SCOPUA Journal of Applied Statistical Research (JASR)

Vol.2, Issue 1, March 2026
DOI: https://doi.org/10.64060/JASR.v2i1.2

Assessment of the performance of the Bayesian Forecasting for Vector ARMA Processes

Emad E. A. Soliman [1*], Sherif S. Ali [1]

[1]Department of Statistics, King Abdulaziz University, Jeddah 21589, Saudi Arabia

* Corresponding Email: ehegazy@kau.edu.sa

Received: 29 November 2025 / Revised: 03 December 2025 / Accepted: 04 December 2025 / Published online: 10 December 2025

ABSTRACT

Multivariate time series are widely observed in numerous domains. In economics, for example, you can monitor the annual savings in conjunction with the real interest rate. Such variables are jointly analyzed to understand the dynamic interactions that exist between them, thus improving the precision of forecasts. Enhanced forecasting is achievable when the series are examined together, especially when one series holds information about another one. An approximate Bayesian analytical method to estimate and forecast vector autoregressive moving average (Vector ARMA) processes was introduced by Shaarawy (1989). A basic goal for the research in hand is the numerical assessment of the proposed approach in tackling forecasting difficulties associated with Vector ARMA processes through a comprehensive simulation study. Furthermore, the research checks how the performance of the suggested method fluctuates when varying parameter values and time series lengths. The findings of the numerical study demonstrated that the methodology was effective in accurately forecasting future observations for Vector ARMA processes across various values of the parameter and different time series lengths.

Keywords*:* Vector autoregressive moving average processes; Multivariate t distribution; Bayesian forecasting; Prediction; Matrix Normal-Wishart prior; Jeffreys' prior

1. Introduction

Multivariate time series are encountered in various areas of scientific and applied research. Regarding economics, researchers can investigate the relationships among production cost, product's sales price, sales volume, and advertising expenses using a four-dimensional framework. In meteorology studies, a bivariate technique may be used to estimate daily maximum temperatures in addition to humidity. According to the works of Tiao and Box (1981), Liu (2009), Box et al. (2016), Chatfield (2019), and Wei (2019), the above mentioned and other examples are examined and forecasted collectively using multivariate models, which seeks to reveal the interaction between different phenomena in order to improve forecasts' goodness.

In a variety of application areas, one encounters several time series data sets that can be successfully analyzed, modeled, and forecasted using the class of multivariate (vector) autoregressive moving average models, or VARMA models. The ultimate goal of this analysis is forecasting (or prediction), which is thought to be the most essential stage in the analysis of time series. The used prediction approach determines how reliable the forecasts are.

Without a doubt, VARMA class of models is the best group for examining multivariate (vector) time series that appear in different domains because of a variety of causes. First of all, according to Lütkepohl (2007), this category is the most frugal. Second, as noted by Lütkepohl (2007), it stays a closed group under both linear transformation and marginalization. Thirdly, this category's estimations and predictions are more accurate than those made by other models such as VAR models (see Raghavan et al. (2009) and Kascha (2012)). Finally, compared to the pure VAR category, it is more consistent with economic theory (see Cooley and Dwyer (1998)).

In spite of these benefits, there are two main reasons why VARMA processes are scarcely applied in real-world. The likelihood function is analytically intractable due to the model errors' nonlinearity in the model's coefficients; as a result,

both parameter estimation and the prediction of future observations require numerical methods. Therefore, even with high-speed computers, these procedures become more difficult as the sample size grows. The second explanation has to do with how complex the VARMA models' identification phase is. The prevalence of pure vector autoregressive models, which facilitate identification, estimation, and forecasting, has been influenced by these two considerations.

Regarding the Bayesian point of view, the analysis of VARMA models' remains relatively obscure. Analytical Bayesian techniques were first used by Shaarawy (1989) to predict upcoming observations of the model. A multivariate t-distribution is used to create an approximate posterior predictive distribution for upcoming observations of VARMA processes. This article's main goal is to examine the Bayesian approximate prediction methodology, which was first presented by Shaarawy (1989), and its numerical effectiveness. There are a number of reasons to investigate this approach. First, it is the only analytical Bayesian method that we are aware of in the literature. The second is that this approach is simply automated. Third, the approach is based on multivariate t-distribution, which has well-known statistical properties explained by Box and Tiao (1973). Fourth, although this methodology is highly relevant, neither its theoretical nor numerical properties have been thoroughly studied before. Finally, this approach can support a complete Bayesian analysis of VARMA processes, alike the one suggested by Broemeling and Shaarawy (1988) for ARMA processes. Comprehensive Monte Carlo simulations are applied to assess the approach numerically. For bivariate ARMA processes, the numerical goodness will be evaluated with respect to the values of the parameters and the sample size. The numerical analysis will expect the smallest time series length required to produce accurate forecasts.

In several disciplines, including economics, finance, and environmental studies, time series data have been successfully modeled using ARMA processes. Box et al. (2016) have provided comprehensive theoretical and methodological manipulations for these processes. Box and Jenkins methodology is thought to be the most beneficial non-Bayesian method for modelling time series data using ARMA process through the four steps: identification, estimation, diagnostic checking, and forecasting. Tiao and Box's (1981) expansion of this methodology to include multivariate time series is a major advantage. Several academics have explained this methodology, such as Wei (2005) and Chatfield (2019).

Since numerous applied researchers, especially in the fields of environmental studies and economics, have not widely adopted the statistical inference classical interpretation in the domain of time series, the Bayesian approach to prediction of univariate time series has gained acceptance in recent decades. In 1971, Zellner studied the first and second order autoregressive (AR) models and achieved their posterior and predictive distributions using Jeffreys' improper prior density. Monahan (1983) modelled and forecasted ARMA models using a numerical integration technique. Numerical Bayesian time series analysis has been made easier over the last three decades by the use of the Gibbs sampling algorithm in the Markov Chain Monte Carlo simulation (MCMC). Readers are referred to the publications of Chib and Greenberg (1994), Marriott et al. (1996) and Amin (2019) for additional information on MCMC approaches.

