Communications for Statistical Applications and Methods
2015, Vol. 22, No. 6, 557–573

DOI: http://dx.doi.org/10.5351/CSAM.2015.22.6.557
Print ISSN 2287-7843 / Online ISSN 2383-4757

Gibbs Sampling for Double Seasonal Autoregressive Models

Ayman A. Amin[1,a], Mohamed A. Ismail[b]

[a]Department of Mathematics, Statistics and Insurance, Munofia University, Egypt
[b]Department of Statistics, Faculty of Economics and Political Science, Cairo University, Egypt

Abstract

In this paper we develop a Bayesian inference for a multiplicative double seasonal autoregressive (DSAR) model by implementing a fast, easy and accurate Gibbs sampling algorithm. We apply the Gibbs sampling to approximate empirically the marginal posterior distributions after showing that the conditional posterior distribution of the model parameters and the variance are multivariate normal and inverse gamma, respectively. The proposed Bayesian methodology is illustrated using simulated examples and real-world time series data.

Keywords: multiplicative seasonal autoregressive, double seasonality, Bayesian analysis, Gibbs sampler, internet traffic data

1. Introduction

Many time series are observed at small time units (such as minutes and hours) with high frequencies. These high frequency time series are characterized by exhibiting complex and multiple seasonal patterns of any periodic pattern of fixed length. For example, hourly electricity load data can exhibit intraday and intraweek seasonal patterns. Other examples where multiple seasonal patterns can occur include daily hospital admissions, hourly volumes of call center arrivals, half-hourly demand for public transportation, hourly traffic on a road, requests for cash at ATM machines every 5 minutes, hourly access to computer web sites and daily usage of water and natural gas.

Seasonal autoregressive integrated moving average (SARIMA) models have been widely and successfully applied to analyze time series data with single seasonal pattern in different disciplines. However, these models need to be modified or extended to accommodate multiple seasonalities. Indeed, the initial notion of modelling multiple seasonalities 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. Several years later, Box *et al.* (1994, p.333) suggested that SARIMA model could be extended to capture multiple seasonality patterns, and Taylor (2003) explicitly stated the multiplicative double SARIMA model. In addition to SARIMA models, other techniques have been extended to fit multiple seasonal time series, which include Neural Network, Exponential Smoother, Innovation State model and Transfer model. A quick review of these techniques can be found in Feinberg and Genethliou (2005, ch.12).

In particular, the multiplicative double SARIMA models have been the subject of interest of many researchers and extensively studied and employed in modeling and forecasting double seasonal time

[1] Corresponding author: Department of Mathematics, Statistics and Insurance, Faculty of Commerce, Munofia University, Egypt. E-mail: aymanaamin@gmail.com

Published 30 November 2015 / journal homepage: http://csam.or.kr

series data. 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 six univariate models including multiplicative SARIMA for electricity demand forecasting in Brazil and in England and Wales. Cruz *et al.* (2011) empirically compared the predictive accuracy of a set of methods for day-ahead spot price forecasting in the Spanish electricity market. Other references can include Au *et al.* (2011), Baek (2010), Caiado (2008), Kim (2013), Mohamed *et al.* (2010, 2011) and Taylor (2008a, 2008b) and references, among others.

Bayesian analysis of SARIMA 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 SARIMA model as an additive not a multiplicative model that can increase the number of unnecessary parameters. 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, 2003b) used Gibbs sampling algorithm to achieve Bayesian analysis for multiplicative seasonal autoregressive (SAR) and seasonal moving average (SMA) models. This work is extended by Ismail and Amin (2010) to a multiplicative SARIMA model. As far as the authors know, the Bayesian analysis of double SARIMA model has not been addressed in the literature except for Ismail and Zahran (2014) who recently developed a Bayesian analysis based on analytical approximation to (Additive) Double Seasonal Autoregressive (DSAR) model. In the current paper, we develop a Bayesian analysis based on Gibbs sampling algorithm to multiplicative DSAR model, which has the advantage that is unconditional on initial values.

The remainder of this paper is organized as: Section 2 presents multiplicative DSAR. Section 3 is devoted to summarizing posterior analysis and the full conditional posterior distributions of the parameters. Section 4 provides the implementation details of the proposed algorithm (including convergence monitoring). The proposed methodology is illustrated in Section 5 using simulated examples and real data sets. Finally, the conclusions are given in Section 6.

2. The Multiplicative Double Seasonal Autoregressive Model (DSAR)

A time series $\{y_t\}$ is said to be generated by a multiplicative seasonal autoregressive model of orders p, P_1, and P_2, denoted by DSAR$(p)(P_1)_{s_1}(P_2)_{s_2}$, if it satisfies

$$\phi_p(B)\Phi_{P_1}(B^{s_1})\Pi_{P_2}(B^{s_2})y_t = \varepsilon_t, \tag{2.1}$$

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 autoregressive polynomial is $\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)$ with order p. In addition, the seasonal autoregressive polynomials are $\Phi_{P_1}(B^{s_1}) = (1 - \Phi_1 B^{s_1} - \Phi_2 B^{2s_1} - \cdots - \Phi_{P_1} B^{P_1 s_1})$ with order P_1 and $\Pi_{P_2}(B^{s_2}) = (1 - \Pi_1 B^{s_2} - \Pi_2 B^{2s_2} - \cdots - \Pi_{P_2} B^{P_2 s_2})$ with order P_2. Finally, the non seasonal and seasonal autoregressive coefficients are $\phi = (\phi_1, \phi_2, \ldots, \phi_p)^T$, $\Phi = (\Phi_1, \Phi_2, \ldots, \Phi_{P_1})^T$ and $\Pi = (\Pi_1, \Pi_2, \ldots, \Pi_{P_2})^T$, respectively. The time series $\{y_t\}$ is assumed to start at time $t = 1$ with unknown starting observations $\mathbf{y_0} = (y_0, y_{-1}, \ldots, y_{1-p-Ps})$.

It should be noted that the DSAR model (2.1) has an extra terms compared with the usual multiplicative single SAR model. The new term is $\Pi_{P_2}(B^{s_2})$ that accommodates the second seasonal pattern.

