# Bayesian Analysis of the Seasonal Moving Average Model: A Gibbs Sampling Approach

## M.A. Ismail

United Arab Emirates University, P.O. Box 17555, Al-Ain, United Arab Emirates

#### **Abstract**

This paper develops a Bayesian analysis for the multiplicative seasonal moving average model by implementing a fast, easy and accurate Gibbs sampling algorithm. The proposed algorithm does not involve any Metropolis-Hastings generation but is generated from normal and inverse gamma distributions. The problem of forecasting multiple future observations is considered. The proposed algorithm is illustrated using a simulated example and airline data. Unlike the classical approach, by using the airline data, the proposed algorithm is easily used to test the significance of the interaction parameter.

Key words: airline data, multiplicative model, posterior distribution, prediction

(Mohamed A. Ismail).

## 1. INTRODUCTION

Seasonal moving average modeling of time series has been successfully applied in a great number of fields including economic forecasting. Bayesian analysis of moving average type models is difficult even for nonseasonal models since the likelihood function is nonlinear in the parameters, which causes problems in prior specification and posterior analysis. Several authors have considered Bayesian analysis of moving average models, e.g. Newbold (1973), Monahan (1983), Broemeling and Shaarawy (1984), Shaarawy and Ismail (1987) and Marriott and Smith (1992), among others.

Recently, Bayesian time series analysis has been advanced by the emergence of Markov Chain Monte Carlo (MCMC) methods, especially the Gibbs sampling method. Assuming a prior distribution on the initial observations and initial errors, Chib and Greenberg (1994) and Marriott et al. (1996) developed a Bayesian analysis for autoregressive moving average (ARMA) models using MCMC technique. Barnett et al. (1996, 1997) used MCMC methodology to estimate the multiplicative seasonal autoregressive and autoregressive moving average model. Their algorithm was based on sampling functions of the partial autocorrelations. However, the algorithm of Barnett et al. is complicated and does not deal with the forecasting problem. Chen (1992) employed Gibbs sampling to develop Bayesian analysis of bilinear time series models. Her approach was based on approximating the likelihood function via estimating the unobserved errors. The approximate

likelihood is then used to derive the conditional distributions required for implementing the Gibbs sampler. A similar approach to that of Chen (1992) is used in this study.

In this paper, we present a simple and fast algorithm to estimate the parameters and forecast future values of multiplicative seasonal moving average models. The proposed algorithm does not involve any Metropolis-Hastings iteration which is an advantage over the existing algorithms in the literature. Our analysis is unconditional on the initial errors; that is, we assume that the series starts at time t=1 with unknown initial errors. The Bayesian analysis of the seasonal moving average model does not enforce invertibility of the model. However, for many practical reasons, one may wish to assume the invertibility of the model. One way to do so is to restrict the parameter space to satisfy invertibility conditions as in Barrnett et al. (1997) and Marriott et al. (1996), among others. Rather than restricting the parameter space to satisfy invertibility conditions, the process could be made invertible by choosing hyperparameters which ensure that the prior for the model coefficients lies in the invertibility region. The latter approach was used by Broemeling (1985) and McCulloch and Tsay (1994), among others, and is going to be used in this paper.

The Bayesian approach for analyzing seasonal MA models has several advantages over the classical approach, the obvious one being pedagogical. In contrast to the classical approach, where a large variety of ad hoc sampling theory techniques are needed, the Bayesian approach unifies and simplifies the analysis. For example, there is no discontinuity of the analysis between invertibility and noninvertibility processes. Moreover, the Bayesian approach allows for incorporating uncertainty about the parameters and initial errors. In addition, various features of the seasonal MA models, which may be complicated to check in the classical framework, may be routinely tested in the sampling based Bayesian framework. As an example, there is often interest in testing the significance of interaction parameters which are the product of the nonseasonal and seasonal coefficients in the model. The proposed algorithm can easily construct confidence intervals for interaction parameters and therefore test their significance.

The paper is organized as follows. Section 2 briefly describes the multiplicative seasonal moving average model. Section 3 is devoted to summarizing posterior analysis and the full conditional posterior distributions of the parameters. In Section 4, the multiple forecast problem is discussed. The implementation details of the proposed algorithm, including convergence monitoring, are given in Section 5. The proposed methodology is illustrated in Section 6 using a simulated example and the airline data presented in Box and Jenkins (1976). Moreover, the significance of the interaction parameter in the identified model is tested using the suggested algorithm. Finally, conclusions are given in Section 7.

## 2. THE MULTIPLICATIVE SEASONAL MOVING AVERAGE MODEL

A time series  $\{y_t\}$  is said to be generated by a multiplicative seasonal moving average model of orders q and Q, denoted by  $SMA(q, Q)_s$ , if it satisfies

$$y_t = \theta_o(B)\Theta_O(B^s)\varepsilon_t,\tag{1}$$

where  $\{\varepsilon_t\}$  is a sequence of independent normal variates with zero mean and variance  $\sigma^2$ . The backshift operator B is defined as  $By_t = y_{t-1}$ , and s is the seasonal period. The nonseasonal moving average polynomial is  $\theta_q(B) = \left(1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q\right)$  with order q and  $\Theta_Q(B^s) = \left(1 + \Theta_1 B^s + \Theta_2 B^{2s} + \dots + \Theta_Q B^{Qs}\right)$  is the seasonal moving average polynomial with order Q. The nonseasonal and seasonal moving average coefficients are  $\theta = \left(\theta_1, \theta_2, \dots, \theta_q\right)^T$  and  $\Theta = \left(\Theta_1, \Theta_2, \dots, \Theta_Q\right)^T$ . The time series  $\{y_t\}$  is assumed to start at time t = 1 with unknown starting errors  $\varepsilon_0 = \left(\varepsilon_0, \varepsilon_{-1}, \dots, \varepsilon_{1-q-Qs}\right)$ .