Analytically speaking, it is widely recognized that AR processes can be efficiently inferenced and forecasted using a posterior analysis based on the normal gamma prior; the main difficulty, however, is in accomplishing an investigation of moving average (MA) and mixed ARMA processes analytically. Analytically intractable, the likelihood function's complexity is the main obstacle to evaluating these models. As a result, or any given prior distribution both the posterior density and the predictive density do not fit standard forms, and both estimation and prediction must be done numerically. Without developing an approach to display the likelihood function in an appropriate way that produces posterior and predictive densities that are analytically tractable, Bayesian prediction of ARMA models would be analytically impossible. Accordingly, the use of Taylor's extension for the disturbances as linear functions in the model’s coefficients was done by Newbold (1973), who applied a Bayesian analysis for transfer function models approximately. Note that, ARMA processes are particular examples of transfer function models. The foundation for Newbold's conclusions is approximating the posterior distribution by t-distribution. Similar to this, Zellner and Reynolds (1978) used the non-linear least squares method (NLSEs) to approximate the errors and expanded their sum of squares as a quadratic function in the model’s coefficients through Taylor's expansion in order to approximate the posterior distribution of the coefficients in ARMA processes by t-distribution. A numerical method for carrying out the identification, estimation and forecasting processes for ARMA models with low-orders was developed by Monahan (1983). This was a significant contribution to the field as it was considered the first attempt to employ a full Bayesian numerical analysis of time series. By creating an analytical approximate complete analysis of ARMA models using the Bayesian approach based on both multivariate and univariate t-distributions, Broemeling and Shaarawy (1988) carried the discussion even further. Their approach was easy to use and analytically tractable. In order to identify MA processes, Shaarawy et al. (2007) took advantage of their methodology. Finally, Broemeling and Shaarawy's approach was extended by Amin (2019) to include double seasonal AR processes.

Since the likelihood function is complex, many theoretical statisticians and a small number of practitioners have attempted to solve the prediction problems of VARMA models from a non-Bayesian point of view. A methodology to employ a thorough analysis of VARMA models using a non-Bayesian approach was proposed by Tiao and Box (1981). Maximum Likelihood Estimates (M.L.E.s) are divided into two types, exact estimates and conditional ones, which are comparable to

the univariate case. The exact M.L.E.s were proposed by Tiao and Box (1981), Hillmer and Tiao (1979), Mauricio (1995), and others, whereas, Tunnicliffe (1973) introduced conditional ones.

However, it is easily observable that the prediction problems with VARMA processes are too difficult and un-useful computationally. As a result, several studies attempted to facilitate the computations for the corresponding likelihood function and obtained M.L.E.s. These include works by Hall and Nicholls (1980), Shea (1987), Mauricio (1995) and others.

As noted earlier, the application of Bayesian analysis to VARMA models has not been extensively investigated. The majority of contributions from Bayesian methods to prediction challenges have focused on AR models. Readers interested in vector AR can consult works of Koop (2013) and Ravishanker and Ray (1997). In their analysis of VARMA models, Marriott et al. (1996) applied sampling-based methods. A Gibbs sampler technique was developed by Chan and Eisenstat (2015) for VARMA processes. To the best of our knowledge, Shaarawy's (1989) analytical method is the only one method identified in the time series literature that emphasizes the prediction stage of VARMA processes. A multivariate t-distribution was used, which has good qualities, to approximate the posterior predictive density of the vector of upcoming observations. Box and Tiao (1973) provided a comprehensive analysis of the multivariate t-distribution's properties. The ultimate goal of the current work is to evaluate the efficiency of Shaarawy's approximate technique numerically, as neither its theoretical nor numerical characteristics have been thoroughly examined.

Ali (2015) investigated how well Shaarawy's Bayesian approach predicts upcoming multivariate moving average process data. Shaarawy (2023) identified, estimated, and predicted a number of numerical examples for VARMA models using his approximate Bayesian approach. Efficiency of Shaarawy's Bayesian approach for the estimation of VARMA parameters was examined by Albassam et al. in 2023. Shaarawy et al. (2024) developed the Bayesian identification technique of seasonal VARMA models.

This article is the first attempt to study the numerical effectiveness, using a large-scale simulation study, of the Bayesian prediction technique proposed by Shaarawy (1989). No previous study has manipulated this aspect.

The following is the format of the sections that follow: The VARMA models are described in Section 2, along with the main presumptions that must be made in order to develop the suggested prediction methodology related to these models. The estimated posterior predictive density for upcoming observations of VARMA processes is presented in Section 3. It also describes the procedure for producing point and interval predictions for future data. The numerical effectiveness of the suggested approach in resolving the prediction issues related to bivariate autoregressive moving average processes is examined in Section 4. Finally, the study's findings and conclusions are summarized in Section 5.