Accordingly, the model (2.1) can be written as

$$\begin{aligned} y_t &= \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{P_1} \Phi_j y_{t-js_1} + \sum_{\tau=1}^{P_2} \Pi_\tau y_{t-\tau s_2} - \sum_{i=1}^{p}\sum_{j=1}^{P_1} \phi_i \Phi_j y_{t-i-js_1} - \sum_{i=1}^{p}\sum_{\tau=1}^{P_2} \phi_i \Pi_\tau y_{t-i-\tau s_2} \\ &- \sum_{j=1}^{P_1}\sum_{\tau=1}^{P_2} \Phi_j \Pi_\tau y_{t-js_1-\tau s_2} + \sum_{i=1}^{p}\sum_{j=1}^{P_1}\sum_{\tau=1}^{P_2} \phi_i \Phi_j \Pi_\tau y_{t-i-js_1-\tau s_2} + \varepsilon_t \\ &= X_t\beta + \varepsilon_t, \end{aligned} \tag{2.2}$$

where

$$\begin{aligned} X_t = \Big(& y_{t-1}, \ldots, y_{t-p}, y_{t-s_1}, y_{t-s_1-1}, \ldots, y_{t-s_1-p}, \ldots\ldots, y_{t-P_1s_1}, y_{t-P_1s_1-1}, \ldots, y_{t-P_1s_1-p}, y_{t-s_2}, \\ & y_{t-s_2-1}, \ldots, y_{t-s_2-p}, y_{t-s_1-s_2}, y_{t-s_1-s_2-1}, \ldots, y_{t-s_1-s_2-p}, \ldots\ldots, y_{t-P_1s_1-s_2}, y_{t-P_1s_1-s_2-1}, \\ & \ldots, y_{t-P_1s_1-s_2-p}, \ldots\ldots\ldots, y_{t-P_2s_2}, y_{t-P_2s_2-1}, \ldots, y_{t-P_2s_2-p}, y_{t-s_1-P_2s_2}, y_{t-s_1-P_2s_2-1}, \ldots, \\ & y_{t-s_1-P_2s_2-p}, \ldots\ldots, y_{t-P_1s_1-P_2s_2}, y_{t-P_1s_1-P_2s_2-1}, \ldots, y_{t-P_1s_1-P_2s_2-p}\Big), \\ \beta = \Big(& \phi_1, \ldots, \phi_p, \Phi_1, -\phi_1\Phi_1, \ldots, -\phi_p\Phi_1, \ldots\ldots, \Phi_{P_1}, -\phi_1\Phi_{P_1}, \ldots, -\phi_p\Phi_{P_1}, \Pi_1, -\phi_1\Pi_1, \ldots, \\ & -\phi_p\Pi_1, -\Phi_1\Pi_1, \phi_1\Phi_1\Pi_1, \ldots, \phi_p\Phi_1\Pi_1, \ldots\ldots, -\Phi_{P_1}\Pi_1, \phi_1\Phi_{P_1}\Pi_1, \ldots, \phi_p\Phi_{P_1}\Pi_1, \ldots\ldots\ldots, \\ & \Pi_{P_2}, -\phi_1\Pi_{P_2}, \ldots, -\phi_p\Pi_{P_2}, -\Phi_1\Pi_{P_2}, \phi_1\Phi_1\Pi_{P_2}, \ldots, \phi_p\Phi_1\Pi_{P_2}, \ldots\ldots, -\Phi_{P_1}\Pi_{P_2}, \phi_1\Phi_{P_1}\Pi_{P_2}, \\ & \ldots, \phi_p\Phi_{P_1}\Pi_{P_2}\Big)^T. \end{aligned} \tag{2.3}$$

Equation (2.2) shows that the multiplicative DSAR model can be written as an autoregressive model of order $(1 + p)(1 + P_1)(1 + P_2) - 1$ with some coefficients that are products of nonseasonal and seasonal coefficients. Therefore, the model is nonlinear in ϕ, Φ, and Π which complicates the Bayesian analysis. However, the following sections explain how Gibbs sampling technique can facilitate the analysis. The DSAR model (2.2) is stationary if the roots of the polynomials $\phi(B)$, $\Phi_{P_1}(B^{s_1})$ and $\Pi_{P_2}(B^{s_2})$ lie outside the unit circle. For more details about the properties of seasonal AR models see Box *et al.* (1994).

3. Posterior Analysis

3.1. Likelihood function

Suppose $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ is a realization of the DSAR model (2.2), 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 $L(\phi, \Phi, \Pi, \Psi, \sigma^2, \mathbf{y_0} \mid \mathbf{y}) = l$ is given by

$$\begin{aligned} l &\propto \left(\sigma^2\right)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{n}\varepsilon_t^2\right\} \\ &= \left(\sigma^2\right)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - X\beta)^T(\mathbf{y} - X\beta)\right\}, \end{aligned} \tag{3.1}$$

where X and β are defined in (2.3), and

$$\varepsilon_t = y_t - \sum_{i=1}^{p} \phi_i y_{t-i} - \sum_{j=1}^{P_1} \Phi_j y_{t-js_1} - \sum_{\tau=1}^{P_2} \Pi_\tau y_{t-\tau s_2} + \sum_{i=1}^{p}\sum_{j=1}^{P_1} \phi_i \Phi_j y_{t-i-js_1} + \sum_{i=1}^{p}\sum_{\tau=1}^{P_2} \phi_i \Pi_\tau y_{t-i-\tau s_2}$$
$$+ \sum_{j=1}^{P_1}\sum_{\tau=1}^{P_2} \Phi_j \Pi_\tau y_{t-js_1-\tau s_2} - \sum_{i=1}^{p}\sum_{j=1}^{P_1}\sum_{\tau=1}^{P_2} \phi_i \Phi_j \Pi_\tau y_{t-i-js_1-\tau s_2}. \tag{3.2}$$

It is obvious that this likelihood function is a complicated function in ϕ, Φ, Π and $\mathbf{y_0}$.

3.2. Prior specification

Assuming that the parameters ϕ, Φ, Π and $\mathbf{y_0}$ are independent apriori, given the error variance parameter σ^2, the joint prior distribution is

$$\begin{aligned}
&\zeta\left(\phi, \Phi, \Pi, \sigma^2, \mathbf{y_0}\right) \\
&= \zeta\left(\phi \mid \sigma^2\right) \times \zeta\left(\Phi \mid \sigma^2\right) \times \zeta\left(\Pi \mid \sigma^2\right) \times \zeta\left(\mathbf{y_0} \mid \sigma^2\right) \times \zeta\left(\sigma^2\right) \\
&= N_p\left(\mu_\phi, \sigma^2\Sigma_\phi\right) \times N_{P_1}\left(\mu_\Phi, \sigma^2\Sigma_\Phi\right) \times N_{P_2}\left(\mu_\Pi, \sigma^2\Sigma_\Pi\right) \times N_{p^\star}\left(\mu_{\mathbf{y_0}}, \sigma^2\Sigma_{y_0}\right) \times \mathrm{IG}\left(\frac{\nu}{2}, \frac{\lambda}{2}\right),
\end{aligned} \tag{3.3}$$

