of the main objectives of this research is to study the effectiveness of the proposed Bayesian methodology in forecasting problems of multivariate (vector) pure moving average processes. In order to achieve this objective, four simulation studies have been conducted. The proposed forecasting Bayesian methodology is employed to forecast the future observations of $MA_2(1)$ process with various parameters values. All computations were performed on a PC using SCA packages.

The simulation process has the following general design: first, a time series is generated from a given bivariate moving average model of the first order with certain parameters. Second, the generated data are used to evaluate the predictive density of the first future vector of observation $Y(n+1)$. Third, performance criteria are

calculated for the predictive density. Fourth, 500 replications of the above three steps are done. Finally, the output is summarized in tables.

Generally, the generation process begins by generating 500 data sets of bivariate normal variates, each of size 501, to represent the noise $\varepsilon(t)$, which is assumed to follow $N_2(0,V)$ .These data sets are then used recursively to generate 500 realizations, each of size 501, from $MA_2(1)$ process with certain parameters. The initial values of $\varepsilon(t)=[\varepsilon(t,1)\ \ \varepsilon(t,2)]$ are considered to equal their unconditional mean, namely zero. The first 200 observation are deleted in order to remove the initialization effect and hence we get 500 time series each of length 301. From the 301 observation, a bivariate time series of the desired length is used to estimate the predictive density of the first future observation $Y(n+1)=[y(n+1,1)\ \ y(n+1,2)]'$ using the proposed Bayesian methodology. In our simulation studies, the time series length are taken to be 30, 50, 100, 150, 200, and 300. Each simulation study correspond to specific mean, coefficients and covariance matrix of the noise. The selected coefficients are selected to represent different positions in the invertibility domain of the $MA_2(1)$ model. It might be important here to emphasis that the Jefferys' non-informative prior is used to conduct all the simulation studies and that the computer program used to execute the simulations is SCA package.

Our main concern is to study the numerical efficiency of the proposed Bayesian forecasting methodology by calculating two groups of criteria: The moments' group and discrepancy measures' group. The moments' group includes the posterior mean and variance of the future observations. The discrepancy measures' group measures the closeness of the estimated observations to the exact ones; this group contains the measures P^* , MAD and MAPE. The measure P^*checks the goodness of interval forecasts calculated form a specified predictive density. Defining 95% highest predictive Density (HPD) region as the interval having probability 0.95 centered at the mean of the predictive density. The percentage P^* of time series for which the HPD region of the predictive density contains the true future observation is defined as

$$P^* = (n^*/500)*100$$

where n^* is the number of time series where the HPD region includes the true value. P^* is evaluated such that the higher the value P^*, The better the performance of the

predictive density in forecasting. It should be noted that according to P^* a certain observation may be ruled as belonging to the HPD region or not. However, P^* does not account the distance of this observation from the center of the region or its boundaries. Therefore, two other measures, named as MAD and MAPE are provided. The MAD stands for the mean absolute deviation of the future observation from the location parameter and is defined by

$$\text{MAD}= \sum_{j=1}^{500} | y_j(n+1,k) - E_j(k) | / 500 \quad , k = 1,2$$

Where $y_j(n+1,k)$ and $E_j(k)$ are the k^{th} component of the first future observation and location parameter, respectively, of the j^{th} simulated series. The MAPE stands for the mean absolute percentage deviation of the future observation from the location parameter and is defined as

$$\text{MAPE} = \left(\sum_{j=1}^{500} | \frac{y_j(n+1,k) - E_j(k)}{y_j(n+1,k)} | / 500 \right), k = 1, 2$$

Where $y_j(n+1,k)$ and $E_j(k)$ are defined as above.