2. Describing VARMA Processes

Several references in the time series literature have defined the VARMA processes. Yet, we will follow Albassam et al. (2023) in their definition of the process. Given a set of random vectors $\{y(t)\}$ of real observable values with dimension $k\times1$, and a set of unobservable random vectors $\{\varepsilon(t)\}$ with dimension $k\times1$ which is independent and normally distributed having a mean zero and an unknown precision matrix with dimension $k\times k$. Where k is the dimension of the multivariate time series, such that, $k \in \{2,3,\dots\}$. They also defined t, p and q to be sequences of positive integers. In addition, assume that the $k\times k$ matrices of unknown constants ϕ_i (i=1, 2,…, p) are called autoregressive coefficients, and the $k\times k$ matrices of unknown constants θ_i (i=1, 2,…, q) are called moving average coefficients. Thus, the VARMA$_k$(p, q) might be written as follows

$$\Phi_p(B)\, y\,(t) = \Theta_q(B)\, \varepsilon\,(t) \tag{1}$$

Where, the symbols in equation (1) above can be defined as follows,

$$\Phi_p(B) = I_k - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p, \qquad y(t) = [y(t,1) \quad y(t,2) \quad \dots \quad y(t,k)]',$$
$$\Theta_q(B) = I_k - \Theta_1 B - \Theta_2 B^2 - \dots - \Theta_q B^q \text{ and } \varepsilon(t) = [\varepsilon(t,1) \quad \varepsilon(t,2) \quad \dots \quad \varepsilon(t,k)]'.$$

The squared matrix with order k, I_k is well known as the identity matrix. The symbol B represents the, so called, backward shift operator. $\Phi_p(B)$ is a $k\times k$ matrix polynomial having degree p in B, the backward shift operator, and called an autoregressive operator of order p, whereas, $\Theta_q(B)$ is a $k\times k$ matrix polynomial having degree q in B, called a moving average operator of order q. VARMA process is considered stationary and invertible when the roots of both the two determinantal equations $|\Phi_p(B)| = 0$ and $|\Theta_q(B)| = 0$, all lie outside the unite circle, respectively.

VARMA$_k$(p, q) model defined in equation (1) can be written in its general form after conditioning the process on the earliest p vectors $y(t)$ of the observations matrix Y,

$$Y = X\,\Gamma + U \tag{2}$$

Such that, the matrix Y is of order $(n\text{-}p)\times k$, where, its ij-th element is $y(p+i, j)$, given that, i = 1, 2…, n-p; and j = 1, 2, …, k. i.e.

$$Y = Y_{(n-p)\times k} = [\, y(p+1) \quad y(p+2) \quad \dots \quad y(n)\,]' \tag{3}$$

Moreover, X is the regressors matrix with order $(n\text{-}p)\times kh$, where, $h = p + q$, has the form

$$X_{(n-p)\times kh} = [X_1 \quad X_2] \tag{4}$$

Where

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

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

In addition, the coefficients matrix Γ has the order $kh \times k$ can be defined as follows

$$\Gamma = \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} \quad , \tag{5}$$

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

where

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

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

3. Posterior and Predictive Treatment for VARMA$_k$(p,q) Models

Based on a VARMA$_k$(p,q) model as defined in equation (2) and having n vectors $y = [y(1) \quad y(2) \quad \dots \quad y(n)]'$ of the time series observations, Albassam et al. (2023) defined the corresponding n vectors of residuals as follows

$$\varepsilon'(t) = y'(t) - X'(t-1)\Gamma \ , \ t = 1, 2, \dots, n \tag{6}$$

Where the regressors vector has the form,

$$X'(t-1) = [y'(t-1) \quad y'(t-2) \quad \dots \quad y'(t-p) \quad -\varepsilon'(t-1) \quad -\varepsilon'(t-2) \quad \dots \quad -\varepsilon'(t-q)]$$

The following analysis is Conditional on the values of the earliest p vectors $y(t)$ of observations and on an assumption that $\varepsilon(p) = \varepsilon(p-1) = \dots = \varepsilon(p-q-1) = 0$ (for more details see Tiao and Box 1981). Based on the assumptions above, one can define for the model parameters Γ and T a likelihood function as follows

$$L(\Gamma, T|Y) \propto |T|^{\frac{n-p}{2}} exp\left(-\frac{1}{2} tr\{\sum_{t=p+1}^{n} \varepsilon(t)\varepsilon'(t)T\}\right), \Gamma \in R^{k(p+q)\times k} \quad \text{and } T > 0 \tag{7}$$

Regarding equation (6), it is considered a recurrence relation in the errors vectors, and the m-th element of the errors' vector $\varepsilon(t)$ can be defined as,

$$\varepsilon(t,m) = y(t,m) - \sum_{i=1}^{p}\sum_{j=1}^{k} \varphi_{i.mj} y(t-i,j) + \sum_{i=1}^{q}\sum_{j=1}^{k} \theta_{i.mj}\varepsilon(t-i,j)$$

$t = 1,2,\dots,n$; $m = 1,2,\dots,k$ (8)

Generally speaking, equation (7) is sophisticated and analytically intractable. This is due to the recurrence equation (8). According to equation (8) the error's vector $\varepsilon(t)$ appears to be a nonlinear function in the coefficients' matrix Γ. This fact is the main reason of the problem encountered in developing an exact manipulation for VARMA models. Yet, the recurrence formula can be employed to estimate the errors recursively if the matrix Γ is known and one can determine the errors' initial values. A Bayesian approximation, suggested by Shaarawy (1989), relies on substituting the exact errors with their estimates using the least squares method, while presuming that initial errors are equivalent to their unconditional means, which is zero. That is, the errors are estimated recursively using the form

$$\hat{\varepsilon}(t,m) = y(t,m) - \sum_{i=1}^{p}\sum_{j=1}^{k} \hat{\varphi}_{i.mj} y(t-i,j) + \sum_{i=1}^{q}\sum_{j=1}^{k} \hat{\theta}_{i.mj}\hat{\varepsilon}(t-i,j)$$

$t = 1,2,\dots,n$; $m = 1,2,\dots,k$ (9)

The symbols $\hat{\varphi}_{i.mj}$ in addition to $\hat{\theta}_{i.mj}$ stand for the non-linear least squares estimates for the coefficients $\varphi_{i.mj}$ and $\theta_{i.mj}$. Depending on the residuals, the estimates of the errors, one can approximate the conditional likelihood function (CLF) by

$$L^*(\Gamma, T|Y) \propto |T|^{\frac{n-p}{2}} exp \ \left(-\frac{1}{2} tr\{\sum_{t=1}^{n} [y(t) - \Gamma'\hat{X}(t-1)]\,[y(t) - \Gamma'\hat{X}(t-1)]'\, T\}\right) \tag{10}$$

Such that the matrix $\hat{X}(t-1)$ is the estimate of $X(t-1)$ obtained by substituting the exact errors by the corresponding residuals.