where $p^\star = p + P_1 s_1 + P_2 s_2$, $N_r(\mu, \Delta)$ is the r-variate normal distribution with mean μ and variance Δ, and $\mathrm{IG}(a, b)$ is the inverse gamma distribution with parameters a and b. Therefore, the joint prior distribution can be written as follows

$$\zeta\left(\phi, \Phi, \Pi, \sigma^2, y_0\right) \propto \left(\sigma^2\right)^{-\left(\frac{\nu^\star}{2}+1\right)} \exp\left\{-\frac{1}{2\sigma^2}\left[\lambda + \left(\phi - \mu_\phi\right)^T \Sigma_\phi^{-1}\left(\phi - \mu_\phi\right) + \left(\Phi - \mu_\Phi\right)^T \Sigma_\Phi^{-1}\left(\Phi - \mu_\Phi\right)\right.\right.$$
$$\left.\left. + \left(\Pi - \mu_\Pi\right)^T \Sigma_\Pi^{-1}\left(\Pi - \mu_\Pi\right) + \left(\mathbf{y_0} - \mu_{\mathbf{y_0}}\right)^T \Sigma_{y_0}^{-1}\left(\mathbf{y_0} - \mu_{\mathbf{y_0}}\right)\right]\right\}, \tag{3.4}$$

where $\nu^\star = \nu + 2p + P_1(1 + s_1) + P_2(1 + s_2)$. The prior distribution (3.4) is chosen for several reasons, especially it is a conjugate prior and thus facilitates the mathematical calculations. Multiplying the joint prior distribution (3.4) by the approximate likelihood function (3.1) results in the joint posterior $\zeta(\phi, \Phi, \Pi, \sigma^2, y_0 \mid \mathbf{y})$ which may be written as

$$\zeta\left(\phi, \Phi, \Pi, \sigma^2, y_0 \mid \mathbf{y}\right) \propto \left(\sigma^2\right)^{-\left(\frac{n+\nu^\star}{2}+1\right)} \exp\left\{-\frac{1}{2\sigma^2}\left[\lambda + \left(\phi - \mu_\phi\right)^T \Sigma_\phi^{-1}\left(\phi - \mu_\phi\right) + \left(\Phi - \mu_\Phi\right)^T\right.\right.$$
$$\Sigma_\Phi^{-1}\left(\Phi - \mu_\Phi\right) + \left(\Pi - \mu_\Pi\right)^T \Sigma_\Pi^{-1}\left(\Pi - \mu_\Pi\right) + \left(\mathbf{y_0} - \mu_{\mathbf{y_0}}\right)^T \Sigma_{y_0}^{-1}\left(\mathbf{y_0} - \mu_{\mathbf{y_0}}\right)$$
$$\left.\left. + \left(\mathbf{y} - X\beta\right)^T\left(\mathbf{y} - X\beta\right)\right]\right\}. \tag{3.5}$$

3.3. Full conditional distributions

The conditional posterior distribution for each of the unknown parameters is obtained from the joint posterior distribution (3.5) 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 the conditional posteriors that are obtained for the unknown parameters are normal and inverse gamma distributions for which sampling techniques exist.

3.3.1. The conditional posterior of ϕ

We obtained the conditional posterior of ϕ by finding out $\zeta(\phi \mid \mathbf{y}, \Phi, \Pi, \sigma^2, \mathbf{y_0})$ that we proved to be a multivariate normal $N(\mu_\phi^\star, v_\phi^\star)$ with

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

and

$$v_\phi^\star = \sigma^2\left(H_\phi^T H_\phi + \Sigma_\phi^{-1}\right)^{-1}.$$

Where H_ϕ is a $(n \times p)$ matrix with the ti^{th} element:

$$H_{\phi_{ti}} = \left(y_{t-i} - \sum_{j=1}^{P_1}\Phi_j y_{t-i-js_1} - \sum_{\tau=1}^{P_2}\Pi_\tau y_{t-i-\tau s_2} + \sum_{j=1}^{P_1}\sum_{\tau=1}^{P_2}\Phi_j\Pi_\tau y_{t-i-js_1-\tau s_2}\right),$$

L_ϕ is a $n \times ((1+P_1)(1+P_2)-1)$ matrix with the t^{th} element:

$$L_{\phi_t} = \left(y_{t-s_1}, \ldots, y_{t-P_1s_1}, y_{t-s_2}, y_{t-s_1-s_2}, \ldots, y_{t-P_1s_1-s_2}, \ldots\ldots, y_{t-P_2s_2}, y_{t-s_1-P_2s_2}, \ldots, y_{t-P_1s_1-P_2s_2}\right),$$

and β_ϕ is a column vector of order $(1+P_1)(1+P_2)-1$ written as:

$$\beta_\phi = \left(\Phi_1, \ldots, \Phi_{P_1}, \Pi_1, -\Phi_1\Pi_1, \ldots, -\Phi_{P_1}\Pi_1, \ldots\ldots, \Pi_{P_2}, -\Phi_1\Pi_{P_2}, \ldots, -\Phi_{P_1}\Pi_{P_2}\right)^T.$$

The conditional posterior of ϕ is

$$\phi^j \sim \zeta\left(\phi^j \mid \mathbf{y}, \left(\sigma^2\right)^{j-1}, \Phi^{j-1}, \Pi^{j-1}, \theta^{j-1}, \Theta^{j-1}, \Psi^{j-1}, \mathbf{y_0}^{j-1}, \boldsymbol{\varepsilon_0}^{j-1}\right) = N\left(\mu_\phi^\star, v_\phi^\star\right),$$

where

$$\mu_\phi^\star = \left[\left(H_\phi^T H_\phi + \Sigma_\phi^{-1}\right)^{-1}\left(\Sigma_\phi^{-1}\mu_\phi + H_\phi^T\left(\mathbf{y} - L_\phi\beta_\phi - \hat{\Lambda}\beta_2\right)\right)\right], \qquad v_\phi^\star = \sigma^2\left(H_\phi^T H_\phi + \Sigma_\phi^{-1}\right)^{-1}.$$

H_ϕ is a $(n \times p)$ matrix with ti^{th} element $H_{\phi_{ti}}2 = (y_{t-i} - \sum_{j=1}^{P}\Phi_j y_{t-js-i})$ and L is a $(n \times P)$ matrix with tj^{th} element $L_{tj} = (y_{t-js})$.