The model (1) can be written as

$$y_{t} = \sum_{i=1}^{q} \theta_{i} \varepsilon_{t-i} + \sum_{j=1}^{Q} \Theta_{js} \varepsilon_{t-js} + \sum_{i=1}^{q} \sum_{j=1}^{Q} \theta_{i} \Theta_{j} \varepsilon_{t-i-js} + \varepsilon_{t}$$

$$= \Lambda_{t} \beta + \varepsilon_{t}, \qquad (2)$$

where

$$\Lambda_{t} = \left(\varepsilon_{t-1}, \dots, \varepsilon_{t-q}; \mathbf{0}; \varepsilon_{t-s}, \varepsilon_{t-s-1}, \dots, \varepsilon_{t-s-q}; \mathbf{0}; \dots; \varepsilon_{t-Qs}, \varepsilon_{t-Qs-1}, \dots, \varepsilon_{t-Qs-q}\right),$$

$$\beta = \left(\theta_{1}, \dots, \theta_{q}; \mathbf{0}; \Theta_{1}, \theta_{1}\Theta_{1}, \dots, \theta_{q}\Theta_{1}; \mathbf{0}; \dots; \Theta_{Q}, \theta_{1}\Theta_{Q}, \dots, \theta_{q}\Theta_{Q}\right)^{T}, \tag{3}$$

and  $\mathbf{0}$  is a (s-q-1) row vector of zeros. Model (2) shows that the multiplicative seasonal moving average model can be written as a moving average model of order q+Qs with some zero coefficients and some coefficients that are products of non seasonal and seasonal coefficients. Thus, the model is nonlinear in  $\theta$  and  $\Theta$  which complicates the Bayesian analysis. However, the following sections explain how the Gibbs sampling technique can facilitate the analysis. The seasonal moving average model (2) is invertible if the roots of the polynomials  $\theta_q(B)$  and  $\Theta_Q(B^s)$  lie outside the unit circle. For more details about the properties of seasonal moving average models see Box and Jenkins (1976).

## 3. POSTERIOR ANALYSIS

## 3.1. Likelihood Function

Suppose  $\mathbf{y} = (y_1, y_2, \dots, y_n)$  is a realization of the seasonal moving average model (2). Assuming that  $f(\varepsilon_t) = \mathbf{N}$  (0,  $\sigma^2$ ) and employing a straightforward random variable transformation from  $\varepsilon_t$  to  $y_t$ , the likelihood function  $L(\theta, \Theta, \sigma^2, \varepsilon_0 \mid \mathbf{y}) = l$  is given by

$$l \propto (\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{t=1}^n \left( y_t - \sum_{i=1}^q \theta_i \varepsilon_{t-i} - \sum_{j=1}^Q \Theta_{js} \varepsilon_{t-js} - \sum_{i=1}^q \sum_{j=1}^Q \theta_i \Theta_j \varepsilon_{t-i-js} \right)^2 \right\}. \tag{4}$$

The likelihood function (4) is a complicated function in  $\theta$ ,  $\Theta$  and  $\varepsilon_0$ . Suppose the errors are estimated recursively as

$$e_{t} = y_{t} - \sum_{i=1}^{q} \hat{\theta}_{i} e_{t-i} - \sum_{j=1}^{Q} \hat{\Theta}_{js} e_{t-js} - \sum_{i=1}^{q} \sum_{j=1}^{Q} \hat{\theta}_{i} \hat{\Theta}_{js} e_{t-i-js},$$
 (5)

where  $\hat{\theta}_i \in R^q$ ,  $\hat{\Theta}_j \in R^Q$  are sensible estimates and  $e_{1-q-Qs} = \cdots = e_{-1} = e_0 = 0$ . Several estimates

tion methods, such as the Innovations Substitution (IS) method proposed by Koreisha and Pukkila (1990), give consistent estimates for  $\theta$  and  $\Theta$ . The idea of the IS method is to fit a long autoregressive model to the series and obtain the residuals. Then appropriate lagged residuals are substituted into the SMA model (2). Lastly, the parameters are estimated using ordinary least squares.

Substituting the residuals in (4) results in an approximate likelihood function

$$l^{\star} \propto (\sigma^{2})^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^{2}} \sum_{t=1}^{n} \left(y_{t} - \sum_{i=1}^{q} \theta_{i} e_{t-i} - \sum_{j=1}^{Q} \Theta_{js} e_{t-js} - \sum_{i=1}^{q} \sum_{j=1}^{Q} \theta_{i} \Theta_{j} e_{t-i-js}\right)^{2}\right\}$$

$$= (\sigma^{2})^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^{2}} (\mathbf{y} - X\beta)^{T} (\mathbf{y} - X\beta)\right\},$$
(6)

where  $\beta$  is defined in (3) and X is an  $n \times (q + Qs)$  matrix with the t-th row

$$X_{t} = (e_{t-1}, \dots, e_{t-q}; \mathbf{0}; e_{t-s}, e_{t-s-1}, \dots, e_{t-s-q}; \mathbf{0}; \dots; e_{t-Qs}, e_{t-Qs-1}, \dots, e_{t-Qs-q}),$$

( $\mathbf{0}$  is a (s-q-1) row vector of zeros).

## 3.2. Prior Specification

Suppose that, given the error variance parameter  $\sigma^2$ , the parameters  $\theta$ ,  $\Theta$  and  $\varepsilon_0$  are independent apriori, i.e.

$$\begin{split} \zeta\left(\theta,\boldsymbol{\Theta},\sigma^{2},\varepsilon_{\boldsymbol{0}}\right) &= \zeta\left(\theta\mid\sigma^{2}\right)\times\zeta\left(\boldsymbol{\Theta}\mid\sigma^{2}\right)\times\zeta\left(\varepsilon_{0}\mid\sigma^{2}\right)\times\zeta\left(\sigma^{2}\right) \\ &= N_{q}(\mu_{\theta},\sigma^{2}\Sigma_{\theta})\times N_{Q}(\mu_{\boldsymbol{\Theta}},\sigma^{2}\Sigma_{\boldsymbol{\Theta}})\times N_{q+Qs}(\mu_{\boldsymbol{0}},\sigma^{2}\Sigma_{0})\times IG(\frac{\nu}{2},\frac{\lambda}{2}), \end{split}$$

where  $N_r(\mu, \Delta)$  is the r-variate normal distribution with mean  $\mu$  and variance  $\Delta$  and IG(a, b) is the inverse gamma distribution with parameters a and b. The joint prior distribution can then be written as

$$\zeta\left(\theta, \mathbf{\Theta}, \sigma^{2}, \varepsilon_{0}\right) \propto (\sigma^{2})^{-\left(\frac{\gamma+2q+Q+Q+Q}{2}+1\right)} \exp\left\{-\frac{1}{2\sigma^{2}}\left[\lambda + (\theta - \mu_{\theta})^{T} \Sigma_{\theta}^{-1} (\theta - \mu_{\theta})\right] + (\mathbf{\Theta} - \mu_{\Theta})^{T} \Sigma_{\Theta}^{-1} (\mathbf{\Theta} - \mu_{\Theta}) + (\varepsilon_{0} - \mu_{0})^{T} \Sigma_{0}^{-1} (\varepsilon_{0} - \mu_{0})\right]\right\}.$$
(7)

The prior distribution (7) is chosen for several reasons. This class of priors is flexible enough to be used in numerous applications. It also facilitates the mathematical calculations. It is, at least conditionally, the conjugate prior for the approximate conditional likelihood.