On the other hand, Shaarawy (1989) suggested that a suitable selection for, the model's parameters Γ and T, joint prior density is a matrix normal-wishart prior defined as follows

$$\xi(\Gamma, T) = \xi_1(\Gamma|T)\,\xi_2(T) \tag{11}$$

Where

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

And

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

Parameters of the matrix normal-wishart density, called hyper-parameters, are defined as follows,

D is a matrix of real numbers with order $kh \times k$, the positive definite matrix W is a $kh \times kh$ one with $h = p+q$, Ψ is a positive definite matrix with order $k \times k$ and, finally, the constant a is a positive scalar. Moreover, Jeffreys' vague prior can be used in the cases where little or no information are available regarding the parameters beforehand (see Shaarawy (1989))

$$\xi(\Gamma, T) \propto |T|^{\frac{-(k+1)}{2}}, \quad \Gamma \in R^{kh \times k}, \quad T > 0 \tag{12}$$

Depending on the above-mentioned likelihood function given in (10), and both the matrix normal-Wishart prior (11) and the Jeffreys' vague prior (12), the following two theorems and corollaries has been proved by Shaarawy (1989):

Theorem 3.1: Combining the approximate CLF in (10) and the matrix normal-Wishart prior in (11), one obtains, for the coefficients matrix Γ, a matrix-t marginal posterior density function. The parameters of the distribution are ($A^{-1}B$, A^{-1}, C-$B'A^{-1}B$, υ) such that

$$A = W + \sum_{t=p+1}^{n} \hat{X}(t-1)\hat{X}'(t-1), \qquad B = WD + \sum_{t=p+1}^{n} \hat{X}(t-1)y'(t),$$
$$C = D'WD + \Psi + \sum_{t=p+1}^{n} y(t)y'(t) \text{ and } \qquad \upsilon = n - p + a - k + 1$$

In addition, for the parameter T, the marginal posterior density function is a Wishart distribution. The parameters of it are ($n + a$, C-$B'A^{-1}B$).

Corollary 3.1: Combining the approximate CLF in (10) and Jeffreys' prior in (12), one obtains, for the coefficients matrix Γ, a matrix-t marginal posterior density function. The parameters of the distribution are ($A^{-1}B$, A^{-1}, C-$B'A^{-1}B$, υ). yet, A, B, C and υ would be adjusted as follows:

$a \rightarrow -kh$, $W \rightarrow 0$ with order ($kh \times kh$) and finally $\Psi \rightarrow 0$ with order **(**$k \times k$**)**. In addition, for the parameter T, the marginal posterior density function would be a Wishart distribution. The parameters of it are ($n + a$, C-$B'A^{-1}B$).

Theorem 3.2: Combining the approximate CLF in (10) and the matrix normal-Wishart prior in (11), one obtains the posterior predictive distribution of $Y(n+1)$ which has the form of multivariate t-distribution of dimension k having degrees of freedom = υ, location parameter = $F^{-1}E'$, and precision matrix = $F[L - E'F^{-1}E]^{-1}\upsilon$. Such that,

$$F = 1 - \hat{x}'(n)[A + \hat{x}(n)\hat{x}'(n)]^{-1}\hat{x}(n),$$
$$E = \hat{x}'(n)[A + \hat{x}(n)\hat{x}'(n)]^{-1}B, \text{ and}$$
$$L = C - B'[A + \hat{x}(n)\hat{x}'(n)]^{-1}B,$$

Where, the quantities A, B, C and υ were previously defined in theorem 3.1 above.

Corollary 3.2: Combining the approximate CLF in (10) and Jeffreys' prior (12), one obtains the posterior predictive distribution for $Y(n+1)$ which has the form of multivariate t-distribution of dimension k with degrees of freedom = υ, location parameter = $F^{-1}E'$ and precision matrix $F[L - E'F^{-1}E]^{-1}\upsilon$.

Such that, F, E and L are as defined in theorem 3.2. Yet, the quantities A, B, C and υ are adjusted as follows: $a \rightarrow -kh$, $W \rightarrow 0$ with order ($kh \times kh$) and finally $\Psi \rightarrow 0$ with order **(**$k \times k$**)**.

The final step in the analysis of multivariate (vector) time series is to forecast (predict) the upcoming observations of the series. When succeeding to pass the first three steps in time series analysis, one will be able to predict the upcoming values of $Y(n+1)$, $Y(n+2)$, … etc. The Bayesian instruments for addressing forecasting problems is the posterior predictive densities for upcoming observations. A primary focus is the forecasting of the subsequent value of $Y(n+1)$. Theorem 3.2 in addition to corollary 3.2 can be used to predict the upcoming value for $Y(n+1)$. To obtain a point forecast for $Y(n+1)$, $F^{-1}E'$, the mean of the posterior predictive density, is used. On the other hand, for an interval forecast for $Y(n+1)$ one might apply the form of the (1 - α) HPD region of the posterior predictive density:

$$R_\alpha\big(Y(n+1)\big) = \big\{y(n+1): [y(n+1) - F^{-1}E']'[L - E'F^{-1}E]^{-1}[y(n+1) - F^{-1}E']F\upsilon \le kF_{\alpha,k,\upsilon}\big\} \tag{13}$$

In order to forecast the r-steps ahead upcoming observation $Y(n+r)$, Theorem 3.2 might be utilized approximately to get the mean of the conditional predictive density for $Y(n+r)$ conditioned on its preceding future observations $Y(n+1)$, $Y(n+2)$, … up to $Y(n + r - 1)$ such that r should be greater than 1.

