

Bayesian Inference for Double Seasonal Moving Average Models: A Gibbs Sampling Approach

Ayman A Amin

Department of Statistics, Mathematics, and Insurance

Faculty of Commerce, Menoufia University, Egypt

ayman.aamin@gmail.com

Abstract

In this paper we use the Gibbs sampling algorithm to develop a Bayesian inference for multiplicative double seasonal moving average (DSMA) models. Assuming the model errors are normally distributed and using natural conjugate priors, we show that the conditional posterior distribution of the model parameters and variance are multivariate normal and inverse gamma respectively, and then we apply the Gibbs sampling to approximate empirically the marginal posterior distributions. The proposed Bayesian methodology is evaluated using simulation study.

Keywords: Bayesian analysis, Double seasonality, Gibbs sampler, Multiplicative seasonal moving average.

1. Introduction

High frequency time series that are observed at small time units may be characterized by exhibiting multiple seasonal patterns. For example, hourly electricity load data can exhibit intraday and intraweek seasonal patterns. Other examples of high frequency time series contain multiple seasonal patterns include daily hospital admissions, daily usage of water and natural gas, hourly volumes of call center arrivals, hourly traffic on a road, hourly access to computer web sites, and half-hourly demand for public transportation. The notion of modelling multiple seasonalities is not new and it can be traced back to 1971 when Thompson and Tiao (1971) showed that monthly disconnections of the Wisconsin telephone company have annual and quarterly (double) seasonal patterns. Accordingly, seasonal autoregressive moving average (SARMA) models being widely applied to analyze time series with single seasonal pattern need to be modified and extended to accommodate multiple seasonalities, see for example Box et al. (1994) and Taylor (2003). In addition to SARMA models, other techniques have been extended to fit multiple seasonal time series, which include neural networks, exponential smoothing methods and innovation state models. A quick review of these techniques can be found in Feinberg & Genethliou (2005).

In particular, the multiplicative double SARMA (DSARMA) models have been the subject of interest of many researchers and extensively studied and employed in modeling and forecasting time series data exhibiting double seasonal patterns. Taylor (2003) showed that electricity load in England and Wales features daily (within day) and weekly (within week) seasonal patterns. Taylor et al. (2006) compared the forecast accuracy of some univariate models including double SARMA for electricity demand forecasting in Brazil and in England and Wales. Cruz et al. (2011) compared the forecasting accuracy of a set of methods for day-ahead spot price forecasting in the Spanish electricity market.

Other references may include Taylor (2008a&b), Caiado (2008), Baek (2010), Mohamed et al. (2010), Au et al. (2011), Mohamed et al. (2011) and Kim (2013).

Bayesian analysis of SARMA model for single seasonality has been well established, and different approaches have been developed in literature. Analytical approximation is one of these approaches, which simply approximates the posterior and predictive densities to be standard closed-form distributions that are analytically tractable, see for example Shaarawy and Ismail (1987). However, this approach is conditioning on the initial values leading to waste observations, and treats SARMA model as an additive not a multiplicative model which can introduce new unnecessary coefficients in the model. To address the limitations of analytical approximation, in recent years MCMC methods, especially Gibbs sampling algorithm, have been proposed to ease the Bayesian time series analysis. Ismail (2003a&b) used Gibbs sampling algorithm to achieve Bayesian analysis for multiplicative seasonal moving average (SMA) and seasonal autoregressive (SAR) models. This work was extended by Ismail and Amin (2014) to multiplicative SARMA model. The literature in Bayesian analysis of DSARMA models is still immature. Amin and Ismail (2015) have used Gibbs sampling algorithm to develop a Bayesian analysis to multiplicative double SAR models. In the current paper, we extend this work to develop a Bayesian analysis to multiplicative DSMA models based on Gibbs sampling algorithm. The initial idea of this work was presented in the 60th ISI World Statistics Congress 2015, and the main advantages of the proposed methodology are that the Bayesian analysis is unconditional on the initial values of errors and it treats DSMA model as a multiplicative not an additive model to achieve parsimonious property.

The remainder of this paper is organized as follows: Section 2 presents multiplicative DSMA models. Section 3 summarizes the posterior analysis and full conditional posterior distributions of the parameters. The implementation details of the proposed algorithm including convergence monitoring are presented in Section 4. The proposed methodology is illustrated in Section 5 using several simulated examples. Finally, the conclusions are given in Section 6.

2. Double Seasonal Moving Average (DSMA) Model

A time series $\{y_t\}$ is said to be generated by a multiplicative seasonal moving average model of orders q , Q_1 , and Q_2 , denoted by $DSMA(q)(Q_1)_{s_1}(Q_2)_{s_2}$, if it satisfies

$$y_t = \theta_q(B)\Theta_{Q_1}(B^{s_1})\Psi_{Q_2}(B^{s_2})\varepsilon_t \quad (1)$$

where $\{\varepsilon_t\}$ is a sequence of independent normal variates with zero mean and variance σ^2 . The backshift operator B is defined as $B^k y_t = y_{t-k}$, s_1 and s_2 are the seasonal periods. The non seasonal moving average polynomial is $\theta_q(B) = (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q)$ with order q . In addition, the seasonal moving average polynomials are $\Theta_{Q_1}(B^{s_1}) = (1 + \Theta_1 B^{s_1} + \Theta_2 B^{2s_1} + \dots + \Theta_{Q_1} B^{Q_1 s_1})$ with order Q_1 and $\Psi_{Q_2}(B^{s_2}) = (1 + \Psi_1 B^{s_2} + \Psi_2 B^{2s_2} + \dots + \Psi_{Q_2} B^{Q_2 s_2})$ with order Q_2 . Finally, the non seasonal and seasonal moving average coefficients are $\theta = (\theta_1, \theta_2, \dots, \theta_q)^T$, $\Theta = (\Theta_1, \Theta_2, \dots, \Theta_{Q_1})^T$ and $\Psi = (\Psi_1, \Psi_2, \dots, \Psi_{Q_2})^T$. The time series $\{y_t\}$ is assumed to start at time $t = 1$ with unknown starting errors $\varepsilon_0 = (\varepsilon_0, \varepsilon_{-1}, \dots, \varepsilon_{1-q-s})$.

It should be noted that the DSMA model (1) has an extra terms compared with the usual multiplicative single SMA model. The new term is $\Psi_{Q_2}(B^{s_2})$ that accommodates the second seasonal pattern. Accordingly, the model (1) can be written as

$$\begin{aligned}
 y_t = & \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \sum_{j=1}^{Q_1} \Theta_j \varepsilon_{t-j s_1} + \sum_{\tau=1}^{Q_2} \Psi_{\tau} \varepsilon_{t-\tau s_2} + \sum_{i=1}^q \sum_{j=1}^{Q_1} \theta_i \Theta_j \varepsilon_{t-i-j s_1} \\
 & + \sum_{i=1}^q \sum_{\tau=1}^{Q_2} \theta_i \Psi_{\tau} \varepsilon_{t-i-\tau s_2} + \\
 & \sum_{j=1}^{Q_1} \sum_{\tau=1}^{Q_2} \Theta_j \Psi_{\tau} \varepsilon_{t-j s_1-\tau s_2} + \sum_{i=1}^q \sum_{j=1}^{Q_1} \sum_{\tau=1}^{Q_2} \theta_i \Theta_j \Psi_{\tau} \varepsilon_{t-i-j s_1-\tau s_2} + \varepsilon_t \\
 = & \Lambda_t \beta + \varepsilon_t
 \end{aligned} \tag{2}$$