Multiplying the joint prior distribution (7) by the likelihood function (6) results in the joint posterior  $\zeta(\theta, \Theta, \sigma^2, \varepsilon_0 \mid \mathbf{y})$ , which may be written as

$$\zeta\left(\theta, \mathbf{\Theta}, \sigma^{2}, \varepsilon_{\mathbf{0}} \mid \mathbf{y}\right) \propto (\sigma^{2})^{-\left(\frac{n+\nu+2q+Q+Qs}{2}+1\right)} \exp\left\{-\frac{1}{2\sigma^{2}}\left[\lambda + (\theta - \mu_{\theta})^{T} \Sigma_{\theta}^{-1} (\theta - \mu_{\theta}) + (\mathbf{\Theta} - \mu_{\theta})^{T} \Sigma_{\mathbf{\Theta}}^{-1} (\theta - \mu_{\theta}) + (\varepsilon_{\mathbf{0}} - \mu_{\mathbf{0}})^{T} \Sigma_{\mathbf{0}}^{-1} (\varepsilon_{\mathbf{0}} - \mu_{\mathbf{0}}) + (\mathbf{y} - X\beta)^{T} (\mathbf{y} - X\beta)\right]\right\}.$$
(8)

#### 3.3. Full Conditional Distributions

The posterior conditional distribution for each of the unknown parameters is obtained from the joint posterior distribution (8) by grouping together terms in the joint posterior that depend on this parameter, and finding the appropriate normalizing constant to form a proper density. In this study all posterior conditionals are normal and inverse gamma distributions for which sampling techniques exist.

# a. The conditional posterior of $\theta$

$$\theta^{j} \sim \zeta\left(\theta^{j} \mid \mathbf{y}, (\sigma^{2})^{j-1}, \Theta^{j-1}, \varepsilon_{\mathbf{0}}^{j-1}\right) = N(\mu_{\theta}^{\star}, \sigma^{2} v_{\theta}^{\star}) ,$$

where

$$\mu_{\theta}^{\star} = \left(H^T H + \Sigma_{\theta}^{-1}\right)^{-1} \left(H^T \mathbf{y}\right), \quad v_{\theta}^{\star} = \left(H^T H + \Sigma_{\theta}^{-1}\right)^{-1}$$

and H is a  $(n \times q)$  matrix with ti-th element  $H_{ti} = \left(e_{t-i} + \sum_{j=1}^{Q} \Theta_{j} e_{t-js-i}\right)$ .

# b. The conditional posterior of $\Theta$

$$\Theta^{j} \sim \zeta \left( \Theta^{j} \mid \mathbf{y}, (\sigma^{2})^{j-1}, \theta^{j-1}, \varepsilon_{\mathbf{0}}^{j-1} \right) = N(\mu_{\Theta}^{\star}, \sigma^{2} v_{\Theta}^{\star}) ,$$

where

$$\boldsymbol{\mu}_{\Theta}^{\star} = \left(\boldsymbol{G}^T \boldsymbol{G} + \boldsymbol{\Sigma}_{\Theta}^{-1}\right)^{-1} \left(\boldsymbol{G}^T \mathbf{y}\right) \;, \;\; \boldsymbol{v}_{\Theta}^{\star} = \left(\boldsymbol{G}^T \boldsymbol{G} + \boldsymbol{\Sigma}_{\Theta}^{-1}\right)^{-1}$$

and G is a  $(n \times Q)$  matrix with elements  $G_{ti} = (e_{t-is} + \sum_{i=1}^{q} \theta_i e_{t-is-j})$ .

# c. The conditional posterior of $\sigma^2$

$$(\sigma^2)^j \sim \zeta\left((\sigma^2)^j \mid \mathbf{y}, \theta^{j-1}, \Theta^{j-1}, \varepsilon_{\mathbf{0}}^{j-1}\right) = IG\left(\frac{\nu + n}{2}, \frac{\lambda + n(S^2)^j}{2}\right) ,$$

where

$$\begin{split} n(S^2)^j &= \left(\mathbf{y} - X\beta^j\right)^T (\mathbf{y} - X\beta^j) \left(\theta^j - \mu_\theta\right)^T \Sigma_\theta^{-1} \left(\theta^j - \mu_\theta\right) + \left(\mathbf{\Theta}^j - \mu_\Theta\right)^T \Sigma_\Theta^{-1} \left(\mathbf{\Theta}^j - \mu_\Theta\right) \\ &+ \left(\varepsilon_\mathbf{0}^j - \mu_\mathbf{0}\right)^T \Sigma_0^{-1} \left(\varepsilon_\mathbf{0}^j - \mu_\mathbf{0}\right). \end{split}$$

Thus, the parameter  $(\sigma^2)^j$  can be sampled from the Chi-square distribution using the transformation  $\frac{\lambda + n(S^2)^j}{(\sigma^2)^j} \sim \chi^2_{\gamma + n}$ .