4. Numerical Studies

The primary goal of the current section is to evaluate the numerical goodness of the suggested Bayesian prediction approach. To reach this goal, four simulated experiments were employed to forecast future observations of the bivariate ARMA process of orders (1,1), known as $VARMA_2(1,1)$, using different parameter values. All calculations were carried out on a personal computer using the R programming language.

The design of the simulated experiments contains the next steps: Initially, generating a time series following $VARMA_2(1,1)$ model having specific parameters. The data set is then used to estimate the predictive distribution of the model's first upcoming value, namely $y(n+1)$. The predictive density's performance measures are then computed. In the fourth step, the previous three processes are repeated 500 times. Finally, the findings are displayed in tabular form.

The simulation procedure starts by generating 500 data sets from a bivariate normal distribution $N_2(0,V)$. Each data set includes 501 elements, representing the error term $\varepsilon(t)$. The 500 data sets are applied to a recursive form to generate 500 realizations, each one has 501 elements, created from a specific $VARMA_2(1,1)$ model with specific parameters' values. It is assumed that the errors have zero initial values. To avoid the initialization effect, the first 200 pairs of data are removed, yielding 500 data sets of bivariate time series observations. Each data set has a length of 301. Using the proposed methodology, a specific number of bivariate time series values is considered from the generated 301 data values to assess the performance the predictive distribution of the one step ahead value $y(n+1)$. The lengths of the bivariate time series are set to 50, 100, 150, 200, and 300. Each simulation experiment has certain coefficients. Whereas, a fixed variance-covariance matrix for the error term is assumed for all simulation examples

$$V(\varepsilon) = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \tag{14}$$

The coefficients are chosen to represent a variety of places inside the $VARMA_2(1,1)$ model's stationarity and invertibility regions. It is important to note that the Jeffreys' prior will be applied in all simulated experiments.

The major objective is to assess the performance of the suggested Bayesian prediction technique using four effectiveness criteria: P^*, MAPE, MAD and RMSE. The percentage P^* evaluates the adequacy of the highest predictive density (HPD) region obtained from a given predictive density for the future value $y(n+1)$. For each considered realization, we define a 95% HPD region for $y(n+1)$, then the computed percentage P^* of cases in which the computed 95% HPD region includes the actual value of $y(n+1)$ is utilized as an indicator of successful results. The predictive density performs better in the prediction process as the P^* value increases. In addition, MAPE stands for the mean absolute percentage error is calculated using the equation,

$$\text{MAPE} = \left(\left(\sum_{j=1}^{500} \left| \frac{y_{n+1,j} - E_j}{y_{n+1,j}} \right| \right) / 500 \right) * 100, \tag{15}$$

whereas, the MAD, the mean absolute deviation, is given by,

$$\text{MAD} = \sum_{j=1}^{500} \left| y_{n+1,j} - E_j \right| / 500, \tag{16}$$

It measured the distance of the value of $y(n+1)$ from its mean of the predictive density, E (Wei, 2005). Moreover, we compute the square root of the mean squared error (RMSE), which is provided by

$$\text{RMSE} = \sqrt{\sum_{j=1}^{500} \left(y_{n+1,j} - E_j \right)^2 / 500}, \tag{17}$$

The numerical effectiveness of the suggested prediction process will be evaluated in terms of both themodel parameters and the length of the time series (n). For illustration, Simulation example 1 starts by generating 500 data sets each follows the bivariate normal distribution and has 501 variates $\varepsilon(t,1)$ and $\varepsilon(t,2)$ representing the errors. Then, using these sets of data, 500 realization of the pairs $y(t,1)$ and $y(t,2)$, everyone has 501 observations, are generated via the $VARMA_2(1,1)$ process having the coefficients.

$$\Phi = \begin{bmatrix} -0.4 & 0.5 \\ 0.4 & -0.5 \end{bmatrix} \text{ and } \Theta = \begin{bmatrix} 0.5 & -0.4 \\ -0.3 & 0.2 \end{bmatrix}$$

Given that all of the initial values are zero, that is, $y(0,1) = y(0,2) = \varepsilon(0,1) = \varepsilon(0,2) = 0$. In order to omit the impact of the initial values, 200 observations in the beginning of each realization are deleted. After that, we have 500 realizations of the pairs $y(t,1)$ and $y(t,2)$, each one with 301 data points.

The next step is to perform the required calculations to determine the posterior mean's value and the 95% HPD region of the predictive distribution of $y(n+1)$ for each realization using Jefferys' prior. Then, based on the true value of $y(n+1)$, one computes P^*, MAPE, MAD and RMSE measures. These calculations are carried out using the first n observations from each realization for a certain time series size. For each selected time series size, this last step is re-applied. Table 1 displays the findings of Simulation 1.

Simulation 2 is employed in a similar manner using the coefficients

$$\Phi = \begin{bmatrix} -0.4 & 0.5 \\ 0.4 & -0.5 \end{bmatrix} \text{ and } \Theta = \begin{bmatrix} -0.2 & -0.2 \\ -0.2 & -0.2 \end{bmatrix}.$$

Simulation 3 is employed similarly using the coefficients

$$\Phi = \begin{bmatrix} -0.2 & -0.2 \\ -0.2 & -0.2 \end{bmatrix} \text{ and } \Theta = \begin{bmatrix} 0.5 & -0.4 \\ -0.3 & 0.2 \end{bmatrix}.$$

Simulation 4 is employed like the above simulations but using

$$\Phi = \begin{bmatrix} -0.4 & 0.5 \\ 0.4 & -0.5 \end{bmatrix} \text{ and } \Theta = \begin{bmatrix} 0.9 & -0.2 \\ 0.9 & -0.2 \end{bmatrix}.$$

The findings of these simulation studies are shown in tables 2, 3, and 4, respectively.