3.3.2. The conditional posterior of Φ

We obtained the conditional posterior of Φ by finding out $\zeta(\Phi \mid \mathbf{y}, \phi, \Pi, \sigma^2, \mathbf{y_0})$ that we proved to be a multivariate normal $N(\mu_\Phi^\star, v_\Phi^\star)$ with

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

and

$$v_\Phi^\star = \sigma^2\left(H_\Phi^T H_\Phi + \Sigma_\Phi^{-1}\right)^{-1}.$$

Where H_Φ is a $(n \times P_1)$ matrix with the tj^{th} element:

$$H_{\Phi_{tj}} = \left(y_{t-js_1} - \sum_{i=1}^{p} \phi_i y_{t-i} - \sum_{\tau=1}^{P_2} \Pi_\tau y_{t-js_1-\tau s_2} + \sum_{i=1}^{p} \sum_{\tau=1}^{P_2} \phi_i \Pi_\tau y_{t-i-js_1-\tau s_2} \right),$$

L_Φ is a $n \times ((1+p)(1+P_2)-1)$ matrix with the t^{th} element:

$$L_{\Phi_t} = \left(y_{t-1}, \ldots, y_{t-p}, y_{t-s_2}, y_{t-1-s_2}, \ldots, y_{t-p-s_2}, \ldots\ldots, y_{t-P_2 s_2}, y_{t-1-P_2 s_2}, \ldots, y_{t-p-P_2 s_2} \right),$$

and β_Φ is a column vector of order $(1+p)(1+P_2)-1$ written as:

$$\beta_\Phi = \left(\phi_1, \ldots, \phi_p, \Pi_1, -\phi_1 \Pi_1, \ldots, -\phi_p \Pi_1, \ldots\ldots, \Pi_{P_2}, -\phi_1 \Pi_{P_2}, \ldots, -\phi_p \Pi_{P_2} \right)^T.$$

3.3.3. The conditional posterior of Π

We obtained the conditional posterior of Π by finding out $\zeta(\Pi \mid \mathbf{y}, \phi, \Phi, \sigma^2, \mathbf{y_0})$ that we proved to be a multivariate normal $N(\mu_\Pi^\star, v_\Pi^\star)$ with

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

and

$$v_\Pi^\star = \sigma^2 \left(H_\Pi^T H_\Pi + \Sigma_\Pi^{-1} \right)^{-1}.$$

Where H_Π is a $(n \times P_2)$ matrix with the $t\tau^{th}$ element:

$$H_{\Pi_{t\tau}} = \left(y_{t-\tau s_2} - \sum_{i=1}^{p} \phi_i y_{t-i} - \sum_{j=1}^{P_1} \Phi_j y_{t-js_1-\tau s_2} + \sum_{i=1}^{p} \sum_{j=1}^{P_1} \phi_i \Phi_j y_{t-i-js_1-\tau s_2} \right),$$

L_Π is a $n \times ((1+p)(1+P_1)-1)$ matrix with the t^{th} element:

$$L_{\Pi_t} = \left(y_{t-1}, \ldots, y_{t-p}, y_{t-s_1}, y_{t-1-s_1}, \ldots, y_{t-p-s_1}, \ldots\ldots, y_{t-P_1 s_1}, y_{t-1-P_1 s_1}, \ldots, y_{t-p-P_1 s_1} \right),$$

and β_Π is a column vector of order $(1+p)(1+P_1)-1$ written as:

$$\beta_\Pi = \left(\phi_1, \ldots, \phi_p, \Phi_1, -\phi_1 \Phi_1, \ldots, -\phi_p \Phi_1, \ldots\ldots, \Phi_{P_1}, -\phi_1 \Phi_{P_1}, \ldots, -\phi_p \Phi_{P_1} \right)^T.$$

3.3.4. The conditional posterior of σ^2

We obtained the conditional posterior of σ^2 by finding out $\zeta(\sigma^2 \mid \mathbf{y}, \phi, \Phi, \Pi, \mathbf{y_0})$ that we proved to be an inverse gamma $IG((n+v^\star)/2, \{\lambda + n(S^2)\}/2)$, where $v^\star = v + 2p + P_1(1+s_1) + P_2(1+s_2)$ and

$$n(S^2) = \Big[\left(\phi - \mu_\phi \right)^T \Sigma_\phi^{-1} \left(\phi - \mu_\phi \right) + (\Phi - \mu_\Phi)^T \Sigma_\Phi^{-1} (\Phi - \mu_\Phi) + (\Pi - \mu_\Pi)^T \Sigma_\Pi^{-1} (\Pi - \mu_\Pi)$$
$$+ \left(\mathbf{y_0} - \mu_{\mathbf{y_0}} \right)^T \Sigma_{y_0}^{-1} \left(\mathbf{y_0} - \mu_{\mathbf{y_0}} \right) + (\mathbf{y} - X\beta)^T (\mathbf{y} - X\beta) \Big].$$

To ease the Gibbs sampling algorithm process, the parameter σ^2 can be sampled from the Chi square distribution using the transformation $\{\lambda + n(S^2)\}/\sigma^2 \sim \chi^2_{(n+v^\star)}$.

3.3.5. The conditional posterior of $\mathbf{y_0}$

In the beginning, we write explicitly the elements of $\mathbf{y_0}$ using the model (2.2) as:

$$F\ \mathbf{y_{p^\star}} = D\ \mathbf{y_0} + \varepsilon_{\mathbf{p^\star}}, \tag{3.6}$$

where

$$F = \begin{pmatrix} 1 & 0 & 0 & \cdots & \cdots & 0 & 0 \\ -\gamma_1 & 1 & 0 & \cdots & \cdots & 0 & 0 \\ -\gamma_2 & -\gamma_1 & 1 & \cdots & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ -\gamma_{p^\star-1} & -\gamma_{p^\star-2} & \cdots & \cdots & -\gamma_2 & -\gamma_1 & 1 \end{pmatrix}_{(p^\star)\times(p^\star)},$$

$$D = \begin{pmatrix} \gamma_1 & \gamma_2 & \gamma_3 & \cdots & \cdots & \gamma_{p^\star-1} & \gamma_{p^\star} \\ \gamma_2 & \gamma_3 & \gamma_4 & \cdots & \cdots & \gamma_{p^\star} & 0 \\ \gamma_3 & \gamma_4 & \gamma_5 & \cdots & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \gamma_{p^\star-1} & \gamma_{p^\star} & 0 & \cdots & \cdots & 0 & 0 \\ \gamma_{p^\star} & 0 & 0 & \cdots & \cdots & 0 & 0 \end{pmatrix}_{(p^\star)\times(p^\star)},$$

$p^\star = p + P_1 s_1 + P_2 s_2$ and $\mathbf{y_{p^\star}} = (y_1, y_2, \ldots, y_{p^\star})^T$ that has the $p^\star$ multivariate normal distribution with zero mean and variance $(\sigma^2 I_{p^\star})$, where $I_{p^\star}$ is the unit matrix of order $p^\star$.