## d. The conditional Posterior for $\varepsilon_0$

Using the model (2), the equations for the elements of  $\varepsilon_0$  can be written as

$$\mathbf{y}_{\mathbf{q}+\mathbf{O}\mathbf{s}} = D \, \boldsymbol{\varepsilon}_{\mathbf{0}} + F \, \boldsymbol{\varepsilon}_{\mathbf{q}+\mathbf{O}\mathbf{s}},$$

where

Bayesian Analysis of the Seasonal Moving Average Model: A Gibbs Sampling Approach

$$D = \begin{pmatrix} \beta_1 & \beta_2 & \beta_3 & \cdots & \beta_{q+Qs-1} & \beta_{q+Qs} \\ \beta_2 & \beta_3 & \beta_4 & \cdots & \beta_{q+Qs} & 0 \\ \beta_3 & \beta_4 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ \beta_{q+Qs-1} & \beta_{q+Qs} & 0 & \cdots & 0 & 0 \\ \beta_{q+Qs} & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}_{(q+Qs)\times(q+Qs)}$$

and

$$F = \begin{pmatrix} 1 & 0 & 0 & \cdots & \cdots & 0 & 0 \\ \beta_1 & 1 & 0 & \cdots & \cdots & 0 & 0 \\ \beta_2 & \beta_1 & 1 & \cdots & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ \beta_{q+Qs-1} & \beta_{q+Qs-2} & \cdots & \cdots & \beta_2 & \beta_1 & 1 \end{pmatrix}_{(q+Qs)\times(q+Qs)}.$$

 $\mathbf{y_{q+Qs}} = \left(y_1, y_2, \cdots, y_{q+Qs}\right)^T$  and  $\varepsilon_{\mathbf{q+Qs}} = \left(\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_{q+Qs}\right)^T$ , which has the (q+Qs) normal distribution with zero mean and variance  $(\sigma^2 I_{q+Qs})$  where  $I_{q+Qs}$  is the unit matrix of order (q+Qs). Using linear regression results and standard Bayesian techniques, the conditional posterior of  $\varepsilon_0$  is

$$\varepsilon_{\mathbf{0}}^{j} \sim \zeta\left(\varepsilon_{\mathbf{0}}^{j} \mid \mathbf{y}, (\sigma^{2})^{j-1}, \theta^{j-1}, \Theta^{j-1}\right) = N(\mu_{0}^{\star}, \sigma^{2}v_{0}^{\star}) \ ,$$

where

$$\mu_0^{\star} = \left[ D^T (FF^T)^{-1} D + \Sigma_0^{-1} \right]^{-1} \left[ D^T (FF^T)^{-1} \mathbf{y_{q+Qs}} \right] \; , \; \; v_0^{\star} = \left[ D^T (FF^T)^{-1} D + \Sigma_0^{-1} \right]^{-1} \; .$$

## 4. PREDICTIVE ANALYSIS

In this section we obtain Bayesian forecasts for k periods ahead. Let  $\mathbf{y}_f = (y_{n+1}, \dots, y_{n+k})$  be the forecasts vector. The predictive distribution is

$$\zeta(\mathbf{y}_f \mid \mathbf{y}) = \int_{\theta} \int_{\Theta} \int_{\sigma^2} \int_{\varepsilon_0} \zeta(\mathbf{y}_f, \theta, \Theta, \sigma^2, \varepsilon_0 \mid \mathbf{y}) \ d\theta \ d\Theta \ d\sigma^2 \ d\varepsilon_0.$$

This predictive distribution does not have a closed analytical form and it is also hard to calculate the integration numerically. However, the Gibbs sampling technique can be employed to compute the predictive distribution. Now, the equations for the k future observations can be written as

$$\mathbf{y}_f = R \, \mathbf{e}_L + E \, \varepsilon_f,$$

where

$$R = \begin{pmatrix} \beta_1 & \beta_2 & \beta_3 & \cdots & \cdots & \beta_{q+Qs-1} & \beta_{q+Qs} \\ \beta_2 & \beta_3 & \beta_4 & \cdots & \cdots & \beta_{q+Qs} & 0 \\ \beta_3 & \beta_4 & \cdots & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \cdots & \vdots & \vdots \\ \beta_k & \beta_{k+1} & \cdots & \beta_{q+Qs} & 0 & \cdots & 0 \end{pmatrix}_{k \times (q+Qs)}$$

and

$$E = \begin{pmatrix} 1 & 0 & 0 & \cdots & \cdots & 0 & 0 \\ \beta_1 & 1 & 0 & \cdots & \cdots & 0 & 0 \\ \beta_2 & \beta_1 & 1 & \cdots & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \cdots & \vdots & \vdots \\ \beta_{k-1} & \beta_{k-2} & \cdots & \cdots & \beta_2 & \beta_1 & 1 \end{pmatrix}_{k \times k},$$

with 
$$\mathbf{e}_L = (e_n, e_{n-1}, \dots, e_{n-q-Qs+1})^T$$
 and  $\varepsilon_f = (\varepsilon_{n+1}, \dots, \varepsilon_{n+k})^T$ .

It should be noted that writing the future values in this way facilitates writing the predictive distribution for any number of future values conditional on the model parameters and initial errors. This enables the generation of any number of future realizations from the seasonal model that is needed for the Gibbs sampler. Since  $\varepsilon_f \sim N(0, \sigma^2 I_k)$  with  $\varepsilon_f$  and  $\varepsilon$  being independent, the conditional predictive density  $\zeta(\mathbf{y}_f \mid \mathbf{y}, \theta, \Theta, \sigma^2, \varepsilon_0)$  is

$$\zeta(\mathbf{y}_f \mid \mathbf{y}, \theta, \Theta, \sigma^2, \varepsilon_0) = N(R \mathbf{e}_L, \sigma^2(EE^T)).$$
(9)

The Bayesian forecasts of the future values can be obtained via extending the posterior analysis directly by stretching the parameter vector to include  $\mathbf{y}_f$ . That is, the above posterior analysis will apply conditionally on the future observations  $\mathbf{y}_f$  where the elements of  $\mathbf{y}_f$  are sampled from the conditional predictive distribution (9).