where

$$\begin{aligned}
 \Lambda_t = & (\varepsilon_{t-1}, \dots, \varepsilon_{t-q}, \varepsilon_{t-s_1}, \varepsilon_{t-s_1-1}, \dots, \varepsilon_{t-s_1-q}, \dots, \varepsilon_{t-Q_1 s_1}, \varepsilon_{t-Q_1 s_1-1}, \dots, \varepsilon_{t-Q_1 s_1-q}, \varepsilon_{t-s_2}, \\
 & \varepsilon_{t-s_2-1}, \dots, \varepsilon_{t-s_2-q}, \varepsilon_{t-s_1-s_2}, \varepsilon_{t-s_1-s_2-1}, \dots, \varepsilon_{t-s_1-s_2-q}, \dots, \varepsilon_{t-Q_1 s_1-s_2}, \varepsilon_{t-Q_1 s_1-s_2-1}, \\
 & \dots, \varepsilon_{t-Q_1 s_1-s_2-q}, \dots, \varepsilon_{t-Q_2 s_2}, \varepsilon_{t-Q_2 s_2-1}, \dots, \varepsilon_{t-Q_2 s_2-q}, \varepsilon_{t-s_1-Q_2 s_2}, \varepsilon_{t-s_1-Q_2 s_2-1}, \dots, \\
 & \varepsilon_{t-s_1-Q_2 s_2-q}, \dots, \varepsilon_{t-Q_1 s_1-Q_2 s_2}, \varepsilon_{t-Q_1 s_1-Q_2 s_2-1}, \dots, \varepsilon_{t-Q_1 s_1-Q_2 s_2-q})
 \end{aligned}$$

$$\begin{aligned}
 \beta = & (\theta_1, \dots, \theta_q, \Theta_1, \theta_1 \Theta_1, \dots, \theta_q \Theta_1, \dots, \Theta_{Q_1}, \theta_1 \Theta_{Q_1}, \dots, \theta_q \Theta_{Q_1}, \Psi_1, \theta_1 \Psi_1, \dots, \theta_q \Psi_1, \Theta_1 \Psi_1, \\
 & \theta_1 \Theta_1 \Psi_1, \dots, \theta_q \Theta_1 \Psi_1, \dots, \Theta_{Q_1} \Psi_1, \theta_1 \Theta_{Q_1} \Psi_1, \dots, \theta_q \Theta_{Q_1} \Psi_1, \dots, \Psi_{Q_2}, \theta_1 \Psi_{Q_2}, \dots, \\
 & \theta_q \Psi_{Q_2}, \Theta_1 \Psi_{Q_2}, \theta_1 \Theta_1 \Psi_{Q_2}, \dots, \theta_q \Theta_1 \Psi_{Q_2}, \dots, \Theta_{Q_1} \Psi_{Q_2}, \theta_1 \Theta_{Q_1} \Psi_{Q_2}, \dots, \theta_q \Theta_{Q_1} \Psi_{Q_2})^T.
 \end{aligned} \tag{3}$$

Equation (2) shows that the multiplicative DSMA model can be written as a moving average model of order $(1+q)(1+Q_1)(1+Q_2)-1$ with some coefficients are products of nonseasonal and seasonal coefficients. Therefore, the DSMA model is nonlinear in the coefficients θ , Θ , and Ψ which complicates its Bayesian analysis. In the following sections we explain how to apply the Gibbs sampling to facilitate the analysis. The DSMA model (2) is invertible if the roots of the polynomials $\theta_q(B)$, $\Theta_{Q_1}(B^{s_1})$ and $\Psi_{Q_2}(B^{s_2})$ lie outside the unit circle. For more details about the properties of SARMA models see Box et al. (1994).

3. Posterior Analysis

3.1 Likelihood Function

Suppose $y = (y_1, y_2, \dots, y_n)$ is a realization of the DSMA model (2), knowing that $\varepsilon_t \sim N(0, \sigma^2)$ and employing a straightforward random variable transformation from ε_t to y_t , the likelihood function is given by

$$L(\theta, \Theta, \Psi, \sigma^2, \varepsilon_0 | y) \propto (\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{t=1}^n \varepsilon_t^2 \right\} \tag{4}$$

where the errors ε_t is computed directly from the model (2). It is obvious that this likelihood function is a complicated function in θ , Θ , Ψ and ε_0 . However, the errors ε_t can be estimated recursively as follows

$$\begin{aligned} e_t = & y_t - \sum_{i=1}^q \hat{\theta}_i e_{t-i} - \sum_{j=1}^{Q_1} \hat{\Theta}_j e_{t-j s_1} - \sum_{\tau=1}^{Q_2} \hat{\Psi}_\tau \varepsilon_{t-\tau s_2} - \sum_{i=1}^q \sum_{j=1}^{Q_1} \hat{\theta}_i \hat{\Theta}_j e_{t-i-j s_1} - \\ & \sum_{i=1}^q \sum_{\tau=1}^{Q_2} \hat{\theta}_i \hat{\Psi}_\tau e_{t-i-\tau s_2} - \sum_{j=1}^{Q_1} \sum_{\tau=1}^{Q_2} \hat{\Theta}_j \hat{\Psi}_\tau e_{t-j s_1-\tau s_2} - \\ & \sum_{i=1}^q \sum_{j=1}^{Q_1} \sum_{\tau=1}^{Q_2} \hat{\theta}_i \hat{\Theta}_j \hat{\Psi}_\tau e_{t-i-j s_1-\tau s_2} \end{aligned} \quad (5)$$

where $\hat{\theta}_i \in R^q$, $\hat{\Theta}_j \in R^{Q_1}$, and $\hat{\Psi}_\tau \in R^{Q_2}$ are consistent estimates obtained by the nonlinear least square (NLS) estimation method. Substituting the estimated errors in the likelihood function (3) results in an approximate likelihood function:

$$\begin{aligned} L^*(\theta, \Theta, \Psi, \sigma^2, \varepsilon_0 | y) & \propto (\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{t=1}^n \varepsilon_t^{*2} \right\} \\ & = (\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} (y - \hat{\Lambda}\beta)^T (y - \hat{\Lambda}\beta) \right\}. \end{aligned} \quad (6)$$

where, β is defined in (3), and $\hat{\Lambda}$ is a $n \times ((1+q)(1+Q_1)(1+Q_2) - 1)$ matrix with the t^{th} row:

$$\begin{aligned} \hat{\Lambda}_t = & (e_{t-1}, \dots, e_{t-q}, e_{t-s_1}, e_{t-s_1-1}, \dots, e_{t-s_1-q}, \dots, e_{t-Q_1 s_1}, e_{t-Q_1 s_1-1}, \dots, e_{t-Q_1 s_1-q}, e_{t-s_2}, \\ & e_{t-s_2-1}, \dots, e_{t-s_2-q}, e_{t-s_1-s_2}, e_{t-s_1-s_2-1}, \dots, e_{t-s_1-s_2-q}, \dots, e_{t-Q_1 s_1-s_2}, e_{t-Q_1 s_1-s_2-1}, \\ & \dots, e_{t-Q_1 s_1-s_2-q}, \dots, e_{t-Q_2 s_2}, e_{t-Q_2 s_2-1}, \dots, e_{t-Q_2 s_2-q}, e_{t-s_1-Q_2 s_2}, e_{t-s_1-Q_2 s_2-1}, \dots, \\ & e_{t-s_1-Q_2 s_2-q}, \dots, e_{t-Q_1 s_1-Q_2 s_2}, e_{t-Q_1 s_1-Q_2 s_2-1}, \dots, e_{t-Q_1 s_1-Q_2 s_2-q}). \end{aligned} \quad (7)$$