Using the above defined matrices and the standard Bayesian techniques, we obtained the conditional posterior of $\mathbf{y_0}$ by finding out $\zeta(\mathbf{y_0} \mid \mathbf{y}, \phi, \Phi, \Pi, \sigma^2)$ that we proved to be a multivariate normal $N(\mu^\star_{y_0}, v^\star_{y_0})$ with

$$\mu^\star_{y_0} = \left[D^T D + \Sigma^{-1}_{y_0}\right]^{-1} \left[\Sigma^{-1}_{y_0}\mu_{y_0} + D^T \left(F\ \mathbf{y_{p^\star}}\right)\right]$$

and

$$v^\star_{y_0} = \sigma^2 \left(D^T D + \Sigma^{-1}_{y_0}\right)^{-1}.$$

4. The Proposed Gibbs Sampler

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

1. Specify starting values ϕ^0, $\mathbf{\Phi}^0$, $\mathbf{\Pi}^0$, $(\sigma^2)^0$, and $\mathbf{y}_0^0$ 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 (3.2) and the IS parameter estimates.

3. Simulate
   - $\phi^j \sim \zeta\left(\phi^j \mid \mathbf{y}, \left(\sigma^2\right)^{j-1}, \Phi^{j-1}, \Pi^{j-1}, \mathbf{y_0}^{j-1}\right)$,
   - $\Phi^j \sim \zeta\left(\Phi^j \mid \mathbf{y}, \left(\sigma^2\right)^{j-1}, \phi^j, \Pi^{j-1}, \mathbf{y_0}^{j-1}\right)$,
   - $\Pi^j \sim \zeta\left(\Pi^j \mid \mathbf{y}, \left(\sigma^2\right)^{j-1}, \phi^j, \Phi^j, \mathbf{y_0}^{j-1}\right)$,

- $(\sigma^2)^j \sim \zeta\left(\left(\sigma^2\right)^j \mid \mathbf{y}, \phi^j, \Phi^j, \Pi^j, \mathbf{y_0}^{j-1}\right)$,
- $\mathbf{y_0}^j \sim \zeta\left(\mathbf{y_0}^j \mid \mathbf{y}, \left(\sigma^2\right)^j, \phi^j, \Phi^j, \Pi^j\right)$.

4. set $j = j + 1$ and go to 3.

This algorithm gives the next value of the Markov chain $\{\phi^{j+1}, \mathbf{\Phi}^{j+1}, \mathbf{\Pi}^{j+1}, (\sigma^2)^{j+1}, \mathbf{y}_0^{j+1}\}$ 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 $\{\phi^{j+1}, \mathbf{\Phi}^{j+1}, \mathbf{\Pi}^{j+1}, (\sigma^2)^{j+1}, \mathbf{y}_0^{j+1}\}$ 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 Markov chain can be monitored by three groups of diagnostics (autocorrelation estimates, Raftery and Lewis diagnostics, and Geweke diagnostics). First, autocorrelation estimates indicate how much independence exists in the sequence of each parameter draws. A high degree of autocorrelation indicates that more draws are needed to obtain accurate posterior estimates. Second, diagnostics proposed by Raftery and Lewis (1992, 1995) include (1) Burn: number of draws used as initial draws or "burn-in" before starting to sample the draws for purpose of posterior inference, (2) Total: total number of draws needed to achieve desired level of accuracy, (3) Nmin: number of draws that would be needed if the draws represented an iid chain, and (4) I-stat: the ratio of the (Total) to (Nmin). Raftery and Lewis suggested that convergence problem may be indicated when values of I-stat exceed 5. Third, diagnostics proposed by Geweke (1992), which includes two groups:

1. 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 required number of draws to produce the same numerical accuracy when iid sample is drawn directly from the posterior distribution.

2. The second group of diagnostics includes a test of whether the sampler has attained an equilibrium state. This is done by carrying out 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 the calculations of the above convergence measures using MATLAB package. These diagnostics will be used in Section 5 to monitor the convergence of the proposed algorithm.

5. Illustrative Examples

In this section we investigate the efficiency of the proposed estimation method based on simulated and real-world datasets.

5.1. Simulated examples

In this subsection we present three examples of DSAR models with simulated data. Table 1 shows these selected DSAR models and their corresponding true parameters values. By these three examples we try to represent different pairs of the season periods that are chosen to cover different seasonality patterns with different data type.

Table 1: Simulated examples design

Model	ϕ_1	ϕ_2	Φ_1	Φ_2	Π_1	Π_2	σ^2
I. DSAR(1)(1)$_3$(1)$_{12}$	0.60	-	0.40	-	−0.30	-	1.00
II. DSAR(1)(1)$_{13}$(1)$_{52}$	0.30	-	−0.60	-	0.40	-	0.50
III. DSAR(2)(2)$_{13}$(2)$_{52}$	0.65	−0.30	0.40	0.15	−0.15	0.10	0.75

Table 2: Bayesian results for Example I

Parameter	True values	Mean	Std. Dev.	Lower 95 % limit	Median	Upper 95 % limit
ϕ	0.60	0.64	0.02	0.59	0.64	0.68
Φ	0.40	0.39	0.03	0.33	0.39	0.44
Π	−0.30	−0.29	0.03	−0.35	−0.29	−0.23
σ^2	1.00	1.02	0.04	0.94	1.02	1.11

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
ϕ	−0.02	−0.02	0.01	−0.03	2	948	994	0.95
Φ	−0.09	0.04	−0.04	−0.04	2	1028	994	1.03
Π	0.02	0.01	−0.04	0.01	2	1028	994	1.03
σ^2	−0.03	−0.05	−0.02	−0.03	2	1028	994	1.03

Table 4: Geweke diagnostics for Example I

	NSE iid	RNE iid	NSE 4%	RNE 4%	NSE 8%	RNE 8%	NSE 15%	RNE 15%	χ^2
ϕ	0.0010	1	0.0010	1.0149	0.0009	1.2064	0.0009	1.4460	0.7383
Φ	0.0010	1	0.0008	1.7299	0.0007	2.2224	0.0006	2.7406	0.8864
Π	0.0010	1	0.0009	1.1668	0.0008	1.6617	0.0007	1.8472	0.6076
σ^2	0.0014	1	0.0012	1.2637	0.0011	1.4413	0.0011	1.4780	0.3342

Once the time series datasets are generated from these three selected DSAR models, the Bayesian analysis is performed by assuming a non informative prior for the parameters ϕ, Φ, Π and σ^2 and a normal prior with zero mean and variance $\sigma^2 I_{p^\star}$ for the initial observations $\mathbf{y_0}$. To apply the proposed Gibbs sampler, the starting values for the parameters ϕ, Φ, Π and σ^2 are obtained using IS method, and the starting values for $\mathbf{y_0}$ 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. All posterior estimates are computed directly as sample averages of the Gibbs sampler draws. In the following, we discuss the results of the proposed Gibbs sampler and investigate the convergence diagnostics.

