A BAYESIAN PROCEDURE TO IDENTIFY THE ORDERS OF VECTOR MOVING AVERAGE PROCESSES WITH SEASONALITY

AUTHORS

SAMIR M. SHAARAWY
ssamir50@hotmail.com
College of Business Administration
Department of Quantitative Methods and Information Systems
Kuwait University, Kuwait

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

EMAD E. A. SOLIMAN
easalam@hotmail.com
Faculty of Science, Department of Statistics, King Abdul-Aziz University, KSA.
Permanent Address: Faculty of Economics and Political Science, Department of Statistics, Cairo University, Egypt.

ABSTRACT

This article develops an approximate Bayesian procedure to identify the orders of vector moving average processes with seasonality. The proposed procedure is based on approximating the likelihood function by a matrix normal–Wishart on the parameter space. Combining the approximate likelihood function with a matrix normal–Wishart or Jeffreys' vague prior and using an indirect Bayesian technique to estimate initial values for the orders, the joint posterior mass function of the orders is developed in a convenient form. Then one may examine the posterior probabilities over the grid of the orders and select the orders at which the posterior joint mass function attains its maximum to be the identified orders. Five simulation studies, with three different prior distributions for the orders, are conducted to demonstrate the performance of the proposed procedure and check its adequacy and applicability in solving the identification problem. The numerical results support using the proposed procedure to identify the orders of vector moving average processes with seasonality.

Keywords: Identification; Seasonal vector moving average processes; Matrix normal–Wishart distribution; Matrix t distribution.

1. INTRODUCTION

Time series analysis of vector moving average processes with seasonality has been shown effective in modeling time series data arise in many fields of application, specially in economics and business. In economics, as an example, one may record quarterly money supply y(t,1), real interest rate y(t,2) and gross national products y(t,3). In business, as another example, one may observe monthly single family–housing starts y(t,1) and houses sold y(t,2) in U.S.A. These variables are modeled and forecasted jointly using a vector model in order to have an insight into the dynamic interrelationship among the variables and increase the precision of the estimates of the parameters and forecasts.

One of the most important phases of vector time series analysis is determining the system mechanism which relates the 'inputs' of the time series with their 'outputs'. In many applications, the system mechanism can be presented by two different matrix polynomials with orders q and Q. The first order q is called the non-seasonal (regular) moving average order, while the second order Q is called the seasonal moving average order. Tiao et al. (1979), Tiao and Box (1981) and Liu (2006) have shown that the seasonal vector moving average processes, denoted by SVMA$_k(q, Q)$ for short, are quite useful in modeling time series data. In practice the orders q and Q are usually unknown and should be identified or estimated.

With regard to the seasonal univariate moving average processes, Box and Jenkins (1970) have presented a non-Bayesian methodology to estimate the orders q and Q by matching the sample autocorrelation function with its theoretical counterpart. Their methodology has grown in popularity and is today the prevailing procedure of time series analysis. For more details about the Box and Jenkins methodology, the reader is referred to Harvey (1981,1993), Priestley (1981), Tsay (1984), Wei (2005), Box et al. (2016) and Brockwell and Davis (2016). The second non-Bayesian procedure to estimate the moving average orders is known as the automatic or exploratory approach. Assuming the maximum orders are known, the foundation of this approach is to fit all possible moving average models and compute a specific criterion for each model; then one may choose the model for which the proposed criterion attains its optimal value. However, there is no complete agreement on the form of the criterion to be optimized. The most favorable automatic criterion AIC or Akaike's Information criterion was introduced by Akaike (1974). For more details about the automatic approach, the reader is referred to Rissanen

(1978), Hannan and Quinn (1979), Mills and Prasad (1992) and Beveridge and Oickle (1994). However, one may notice that using any automatic procedure with seasonal models is time consuming and costly.

Regarding the Bayesian approach to estimate the orders of univariate time series, Monahan (1983) has given a numerical procedure to the estimation problem of non-seasonal processes with low orders. Broemeling and Shaarawy (1988) have developed an approximate procedure to identify the orders of non-seasonal processes. Shaarawy and Ali (2003) have initiated a Bayesian solution to estimate the orders of the seasonal autoregressive processes. Recently, Shaarawy et al. (2007) have developed an approximate procedure to estimate the orders of non–seasonal moving average processes.

With regard to vector version, Tiao and Box (1981), Tiao and Tsay (1983) and others have studied the problem of estimating the orders of the process, from non-Bayesian viewpoint, by matching the cross correlation functions computed from the data with their theoretical counterparts. On the other hand, the Bayesian methods of estimating the orders of vector processes have recently been studied. Shaarawy and Ali (2008) have initiated the Bayesian solution of estimating the orders of vector autoregressive processes. In 2012, they developed an approximate procedure to estimate the orders of vector moving average processes. Most recently, Shaarawy and Ali (2015) have initiated an approximate procedure to estimate the orders of seasonal vector autoregressive processes. For well-understood reasons, one may say that a purely Bayesian procedure to identify the orders of vector moving average processes with seasonality has not been explored yet.

Using a matrix normal–Wishart or Jeffreys' vague prior, the main objective of this article is to develop an approximate joint posterior probability mass function for the orders of seasonal vector moving average processes in a convenient form. Then one may examine the behavior of the joint posterior probability mass function over the grid of the orders and choose the orders at which the mass function attains its maximum to be the identified orders. In order to determine the effectiveness of the proposed Bayesian procedure, five simulation studies with three different prior distributions for the orders are conducted to identify the orders of bivariate moving average processes with seasonality.

The rest of the paper is organized as follows: Section 2 introduces the definition of the vector moving average processes with seasonality. Section 3 introduces an indirect Bayesian procedure to identify the orders of the seasonal vector moving average processes.

This procedure will be used later to approximate the likelihood function. Section 4 is devoted to develop an approximate joint posterior probability mass function of the model orders and explain the proposed Bayesian identification procedure. Section 5 is devoted to examine and assess the numerical effectiveness of the proposed Bayesian procedure in solving the identification problem of seasonal vector moving average processes.