3.2 Prior Specification

Assuming that the parameters θ , Θ , Ψ and ε_0 are independent apriori, given the error variance parameter σ^2 , the joint prior distribution is

$$\begin{aligned} \zeta(\theta, \Theta, \Psi, \sigma^2, \varepsilon_0) & = \zeta(\theta | \sigma^2) \times \zeta(\Theta | \sigma^2) \times \zeta(\Psi | \sigma^2) \times \zeta(\varepsilon_0 | \sigma^2) \times \zeta(\sigma^2) \\ & = N_q(\mu_\theta, \sigma^2 \Sigma_\theta) \times N_{Q_1}(\mu_\Theta, \sigma^2 \Sigma_\Theta) \times N_{Q_2}(\mu_\Psi, \sigma^2 \Sigma_\Psi) \times N_{q^*}(\mu_{\varepsilon_0}, \sigma^2 \Sigma_{\varepsilon_0}) \times IG\left(\frac{\nu}{2}, \frac{\lambda}{2}\right), \end{aligned} \quad (8)$$

where $q^* = q + Q_1 s_1 + Q_2 s_2$, $N_r(\mu, \Delta)$ is the r-variate normal distribution with mean μ and variance Δ , and $IG(a, b)$ is the inverse gamma distribution with parameters a and b. Therefore, the joint prior distribution can be written as follows

$$\begin{aligned} \zeta(\theta, \Theta, \Psi, \sigma^2, \varepsilon_0) & \propto (\sigma^2)^{-\left(\frac{\nu^*}{2} + 1\right)} \exp \left\{ -\frac{1}{2\sigma^2} [\lambda + (\theta - \mu_\theta)^T \Sigma_\theta^{-1} (\theta - \mu_\theta) \right. \\ & \quad \left. + (\Theta - \mu_\Theta)^T \Sigma_\Theta^{-1} (\Theta - \mu_\Theta) + (\Psi - \mu_\Psi)^T \Sigma_\Psi^{-1} (\Psi - \mu_\Psi) + (\varepsilon_0 - \mu_{\varepsilon_0})^T \Sigma_{\varepsilon_0}^{-1} (\varepsilon_0 - \mu_{\varepsilon_0}) \right\}, \end{aligned} \quad (9)$$

where

$$\nu^* = \nu + 2q + Q_1(1 + s_1) + Q_2(1 + s_2). \quad (10)$$

The prior distribution (9) is chosen for several reasons, especially it is a conjugate prior and thus facilitates the mathematical calculations. Multiplying the joint prior distribution

(9) by the approximate likelihood function (6) results in the joint posterior $\zeta(\theta, \Theta, \Psi, \sigma^2, \varepsilon_0|y)$ which may be written as

$$\begin{aligned} \zeta(\theta, \Theta, \Psi, \sigma^2, \varepsilon_0|y) \propto (\sigma^2)^{-\left(\frac{n+\nu^*}{2}+1\right)} \exp \left\{ -\frac{1}{2\sigma^2} [\lambda + (\theta - \mu_\theta)^T \Sigma_\theta^{-1} (\theta - \mu_\theta) \right. \\ \left. + (\Theta - \mu_\Theta)^T \Sigma_\Theta^{-1} (\Theta - \mu_\Theta) + (\Psi - \mu_\Psi)^T \Sigma_\Psi^{-1} (\Psi - \mu_\Psi) \right. \\ \left. + (\varepsilon_0 - \mu_{\varepsilon_0})^T \Sigma_{\varepsilon_0}^{-1} (\varepsilon_0 - \mu_{\varepsilon_0}) + (y - \hat{\Lambda}\beta)^T (y - \hat{\Lambda}\beta) \right\} \end{aligned} \quad (11)$$

3.3 Full Conditional Posterior Distributions

The conditional posterior distribution for each one of the DSMA parameters is obtained from the joint posterior distribution (11) by first grouping together terms in the joint posterior that depend on this parameter, and then finding the appropriate normalizing constant to form a proper and closed-form density. In this study all the conditional posteriors obtained for the unknown parameters are normal and inverse gamma distributions for which sampling techniques exist.

3.3.1 The conditional posterior of θ

We obtained the conditional posterior of θ by finding out $\zeta(\theta|y, \Theta, \Psi, \sigma^2, \varepsilon_0)$ that we proved to be a multivariate normal $N(\mu_\theta^*, v_\theta^*)$ with

$$\mu_\theta^* = \left[(H_\theta^T H_\theta + \Sigma_\theta^{-1})^{-1} (\Sigma_\theta^{-1} \mu_\theta + H_\theta^T (y - L_\theta \beta_\theta)) \right]$$

and

$$v_\theta^* = \sigma^2 (H_\theta^T H_\theta + \Sigma_\theta^{-1})^{-1}.$$

Where, H_θ is a $(n \times q)$ matrix with the ti^{th} element:

$$H_{\theta_{ti}} = \left(e_{t-i} + \sum_{j=1}^{Q_1} \Theta_j e_{t-i-j s_1} + \sum_{\tau=1}^{Q_2} \Psi_\tau e_{t-i-\tau s_2} + \sum_{j=1}^{Q_1} \sum_{\tau=1}^{Q_2} \Theta_j \Psi_\tau e_{t-i-j s_1 - \tau s_2} \right),$$

L_θ is a $n \times ((1 + Q_1)(1 + Q_2) - 1)$ matrix with the t^{th} row:

$$L_{\theta_t} = (e_{t-s_1}, \dots, e_{t-Q_1 s_1}, e_{t-s_2}, e_{t-s_1-s_2}, \dots, e_{t-Q_1 s_1-s_2}, \dots, e_{t-Q_2 s_2}, e_{t-s_1-Q_2 s_2}, \dots, e_{t-Q_1 s_1-Q_2 s_2}),$$

and β_θ is a column vector of order $(1 + Q_1)(1 + Q_2) - 1$ written as:

$$\beta_\theta = (\Theta_1, \dots, \Theta_{Q_1}, \Psi_1, \Theta_1 \Psi_1, \dots, \Theta_{Q_1} \Psi_1, \dots, \Psi_{Q_2}, \Theta_1 \Psi_{Q_2}, \dots, \Theta_{Q_1} \Psi_{Q_2})^T.$$

3.3.2 The conditional posterior of Θ

We obtained the conditional posterior of Θ by finding out $\zeta(\Theta|y, \theta, \Psi, \sigma^2, \varepsilon_0)$ that we proved to be a multivariate normal $N(\mu_\Theta^*, v_\Theta^*)$ with

$$\mu_\Theta^* = \left[(H_\Theta^T H_\Theta + \Sigma_\Theta^{-1})^{-1} (\Sigma_\Theta^{-1} \mu_\Theta + H_\Theta^T (y - L_\Theta \beta_\Theta)) \right]$$

and

$$v_\Theta^* = \sigma^2 (H_\Theta^T H_\Theta + \Sigma_\Theta^{-1})^{-1}.$$

Where H_{Θ} is a $(n \times Q_1)$ matrix with the tj^{th} element:

$$H_{\Theta_{tj}} = \left(e_{t-j_{S_1}} + \sum_{i=1}^q \theta_i e_{t-i-j_{S_1}} + \sum_{\tau=1}^{Q_2} \Psi_{\tau} e_{t-j_{S_1}-\tau_{S_2}} + \sum_{i=1}^q \sum_{\tau=1}^{Q_2} \theta_i \Psi_{\tau} e_{t-i-j_{S_1}-\tau_{S_2}} \right),$$

L_{Θ} is a $n \times ((1+q)(1+Q_2) - 1)$ matrix with the t^{th} row:

$$L_{\Theta_t} = (e_{t-1}, \dots, e_{t-q}, e_{t-S_2}, e_{t-1-S_2}, \dots, e_{t-q-S_2}, \dots, e_{t-Q_2 S_2}, e_{t-1-Q_2 S_2}, \dots, e_{t-q-Q_2 S_2}),$$

and β_{Θ} is a column vector of order $(1+q)(1+Q_2) - 1$ written as:

$$\beta_{\Theta} = (\theta_1, \dots, \theta_q, \Psi_1, \theta_1 \Psi_1, \dots, \theta_q \Psi_1, \dots, \Psi_{Q_2}, \theta_1 \Psi_{Q_2}, \dots, \theta_q \Psi_{Q_2})^T.$$

3.3.3 The conditional posterior of Ψ

We obtained the conditional posterior of Ψ by finding out $\zeta(\Psi|y, \theta, \Theta, \sigma^2, \varepsilon_0)$ that we proved to be a multivariate normal $N(\mu_{\Psi}^*, v_{\Psi}^*)$ with

$$\mu_{\Psi}^* = \left[(H_{\Psi}^T H_{\Psi} + \Sigma_{\Psi}^{-1})^{-1} (\Sigma_{\Psi}^{-1} \mu_{\Psi} + H_{\Psi}^T (y - L_{\Psi} \beta_{\Psi})) \right]$$

and

$$v_{\Psi}^* = \sigma^2 (H_{\Psi}^T H_{\Psi} + \Sigma_{\Psi}^{-1})^{-1}.$$

Where H_{Ψ} is a $(n \times Q_2)$ matrix with the $t\tau^{th}$ element:

$$H_{\Psi_{t\tau}} = \left(e_{t-\tau_{S_2}} + \sum_{i=1}^q \theta_i e_{t-i-\tau_{S_2}} + \sum_{j=1}^{Q_1} \Theta_j e_{t-j_{S_1}-\tau_{S_2}} + \sum_{i=1}^q \sum_{j=1}^{Q_1} \theta_i \Theta_j e_{t-i-j_{S_1}-\tau_{S_2}} \right),$$