The numerical efficiency of the proposed forecasting procedure will be examined with respect the time series length (n) as well as the parameters of the selected model. Simulation I, as an illustration, begins with the generation of 500 data sets of bivariate normal varieties, each with 600 observations 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 300, from $VMA_2(1)$ process with

$$\mu = 10, \ \Theta = \begin{bmatrix} 0.9 & -0.2 \\ 1.1 & -0.9 \end{bmatrix} \text{ and } Cov(\varepsilon) = \begin{bmatrix} 0.9 & -0.2 \\ 1.1 & -0.9 \end{bmatrix}$$

Assuming the starting values are zero and Jefferys' prior, the second step is to carry out all computations required to estimate the predictive density of the first next observation $Y(n+1)$ of each of the 500 realizations and compute the P^*, MAD and MAPA values. Such computations are done for a specific time series length using the first n observations of each generated set. This second step is repeated for each chosen time series length. The results of Simulation I are summarized in table (5.1) which consists of two parts. The first part consists of six rows and four columns. Each row corresponds to a time series length, and the four columns are reserved to the corresponding variance, MAD, MAPE and P^* values for $Y(n+1, 1)$. The second part is

reserved to $Y(n+1,2)$ an explained in similar way to the first part. Simulations II, III and IV are designed in a similar manner but using different coefficients, and their results are reported in tables (5.2), (5.3) and (5.4) respectively.

Table (5.1) $\mu = 10, \Theta = \begin{bmatrix} 0.9 & -0.2 \\ 1.1 & -0.9 \end{bmatrix}, COV(\varepsilon) = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$

	LENGTH	$V[Y(n+1,1)]$	P^*	MAPE	MAD
$Y(n+1,1)$	30	1.6968	93.6	0.1491	1.3877
	50	1.5543	89.4	0.1606	1.4980
	100	1.4636	93.6	0.1314	1.2319
	150	1.4436	94.2	0.1205	1.1276
	200	1.4311	94.2	0.1290	1.1863
	300	1.4250	94.8	0.1254	1.1718
	LENGTH	$V[Y(n+1,2)]$	P^*	MAPE	MAD
$Y(n+1,2)$	30	1.2052	94.0	0.0985	0.9283
	50	1.1027	89.4	0.1111	1.0618
	100	1.0353	93.6	0.0894	0.8596
	150	1.0190	94.2	0.0855	0.8227
	200	1.0103	94.2	0.0836	0.8005
	300	1.0063	94.4	0.0840	0.8076

Table (5.2) μ=100, $\Theta = \begin{bmatrix} 0.5 & -0.4 \\ -0.3 & 0.2 \end{bmatrix}, COV(\varepsilon) = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$

	LENGTH	$V[Y(n+1,1)]$	P^*	MAPE	MAD
$Y(n+1,1)$	30	1.5606	94.8	0.0123	1.2308
	50	1.4839	87.4	0.0176	1.7590
	100	1.4436	92.6	0.0141	1.4122
	150	1.4335	93.8	0.0130	1.3018
	200	1.4288	92.6	0.0186	1.8495
	300	1.4238	94.4	0.0117	1.1646
	LENGTH	$V[Y(n+1,2)]$	P^*	MAPE	MAD
$Y(n+1,2)$	30	1.0535	93.6	0.0084	0.8381
	50	1.0290	87.2	0.0118	1.1800
	100	1.0146	93.4	0.0096	0.9553
	150	1.0095	93.0	0.0090	0.9031
	200	1.0074	92.6	0.0113	1.1258
	300	1.0047	94.8	0.0081	0.8074

Table (5.3) μ=1000, $\Theta = \begin{bmatrix} 0.2 & 0.3 \\ -0.6 & 1.1 \end{bmatrix}, COV(\varepsilon) = \begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix}$

	LENGTH	$V[Y(n+1,1)]$	P^*	MAPE	MAD
$Y(n+1,1)$	30	2.2501	95.6	0.0018	1.7591
	50	2.1585	95.2	0.0017	1.7034
	100	2.1081	94.0	0.0017	1.6664
	150	2.0931	95.4	0.0017	1.6505
	200	2.0835	94.6	0.0017	1.7274
	300	2.0771	95.2	0.0017	1.6827
	LENGTH	$V[Y(n+1,2)]$	P^*	MAPE	MAD
$Y(n+1,2)$	30	1.2715	96.2	0.0009	0.8965
	50	1.1994	96.2	0.0009	0.9153
	100	1.1535	95.0	0.0009	0.9294
	150	1.1400	93.6	0.0009	0.9210
	200	1.1348	95.4	0.0009	0.8945
	300	1.1297	95.0	0.0009	0.9029