2. MOVING AVERAGE PROCESSES WITH SEASONALITY

Let {t} be a sequence of integers, $q \in \{1, 2, \ldots\}$, $Q \in \{1, 2, \ldots\}$, $k \in \{1, 2, \ldots\}$, $\{y(t)\}$ is a sequence of $k\times1$ real observable random vectors, θ_i ($i = 1, 2, \ldots, q$), Θ_i ($i = 1, 2, \ldots, Q$) are $k\times k$ unknown matrices of real constants, s is the number of seasons per time unit, and $\{\varepsilon(t)\}$ is a sequence of independent and normally distributed $k\times1$ unobservable random vectors with zero mean and a $k\times k$ unknown precision matrix T. Then the k–variate seasonal vector moving average process of orders q and Q is defined for n vectors of observations as

$$y(t) = \theta_q(B)\Theta_Q(B^s)\varepsilon(t) , \qquad t = 1, 2, \ldots, n \tag{2.1}$$

Where

$$y(t) = [y(t,1) \quad y(t,2) \quad \cdots \quad y(t,k)]' ,$$

$$\theta_q(B) = I_k - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q ,$$

$$\Theta_Q(B^s) = I_k - \Theta_1 B^s - \Theta_2 B^{2s} - \cdots - \Theta_Q B^{Qs} ,$$

$$\varepsilon(t) = [\varepsilon(t,1) \quad \varepsilon(t,2) \quad \cdots \quad \varepsilon(t,k)]'$$

I_k is the identity matrix of order k, B is the backward shift operator defined by $B^r y(t) = y(t-r)$ and s is the periodicity of the time series. The $k\times k$ matrix polynomial $\theta_q(B)$, of degree q in the backshift operator B, is known as the regular (non-seasonal) moving average operator of order q, while the $k\times k$ matrix polynomial $\Theta_Q(B^s)$, of degree Qs in B, is known as the seasonal moving average operator of order Q. The process $y(t)$ is always stationary and is invertible if all the roots of the two determinantal equations $|\theta_q(B)| = 0$ and $|\Theta_Q(B^s)| = 0$ lie outside the unit circle.

The class of models (2.1) can be expressed in explicit from as

$$Y = X(q,Q)\Gamma(q,Q) + U \qquad (2.2)$$

Where Y is a matrix of order $n\times k$ with ij^{th} element equal to $y(i,j)$ and $X(q,Q)$ is a matrix of order $n\times kh$, $h=q+Q+qQ$, defined by

$$X = \lfloor X_1 \quad X_2 \quad Z_1 \quad Z_2 \quad \cdots \quad Z_Q \rfloor, \text{ where}$$

$$X_1 = \begin{bmatrix} -\varepsilon'(0) & -\varepsilon'(-1) & \cdots & -\varepsilon'(1-q) \\ -\varepsilon'(1) & -\varepsilon'(0) & \cdots & -\varepsilon'(2-q) \\ \vdots & \vdots & & \vdots \\ -\varepsilon'(n-1) & -\varepsilon'(n-2) & \cdots & -\varepsilon'(n-q) \end{bmatrix}, \quad X_2 = \begin{bmatrix} -\varepsilon'(1-s) & -\varepsilon'(1-2s) & \cdots & -\varepsilon'(1-Qs) \\ -\varepsilon'(2-s) & -\varepsilon'(2-2s) & \cdots & -\varepsilon'(2-Qs) \\ \vdots & \vdots & & \vdots \\ -\varepsilon'(n-s) & -\varepsilon'(n-2s) & \cdots & -\varepsilon'(n-Qs) \end{bmatrix}$$

And

$$Z_r = \begin{bmatrix} \varepsilon'(-rs) & \varepsilon'(-rs-1) & \cdots & \varepsilon'(-rs-q+1) \\ \varepsilon'(-rs+1) & \varepsilon'(-rs) & \cdots & \varepsilon'(-rs-q+2) \\ \vdots & \vdots & & \vdots \\ \varepsilon'(n-rs-1) & \varepsilon'(n-rs-2) & \cdots & \varepsilon'(n-rs-q) \end{bmatrix}, \quad r = 1, 2, \ldots, Q$$

This means that the columns of the matrix $X(q, Q)$ consist of the elements of the regressors $\varepsilon(t-1)$, $\varepsilon(t-2)$, …, $\varepsilon(t-Qs-q)$, respectively. Furthermore $\Gamma(q, Q)$ is the $kh\times k$ matrix of coefficients defined by

$$\Gamma(q,Q) = \begin{bmatrix} \overbrace{\theta(q)}^{kq\times k} \\ \cdots\cdots \\ \overbrace{\Theta(Q)}^{kQ\times k} \\ \cdots\cdots \\ \overbrace{\gamma(q,Q)}^{kqQ\times k} \end{bmatrix}$$

Where

$$\theta(q) = \begin{bmatrix} \overbrace{\theta_1'}^{k \times k} \\ \cdots \; \cdots \\ \overbrace{\theta_2'}^{k \times k} \\ \cdots \; \cdots \\ \vdots \\ \cdots \; \cdots \\ \overbrace{\theta_q'}^{k \times k} \end{bmatrix}, \theta_i = \begin{bmatrix} \theta_{i.11} & \theta_{i.12} & \cdots & \theta_{i.1k} \\ \theta_{i.21} & \theta_{i.22} & \cdots & \theta_{i.2k} \\ \vdots & \vdots & & \vdots \\ \theta_{i.k1} & \theta_{i.k2} & \cdots & \theta_{i.kk} \end{bmatrix} \quad i = 1,2,\ldots,q$$