## 5. THE PROPOSED GIBBS SAMPLER

The proposed Gibbs sampling algorithm for the  $SMA(q, Q)_s$  model can be implemented as follows:

- 1. Specify starting values  $\theta^0$ ,  $\Theta^0$ ,  $(\sigma^2)^0$ ,  $\varepsilon_0^0$  and  $\mathbf{y}_f^0$  and set j=0. A set of initial estimates of the model parameters can be obtained using the IS technique of Koreisha and Pukkila (1990).
- 2. Calculate the residuals recursively using (5) and the IS parameter estimates.
- Simulate

$$\bullet \theta^{j} \sim \zeta \left(\theta^{j} \mid \mathbf{y}, (\sigma^{2})^{j-1}, \mathbf{\Theta}^{j-1}, \varepsilon_{\mathbf{0}}^{j-1}, \mathbf{y}_{f}^{j-1}\right),$$

$$\bullet \mathbf{\Theta}^{j} \sim \zeta \left(\mathbf{\Theta}^{j} \mid \mathbf{y}, (\sigma^{2})^{j-1}, \theta^{j-1}, \varepsilon_{\mathbf{0}}^{j-1}, \mathbf{y}_{f}^{j-1}\right),$$

$$\bullet (\sigma^{2})^{j} \sim \zeta \left((\sigma^{2})^{j} \mid \mathbf{y}, \theta^{j-1}, \mathbf{\Theta}^{j-1}, \varepsilon_{\mathbf{0}}^{j-1}, \mathbf{y}_{f}^{j-1}\right),$$

$$\bullet \varepsilon_{\mathbf{0}}^{j} \sim \zeta \left(\varepsilon_{\mathbf{0}}^{j} \mid \mathbf{y}, (\sigma^{2})^{j-1}, \theta^{j-1}, \mathbf{\Theta}^{j-1}, \mathbf{y}_{f}^{j-1}\right),$$
and
$$\bullet \mathbf{y}_{f}^{j} \sim \zeta \left(\mathbf{y}_{f}^{j} \mid \mathbf{y}, (\sigma^{2})^{j-1}, \theta^{j-1}, \mathbf{\Theta}^{j-1}, \varepsilon_{\mathbf{0}}^{j-1}\right).$$
4. Set  $j = j + 1$  and go to 3.

This algorithm gives the next value of the Markov chain  $\{\theta^{j+1}, \mathbf{\Theta}^{j+1}, (\sigma^2)^{j+1}, \varepsilon_0^{j+1}, \mathbf{y}_f^{j+1}\}$  by simulating each of the full conditionals where the conditioning elements are revised during the cycle. It should be noted that the first four posterior conditionals in Step 3 are those obtained in Section 3 with *n* replaced by n+k and the conditional predictive distribution is given by (9). This iterative pro-

cess is repeated for a large number of iterations and convergence is monitored. After the chain has converged, say at  $n_0$  iterations, the simulated values  $\{\theta^{j+1}, \mathbf{\Theta}^{j+1}, (\sigma^2)^{j+1}, \varepsilon_0^{j+1}, \mathbf{y}_f^{j+1}, j > n_0\}$ , are used as a sample from the joint posterior. Posterior estimates of the parameters are computed directly via sample averages of the simulation outputs. In what follows three groups of convergence diagnostics are presented.

- Autocorrelation estimates indicating how much independence exists in the sequence of draws for each parameter. A high degree of autocorrelation indicates that more draws are needed to get accurate posterior estimates.
- 2. Raftery and Lewis diagnostics. Raftery and Lewis(1992,1995) proposed a set of diagnostics that include
  - a thinning ratio (Thin), which is a function of the autocorrelation in the draws,
  - the number of draws (Burn) to use for initial draws or "burn-in" before starting to sample the draws for the purpose of posterior inference,
  - the total number of draws (Total) needed to achieve the desired level of accuracy,
  - the number of draws (Nmin) that would be needed if the draws represented an iid chain, and
  - (I-stat), which is the ratio of (Total) to (Nmin). They suggested that convergence problems may be indicated when values of (I-stat) exceed 5.
- 3. Geweke diagnostics. Geweke (1992) proposed two groups of diagnostics.
- (a) The first group includes the numerical standard errors (NSE) and relative numerical efficiency (RNE). The NSE estimates reflect the variation that can be expected if the simulation were to be repeated. The RNE estimates indicate the number of draws required to produce the same numerical accuracy when an iid sample is drawn directly from the posterior distribution. The estimates of NSE and RNE are based on spectral analysis of time series where two sets of these estimates are obtained. The first set is based on the assumption that the draws come from an iid process. The second set is based on different tapering or truncating of the periodogram window. When there are big differences between the two sets, the second set of estimates would be chosen because it would take into consideration autocorrelations in the draws.
- (b) The second group of diagnostics includes a test of whether the sampler has attained an equilibrium state. This is done by carrying out a Z-test for the equality of the two means of the first and last parts of draws and the Chi-squared marginal probability is obtained. Usually, the first and last parts are the first 20% and the last 50% of the draws.

LeSage (1999) implemented calculations of the above convergence measures using the Matlab package. These diagnostics will be used in Section 6 to monitor the convergence of the proposed algorithm.

## 6. ILLUSTRATIVE EXAMPLES

#### Simulated Data

In this subsection we present a simulated example to evaluate the efficiency of the proposed methodology. The example deals with generating 250 observations from the following  $SMA(1, 1)_4$ :  $y_t = 0.7\varepsilon_{t-1} + 0.4\varepsilon_{t-4} + 0.28\varepsilon_{t-5} + \varepsilon_t$ .

The analysis was implemented using Matlab and run on a Pentium 1000 MHZ PC. Runs took two minutes (on average) to complete. The error variance  $\sigma^2$  was chosen to be 1 and a non informative prior was assumed for  $\theta$ ,  $\Theta$  and  $\sigma^2$  via setting  $\Sigma_{\theta}^{-1} = \Sigma_{\Theta}^{-1} = 0$ ,  $\lambda = S^2$  and  $\nu_0 = 3$ . A normal prior with zero mean and variance  $\sigma^2 I_{q+Qs}$  was used for the initial error vector  $\varepsilon_0$ . The starting values for the parameters  $(\theta, \Theta, \sigma^2)$  were obtained using the IS method. The starting values for  $\varepsilon_0$  and  $y_f$  were assumed to be zero.

The Gibbs sampler was run with 12000 iterations, where the first 2000 draws were ignored and every tenth value in the sequence of the last 10000 draws was recorded, to obtain an approximately independent sample. All posterior and predictive estimates were computed as sample averages of the simulated outputs.