L_{Ψ} is a $n \times ((1+q)(1+Q_1) - 1)$ matrix with the t^{th} row:

$$L_{\Psi_t} = (e_{t-1}, \dots, e_{t-q}, e_{t-S_1}, e_{t-1-S_1}, \dots, e_{t-q-S_1}, \dots, e_{t-Q_1 S_1}, e_{t-1-Q_1 S_1}, \dots, e_{t-q-Q_1 S_1}),$$

and β_{Ψ} is a column vector of order $(1+q)(1+Q_1) - 1$ written as:

$$\beta_{\Psi} = (\theta_1, \dots, \theta_q, \Theta_1, \theta_1 \Theta_1, \dots, \theta_q \Theta_1, \dots, \Theta_{Q_1}, \theta_1 \Theta_{Q_1}, \dots, \theta_q \Theta_{Q_1})^T.$$

3.3.4 The conditional posterior of σ^2

We obtained the conditional posterior of σ^2 by finding out $\zeta(\sigma^2|y, \theta, \Theta, \Psi, \varepsilon_0)$ that we proved to be an inverse gamma $IG(\frac{n+\nu^*}{2}, \frac{\lambda+n(S^2)}{2})$, where ν^* is defined in (10) and

$$n(S^2) = [(\theta - \mu_{\theta})^T \Sigma_{\theta}^{-1} (\theta - \mu_{\theta}) + (\Theta - \mu_{\Theta})^T \Sigma_{\Theta}^{-1} (\Theta - \mu_{\Theta}) + (\Psi - \mu_{\Psi})^T \Sigma_{\Psi}^{-1} (\Psi - \mu_{\Psi}) + (\varepsilon_0 - \mu_{\varepsilon_0})^T \Sigma_{\varepsilon_0}^{-1} (\varepsilon_0 - \mu_{\varepsilon_0}) + (y - \hat{\Lambda} \beta)^T (y - \hat{\Lambda} \beta)].$$

To ease the Gibbs sampling algorithm process, the parameter σ^2 can be sampled from the Chi square distribution using the transformation $\frac{\lambda+n(S^2)}{(\sigma^2)} \sim \chi_{(n+\nu^*)}^2$.

3.3.5 The conditional posterior of ε_0

In the beginning, we write explicitly the elements of ε_0 using the model (2) as follows:

$$y_{q^*} = M \varepsilon_0 + N \varepsilon_{q^*}, \quad (12)$$

where

$$M = \begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \cdots & \cdots & \alpha_{q^*-1} & \alpha_{q^*} \\ \alpha_2 & \alpha_3 & \alpha_4 & \cdots & \cdots & \alpha_{q^*} & 0 \\ \alpha_3 & \alpha_4 & & \cdots & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \cdots & \vdots & \vdots \\ \alpha_{p^*-1} & \alpha_{p^*} & 0 & \cdots & \cdots & 0 & 0 \\ \alpha_{p^*} & 0 & 0 & \cdots & \cdots & 0 & 0 \end{pmatrix}_{(q^*) \times (q^*)},$$

$$N = \begin{pmatrix} 1 & 0 & 0 & \cdots & \cdots & 0 & 0 \\ \alpha_1 & 1 & 0 & \cdots & \cdots & 0 & 0 \\ \alpha_2 & \alpha_1 & 1 & \cdots & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \cdots & \vdots & \vdots \\ \alpha_{p^*-1} & \alpha_{p^*-2} & \cdots & \cdots & \alpha_2 & \alpha_1 & 1 \end{pmatrix}_{(q^*) \times (q^*)},$$

$q^* = q + Q_1s_1 + Q_2s_2$ and $\varepsilon_{p^*} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{q^*})^T$ that has the q^* multivariate normal distribution with zero mean and variance $(\sigma^2 I_{q^*})$, where I_{q^*} is the unit matrix of order q^* .

Using the above defined matrices and the standard Bayesian techniques, we obtained the conditional posterior of ε_0 by finding out $\zeta(\varepsilon_0|y, \theta, \Theta, \Psi, \sigma^2)$ that we proved to be a multivariate normal $N(\mu_{\varepsilon_0}^*, v_{\varepsilon_0}^*)$ with

$$\mu_{\varepsilon_0}^* = [H_{\varepsilon_0}M + \Sigma_{\varepsilon_0}^{-1}]^{-1}[\Sigma_{\varepsilon_0}^{-1}\mu_{\varepsilon_0} + H_{\varepsilon_0}y_{q^*}]$$

and

$$v_{\varepsilon_0}^* = \sigma^2[H_{\varepsilon_0}M + \Sigma_{\varepsilon_0}^{-1}]^{-1},$$

where, $H_{\varepsilon_0} = M^T(NN^T)^{-1}$.

4. The Proposed Gibbs Sampler

The proposed Gibbs sampling algorithm for DSMA model can be implemented as follows:

1. Specify starting values $\theta^0, \Theta^0, \Psi^0, (\sigma^2)^0$ and ε_0^0 and set $j=0$. A set of initial estimates of the model parameters can be obtained using the NLS method.
2. Calculate the residuals recursively using (5) and the NLS estimates.
3. Simulate
 - $\theta^j \sim \zeta(\theta^j|y, (\sigma^2)^{j-1}, \Theta^{j-1}, \Psi^{j-1}, \varepsilon_0^{j-1})$,
 - $\Theta^j \sim \zeta(\Theta^j|y, (\sigma^2)^{j-1}, \phi^j, \Psi^{j-1}, \varepsilon_0^{j-1})$,
 - $\Psi^j \sim \zeta(\Psi^j|y, (\sigma^2)^{j-1}, \theta^j, \Theta^j, \varepsilon_0^{j-1})$,
 - $(\sigma^2)^j \sim \zeta((\sigma^2)^j|y, \theta^j, \Theta^j, \Psi^j, \varepsilon_0^{j-1})$,
 - $\varepsilon_0^j \sim \zeta(\varepsilon_0^j|y, (\sigma^2)^j, \theta^j, \Theta^j, \Psi^j)$,
4. set $j=j+1$ and go to 3.

This algorithm gives the next value of the Markov chain $\{\theta^{j+1}, \Theta^{j+1}, \Psi^{j+1}, (\sigma^2)^{j+1}, \varepsilon_0^{j+1}\}$ by simulating each of the full conditionals where the conditioning elements are revised during the cycle. This iterative process is repeated for a large number of iterations and continuously the convergence is monitored. After the chain has converged, say at n_0 iterations, the simulated values $\{\theta^{j+1}, \Theta^{j+1}, \Psi^{j+1}, (\sigma^2)^{j+1}, \varepsilon_0^{j+1}\}$ are used as a sample from the joint posterior. Posterior estimates of the parameters are computed directly via sample averages of the simulation outputs. The convergence of the Gibbs sampler's Markov chain can be monitored by three groups of diagnostics, which include autocorrelation coefficients estimates, Raftery and Lewis diagnostics, and Geweke diagnostics. First, autocorrelation coefficients estimates indicate how much independence exists in the sequence of each parameter draws. High values of autocorrelation coefficients indicate more draws needed to get accurate posterior estimates. Second, diagnostics proposed by Raftery and Lewis (1992,1995) include (1) Burn: number of draws used as initial draws before starting to sample the draws for posterior inference, (2) Nmin: number of draws that would be needed if the draws represented an iid chain, (3) Total: total number of draws needed to achieve desired level of accuracy, and (4) I-stat: the ratio of the (Total) to (Nmin). Raftery and Lewis suggested that a convergence problem is indicated when values of I-stat exceed 5. Third, diagnostics proposed by Geweke (1992), which includes two groups:

1. The first group includes the relative numerical efficiency (RNE) and numerical standard errors (NSE). The RNE estimates indicate the required number of draws to produce the same numerical accuracy when iid sample is drawn directly from the posterior distribution. The NSE estimates reflect the variation that can be expected if the simulation were to be repeated.
2. The second group of diagnostics includes a test of whether the Gibbs sampler has attained an equilibrium state. This can be achieved by testing 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 these convergence using Matlab package, and we use them in the following section to monitor the convergence of our proposed Gibbs sampling algorithm for DSMA model.