$$\Theta(Q) = \begin{bmatrix} \overbrace{\Theta_1'}^{k \times k} \\ \cdots \; \cdots \\ \overbrace{\Theta_2'}^{k \times k} \\ \cdots \; \cdots \\ \vdots \\ \cdots \; \cdots \\ \overbrace{\Theta_Q'}^{k \times k} \end{bmatrix}, \Theta_i = \begin{bmatrix} \Theta_{i.11} & \Theta_{i.12} \cdots & \Theta_{i.1k} \\ \Theta_{i.21} & \Theta_{i.22} \cdots & \Theta_{i.2k} \\ \vdots & \vdots & \vdots \\ \Theta_{i.k1} & \Theta_{i.k2} \cdots & \Theta_{i.kk} \end{bmatrix}, \quad i = 1,2 \ldots, Q$$

and $\gamma(q, Q) = \gamma(\theta_i, \Theta_j)$, i=1,2, …, q ; j=1, 2, …,Q. Finally, U is the white noise matrix of order $n \times k$ with ijth element equal to $\varepsilon(i, j)$.

It should be born in mind that the dimension of the regressor matrix $X(q, Q)$ depends on the orders q and Q. This means that for each specific order, say (q_0, Q_0) there is a specific matrix $X(q_0,Q_0)$. One may also notice that the parameters $\gamma(q, Q)$ will be treated as free parameters in our proposed Bayesian identification procedure in order to have an intractable likelihood function. If n is sufficiently large, the approximate likelihood function is expected to serve as a good approximation to the exact one.

3. AN INDIRECT BAYESIAN TECHNIQUE

The vector moving average model with seasonality, denoted by SVMA$_k(q, Q)_s$ for short, is quite useful in modeling and forecasting time series data and frequently q and Q are not in excess of three. In practice, the values of the orders q and Q are unknown and one has to identify them using the observed n vectors of observations S_n. The direct Bayesian approach to identify these two values is to find the bivariate posterior probability mass function of the orders q and Q. Then one may inspect the posterior probabilities over the grid of q and Q and choose the values at which the bivariate posterior mass function attains its maximum to be the identified values. The approach here is somewhat different from the direct Bayesian approach. Instead of working directly with the posterior mass function of q and Q, it is proposed to focus on the joint posterior distribution of the coefficients Γ_0 where

$$\Gamma_0' = \left[\theta_1' \ \vdots \ \theta_2' \ \vdots \cdots \vdots \ \theta_m' \quad \Theta_1' \ \vdots \ \Theta_2' \ \vdots \cdots \vdots \ \Theta_r' \ \vdots \ \gamma_0' \right] \tag{3.1}$$

Where θ_i (i= 1, 2, …, m), Θ_i (i= 1, 2, …, r) are as defined in the previous section, and $\gamma_0(m,r) = \gamma_0(\theta_i, \Theta_i)$, i= 1, 2, …, m; j= 1, 2, …, r. The maximum orders m and r are assumed to be known.

Assuming that ε (0) = ε (-1) = … = ε (1-m-rs) = 0, the likelihood function of the parameters Γ_0 and T is

$$L(\Gamma_0, T \mid S_n) \propto |T|^{\frac{n}{2}} \exp\left(-\tfrac{1}{2} tr \left\{ \sum_{t=1}^{n} \varepsilon(t)\varepsilon'(t) T \right\}\right) \tag{3.2}$$

Where $\Gamma_0 \in R^{k(r+m+rm)\times k}$, $T > 0$ and

$$\varepsilon'(t) = y'(t) - x'(t-1)\Gamma_0 \tag{3.3}$$

Where

$$\begin{aligned} x'(t-1) = [&-\varepsilon'(t-1) \ -\varepsilon'(t-2) \cdots -\varepsilon'(t-m) \ -\varepsilon'(t-s) \ -\varepsilon'(t-2s) \cdots -\varepsilon'(t-rs) \\ &\varepsilon'(t-s-1) \ \varepsilon'(t-2s-1) \cdots \varepsilon'(t-rs-1) \ \varepsilon'(t-s-2) \ \varepsilon'(t-2s-2) \cdots \\ &\varepsilon'(t-rs-2) \cdots \varepsilon'(t-s-m) \ \varepsilon'(t-2s-m) \cdots \varepsilon'(t-rs-m)] \end{aligned}$$

The expression (3.3) is a recurrence relation for the residuals. This recurrence causes the main problem in developing an exact analysis of seasonal vector moving average processes. However, this recurrence may be used to evaluate the residuals recursively if one knows Γ_0 and the initial values of the residuals. The proposed approximation is based on replacing the exact residuals $\varepsilon(t\text{-}j)$ by their least squares estimates and assuming the initial values of the residuals equal to their unconditional means, namely zero. Thus, we estimate the residuals recursively by

$$\hat{\varepsilon}'(t) = y'(t) - \hat{x}'(t-1)\hat{\Gamma}_0$$

Where, the matrix $\hat{\Gamma}_0$ is the non-linear least squares estimate of the coefficient matrix Γ_0 and $\hat{x}(t-1)$ is the same as $x(t-1)$ but using the estimated residuals instead of the exact ones. Using the estimates of the residuals, one may write the likelihood function (3.2) approximately as

$$L^*(\Gamma_0, T|S_n) \propto |T|^{\frac{n}{2}} \exp(-\tfrac{1}{2} tr \sum_{t=1}^{n} [y(t) - \Gamma_0' \hat{x}(t-1)]' [y(t) - \Gamma_0' \hat{x}(t-1)] T) \quad (3.4)$$

An adequate choice of prior distribution of the parameters Γ_0 and T is a matrix normal–Wishart distribution, i.e.