Table 1 presents the true values and Bayesian estimates of the parameters and the next four future values. Moreover, a 95% confidence interval using the 0.025 and 0.975 percentiles of the simulated draws is constructed for every parameter and future value. From Table 1, it is clear that Bayesian estimates are close to the true values and the 95% confidence interval includes the true value for every parameter and future value. Sequential plots of the parameter generated sequences together with marginal densities are displayed in Figure 1. The marginal predictive densities for the next four observations are displayed in Figure 2. The marginal densities are computed using the non-parametric technique with Gaussian kernel. Figure 1 shows that the proposed algorithm is stable and fluctuates in the neighborhood of the true values. In addition, the marginal densities show that the true value of each parameter (which is indicated by the vertical line) falls in the constructed 95% confidence interval. Similar conclusions for the predictive performance of the proposed algorithm may be obtained from Figure 2.

The convergence of the proposed algorithm is monitored using the diagnostic measures explained in Section 5. The autocorrelations and Raftery-Lewis diagnostics are displayed in Table 2; Table 3

	True		Std.	Lower		Upper
Parameter	values	Mean	Dev.	95 % limit	Median	95 % limit
θ	0.7	0.6807	0.0651	0.5586	0.6794	0.8004
Θ	0.4	0.3836	0.0555	0.2768	0.3857	0.4963
$\sigma^{-2}$	1.0	1.0134	0.0937	0.8465	1.0070	1.2200
<b>y</b> 251	-1.0134	-1.5279	1.0273	-3.5631	-1.5006	0.4842
<b>Y</b> 252	0.4168	-0.5912	1.2729	-3.1678	-0.6019	1.9410
<i>y</i> 253	0.1782	0.0363	1.2030	-2.4596	0.0684	2.2727
<b>Y</b> 254	-0.8742	-0.2967	1.1751	-2.5934	-0.2956	2.0063

Table 1. Bayesian results for the simulated example.

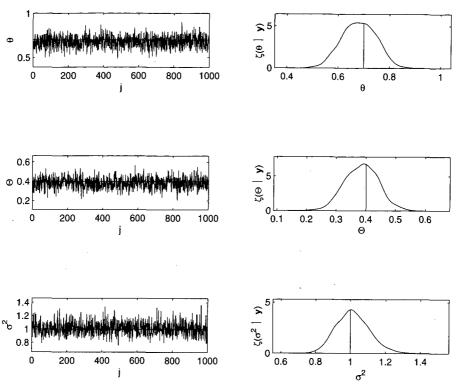


Figure 1. Sequential plots and marginal posterior distributions for the simulated example.

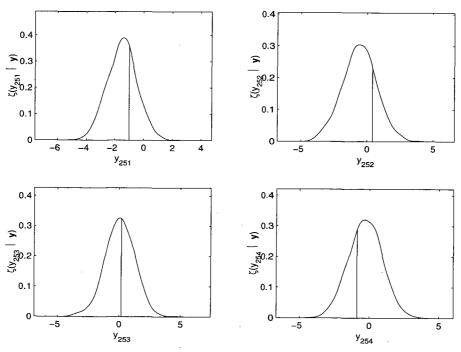


Figure 2. Marginal predictive distributions for the simulated example.

presents the diagnostic measures of Geweke (1992). Table 2 shows that the draws for each parameter or future observation had small autocorrelations at lags 1, 5, 10 and 50, which indicates no convergence problem. This conclusion was confirmed by the the diagnostic measures of Raftery and Lewis where the thining estimate (Thin) was 1, the reported (Nmin) was 937, which is close to the 1000 draws we used, and I-stat value was 1.0342, which is less than 5. Scanning the entries of Table 3 confirms the convergence of the proposed algorithm where the Chi-squared probabilities show that the equal-means hypothesis can not be rejected and no dramatic differences between the NSE estimates were found. In addition, the RNE estimates were close to 1, which indicates the iid nature of the output sample.

	Autocorrelations					Raftery-Lewis diagnostics					
parameter	Lag 1	Lag 5	Lag 10	Lag 50	Thin	Burn	Total(N)	(Nmin)	I-stat		
$\theta$	-0.0198	-0.0250	-0.0050	0.0396	1	2	969	937	1.0342		
Θ	-0.0207	0.0334	0.0217	0.0536	1	2	969	937	1.0342		
$\sigma^2$	-0.0005	-0.0105	0.0256	0.0015	1	2	969	937	1.0342		
<i>y</i> 201	0.0119	-0.0340	0.0813	-0.0013	1	2	969	937	1.0342		
<i>y</i> 202	-0.0344	0.0435	0.0376	-0.0745	1	2	969	937	1.0342		
<i>y</i> 203	0.0128	-0.0064	0.0184	-0.0760	1	2	969	937	1.0342		
<i>Y</i> 204	-0.0420	0.0249	0.0082	-0.0464	1	2	969	937	1.0342		

Table 2. Autocorrelations and Raftery-Lewis diagnostics for the simulated example.

Table 2	Correles	diagnostics	fan tha	aimulated	ovomnlo
Ianie 3	Leweke	diagnostics	tor the	similiared	example.

					•				
parameter	NSE	RNE	NSE	RNE	NSE	RNE	NSE	RNE	$\chi^2$
	iid	iid	4%	4%	8%	8%	15%	15%	
θ .	0.0021	1	0.0022	0.8655	0.0024	0.7487	0.0022	0.8409	0.3303
Θ	0.0018	1	0.0019	0.8936	0.0018	0.9652	0.0016	1.1613	0.9197
$\sigma^2$	0.0030	1	0.0026	1.3391	0.0023	1.6984	0.0022	1.8247	0.7093
<i>y</i> 251	0.0325	1	0.0334	0.9458	0.0299	1.1786	0.0231	1.9742	0.9799
<i>y</i> 252	0.0402	1	0.0454	0.7870	0.0436	0.8519	0.0365	1.2168	0.4001
<i>y</i> 253	0.0380	1	0.0341	1.0400	0.0259	1.1485	0.0209	1.3203	0.5363
<i>y</i> 254	0.0371	1	0.0420	0.7813	0.0404	0.8434	0.0365	1.0347	0.7357

## Airline Data

The airline data consist of 144 monthly values of the number of passengers in international air travel. The data were first considered by Box and Jenkins (1976) and became a standard example of seasonal time series analysis. Several authors identified different models for these data. Box and Jenkins (1976) proposed the following model for the natural logarithm:

$$w_t = (1 + \theta B) \left( 1 + \Theta B^{12} \right) \varepsilon_t, \tag{10}$$

where  $w_t = (1 - B)(1 - B^{12})\ln(y_t)$ .