5. Simulation Study

In this section we show the efficiency of the proposed estimation method using simulation study for five examples of DSMA models. Table 0 shows these selected DSMA models and their true parameters values used in the simulation. By these five examples we try to represent different pairs of the season periods that are chosen to cover different seasonality patterns with different data types.

Table 1: Simulated examples design.

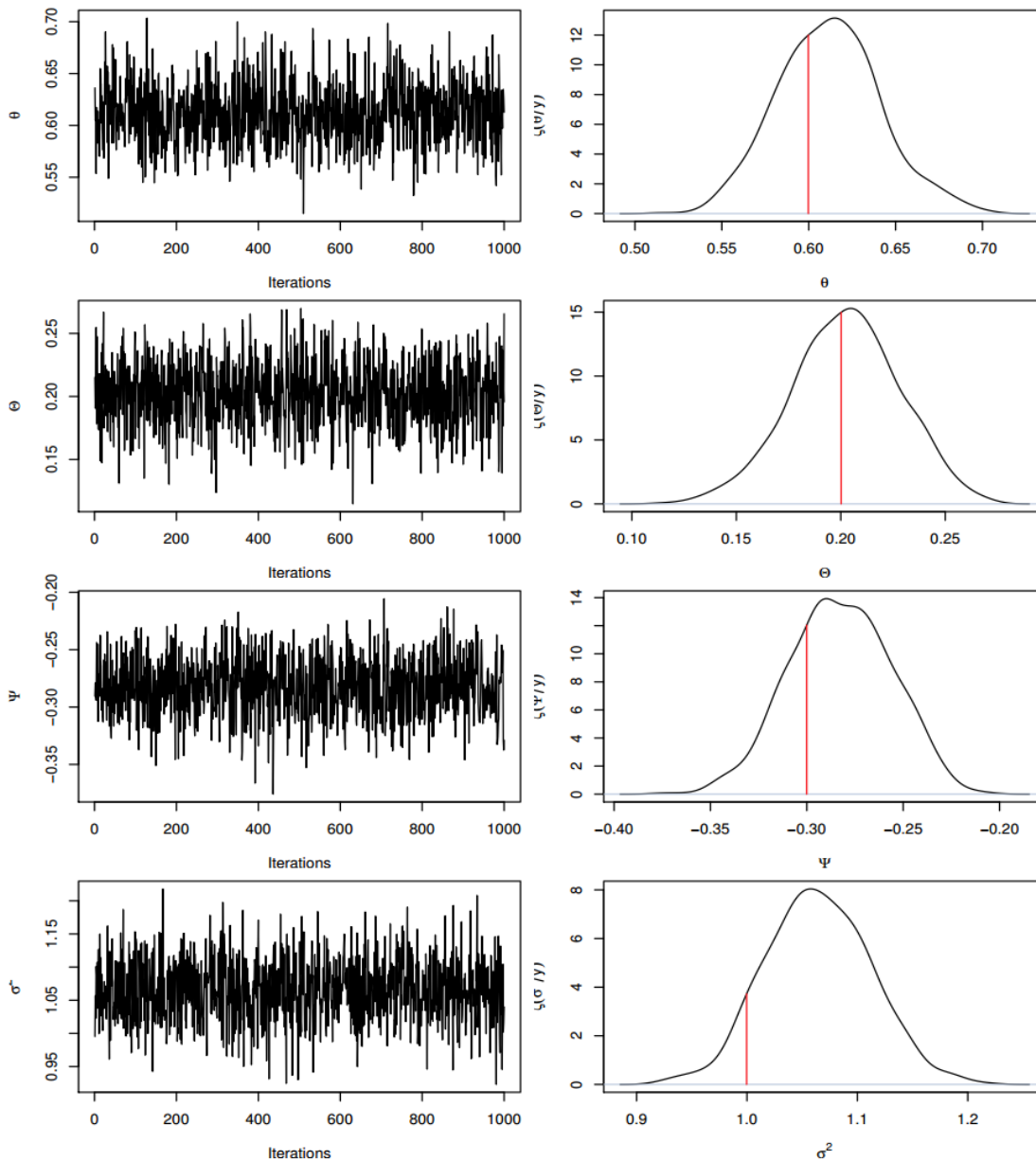
Model	θ_1	θ_2	Θ_1	Θ_2	Ψ_1	Ψ_2	σ^2
I. DSMA(1)(1) ₃₍₁₎₁₂	0.6	-	0.2	-	-0.3	-	1.0
II. DSMA(1)(1) ₃₍₁₎₂₁	0.5	-	-0.4	-	0.1	-	2.0
III. DSMA(1)(1) ₁₃₍₁₎₅₂	0.2	-	-0.3	-	0.4	-	0.5
IV. DSMA(2)(1) ₃₍₁₎₁₂	0.6	-0.1	0.5	-	-0.2	-	1.0
V. DSMA(2)(2) ₃₍₂₎₁₂	0.5	0.2	-0.4	0.3	0.2	0.3	1.0

Once the time series datasets of size $n = 1,000$ are generated from these selected DSMA models, the Bayesian analysis is performed by assuming a non informative prior for the parameters θ , Θ , Ψ and σ^2 and a normal prior with zero mean for initial errors ε_0 with variance $\sigma^2 I_{q^*}$. To apply the proposed Gibbs sampler, the starting values for the parameters θ , Θ , Ψ and σ^2 are obtained using NLS method, and the starting values for ε_0 are assumed to be zeros. For each dataset, the Gibbs sampler was run 11,000 iterations where the first 1,000 draws are ignored and every tenth value in the sequence of the last 10,000 draws is recorded to have an approximately independent sample of 1,000 draws. Accordingly, all posterior estimates of the parameters are computed directly as sample averages of the 1,000 Gibbs sampler draws. In the following, we discuss the results of the proposed algorithm and investigate its convergence.

Table 1 presents the true values and the Bayesian estimates of the parameters for example I. In addition, a 95% confidence interval using the 0.025 and 0.975 percentiles of the simulated draws is constructed for every parameter. From Table 1, it is clear that the Bayesian estimates are close to the true values and the 95% confidence interval includes the true value for each parameter. Sequential plots of the parameters generated sequences together with marginal densities are displayed in Figure 1. The marginal densities are computed using 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.

Table 2: Bayesian results for example I.

Parameter	True Values	Mean	Std. Dev.	Lower 95 % limit	Median	Upper 95 % limit
θ	0.60	0.61	0.03	0.55	0.61	0.67
Θ	0.20	0.20	0.03	0.15	0.20	0.25
Ψ	-0.30	-0.28	0.03	-0.33	-0.28	-0.23
σ^2	1.00	1.06	0.05	0.98	1.06	1.15

Figure 1: Sequential plots and marginal posterior distributions of example I

The convergence of the proposed algorithm is monitored using the diagnostic measures explained in Section 4. The autocorrelations and Raftery and Lewis diagnostics are displayed in Table 2 whereas Table 3 presents the diagnostic measures of Geweke (1992). Table 2 shows that the draws for each of the parameter have small autocorrelations at lags 1, 5, 10 and 50 which indicates that there is no convergence problem. This conclusion was confirmed by the the diagnostic measures of Raftery and Lewis where the reported (Nmin) is 994 which is close to the 1,000 draws we used and I-stat value is about 1 which is less than 5. Geweke diagnostics in Table 3 confirm the convergence of the proposed algorithm where Chi squared probabilities show that the

equal means hypothesis can not be rejected and no dramatic differences between the NSE estimates are found. In addition, the RNE estimates are close to 1 which indicates the iid nature of the output sample.