$$\xi(\Gamma_0, T) = \xi_1(\Gamma_0 | T)\, \xi_2(T) \quad (3.5)$$

Where

$$\xi_1(\Gamma_0 | T) \propto |T|^{\frac{k(m+r+mr)}{2}} \exp(-\tfrac{1}{2} tr(\Gamma_0 - D)' V (\Gamma_0 - D) T)$$

and

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

Where the hyper-parameters $D \in R^{k(m+r+mr)\times k}$, V is a $k(m+r+mr)\times k(m+r+mr)$ positive definite matrix, and ψ is a $k\times k$ positive definite matrix. If one has 'little' information about the parameters, a priori, one may use Jeffrey's vague prior

$$\xi(\Gamma_0, T) \propto |T|^{\frac{-(k+1)}{2}}, \quad \Gamma_0 \in R^{k(m+r+mr)\times k}, T > 0 \quad (3.6)$$

Theorem 3.1

Combining the approximate likelihood function (3.4) with the prior density (3.5), the posterior distribution of Γ_0 is a matrix t distribution with parameters $(A^{-1}B, A^{-1}, C - B'A^{-1}B, \upsilon)$ where

$$A = V + \sum_{t=1}^{n} \hat{x}(t-1)\hat{x}'(t-1),$$

$$B = VD + \sum_{t=1}^{n} \hat{x}(t-1) y'(t),$$

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

and

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

Corollary 3.1

Combining the approximate likelihood function (3.4) with Jeffrey's vague prior (3.6), the posterior distribution of Γ_0 is a matrix t with parameters $(A^{-1}B, A^{-1}, C - B'A^{-1}B, \upsilon)$. However, A, B, C and υ will be modified by letting $V \rightarrow 0_{k(m+r+mr)\times k(m+r+mr)}$, $a \rightarrow -k(m+r+mr)$ and $\psi \rightarrow 0_{k\times k}$.

Since Γ_0 has a matrix t distribution, any subset of k rows has a matrix t distribution. Also, the conditional distribution of a subset of rows given any other subset of rows is a matrix t. Furthermore, one can test any subset of rows to be zero (marginally or conditionally) using an exact F statistic for $k = 1,2$. For $k \geq 3$, one can use an approximate χ^2 statistic. For more details about the form and properties of the matrix t density function, the reader is referred to Box and Tiao (1973).

Instead of working directly with the distribution of the orders q and Q, it is proposed to focus on the posterior distribution of Γ, given by the previous theorems, and do a backward elimination procedure to identify initial values for the orders q and Q as follows:

1. Test H_0: $\Theta_r = 0$ versus $\Theta_r \neq 0$ using the marginal posterior distribution of Θ_r which is a matrix t distribution.
2. If the above H_0 is not rejected, test H_0: $\Theta_{r-1} = 0$ versus $\Theta_{r-1} \neq 0$ using the conditional distribution of Θ_{r-1} given $\Theta_r = 0$ which is also a matrix t distribution.
3. In the above H_0 is not rejected, test H_0: $\Theta_{r-2} = 0$ versus $\Theta_{r-2} \neq 0$ using the conditional distribution of Θ_{r-2} given $\Theta_r = \Theta_{r-1} = 0$ which is also a matrix t distribution.
4. The procedure is continued in this fashion until the hypothesis Θ_{r0} is rejected for some r_0 where $0 < r_0 \leq r$. The value r_0 is the initial indirect Bayesian solution of the seasonal order Q.
5. The four previous steps are repeated for the non-seasonal order q until the hypothesis θ_{m0} is rejected for some m_0 where $0 < m_0 < m$. The value m_0 is the initial indirect Bayesian solution of the regular order q. Then the values m_0 and r_0 are the proposed indirect Bayesian solution of our identification problem and the proposed initial values of our proposed direct Bayesian approach.

Frequently q and Q do not exceed three. Here we focus on how to perform the indirect Bayesian procedure, outlined above, in identifying an initial value for the seasonal order Q assuming $m = r = 3$ using Jeffreys' vague prior. In order to do that, define the following quantities:

$$\Gamma_0 = \begin{bmatrix} \overbrace{\theta}^{3k \times k} \\ \cdots\cdots \\ \overbrace{\Theta}^{3k \times k} \\ \cdots\cdots \\ \overbrace{\gamma}^{9k \times k} \end{bmatrix}, \theta = \begin{bmatrix} \theta_1 \\ \cdots \\ \theta_2 \\ \cdots \\ \theta_3 \end{bmatrix}, \Theta = \begin{bmatrix} \Theta_1 \\ \cdots \\ \Theta_2 \\ \cdots \\ \Theta_3 \end{bmatrix}, \Gamma_1 = \begin{bmatrix} \theta \\ \cdots \\ \Theta \end{bmatrix}$$

$$\hat{\Gamma}_0 = \begin{bmatrix} \overbrace{\hat{\theta}}^{3k \times k} \\ \cdots\cdots \\ \overbrace{\hat{\Theta}}^{3k \times k} \\ \cdots\cdots \\ \overbrace{\hat{\gamma}}^{9k \times k} \end{bmatrix}, \quad \hat{\theta} = \begin{bmatrix} \hat{\theta}_1 \\ \cdots \\ \hat{\theta}_2 \\ \cdots \\ \hat{\theta}_3 \end{bmatrix}, \quad \hat{\Theta} = \begin{bmatrix} \hat{\Theta}_1 \\ \cdots \\ \hat{\Theta}_2 \\ \cdots \\ \hat{\Theta}_3 \end{bmatrix}, \quad \hat{\Gamma}_1 = \begin{bmatrix} \hat{\theta} \\ \cdots \\ \hat{\Theta} \end{bmatrix}$$

and

$$A^{-1} = \begin{bmatrix} \overbrace{A_{11}}^{6k \times 6k} & \vdots & \overbrace{A_{12}}^{6k \times 9k} \\ \cdots\cdots\cdots & \vdots & \cdots\cdots\cdots \\ \overbrace{A_{21}}^{9k \times 6k} & \vdots & \overbrace{A_{22}}^{9k \times 9k} \end{bmatrix}$$

$\Gamma_0 \sim t_{15k \times k}(A^{-1}B, A^{-1}, C - B'A^{-1}B, \upsilon)$

Γ_1 is a subset of $6k$ rows of Γ_0, thus

$\Gamma_1 \sim t_{6k \times k}(\hat{\Gamma}_1, A_{11}, C - B'A^{-1}B, \upsilon)$, see Box and Tiao (1973, pp. 445).