In this subsection, we demonstrate how the proposed Bayesian analysis can be used to analyze the airline data. Using the model  $SMA(1, 1)_{12}$ , the proposed methodology is applied to the first 125 values of the differenced airline series  $w_t$  and the last six values are set aside for forecasting evaluation. The hyperparameters and starting values are chosen as in the previous example. Table 4 summarizes the posterior and predictive results for the differenced airline series together with the corresponding results using the model of Box and Jenkins (1976). From Table 4, all of the true values are within the 95% confidence interval formed by the 0.025 and 0.975 percentiles. Moreover, the variance of our Bayesian forecasts is 0.029, which is equal to the variance of the forecasts of Box and Jenkins (1976). It should be noted that uncertainty about the parameters and initial errors were incorporated. However, there exist some non Bayesian papers that are concerned with the prediction error arising from uncertainty about parameter estimation, such as Yamamoto (1976, 1981).

Table 4. Bayesian results for the differenced airline series using  $SMA(1,1)_{12}$ 

	True		Std.	Lower		Upper	B-J
Parameter	values	Mean	Dev.	95 % limit	Median	95 % limit	estimates
θ .	-	-0.4062	0.0923	-0.5795	-0.4063	-0.2235	-0.39
Θ	-	-0.4789	0.0967	-0.6710	-0.4815	-0.2976	-0.48
$\sigma^2$	-	0.0016	0.0002	0.0013	0.0016	0.0021	0.0014
$w_{126}$	0.0014	-0.0118	0.0394	-0.0874	-0.0132	0.0623	-0.0179
$w_{127}$	-0.0459	-0.0085	0.0432	-0.0915	-0.0096	0.0752	-0.0091
$w_{128}$	0.0120	0.0090	0.0447	-0.0825	0.0085	0.0974	0.0137
$w_{129}$	0.0318	-0.0030	0.0450	-0.0943	-0.0044	0.0856	-0.0019
$w_{130}$	-0.0501	-0.0110	0.0446	-0.0961	-0.0117	0.0769	-0.0117
$w_{131}$	-0.0100	-0.0044	0.0442	-0.0939	-0.0046	0.0833	-0.0045

Sequential plots of the parameter generated sequences together with marginal densities, which are displayed in Figure 3, show that the proposed algorithm is stable. Autocorrelations and Raftery-Lewis diagnostics and the diagnostic measures of Geweke are displayed in Tables 5 and 6. Scanning the diagnostic measures of Tables 5 and 6 indicates no convergence problem.

It is worthwhile to test the significance of the interaction parameters in the seasonal MA model. As an example, consider the  $SMA(1,1)_s$  model

$$y_t = \theta \varepsilon_{t-1} + \Theta \varepsilon_{t-s} + \theta \Theta \varepsilon_{t-s-1}.$$

The interaction parameter is  $\eta = \theta\Theta$ . Although in the classical approach testing the significance of  $\eta$  is complicated or even impossible, it is straightforward in the suggested Bayesian framework. Using the proposed Gibbs sampling algorithm, the marginal posterior distribution of  $\eta$  can be obtained and a confidence interval can be constructed. When the credible interval of  $\eta$  contains zero, the significance hypothesis of  $\eta$  is rejected. The above testing procedure was applied to the airline series, where the marginal posterior of  $\eta$  was obtained and displayed in Figure 4. Moreover a 95% credible interval for  $\eta$  is  $0.08 \le \eta \le 0.3$ , which supports the alternative hypothesis of the interaction in the model.

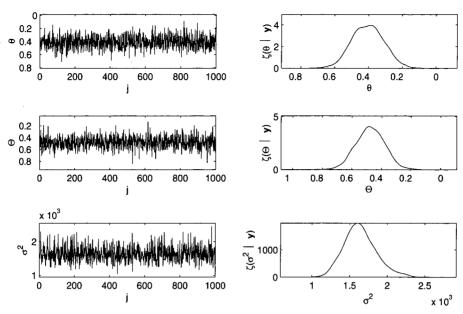


Figure 3. Sequential plots and marginal posterior distributions for the differenced airline series.

 Table 5.
 Autocorrelations and Raftery-Lewis diagnostics for the differenced airline series.

		Autocor	relations		Raftery-Lewis diagnostics				
parameter	Lag 1	Lag 5	Lag 10	Lag 50	Thin	Burn	Total(N)	(Nmin)	I-stat
, θ	0.0273	-0.0328	-0.0059	0.0231	1	2	, 969	937	1.0342
Θ	0.0366	-0.0434	-0.0326	0.0437	1	2	969	937	1.0342
$\sigma^2$	0.0725	-0.0362	-0.0689	-0.0130	1	2	969	937	1.0342
$w_{126}$	0.0249	-0.0299	0.0030	-0.0109	1	2	969	937	1.0342
$w_{127}$	-0.0601	-0.0334	0.0598	-0.0155	1	2	969	937	1.0342
$w_{128}$	0.0430	0.0299	0.0227	-0.0228	1	2	969	937	1.0342
$w_{129}$	-0.0342	-0.0441	0.0269	-0.0296	1	2	969	937	1.0342
$w_{130}$	0.0367	0.0321	-0.0487	0.0102	1	2	969	937	1.0342
$w_{131}$	0.0064	0.0704	-0.0165	-0.0189	- 1	2	969	937	1.0342

Table 6. Geweke diagnostics for the differenced airline series.