Table (5.4) μ=10000, $\Theta = \begin{bmatrix} 0.2 & -0.6 \\ 0.3 & 1.1 \end{bmatrix}, COV(\varepsilon) = \begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix}$

	LENGTH	$V[Y(n+1,1)]$	P^*	MAPE	MAD
$Y(n+1,1)$	30	2.0961	91.0	0.0002	1.7824
	50	2.0719	92.0	0.0002	1.8044
	100	2.0451	93.8	0.0002	1.6658
	150	2.0310	94.4	0.0002	1.6478
	200	2.0224	94.6	0.0002	1.6547
	300	2.0169	94.4	0.0002	1.6442
	LENGTH	$V[Y(n+1,2)]$	P^*	MAPE	MAD
$Y(n+1,2)$	30	1.2902	95.4	0.0001	1.0210
	50	1.1741	93.8	0.0001	0.9703
	100	1.0742	94.8	0.0001	0.8424
	150	1.0488	94.6	0.0001	0.8347
	200	1.0327	95.4	0.0001	0.8024
	300	1.0246	93.8	0.0001	0.8160

Inspection of the numerical results shows that the percentages P^* are very high, being greater than 92%, for sample size 100 or more no matter what the coefficients and covariance matrices are. In addition, the numerical results shows that the numerical efficiency increases as the time series length n increases for all selected parameters. The MAPE values are reasonably small, being less than 14%, for sample size 100 or more no matter what the coefficients and covariance matrices are. Moreover, the variances of the first future observations and MAD values decrease as the length n increases for all selected parameters.

Considering the above comments, one may say that the numerical results support the adequacy of using the proposed Bayesian procedure in forecasting the future observations of vector MA processes for small, moderate and large sample size.

6. SUMMARY AND CONCLUSIONS

The main objective of the article is to develop a convenient Bayesian technique to forecast the future observations of pure vector moving average processes. The predictive density of the first future vector of observations is developed in a convenient form using an approximate likelihood function and a matrix – normal Wishart prior density or Jeffreys' vague prior density. Then one may easily calculate the posterior expectation, precision and HPD region of the future observation.

In order of demonstrate the performance of the proposed Bayesian procedure and test its adequacy in solving the forecasting problems of bivariate processes, four simulation studies with Jeffreys' vague prior density have been conducted. The analysis of the numerical results indicates that the proposed Bayesian procedure can efficiently forecast future observations of $MA_2(1)$ processes with high precision for moderate and large time series length.

REFERENCES

(1) Abdel-Ellah, A. (2009). Bayesian Inference for Seasonal ARMA Models: A Gibbs Sampling Approach. Unpublished M.Sc. Thesis. Dept. of Statistics, Faculty of Economics and Political science, Cairo University.

(2) Bowerman, B and O'Connell (1987). Time Series Forecasting: Unified Concepts and Computer Implementation. Boston: PWS Publishers, Duxbury Press

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

(4) Box, G., Jenkins, G., Reinsel, G. (2008). Time Series Analysis, Forecasting and Control. 4th ed. New York: Wiley.

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

(6) Brockwell, P., J. and Davis, R. (1991). Time Series: Theory and Methods. 2nd edition, Springer-Verlag.

(7) 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.

(8) Chatfield, C. (2001). The Analysis of Time Series: Theory and Practice. London: Chapman and Hall.

(9) Chen, C. (1992)." Bayesian Inferences and Forecasting in Bilinear Time Series Models", Communication in Statistics – Theory and Methods, 21, pp. 1725-1743.

(10) Chow, G. (1975)." Multi-Period Prediction from Stochastic Difference equations by Bayesian Methods", Chapter 8 of Studies in Bayesian Econometrics and Statistics, edited by S.E. Fienberg and A. Zellner. North-Holland, Amsterdam.

(11) Geman, S. and Geman, D. (1984). "Stochastic relaxation and the Bayesian restoration of images", IEEE Transactions on Pattern Analysis and Machine Intelligence, 6(6), 721-741., Gibbs distributions stationary and partially nonstationary vector autoregressive -moving average processes. Biometrika, 3, 555-565.