Similarly, $\Theta \sim t_{3k \times k}(\hat{\Theta}, A^{*}_{22}, C - B'A^{-1}B, \upsilon)$ where

$$A_{11} = \begin{bmatrix} \overbrace{A^*_{11}}^{3k\times 3k} & \vdots & \overbrace{A^*_{12}}^{3k\times 3k} \\ \cdots\cdots & \vdots & \cdots\cdots\cdots \\ \overbrace{A^*_{21}}^{3k\times 3k} & \vdots & \overbrace{A^*_{22}}^{3k\times 3k} \end{bmatrix}$$

Let

$$\Theta = \begin{bmatrix} \overbrace{\Theta_1}^{k\times k} \\ \cdots \\ \overbrace{\Theta^*_2}^{2k\times k} \end{bmatrix}, \quad \Theta^*_2 = \begin{bmatrix} \overbrace{\Theta_2}^{k\times k} \\ \cdots \\ \overbrace{\Theta_3}^{k\times k} \end{bmatrix} \quad \text{and} \quad A^*_{22} = \begin{bmatrix} \overbrace{E_{11}}^{k\times k} & \vdots & \overbrace{E_{12}}^{k\times 2k} \\ \cdots\cdots\cdots & \vdots & \cdots\cdots\cdots \\ \overbrace{E_{21}}^{2k\times k} & \vdots & \overbrace{E_{22}}^{2k\times 2k} \end{bmatrix}$$

Θ^*_2 is a subset of $2k$ rows of Θ, then

$$\Theta^*_2 \sim t_{2k\times k}(\hat{\Theta}^*_2, E_{22}, C - B'A^{-1}B, \upsilon)$$

Thus $\Theta_3 \sim t_{k\times k}(\hat{\Theta}_3, E^*_{22}, C - B'A^{-1}B, \upsilon)$ where

$$E_{22} = \begin{bmatrix} \overbrace{E^*_{11}}^{k\times k} & \vdots & \overbrace{E^*_{12}}^{k\times k} \\ \cdots\cdots & \vdots & \cdots\cdots\cdots \\ \overbrace{E^*_{21}}^{k\times k} & \vdots & \overbrace{E^*_{22}}^{k\times k} \end{bmatrix}$$

Using the result (8.4.32) given by Box and Tiao (1973), we have

$\Theta_2 \mid \Theta_3 \sim t_{k\times k}(\tilde{\Theta}_2, E^*_{22.3}, G, \upsilon + k)$ where

$$\tilde{\Theta}_2 = \hat{\Theta}_2 + E^*_{12} E^{*\,-1}_{22} (\Theta_3 - \hat{\Theta}_3),$$

$$E^*_{22.3} = E^*_{11} - E^*_{12} E^{*-1}_{22} E^*_{21} \text{ And}$$

$$G = C - B'A^{-1}B + (\Theta_3 - \hat{\Theta}_3)' E^{*\,-1}_{22} (\Theta_3 - \hat{\Theta}_3)$$

Test (1): H_0: $\Theta_3 = 0$ versus H_1: $\Theta_3 \neq 0$

For $k = 2$, one may use the test statistic

$$F_{calc.} = \frac{\nu}{2}\left[\left| I_2 + (C - B'A^{-1}B)^{-1} \hat{\Theta}'_3 E^{*-1}_{22} \hat{\Theta}_3 \right|^{\frac{1}{2}} - 1 \right]$$

We reject H$_0$ if $F_{calc.} > F(4, 2\upsilon, \alpha)$, where $F(4, 2\upsilon, \alpha)$ is the upper 100(α%) point of the F distribution with 4 and 2υ degrees of freedom.

For $k \geq 3$, one may use the approximate test statistic

$$\chi^2_{calc.} = -\upsilon[1 + \frac{(2k-3)}{2\upsilon}]log\,|\,I_2 + (C - B'A^{-1}B)^{-1}\hat{\Theta}'_3 E^{*-1}_{22}\hat{\Theta}_3\,|^{-1}$$

We reject H$_0$ if $\chi^2_{calc} > \chi^2(k^2, \alpha)$, where $\chi^2(k^2, \alpha)$ is the upper 100(α%) point of the χ^2 distribution with k^2 degrees of freedom. For more details about this test and the coming down one, the reader is referred to Box and Tiao (1973). If we reject H$_0$, we must stop the testing process and conclude that the initial indirect Bayesian solution of the seasonal order Q is $r_0 = 3$ .If we do not reject H$_0$, we should go to test (2)

Test (2): H$_0$: $\Theta_2 = 0$ given that $\Theta_3 = 0$

For $k = 2$, one may use the test statistic

$$F_{calc.} = (\frac{\upsilon + 2}{2})[\,|\,I_2 + G^{-1}\tilde{\Theta}'_2 E^{*-1}_{22.3}\tilde{\Theta}_2\,|^{\frac{1}{2}} - 1]$$

We reject H$_0$ if $F_{calc.} > F(4, 2(\upsilon + 2), \alpha)$

For $k \geq 3$, one may use the approximate test statistic

$$\chi^2_{calc.} = -(\upsilon + k)[1 + \frac{(2k-3)}{2(\upsilon + k)}]log\,|\,I_k + G^{-1}\tilde{\Theta}'_2 E^{*-1}_{22.3}\tilde{\Theta}_2\,|^{-1}$$

We reject H$_0$ if $\chi^2_{calc.} > \chi^2(k^2, \alpha)$

If we reject H$_0$, we conclude that the initial identified value of the seasonal order Q is 2. If do not reject H$_0$, we conclude that the initial identified value of the seasonal order Q is 1.