4.1 Simulation 1 Results

This subsection is devoted to the first simulation study, in which we have a $VARMA_2(1,1)$ process with parameters:

$$\Phi = \begin{bmatrix} -0.4 & 0.5 \\ 0.4 & -0.5 \end{bmatrix} \text{ and } \Theta = \begin{bmatrix} 0.5 & -0.4 \\ -0.3 & 0.2 \end{bmatrix}$$

The findings of this simulated example are shown via table 1:

Table 1: *Percentages P^*, MAPE, MAD and RMSE for the Posterior Predictive Density for Y(n+1) in Simulation 1.*

	LENGTH	$SE[Y(n+1,1)]$	P^*	MAPE	MAD	RMSE
$Y(n+1,1)$	50	1.7638	64.0	6.1149	3.0558	3.9941
	100	1.7653	61.6	4.2501	3.0947	3.8502
	150	1.6968	57.8	5.0743	3.1490	3.9087
	200	1.7112	62.8	3.2452	3.0055	3.8256
	300	1.6583	59.0	7.0870	3.0861	3.9374
	LENGTH	$SE[Y(n+1,2)]$	P^*	MAPE	MAD	RMSE
$Y(n+1,2)$	50	1.2718	52.4	3.2171	2.8887	3.5835
	100	1.2584	50.8	3.3198	2.9043	3.6845
	150	1.2034	44.2	3.6086	2.9916	3.6775
	200	1.2112	50.6	3.2886	2.8420	3.5848
	300	1.1762	47.0	3.3397	3.0070	3.7621

Table 1 Shows that, for $y(n+1,1)$ the values of the percentage P^* are over 57%. Moreover, the values of the three accuracy measures MAPE, MAD and RMSE are small numbers and appear to be stable when the time series length increases. The estimate of $y(n+1,1)$ has a standard error that decreases as the length increases. For $y(n+1,2)$, similar observation is obtained, where P^* values are over 44%. These results provide evidence about the success of the proposed Bayesian prediction methodology to predict the future observation in bivariate $VARMA_2(1,1)$ process for time series length 50 and more.

4.2 Simulation 2 Results

This subsection is devoted to the second simulation study, in which we have a $VARMA_2(1,1)$ process with coefficients:

$$\Phi = \begin{bmatrix} -0.4 & 0.5 \\ 0.4 & -0.5 \end{bmatrix} \text{ and } \Theta = \begin{bmatrix} -0.2 & -0.2 \\ -0.2 & -0.2 \end{bmatrix}$$

The results are presented in table 2:

Table 2: *Percentages P^*, MAPE, MAD and RMSE for the Posterior Predictive Density for Y(n+1) in Simulation 2.*

	LENGTH	$SE[Y(n+1,1)]$	P^*	MAPE	MAD	RMSE
$Y(n+1,1)$	50	1.6572	78.6	4.4026	2.1122	2.6300
	100	1.5420	78.6	3.3824	2.0065	2.4842
	150	1.5043	76.4	2.2559	1.9552	2.4795
	200	1.4800	74.8	3.0835	2.0476	2.5335
	300	1.4576	76.8	6.5935	1.8495	2.3719
	LENGTH	$SE[Y(n+1,2)]$	P^*	MAPE	MAD	RMSE
$Y(n+1,2)$	50	1.1927	69.0	4.7221	1.8610	2.3393
	100	1.1008	65.0	5.7040	1.7839	2.2251
	150	1.0699	68.8	7.8417	1.7017	2.1861
	200	1.0502	67.4	4.0850	1.7212	2.1417
	300	1.0338	63.4	6.4969	1.7879	2.2706

Table 2 Shows that, for $y(n+1,1)$ P^* values are large being over 74%. Moreover, the values of the three accuracy measures MAPE, MAD and RMSE are small numbers. The MAD and RMSE appear to decrease slightly when the time series length increases. Moreover, standard error of $y(n+1,1)$ estimate also decreases as the length increases. On the other hand, for $y(n+1,2)$, similar observation is obtained, where P^* values are also high being over 63%. These results provide evidence about the success of the adopted Bayesian prediction methodology to predict the future observation in predicting the future observation in this process. Results of simulation 2 are better than those of simulation 1.

4.3 Simulation 3 Results

This subsection is devoted to the third simulation study, in which we have a $VARMA_2(1,1)$ model having parameters:

$$\Phi = \begin{bmatrix} -0.2 & -0.2 \\ -0.2 & -0.2 \end{bmatrix} \text{ and } \Theta = \begin{bmatrix} 0.5 & -0.4 \\ -0.3 & 0.2 \end{bmatrix}$$

The findings are shown in table 3:

Table 3: *Percentages P^*, MAPE, MAD and RMSE for the Posterior Predictive Density for Y(n+1) in Simulation 3.*

	LENGTH	$SE[Y(n+1,1)]$	P^*	MAPE	MAD	RMSE
$Y(n+1,1)$	50	1.6853	90.0	3.4623	1.5796	2.0211
	100	1.5624	90.6	2.4915	1.4917	1.8972
	150	1.5280	91.4	3.4267	1.3946	1.7627
	200	1.4922	92.4	1.9820	1.3429	1.6796
	300	1.4675	92.6	2.4859	1.2732	1.6379
	LENGTH	$SE[Y(n+1,2)]$	P^*	MAPE	MAD	RMSE
$Y(n+1,2)$	50	1.1996	86.8	8.5188	1.2144	1.5266
	100	1.1084	88.6	3.6279	1.1245	1.4065
	150	1.0814	88.0	2.7174	1.0969	1.3978
	200	1.0561	87.8	2.5455	1.1011	1.3775
	300	1.0388	88.2	3.0652	1.0102	1.2891