parameter	NSE	RNE	NSE	RNE	NSE	RNE	NSE	RNE	$\chi^2$
	iid	iid	4%	4%	8%	8%	15%	15%	
θ	0.0029	1	0.0032	0.8301	0.0030	0.9272	0.0032	0.8245	0.0203
Θ	0.0031	1	0.0026	1.3334	0.0024	1.5581	0.0024	1.6098	0.2503
$\sigma^2$	0.0001	1	0.0001	1.1656	0.0000	1.2293	0.0001	2.5331	0.8506
$w_{126}$	0.0012	1	0.0010	1.6280	0.0009	1.8634	0.0008	2.5020	0.3368
$w_{127}$ .	0.0014	1	0.0010	1.7727	0.0009	2.3450	0.0008	2.8541	0.9232
$w_{128}$	0.0014	1	0.0016	0.7488	0.0017	0.7000	0.0019	0.5832	0.0008
$w_{129}$	0.0014	1	0.0013	1.1463	0.0013	1.1928	0.0014	1.1039	0.0908
w <sub>130</sub>	0.0014	1	0.0015	0.8490	0.0015	0.8430	0.0015	0.9068	0.1534
$w_{131}$	0.0014	1	0.0016	0.8050	0.0015	0.8610	0.0015	0.8555	0.2918

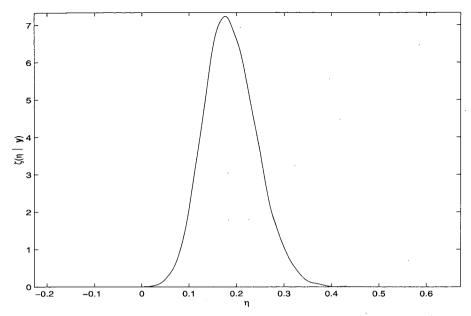


Figure 4. Marginal posterior distribution of the interaction parameter  $\eta$ 

## 7. Concluding Remarks

In this paper we developed a simple and fast Gibbs sampling algorithm for estimating the parameters and forecasting multiple future observations of the multiplicative seasonal MA model. The empirical results of the simulated examples and the airline series showed the accuracy of the proposed methodology. An extensive check of convergence using several diagnostics showed that the convergence of the proposed algorithm was achieved.

Although the employed prior distribution in Section 3 is informative, a noninformative prior is used for the parameters  $\theta$ ,  $\Theta$  and  $\sigma^2$  in the illustrative examples for the sake of the simplicity. However, if one needs to use informative prior, the hyperparameters of the prior distribution must be elicited. One way to elicit the hyperparameters is the training sample approach where the data is divided into two parts. The first part constitutes the training sample and is used to provide proper priors. Then, posterior distributions are obtained by combining these priors with the likelihood based on the second part of the data (non-training sample). The training sample approach is used by Lempers (1971) and Speigelhalter and Smith (1982) among others.

Future work may investigate model identification using stochastic search variable selection, outliers detection, and extension to multivariate seasonal models.

#### References

Barnett, G., Kohn, R., and Sheather, S. (1996): Bayesian estimation of an autoregressive model using Markov Chain Monte Carlo. *Journal of Econometrics* 74, 237–254.

Barnett, G., Kohn, R., and Sheather, S. (1997): Robust Bayesian estimation of an autoregressive moving-average model.

- Journal of Time Series Analysis 18, 11-28.
- Box, G.E.P and Jenkins, G.M. (1976): Time Series Analysis, Forecasting and Control. Holden-Day, San Francisco.
- Broemeling, L.D. (1985): Bayesian Analysis of Linear Models. Marcel Dekker Inc., New York.
- Broemeling, L.D. and Shaarawy, S. (1984): Bayesian inferences and forecasts with moving average processes. *Communications in Statistics: Theory and Methods* 13, 1871–1888.
- Chen, C.W.S. (1992): Bayesian Analysis of Bilinear Time Series Models: A Gibbs Sampling Approach. *Communications in Statistics: Theory and Methods* 21, 3407–3425.
- Chib, S. and Greenberg, E. (1994): Bayes inference in regression models with ARMA(p,q) errors. *Journal of Econometrics* **64**, 183–206.
- Geweke, J. (1992): Evaluating the Accuracy of Sampling-Based Approaches to the Calculations of Posterior Moments. In *Bayesian Statistics 4*, J.M., Bernardo, J.O., Berger, A.P., Dawid, and A.F.M., Smith, (eds), 641–649. Oxford, Clarendon Press
- Koreisha and Pukkila (1990): Linear methods for estimating ARMA and regression models with serial correlation. *Communications in Statistics: Simulation* 19, 71–102.
- Lempers, F.B. (1971): Posterior probabilities of alternative linear models. Rotterdam University press.
- LeSage, J.P. (1999): Applied Econometrics using Matlab. Dept. of Economics, University of Toledo, available at http://www.econ.utoledo.edu.
- 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.
- Marriott, J. and Smith, A.F.M. (1992): Reparameterization aspects of numerical Bayesian methodology for autoregressive moving-average models. *Journal of Time Series Analysis* 13, 327–343.
- McCulloch, R.E. and Tsay, R.S. (1994): Bayesian analysis of autoregressive time series via the Gibbs sampler. *Journal of Time Series Analysis* **15(2)**, 235–250.
- Monahan, J.F. (1983): Fully Bayesian analysis of ARMA time series models. J. Econometrics 19, 147-164.
- Newbold, P. (1973): Bayesian estimation of Box Jenkins transfer function noise models. *Journal of the Royal Statistical Society, Series B* 35, 323–336.
- Raftery, A.E. and Lewis, S. (1992): One long run with diagnostics: Implementation strategies for Markov Chain Monte Carlo. Statistical Science 7, 493-497.
- Raftery, A.E. and Lewis, S. (1995): The number of iterations, convergence diagnostics and generic Metropolis algorithms. In *Practical Markov Chain Monte Carlo*, W.R., Gilks, D.J., Spiegelhalter and S., Richardson, (eds). London, Chapman and Hall
- Shaarawy, S. and Ismail, M.A. (1987): Bayesian Inference for Seasonal ARMA Models. *The Egyptian Statistical Journal* 31, 323-336.
- Spiegelhalter, D.J. and Smith, A.F.M. (1982): Bayes factors for linear and log-linear models with vague prior information. JRSS B44, 377-387.
- Yamamoto, T. (1976): Asymptotic mean square prediction error for an autoregressive model with estimated coefficients. Applied Statist. 25, 123-127.
- Yamamoto, T. (1981): Prediction of multivariate autoregressive moving average models. Biometrika 68, 475-492.

Received January 8, 2003; Received in final form July 10, 2003; Accepted September 1, 2003.