In similar fashion, one can identify an initial value of the regular order (non-seasonal) q. The two initial identified values of the regular and seasonal order, say q_0 and Q_0, will be used later by our proposed direct Bayesian algorithm to develop an approximate posterior probability mass function for the model orders q and Q in an easy and convenient form.

4. A BAYESIAN IDENTI FICATION PROCEDURE

Given the initial values q_0 and Q_0, estimated above by the indirect procedure, the main goal of this section is to develop an approximate direct approach to identify the orders q and Q of the vector moving average processes with seasonality. Unlike the indirect technique, the orders q and Q are assumed to be random variables and the problem is how to find their joint posterior probability mass function in an easy and convenient form. In order to do that, let S_n be the n vectors of observations generated form a seasonal vector moving average process of orders q and Q having the form (2.1) where the orders q and Q are non-negative unknown integers. The likelihood function can be written as

$$L(\Gamma(q,Q),q,Q,\mathrm{T}\mid S_n)\propto|T|^{\frac{n}{2}}\ exp(-\tfrac{1}{2}tr\sum_{t=1}^{n}[y(t)-\Gamma'(q,Q)x_{q,Q}(t-1)]' \\ [y(t)-\Gamma'(q,Q)x_{q,Q}(t-1)]T) \qquad (4.1)$$

Where $\Gamma(q,Q)\in R^{kh\times k}$, $T>0$, $q=1, 2, \ldots, m$; $Q=1, 2, \ldots, r$. Moreover, m and r are the largest possible orders of the process and $X'_{q,Q}(t-1)$ is the t^{th} row of the regressor matrix $X(q, Q)$ defined as

$$X'_{q,Q}(t-1)=[-\varepsilon'(t-1)\ \ -\varepsilon'(t-2)\cdots-\varepsilon'(t-q)\ \ -\varepsilon'(t-s)\ \ -\varepsilon'(t-2s)\cdots-\varepsilon'(t-Qs) \\ \varepsilon'(t-s-1)\ \varepsilon'(t-2s-1)\cdots\varepsilon'(t-Qs-1)\ \varepsilon'(t-s-2)\ \varepsilon'(t-2s-2)\cdots\varepsilon'(t-Qs-2) \\ \cdots\varepsilon'(t-s-q)\ \varepsilon'(t-2s-q)\cdots\varepsilon'(t-Qs-q)]$$

The likelihood function (4.1) is analytically intractable since the errors $\varepsilon(t\text{-}j)$'s are nonlinear functions in the model coefficients θ_i and Θ_i. In order to simplify the likelihood function, we propose to use the initial values q_0 and Q_0 to estimate the residuals $\varepsilon(t\text{-}j)$'s by the recurrence formula

$$\hat{\varepsilon}(t)=y(t)-\hat{\Gamma}\,\hat{X}_{q,Q}(t-1)$$

Where $\hat{\Gamma}$ is the nonlinear least square estimate of the coefficients matrix Γ and $\hat{x}_{q,Q}(t-1)$ is the same as $x_{q,Q}(t-1)$ but using estimated residuals instead of the exact ones. Once the estimates $\hat{\varepsilon}(t-j)'s$ are obtained, they are substituted in the likelihood function (4.1) to get an approximate likelihood function in the form

$$L^*(\Gamma(q,Q),q,Q,T\mid S_n)\propto|T|^{\frac{n}{2}}\ exp(-\tfrac{1}{2}tr\sum_{t=1}^{n}[y(t)-\Gamma'(q,Q)\hat{x}_{q,Q}(t-1)]' \\ [y(t)-\Gamma'(q,Q)\hat{x}_{q,Q}(t-1)]T) \qquad (4.2)$$

~~Where $\Gamma(q,Q)\in R^{kh(q,Q)\times k}$, $T>0$, $q=1,2,\ldots,m$; $Q=1,2,\ldots,r$~~

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

$$\xi_1(\Gamma(q,Q)\mid q,Q,T)=(2\pi)^{-h(q,Q)k}\left|R(q,Q)\right|^{\frac{k}{2}}\left|T\right|^{\frac{kh(q,Q)}{2}} exp(-\tfrac{1}{2}tr \; [\Gamma(q,Q)-D(q,Q)]'R(q,Q)[\Gamma(q,Q)-D(q,Q)]T) \tag{4.3}$$

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

$$\xi_2(T)\propto\left|T\right|^{\frac{a-(k+1)}{2}} exp(-\tfrac{1}{2}tr\,\Psi T) \tag{4.4}$$

Where, Ψ is a $k\times k$ positive definite matrix.

Let β_{ij} be the prior probability that the time series realization S_n is generated from a seasonal vector moving average process of orders i and j ; i.e.

$$\beta_{ij}=P_r[q=i,Q=j],\qquad i = 1, 2, \ldots, m \;;\; j = 1, 2, \ldots, r \tag{4.5}$$

The two maximum orders of the process m and r are assumed to be known.

From (4.3), (4.4) and (4.5), the joint distribution of the parameters $\Gamma(q, Q)$, q, Q and T is

$$\xi(\Gamma(q,Q),q,Q,T)\propto\beta_{ij}(2\pi)^{\frac{h(q,Q)k^2}{2}}\left|R(q,Q)\right|^{\frac{k}{2}}\left|T\right|^{\frac{[kh(q,Q)+a-(k+1)]}{2}} exp(-\tfrac{1}{2}tr \; \{[\Gamma(q,Q)-D(q,Q)]'R(q,Q)[\Gamma(q,Q)-D(q,Q)]+\Psi\}T) \tag{4.6}$$