Table 2 presents the true values and Bayesian estimates of the parameters for Example I. Moreover, a 95% confidence interval using the 0.025 and 0.975 percentiles of the simulated draws is constructed for every parameter. Table 2 shows 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.

The convergence of the proposed algorithm is monitored using the diagnostic measures explained

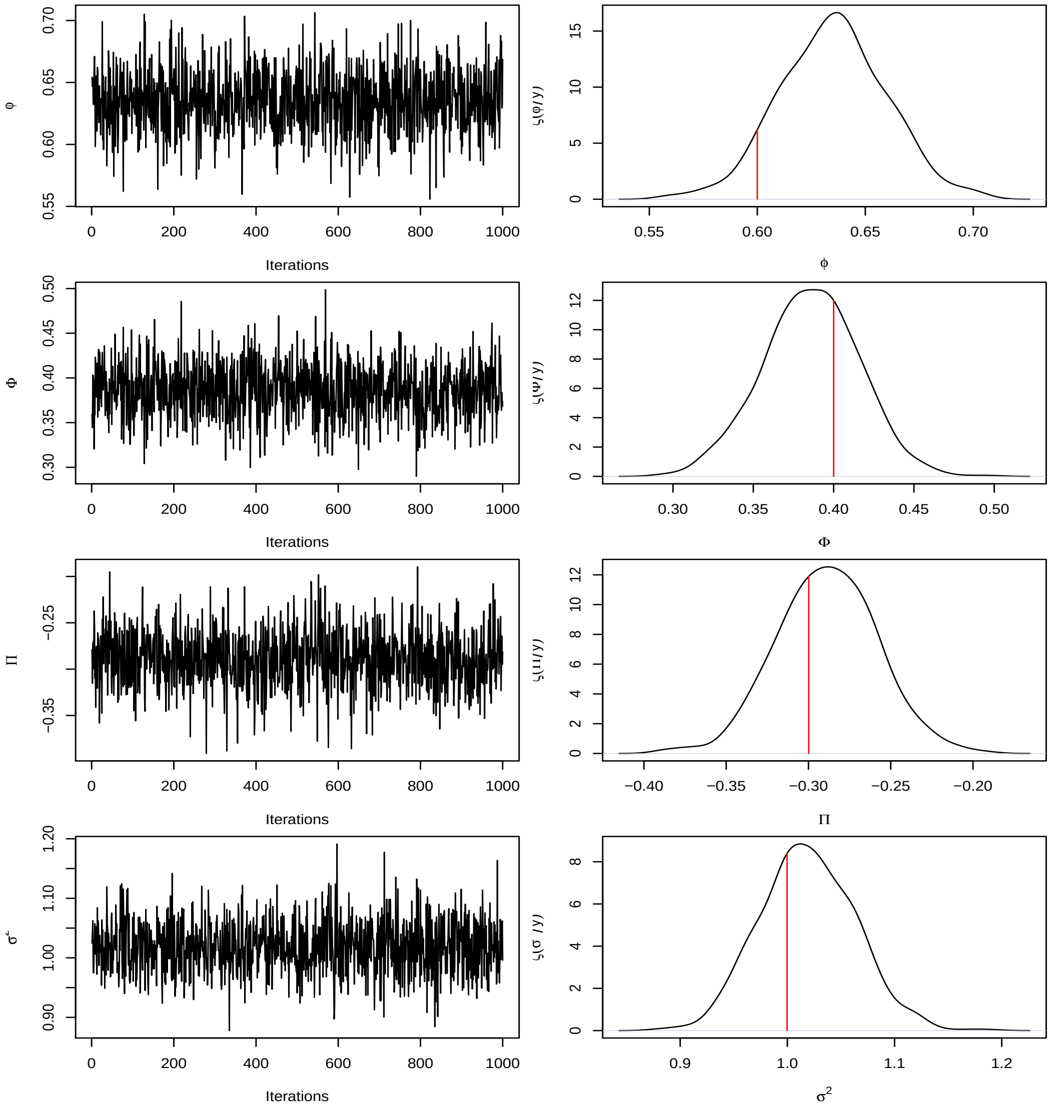

Figure 1: *Sequential plots and marginal posterior distributions of Example I.*

in Section 4. The autocorrelations and Raftery and Lewis diagnostics are displayed in Table 3 whereas Table 4 presents the diagnostic measures of Geweke (1992). Table 3 shows that the draws for each of the parameter have small autocorrelations at lags 1, 5, 10 and 50, which indicates there is no convergence problem. This conclusion was confirmed by the diagnostic measures of Raftery and Lewis where the reported (Nmin) is 994 that is close to the 1000 draws we used and I-stat value is about 1, which is less than 5. Scanning the entries of Table 4 confirms the convergence of the proposed algorithm where Chi squared probabilities show that the equal means hypothesis cannot be rejected and no dramatic differences between the NSE estimates are found. In addition, the RNE estimates are

Table 5: Bayesian results for Example II

Parameter	True values	Mean	Std. Dev.	Lower 95 % limit	Median	Upper 95 % limit
ϕ	0.30	0.29	0.03	0.23	0.28	0.35
Φ	−0.60	−0.61	0.03	−0.67	−0.61	−0.56
Π	0.40	0.35	0.03	0.29	0.35	0.41
σ^2	0.50	0.47	0.02	0.43	0.47	0.51

Table 6: Bayesian results for Example III

Parameter	True values	Mean	Std. Dev.	Lower 95 % limit	Median	Upper 95 % limit
ϕ_1	0.65	0.64	0.03	0.58	0.64	0.69
ϕ_2	−0.30	−0.29	0.03	−0.35	−0.29	−0.24
Φ_1	0.40	0.40	0.03	0.34	0.40	0.46
Φ_2	0.15	0.13	0.03	0.06	0.13	0.19
Π_1	−0.15	−0.17	0.03	−0.23	−0.17	−0.11
Π_2	0.10	0.11	0.03	0.05	0.11	0.17
σ^2	0.75	0.69	0.03	0.64	0.69	0.75

close to 1, which indicates the iid nature of the output sample.

Similarly to Example I, Tables 5 and 6 present the true values and Bayesian estimates of the parameters for Examples II and III. In addition, sequential plots with marginal densities of these two examples are displayed in Figures 2 and 3. Similar conclusions to those of example I are obtained. We have applied the proposed Gibbs sampler to several simulated datasets from other DSAR$(1)(1)_{s1}(1)_{s2}$ and DSAR$(2)(2)_{s1}(2)_{s2}$ models. We found their results are similar to those of presented examples, and therefore they are not presented here.