Table 3: Autocorrelations and Raftery-Lewis diagnostics for example I.

Parameter	Autocorrelations				Raftery-Lewis Diagnostics			
	Lag 1	Lag 5	Lag 10	Lag 50	Burn	Total(N)	(Nmin)	I-stat
θ	0.03	0.04	0.01	-0.01	3	1117	994	1.12
Θ	0.01	0.03	0.01	-0.04	2	1028	994	1.03
Ψ	0.00	0.00	-0.01	0.01	2	1028	994	1.03
σ^2	-0.02	-0.00	-0.02	-0.00	2	948	994	0.95

Table 4: Geweke diagnostics for example I.

Parameter	NSE iid	RNE iid	NSE 4%	RNE 4%	NSE 8%	RNE 8%	NSE 15%	RNE 15%	χ^2
θ	0.0009	1	0.0010	0.93	0.0009	1.07	0.0009	1.17	0.98
Θ	0.0008	1	0.0006	1.78	0.0005	2.31	0.0005	2.99	0.96
Ψ	0.0008	1	0.0008	1.05	0.0008	1.15	0.0007	1.27	0.83
σ^2	0.0015	1	0.0014	1.11	0.0012	1.50	0.0011	2.00	0.63

Similarly to example I, Tables 4-7 present the true values and Bayesian estimates of the parameters for examples II - V. In addition, sequential plots with marginal densities of these examples are displayed in Figures 2 - 5. Similar conclusions to those of example I are obtained. We have applied the proposed Gibbs sampler to several simulated datasets from other DSMA models; and we found their results are similar to those of presented examples, and therefore they are not presented here.

Table 5: Bayesian results for example II

Parameter	True Values	Mean	Std. Dev.	Lower 95 % limit	Median	Upper 95 % limit
θ	0.50	0.50	0.03	0.45	0.50	0.56
Θ	-0.40	-0.39	0.03	-0.45	-0.39	-0.34
Ψ	0.10	0.08	0.03	0.03	0.08	0.13
σ^2	2.00	2.05	0.09	1.88	2.05	2.23

Figure 2: Sequential plots and marginal posterior distributions of example II

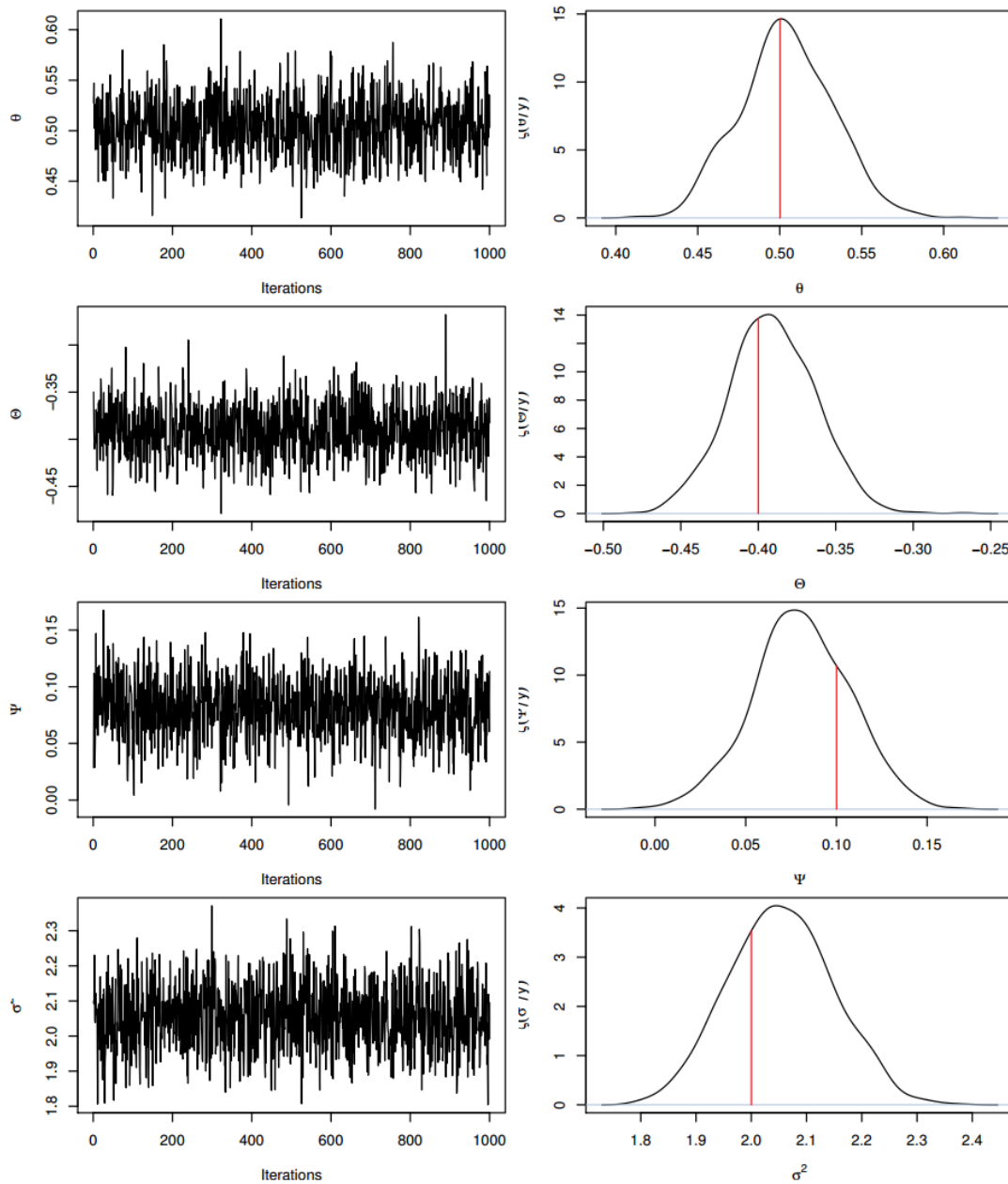


Table 6: Bayesian results for example III

Parameter	True Values	Mean	Std. Dev.	Lower 95 % limit	Median	Upper 95 % limit
θ	0.20	0.22	0.03	0.17	0.22	0.28
Θ	-0.30	-0.27	0.03	-0.32	-0.27	-0.21
Ψ	0.40	0.36	0.03	0.30	0.36	0.42
σ^2	0.50	0.46	0.02	0.42	0.46	0.50