If one is not quite confident about the hyperparameters $D(q, Q)$, $R(q, Q)$, a and Ψ, one might use Jeffrey's vague prior

$$\xi(\Gamma(q,Q),q,Q,T)\propto\left|T\right|^{\frac{-(k+1)}{2}} \tag{4.7}$$

Combining the approximate likelihood function (4.2), via Bayes' theorem, with the prior distribution (4.6), the joint posterior distribution of the parameters $\Gamma(q, Q)$, q, Q and T is

$$f(\Gamma(q,Q),q,Q,T\mid S_n)\propto\beta_{ij}(2\pi)^{\frac{-h(q,Q)k^2}{2}}\left|R(q,Q)\right|^{\frac{k}{2}}\left|T\right|^{\frac{\alpha(q,Q)}{2}} exp(-\tfrac{1}{2}tr \; \{[\Gamma(q,Q)-D(q,Q)]'R(q,Q)[\Gamma(q,Q)-D(q,Q)]\}T) \tag{4.8}$$

Where $\alpha(q,Q)=n+h(q,Q)+a-k-1$

Theorem 4.1:

Using the approximate likelihood function (4.2) and the prior density (4.5), the joint posterior probability mass function of the orders q and Q is

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

Where

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

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

and

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

Theorem (4.1) can be proved by integrating (4.8) with respect to the parameters Γ and T respectively. The integral with respect to Γ is done by completing the squares of the exponent in (4.8) and then applying the matrix normal integral. The integral with respect to the parameter T is done using Wishart density, see Box and Tiao (1973).

Corollary 4.1:

Using the approximate likelihood (4.2) and Jeffrey's vague prior (4.7), the joint posterior probability mass function of the orders q and Q is

$$h_1(q,Q|S_n) \propto (\pi)^{\frac{hk^2}{2}} \, |A^*(q,Q)|^{\frac{-k}{2}} \, |C^*(q,Q)|^{\frac{-[n-hk]}{2}} \prod_{j=1}^{k} \Gamma\left(\frac{n-hk-k+j}{2}\right), \quad n>k+hk\text{-}1$$

Where

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

$$B^*(p,q) = \sum_{t=1}^{n} \hat{x}_{q,Q}(t-1)y'(t)$$

and

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

The form of the joint posterior probability mass function of the model orders is convenient and easy to handle with computer. Then one may calculate and inspect all posterior probabilities over the grid of the orders and choose the values q and Q at which the posterior probability mass function attains its maximum to be the most suitable orders of the vector time series data being analyzed.

5. AN EFECTIVENESS STUDY

The main objective of this section is to assess the performance of the proposed Bayesian procedures in identifying the orders of vector moving average processes with seasonality. In order to achieve this goal, five simulation studies have been conducted. The proposed Bayesian procedure is employed, with three different priors, to identify the orders of $SVMA_2(1,1)_4$ models with various parameter values. The parameters in some cases are chosen to be well inside the invertibility domain while in some other cases they are chosen to be near the boundaries. All computations are performed on Pc using SCA package.

Our main concern is to investigate the effectiveness of the proposed procedures by calculating the percentage of correct identification. Such effectiveness will be examined with respect to time series length as well as the parameters of the selected models. For all sample sizes and parameter sets the covariance matrix of the noise term is fixed at $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$. Simulation 1, for illustration, begins with generating pairs of 500 data sets of bivariate normal variates, each of size 2500, to represent the noise $\varepsilon(t)$. These data sets are then used to generate a pair of 500 realizations, each of size 2000, from $SVMA_2(1,1)$ process with the coefficients' matrices $\theta_1 = \begin{pmatrix} 0.5 & -0.4 \\ -0.3 & 0.2 \end{pmatrix}$ and $\Theta_1 = \begin{pmatrix} 0.5 & -0.4 \\ -0.3 & 0.2 \end{pmatrix}$. Note that the first 500 observations are ignored to remove the initialization effect. For a specific prior, the second step of simulation 1 is to carry out all computations, assuming a specific maximum order (m, r), required to identify each of the 500 realizations and to find the percentage of correct identification. Such computations are done for a specific time series length n using the first n observations of each generated realization. The second step is repeated for each chosen time series and prior combinations. The sample size n is taken to be 200, 400, 600, 800, 1500 and 2000, while the maximum order (m, r) is taken to be (3, 3). With respect to the prior probability mass function of the orders q and Q, which is combined with Jeffreys' vague prior of $\Gamma(q, Q)$ and T, the following three prior distributions are used:

Prior 1:

$$\beta_{ij} = P_r[q=i, Q=j] = 1/9 \qquad i=1,2,3 \ ; \ j=1,2,3$$

Prior 2:

$$\beta_{ij} \propto (0.5)^{i+j} \ , \ i=1,2,3 \ ; \ j=1,2,3$$

Prior 3:

$$\beta_{11} = 0.211, \ \beta_{12} = \beta_{21} = 0.161, \ \beta_{13} = \beta_{31} = \beta_{22} = 0.111,$$
$$\beta_{23} = \beta_{32} = 0.061 \ \ and \ \ \beta_{33} = 0.011$$

The first prior assigns equal probabilities to each combination of orders. The second prior assigns probabilities that decline exponentially with the orders, while the third prior is chosen in such a way to give probabilities that decrease with constant value 0.05 as the order increases.

Simulation 2 is done in a similar way but using $\theta_1 = \begin{pmatrix} -0.2 & -0.2 \\ -0.2 & -0.2 \end{pmatrix}$ and $\Theta_1 = \begin{pmatrix} 0.5 & -0.4 \\ -0.3 & 0.2 \end{pmatrix}$. The other simulations are done in a similar way but using different values of θ_1 and Θ_1 and their result are presented in table 1. The parameters in the simulations have been chosen in such a way to satisfy the invertibility conditions, see Harvey (1981). In some simulations, the parameters are chosen to be nearby the non-invertibility region, see case 5.