5.2. Internet traffic time series

The internet traffic time series represents real-world hourly internet amount of traffic (in bits) dataset collected from private Internet service provider (ISP) with centers in 11 European cities. In particular, Cortez *et al.* (2012) collected this time series dataset from 06:57 hours on 7 June to 11:17 hours on 29 July 2005, which is available online and can be downloaded (http://www3.dsi.uminho.pt/pcortez/series/). Figure 4 shows how the daily and weekly seasonal patterns are represented in the internet traffic time series which reveal clearly the double seasonality. Accordingly, it is strongly recommended to consider two seasonal components for this dataset, where the first seasonal component is due to the intraday cycle ($s_1 = 24$) and the second seasonal component is due to the intraweek cycle ($s_2 = 168$).

To specify the best DSAR model that can fit the internet amount of traffic dataset, Cortez *et al.* (2012) tested a collection of DSAR models using different combinations of the p, P_1, and P_2 values up to a maximum order of 2; and they decided that DSAR$(2)(2)_{24}(2)_{168}$ is the best model that has smallest value of Bayesian information criterion (BIC). Cortez et al. used non-Bayesian least squares method to estimate the model parameters, and their estimates (Table 7).

Using the model DSAR$(2)(2)_{24}(2)_{168}$, the proposed Bayesian estimation methodology is applied to the internet traffic time series. The hyperparameters and starting values are chosen as in the simulated examples in previous subsection. Table 7 summarizes the Bayesian estimates results, where their sequential plots and marginal densities are displayed in Figure 3. The Bayesian estimate of the nonseasonal parameters is close to the corresponding classical estimates, but the estimates of the seasonal parameters are not enjoying that closeness property

It is worthwhile to test the significance of the interaction parameters in the DSAR model. As

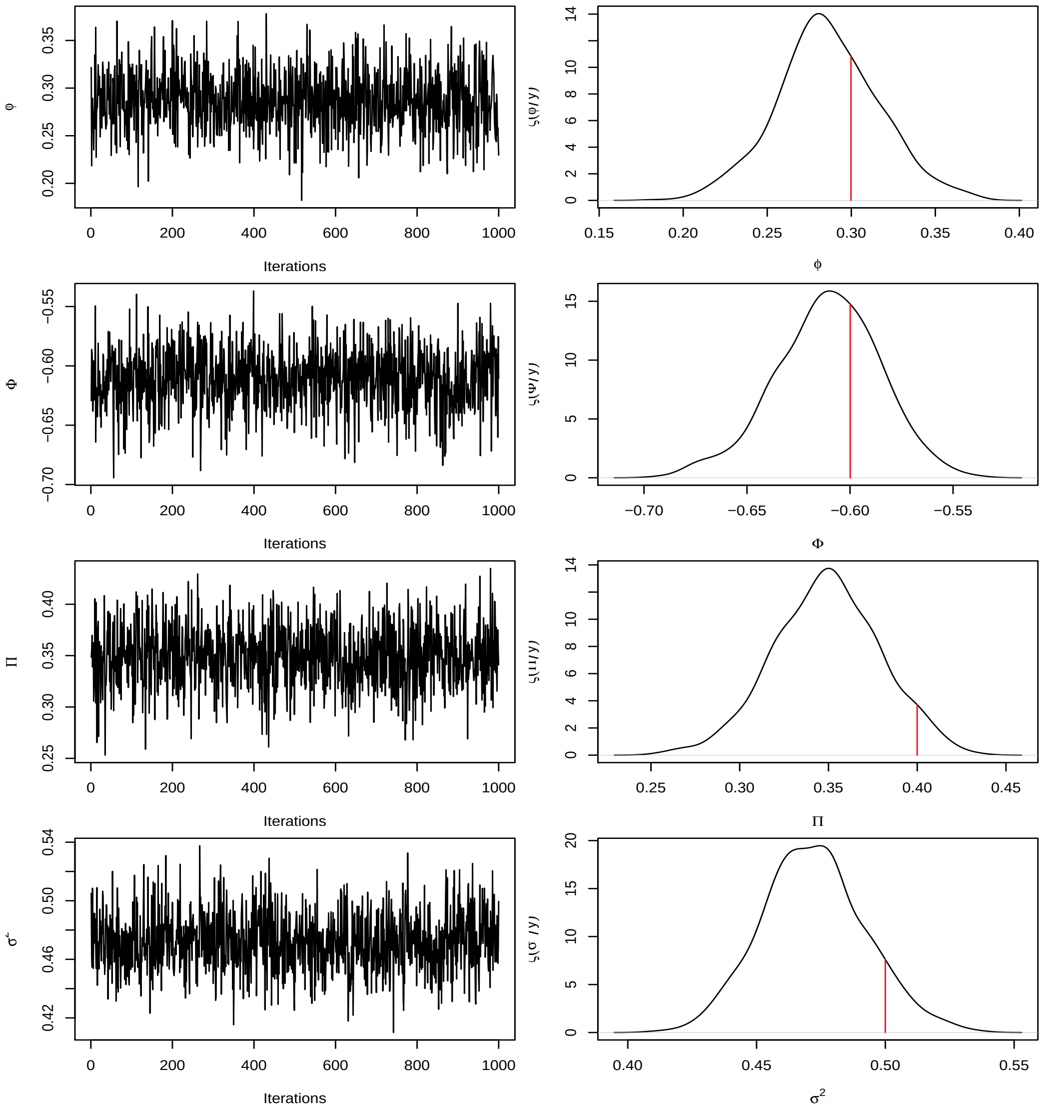

Figure 2: *Sequential plots and marginal posterior distributions of Example II.*

an example, consider the current model DSAR(2)(2)$_{24}$(2)$_{168}$, the interaction parameters include $\phi_i\Phi_j$, $\phi_i\Pi_j$, $\Phi_i\Pi_j$, and $\phi_i\Phi_j\Pi_k$ for $i = 1, 2$; $j = 1, 2$; and $k = 1, 2$. Testing the significance of these interaction parameters in the classical approach is complicated or even impossible, however, it can be achieved straightforwardly in the proposed Bayesian-based Gibbs sampling algorithm.