Regarding table 3, one finds that, for $y(n+1,1)$ P^* values are about 90%. Moreover, the values of the three accuracy measures MAPE, MAD and RMSE are small numbers and appear to decrease slightly when the time series length increases. Alike previous cases, standard error of $y(n+1,1)$ estimate decreases as the series length gets larger. For $y(n+1,2)$, similar observation is obtained, where P^* values are large being over 86%. These results provide evidence about the success of the adopted Bayesian prediction methodology to predict the future observation in this process. Results for this case are better than those of the above two simulations.

4.4 Simulation 4 Results

This subsection is devoted to the fourth simulation study, in which we have a $VARMA_2(1,1)$ process with coefficients:

$$\Phi = \begin{bmatrix} -0.4 & 0.5 \\ 0.4 & -0.5 \end{bmatrix} \text{ and } \Theta = \begin{bmatrix} 0.9 & -0.2 \\ 0.9 & -0.2 \end{bmatrix}$$

The findings are presented in table 4:

Table 4: *Percentages P^*, MAPE, MAD and RMSE for the Posterior Predictive Density for Y(n+1) in Simulation 4.*

	LENGTH	$SE[Y(n+1,1)]$	P^*	MAPE	MAD	RMSE
$Y(n+1,1)$	50	1.5902	82.2	3.2647	1.8751	2.3997
	100	1.4953	80.0	3.1319	1.7887	2.2944
	150	1.4672	78.0	2.2594	1.7798	2.2506
	200	1.4533	80.8	2.5302	1.7811	2.2521
	300	1.4393	84.4	2.3698	1.5586	2.0222
	LENGTH	$SE[Y(n+1,2)]$	P^*	MAPE	MAD	RMSE
	50	1.1381	71.0	2.9696	1.6795	2.1340
$Y(n+1,2)$	100	1.0679	72.2	3.2499	1.6114	2.0142
	150	1.0456	70.0	4.8601	1.5112	1.8859
	200	1.0329	70.8	3.2875	1.5202	1.9093
	300	1.0222	72.0	7.4051	1.5542	1.9698

Regarding table 4 one finds that, for $y(n+1,1)$ P^* values are large being over 78%. Moreover the values of the three accuracy measures MAPE, MAD and RMSE are small numbers. The MAD and RMSE appear to decrease slightly as the time series length increases. Alike previous cases, standard error for $y(n+1,1)$ estimate decreases as the series length gets larger. For $y(n+1,2)$, similar observation is obtained, where P^* values are high being over 70%. These results provide evidence about the success of the adopted Bayesian prediction methodology to predict the future observation in this process. Findings of simulation 4 indicate higher performance than the findings of both simulation 1 and 2 but the results of simulation 3 dominates all.

5. Final Conclusion

Shaarawy (1989) presented the Bayesian technique to addressing prediction issues related with multivariate (vector) autoregressive moving average (VARMA) processes, which is notable because of its analytical character, practicality, and

ease of implementation. Nonetheless, neither the theoretical nor numerical aspects of this technique have received substantial attention. The primary goal of this paper is to investigate numerically the efficiency of Shaarawy's method for forecasting future observations of bivariate ARMA models. We carried out the numerical studies using an extensive Monte Carlo simulations depending on four distinct effectiveness criteria and several parameter values that fall in different regions in the domains of the stationarity and invertibility of the model. Furthermore, the study treated various time series lengths to determine the impact of increasing the length of the time series. The numerical findings for all considered parameters were satisfactory, indicating that the proposed Bayesian prediction approach was successful in forecasting future observations for the VARMA2(1,1). The results of the simulation study are limited to $VARMA_2(1,1)$. The extension to higher order and higher dimensional models meets many difficulties in the generation of data which satisfies the stationarity and invertibility conditions of such models.

It is suggested for future work to search for a real data set to employ a full Bayesian multivariate time series analysis including identification, estimation, diagnosis checking and prediction to apply the proposed technique. It may also be suggested to study the extension of the current article to higher order and higher dimensional VARMA models.

REFERENCES

Albassam, M.S., Soliman E.E.A. and Ali, S.S. (2023). An Effectiveness Study of the Bayesian Inference with Multivariate Autoregressive Moving Average Processes. *Communications in statistics- Simulation and Computation*. 52(10):4773-4788. https://doi.org/10.1080/ 03610918.2021.1967986.

Ali, S.S. (2015). Bayesian Forecasting Of Vector Moving Average Processes. . *The Egyptian Statistical Journal*, 59(1), pp. (68-89). https://doi.org/10.21608/ESJU.2015.314457.

Amin, A. (2019). Gibbs Sampling for Bayesian Prediction of SARMA Processes. *Pakistan Journal of Statistics and Operation Research*. 15(2):483-499. https://doi.org/10.18187/ pjsor.v15i2.2174.

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.

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

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

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

Chan, J. C. and Eisenstat, E. (2015). *Efficient estimation of Bayesian VARMAs with time-varying coefficients [Paper presentation]*. 2015 CAMA Working Paper 19/2015. The Australian National University, Australia.

Chatfield, C. (2019). *The Analysis of Time Series: Theory and Practice. An Introduction with R* (7th ed.). Chapman & Hall.

Chib, S. and Greenberg, E. (1994). Bayes inference in regression models with ARMA (p, q) errors. *Journal of Econometrics*. 64(1-2):183-206. https://doi.org/10.1016/0304-076 (94)90063-9.