(12) Granger, C. and Newbold, P. (1973). Forecasting Economic Time Series. New York: Academic Press

(13) Harison, P. and Stevens, C. (1976). "Bayesian Forecasting (with discussion)", Journal of the Royal Statistical Society, Series B, 38, 205-242.

(14) Harvey, A. C. (1993).Time Series Models. 2nd ed. Cambridge, MA: The MIT Press.

(15) Ismail, M. (2003a)."Bayesian Analysis of Seasonal Autoregressive Models", Journal of Applied Statistical Science, 12(2), pp. 123-136.

(16) Ismail, M. (2003b). Bayesian Analysis of Seasonal Moving Average Model: A Gibbs Sampling Approach. Japanese Journal of Applied Statistics, 32(2), pp. 61-75

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

(18) Luceno, A. (1994). "A fast algorithm for the exact likelihood of stationary and partially nonstationary vector autoregressive -moving average processes. Biometrika, 81, No.3, 555-565.

(19) Mauricio, J. A. (1995). "Exact maximum likelihood estimation of stationary vector ARMA models. Journal of the American Statistical Association, 90, 282-291

(20) Monahan, J. F. (1983). Fully Bayesian Analysis of ARIMA Time Series Models, Journal of Econometrics, 21, 307–331.

(21) Newbold, P. (1973). "Bayesian Estimation of Box and Jenkins Transfer Function Model for Noise Models", JRSSB, 35(2), 323-336.

(22) Philippe, A. (2006). "Bayesian analysis of Autoregressive Moving Average Processes with Unknown Orders", Computational Statistics and Data Analysis, 15, 1904-1923.

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

(24) Rubbin, D. B. (1988). "Using the SIR Algorithm to Simulate the Posterior Distribution", Studies in Bayesian Statistics 3 (eds. Bernardo, J. M. et. al.), Oxford university press, pp. 395-402.

(25) Shaarawy, S. M. and El-Shawadfy, G.A. (1987). "Approximate Bayesian densities of future observations for seasonal models". The 22nd Annual Conference on Statistics, Computer Science and Operation Research, Cairo University, 2, 160-178.

(26) Shaarawy, S. (1989). "Bayesian Inferences and Forecasts with Multiple ARMA Models", Communications in Statistics-Theory and Methods, 18, 1481–1509.

(27) Shaarawy, S. and Broemeling, L. (1984). Bayesian Inference and Forecasts with Moving Average Processes. Communications in Statistics- Theory and Methods, 13, Issue 15, pp. 1871-1888.

(28) Shaarawy, S. M. (1993). Multivariate Time Series: A Complete Bayesian Analysis in Time Domain. The 28th Annual Conference on Statistics, Computer Science and Operation Research, Institute of Statistical Studies and Research, Cairo University.

(29) Shaarawy, S. and Ismail, M. (1987). "Bayesian inference for seasonal ARMA models". The Egyptian Statistical Journal, Institute of Statistical Studies and Research, Cairo, Egypt.

(30) Tanner, M. A. and Wong, W. H. (1987). "The Calculation of Posterior Distributions by Data Augmentation. Journal of the American Statistical Association, 82, pp.528–550.

(31) Tiao, G. C., Box, G. E. P. (1981). Modeling Multiple Time Series with Applications. Journal of the American Statistical Association, 76, 802–816

(32) Wei, W. W. S. (2005). Time Series Analysis: Univariate and Multivariate Methods. Reading, MA: Addison Wesley.

(33) Young, K. , d. s. and Petti, L. I. (2006). "The effect of observation on Bayesian choice of an autoregressive model", Journal of Time Series Analysis, 27 (1), 41-50.

(34) Zellner, A. (1971). An Introduction to Bayesian Inference in Econometrics. John Wiley and Sons, New York.

(33) Young, K. , d. s. and Petti, L. I. (2006). "The effect of observation on Bayesian choice of an autoregressive model", Journal of Time Series Analysis, 27 (1), 41-50.

(34) Zellner, A. (1971). An Introduction to Bayesian Inference in Econometrics. John Wiley and Sons, New York.