Figure 3: Sequential plots and marginal posterior distributions of example III

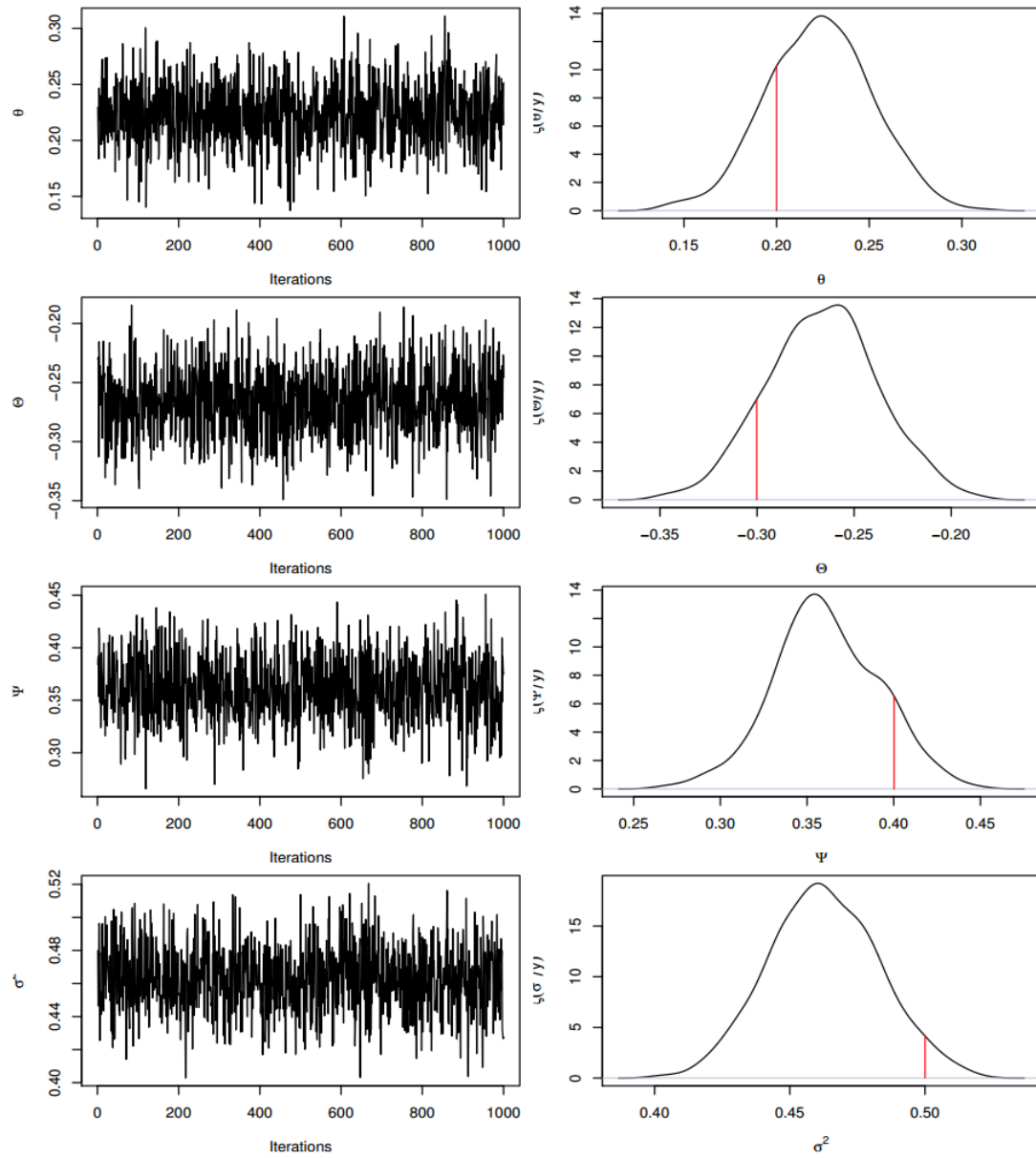
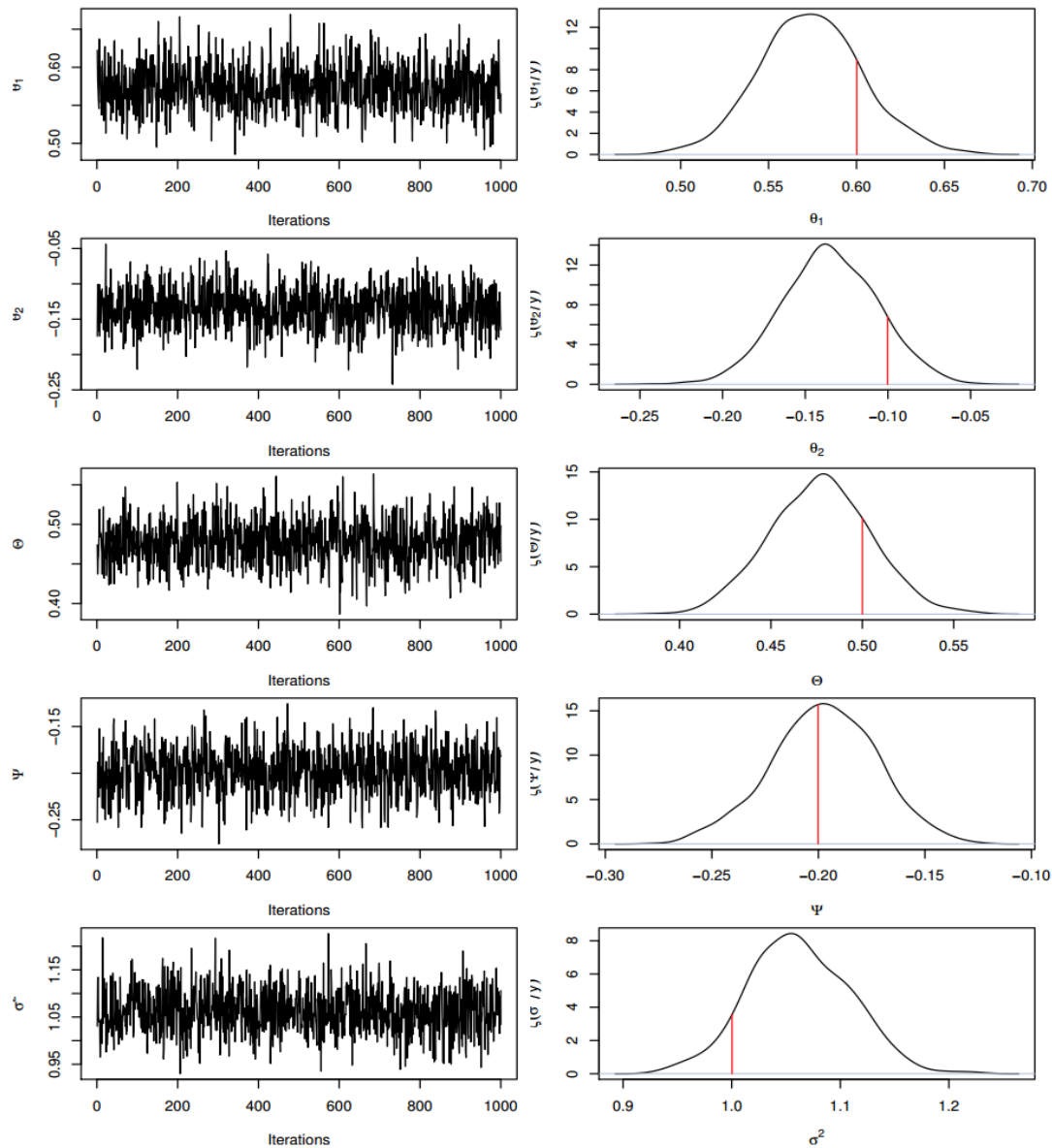


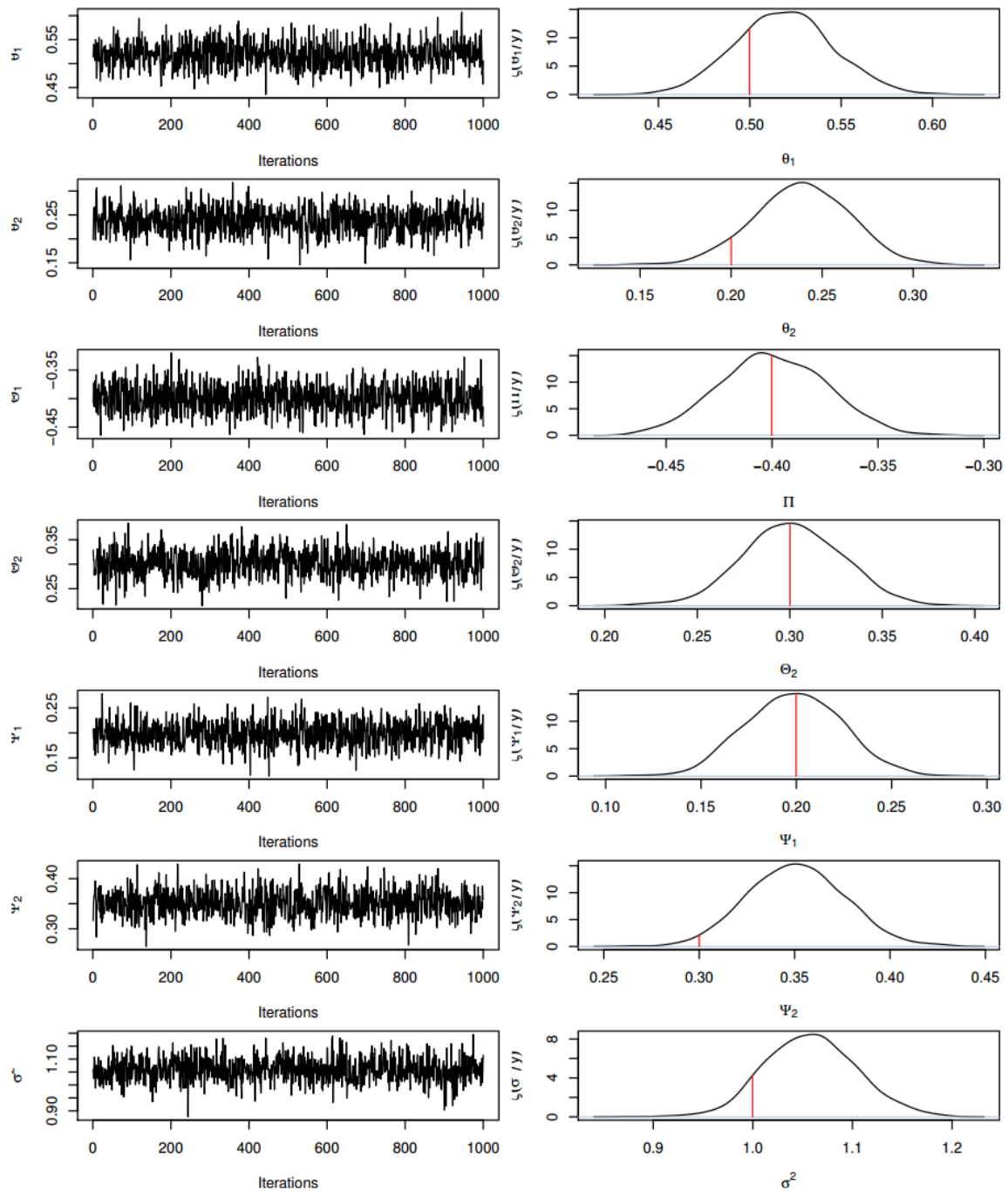
Table 7: Bayesian results for example IV.