Cooley, T. F. and Dwyer, M. (1998). Business cycle analysis without much theory a look at structural VARs. *Journal of Econometrics*. 83(1-2):57–88. https://doi.org/10.1016/S0304-4076(97)00065-1.

Hall, A. D. and Nicholls, D. F. (1980). The evaluation of exact maximum likelihood estimates for VARMA models. *Journal of Statistical Computation and Simulation*. 10(3-4):251-262. https://doi.org/10.1080/00949658008810373.

Hillmer, S. C. and Tiao. G. C. (1979). Likelihood function of stationary multiple autoregressive moving average models. *Journal of the American Statistical Association*. 74(367):652-660. http://doi.org/10.1080/01621459.1979.10481666.

Kascha, C. (2012). A comparison of estimation methods for vector autoregressive moving -average models. *Econometric Reviews*. 31(3):297–324. https://doi.org/10.1080/07474938. 2011.607343.

Koop, G. (2013). Forecasting with Medium and Large Bayesian VARS. *Journal of Applied Econometrics*. 28(2):177-203. https://doi.org/10.1002/jae.1270.

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

Lütkepohl. (2007). *New Introduction to Multiple Time Series Analysis*. Berlin: Springer.

Marriott, J., Ravishanker, N., Gelfand, A. and Pai, J. (1996). Bayesian analysis of ARMA processes: Complete sampling-based inference under full likelihoods. *In Bayesian Statistics and Econometrics: Essays in honor of Arnold Zellner, D., Berry, K., Chaloner, and J., Geweke*; (eds). New York: Wiley.

Mauricio, J. A. (1995). Exact maximum likelihood estimation of stationary vector ARMA models. *Journal of the American Statistical Association*. 90(429):282-291. https://doi.org/ 10.1080/01621459.1995.10476511.

Monahan, J. F. (1983). Fully Bayesian Analysis of ARIMA Time Series Models. *Journal of Econometrics*. 21(3):307–331. https://doi.org/10.1016/0304-4076(83)90048-9.

Newbold, P. (1973). Bayesian Estimation of Box and Jenkins Transfer Function Model for Noise Models. *Journal of the Royal Statistical Society-Series B*. 35(2):323-336.

Raghavan, M., Athanasopoulos, G. and Silvapulle, P. (2009). *VARMA models for Malaysian Monetary Policy Analysis [Paper Presentation]*. 2009 Monash University, Department of Econometrics and Business Statistics Working Papers, Malaysia.

Ravishanker, N., I. and Ray, B. K. (1997). Bayesian analysis of vector ARMA models using Gibbs sampling. *Journal of Forecasting*. 16(3):177–194. https://doi.org/10.1002/(SICI)1099-131X(199705)16:3%3C177::AID-FOR650%3E3.0.CO;2-%23.

Shaarawy, S.M. (l989). Bayesian Inferences and Forecasts with Multiple ARMA Models. *Communications in Statistics-Theory and Methods*. 18(4):1481–1509. https://doi.ord/ 10.1080/ 03610918908812836.

Shaarawy, S.M., Soliman, E.E.A. and Ali, S.S. (2007). Bayesian identification of moving average models. *Communications in Statistics-Theory and Method*. 36(12):2301-2312. https://doi.org/10.1080/03610920701215449.
Shaarawy, S.M. (2023). Bayesian modeling and forecasting of vector autoregressive moving average processes. *Communications in Statistics-Theory and Method*. 52(11):3795-3815. https://doi.org/ 10.1080/03610926.2021.1980047.
Shaarawy, S.M., Ali, S.S. and Soliman, E.E.A. (2024). Bayesian Identification of Seasonal Vector ARMA Processes. *The Egyptian Statistical Journal*, 68(2), pp. (129-145). https://doi.org/10.21608/esju.2024.318938.1044.
Shea, B. L. (1987). Estimation of multivariate time series. *Journal of Time Series Analysis*. 8(1):95-109. https://doi.org/10.1111/j.1467-9892.1987.tb00423.x
Tiao, G. C., Box, G. E. P. (1981). Modeling Multiple Time Series with Applications. *Journal of the American Statistical Association*. 76(376): 802–816. https://doi.org/10.1080/ 01621459.1981.10477728.
Tunnicliffe W. G. (1973). The estimation of parameters in multivariate time series models. *Journal of the Royal Statistical Society-Series B*. 35(1):76-85. https://doi.org/10.1111/j.2517-6161.1973.tb00938.x
Wei, W. W. S. (2005). *Time Series Analysis: Univariate and Multivariate Methods* (2nd ed.). Reading, MA: Addison Wesley.
Wei. W. W. S. (2019). *Multivariate Time Series Analysis and Applications*. John Wiley & Sons Ltd.
Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*. John Wiley and Sons, New York.
Zellner, A. and Reynolds, R. (1978, June 2-3). *Bayesian Analysis of ARMA Models [Paper Presentation]*. 1978 the Sixteenth Seminar on Bayesian Inference in Econometrics.

Declaration

Conflict of Study: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
Author Contribution Statement: Both authors conceived the idea and designed the research; Analyzed and interpreted the data, and wrote the paper.
Funding Statement: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.
Availability of Data and Materials: Data will be made available by the corresponding author on reasonable request.
Consent to Publish: All authors have agreed to publish this manuscript in the SCOPUA Journal of Applied Statistical Research.
Ethical Approval: Not Applicable
Consent to Participate: Not Applicable
Acknowledgment: Not Applicable

Author(s) Bio / Authors' Note

EMAD E. A. SOLIMAN: ehegazy@kau.edu.sa
SHERIF S. ALI: ssali@kau.edu.sa