Table (1): Percentages of Correct Identification for SVMA$_2$(1,1) Processes with maximum order (3, 3)

PARAMETERS θ, Θ	*n*	*PRIOR1*	*PRIOR2*	*PRIOR3*
$\begin{bmatrix} 0.5 & -0.4 \\ -0.3 & 0.2 \end{bmatrix}, \begin{bmatrix} 0.5 & -0.4 \\ -0.3 & 0.2 \end{bmatrix}$	*200*	88.2	98.6	97.4
	400	93.8	100.0	99.8
	600	96.2	100.0	100.0
	800	96.0	100.0	100.0
	1500	98.4	100.0	100.0
	2000	99.0	100.0	100.0
$\begin{bmatrix} -0.2 & -0.2 \\ -0.2 & -0.2 \end{bmatrix}, \begin{bmatrix} 0.5 & -0.4 \\ -0.3 & 0.2 \end{bmatrix}$	*200*	90.6	100.0	98.8
	400	96.8	100.0	100.0
	600	98.8	100.0	100.0
	800	99.8	100.0	100.0
	1500	99.8	100.0	100.0
	2000	99.6	100.0	100.0

$\begin{bmatrix} 0.5 & -0.4 \\ -0.3 & 0.2 \end{bmatrix}, \begin{bmatrix} -0.2 & -0.2 \\ -0.2 & -0.2 \end{bmatrix}$	***200***	90.4	100.0	98.4
	400	97.0	100.0	100.0
	600	98.6	100.0	100.0
	800	99.2	100.0	100.0
	1500	99.6	100.0	100.0
	2000	100.0	100.0	100.0
$\begin{bmatrix} 0.2 & 0.2 \\ 0.2 & 0.2 \end{bmatrix}, \begin{bmatrix} 0.2 & 0.2 \\ 0.2 & 0.2 \end{bmatrix}$	***200***	94.4	100.0	99.8
	400	98.8	100.0	100.0
	600	99.6	100.0	100.0
	800	99.8	100.0	100.0
	1500	100.0	100.0	100.0
	2000	100.0	100.0	100.0
$\begin{bmatrix} 1.1 & 0.3 \\ -0.6 & 0.2 \end{bmatrix}, \begin{bmatrix} -0.2 & -0.2 \\ -0.2 & -0.2 \end{bmatrix}$	***200***	49.2	90.8	76.6
	400	58.2	96.2	88.2
	600	66.0	98.0	91.8
	800	70.8	98.8	93.0
	1500	74.4	99.2	95.2
	2000	78.8	99.2	95.8

Source: Simulated Data

Inspection of the numerical results shows that the percentages of correct identification increase as the time series length n increases for all models and priors. The percentages of correct identification are reasonably high, being greater than 70 % for all models and priors, for time series length 800 or more no matter what the coefficients are. In addition, the percentages of correct identification achieved by prior 3 are higher than the corresponding percentages achieved by prior 1, while the corresponding percentages achieved by prior 2 are the highest. However, for sufficiently large n, one may notice that the differences among the percentages of correct identification achieved by the three priors tend to die down. This means that the proposed Bayesian identification procedure is not very sensitive to the minor changes between the three prior distributions.

Considering the above comments, one may conclude that the numerical results support the adequacy of using the proposed Bayesian procedure in solving the identification problems of vector moving average processes with seasonality.

6. CONCLUSION

The article has proposed an approximate Bayesian procedure to identify the orders of vector moving average processes with seasonality. The joint posterior probability mass function of the model orders has been developed in a convenient from using an approximate likelihood function and a matrix normal–Wishart (or Jeffreys' vague) prior. Then one may easily calculate and investigate the posterior probabilities of all values of the problem. In order to check the performance and quality of the proposed procedure, a simulation study with three different prior distributions has been conducted. The numerical results show that the proposed Bayesian procedure can efficiently identify the orders of bivariate seasonal moving average processes with high precision for moderate and large time series lengths.

REFERENCES

(1) Akaike, H. (1974). A New Look at Statistical Model Identification. *IEEE Transaction on Automatic Control* 19:716–723.

(2) Beveridge, S. and Oickle, C. (1994). A Comparison of Box–Jenkins and Objective Methods for Determining the Order of a Non-Seasonal ARMA Model. *Journal of Forecasting*13:419–434.

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

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

(5) Brockwell, Peter J. and Davis, Richard A. (2016).Introduction To Time Series And Forecasting, 3rd ed. Springer International Publishing

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

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

(8) Hannan, E. J. and Quinn, B. G. (1979). The Determination of an Autoregression. *Journal of the Royal Statistical Society, Series B* 42:190–195.

(9) Harvey, A. C. (1981). Time Series Models. *New York: Wiley.*

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

(11) Liu, L. M. (2006). Time Series Analysis and Forecasting. *2nd ed. Villa Park, IL: Scientific Computing Associates.*

(12) Mills, J. and Prasad, K. (1992). A Comparison of Model Selection criteria. *Econometric Reviews*11:201–233.

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

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

(15) Rissanen, J. (1978). Modeling by Shortest Data Description. *Automatica*, 14:465-471.

(16) Shaarawy, S.M. and Ali, S.S. (2003). Bayesian Identification of Seasonal Autoregressive Models. *Communications in Statistics-Theory and Methods,* 32:1067–1084.

(17) Shaarawy, S.M. and Ali, S.S. (2008). Bayesian Identification of Vector Autoregressive Processes. *Communications in Statistics-Theory and Methods,* 37:791–802.

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

(19) Shaarawy, S. M, and Ali, S. S. (2015). Bayesian Identification of Seasonal Vector Autoregressive. *Communications in Statistics-Theory and Methods*, 44: 823-838.

(20) Shaarawy, S. M., Al Bassam, M. and Ali, S. S. (2006). A Direct Bayesian Procedure to Select the Order of Bivariate Autoregressive Process. *The Egyptian Statistical Journal*, 50(1):1-22.

(21) Shaarawy, S.M., Soliman, E.E.A. and Ali, S.S. (2007). Bayesian Identification of Moving Average Models. *Communications in Statistics- Theory and Methods*, 36(12):2301-2312.

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

(23) Tiao, G. C. and Tsay, R. (1983). Multiple Time Series Modeling and Extended Sample Cross Correlation. *Journal of Business and Economic Statistics,* 1:43–56.