Parameter	True Values	Mean	Std. Dev.	Lower 95 % limit	Median	Upper 95 % limit
θ_1	0.60	0.57	0.03	0.52	0.57	0.64
θ_2	-0.10	-0.14	0.03	-0.19	-0.14	-0.08
θ	0.50	0.48	0.03	0.42	0.48	0.53
Ψ	-0.20	-0.20	0.02	-0.25	-0.20	-0.15
σ^2	1.00	1.06	0.05	0.97	1.06	1.15

Figure 4: Sequential plots and marginal posterior distributions of example IV**Table 8: Bayesian results for example V**

Parameter	True Values	Mean	Std. Dev.	Lower 95 % limit	Median	Upper 95 % limit
θ_1	0.50	0.52	0.03	0.47	0.52	0.57
θ_2	0.20	0.24	0.03	0.19	0.24	0.29
Θ_1	-0.40	-0.40	0.02	-0.45	-0.40	-0.35
Θ_2	0.30	0.30	0.03	0.25	0.30	0.35
Ψ_1	0.20	0.20	0.02	0.15	0.20	0.25
Ψ_2	0.30	0.35	0.03	0.30	0.35	0.40
σ^2	1.00	1.06	0.05	0.97	1.06	1.15

Figure 5: Sequential plots and marginal posterior distributions of example V



6. Conclusion

In this paper we showed that the conditional posterior distribution of the DSMA model parameters and the variance are multivariate normal and inverse gamma, respectively. Exploiting that the conditional posterior densities are standard distributions, we used the Gibbs sampling algorithm to develop a Bayesian method for estimating the parameters of the multiplicative DSMA model. Simply, we applied the Gibbs sampling algorithm to

approximate empirically the marginal posterior distributions along with using several diagnostics that showed the convergence of the proposed algorithm was achieved. Accordingly, we computed directly the posterior estimates of the parameters via sample averages of the simulation outputs. The simulation results confirmed the accuracy of the proposed methodology. Future work may investigate model identification using stochastic search variable selection, outliers detection, and extension to multivariate double seasonal models.

References

1. Amin, A.A. and Ismail, M.A. (2015). Gibbs Sampling for Double Seasonal Autoregressive Models. *Communications for Statistical Applications and Methods*, 22(6), 557-573.
2. Au, T, Ma, G.Q, and Yeung, SN (2011). Automatic Forecasting of Double Seasonal Time Series with Applications on Mobility Network Traffic Prediction. 2011 Joint Statistical Meetings, July 30-August4, Miami Beach, Florida, USA.
3. Baek, M. (2010). Forecasting Hourly Electricity Loads of South Korea: Innovations State Space Modeling Approach. *The Korean Journal of Economics* 17(2), 301-317.
4. Box, G.E.P, Jenkins, G.M and Reinsel, G.C. (1994). *Time Series Analysis: Forecasting and Control*, 3rd ed, Prentice-Hall, Inc.:NJ.
5. Caiado, J (2008). Forecasting Water Consumption in Spain Using Univariate Time Series Models. MPRA Paper no. 6610.
6. Cortez, P., Rio, M., Rocha, M. and Sousa, P. (2012). Multi-scale Internet Traffic Forecasting Using Neural Networks and Time Series Methods. *Expert Systems* 29 (2), 143-155.
7. Cruz, A., Munoz,A., Zamora, J.L., Espinola, R. (2011). The effect of Wind Generation and Weekday on Spanish Electricity Spot price Forecasting. *Electric Power Systems Research* 81(10), 1924-1935.
8. Feinberg, E. and Genethliou, D. (2005). Load forecasting. In F. W. J. Chow, *Applied Mathematics for Restructured Electric Power Systems: Control and Computational Intelligence* (pp. 269-285). Springer.
9. 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.
10. Ismail, M. A. (2003a). Bayesian Analysis of Seasonal Autoregressive Models. *Journal of Applied Statistical science*, 12(2), 123-136.
11. Ismail, M. A. (2003b). Bayesian Analysis of Seasonal Moving Average Model: A Gibbs Sampling Approach. *Japanese Journal of Applied Statistics*, 32(2), 61-75.
12. Ismail, M.A. and Amin, A.A. (2014). Gibbs Sampling for SARMA Models. *Pakistan Journal of Statistics*, 30(2), 153-168.

13. LeSage, J. P. (1999). *Applied Econometrics using Matlab*. Dept. of Economics, University of Toledo, available at <http://www.econ.utoledo.edu>.
14. Raftrey, A. E. and Lewis, S. (1992). One long run with diagnostics: Implementation strategies for Markov Chain Monte Carlo. *Statistical Science*, 7, 493-497.
15. Raftrey, 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.
16. Shaarawy, S. and Ismail, M. A. (1987). Bayesian Inference for Seasonal ARMA Models. *The Egyptian Statistical Journal*, 31, 323-336.
17. Kim, M.S. (2013). Modeling Special-day Effects for Forecasting Intraday Electricity demand, *European Journal of Operational Research* 230(1), 170-180.
18. Mohamed, N., Ahmad, M.H., Ismail, Z. and Suhartono (2010). Double Seasonal ARIMA Model for Forecasting Load Demand. *Matematika* 26 (2), 217-231.
19. Mohamed, N., Ahmad, MH and Suhartono (2011). Forecasting Short Term Demand Using Double Seasonal ARIMA Model. *World Applied Sciences Journal* 13, 27-35.
20. Taylor, J.W. (2003). Short-Term Electricity Demand Forecasting Using Double Seasonal Exponential Smoothing. *The Journal of the Operational Research Society* 54(8), 799-805.
21. Taylor, J. W., de Menezes, L. M. and McSharry, P. A (2006). Comparison of Univariate Methods for Forecasting Electricity Demand up to a Day ahead. *International Journal of Forecasting* 22: 1â€“16.
22. Taylor, J. W. (2008a). An Evaluation of Methods for very Short-term Load Forecasting using Minute-by-Minute British Data. *International Journal of Forecasting* 24: 645â€“ 658.
23. Taylor, J.W. (2008b). A Comparison of Univariate Time Series Methods for Forecasting Intraday Arrivals at a Call Center. *Management Science* 54, 253-265.
24. Thompson, H.E, and Tiao, G.C. (1971). Analysis of Telephone Data. *The Bell Journal of Economics and Management Science* 2, 514-541.