For more illustration, to test the significance of the interaction parameter $\eta = \phi_1\Phi_1$, the marginal posterior distribution of η can be obtained and a confidence interval can be constructed. Accordingly, when the credible interval of η contains zero, the significance hypothesis of η is rejected. Applying this testing procedure to the internet traffic time series, the marginal posterior distribution of η is

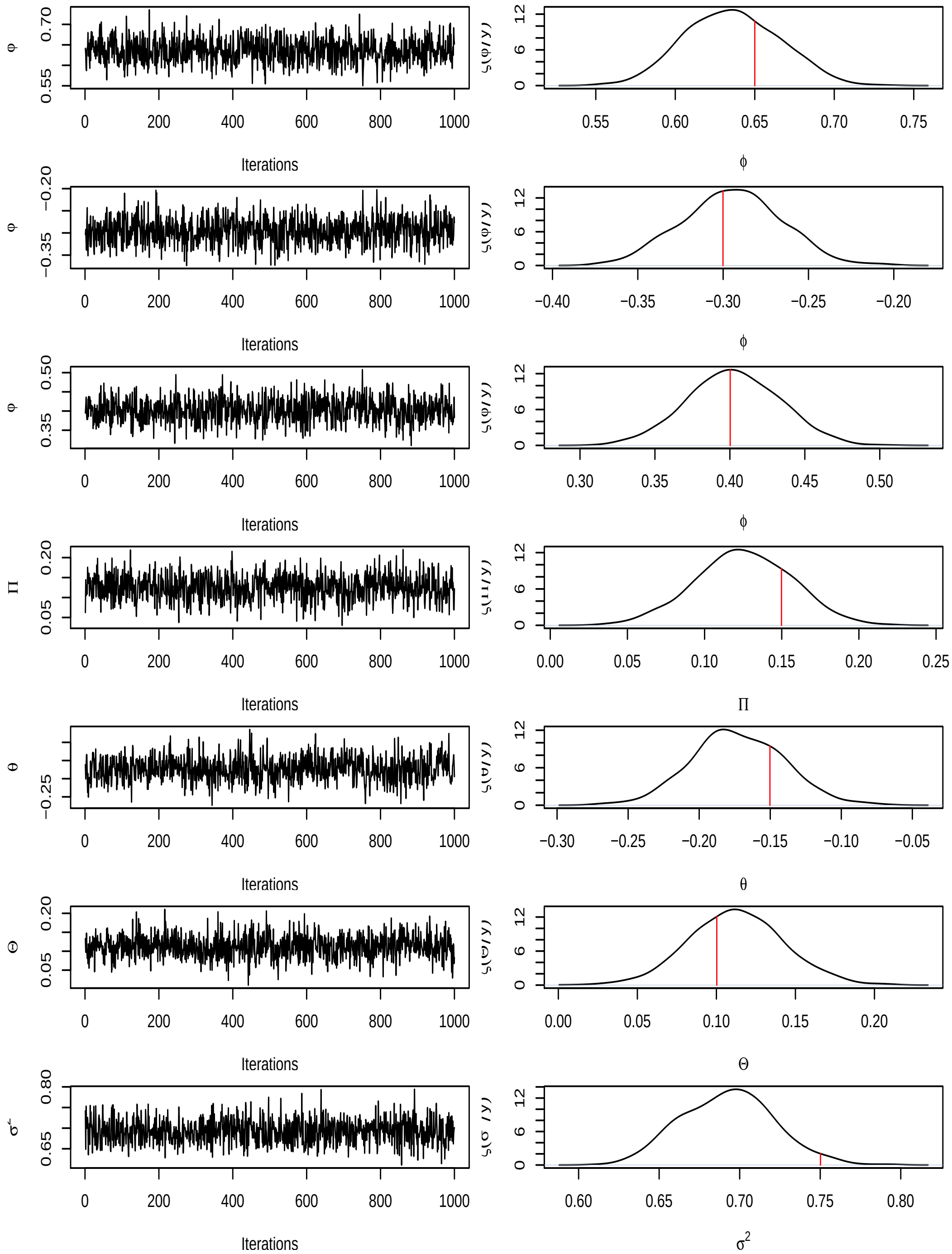

Figure 3: *Sequential plots and marginal posterior distributions of Example III.*

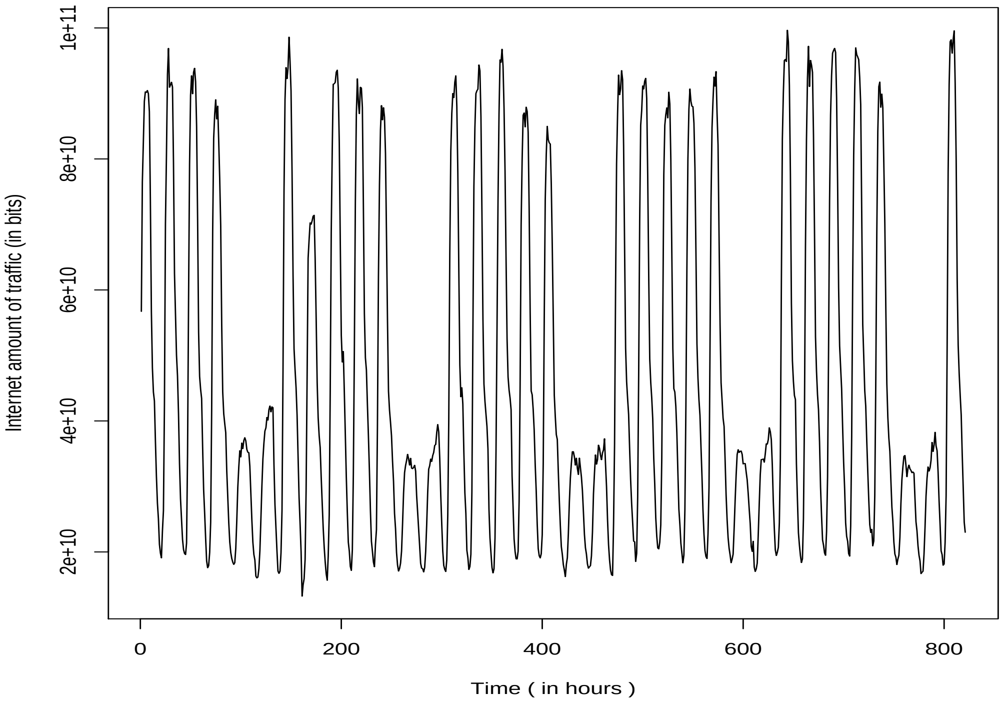

Figure 4: *Time plot of the internet traffic time series.*

Table 7: Bayesian results for the internet traffic time series.

Parameter	Mean	Std. Dev.	Lower 95 % limit	Median	Upper 95 % limit	Cortez *et al.* estimates
ϕ_1	1.40	0.03	1.33	1.40	1.46	1.70
ϕ_2	−0.42	0.03	−0.49	−0.42	−0.36	−0.27
Φ_1	0.23	0.03	0.17	0.23	0.30	0.60
Φ_2	0.07	0.02	0.02	0.06	0.11	0.06
Π_1	0.07	0.02	0.02	0.07	0.12	−0.08
Π_2	−0.01	0.02	−0.05	−0.01	0.04	−0.28
σ^2	$2.41e^{19}$	$1.03e^{19}$	$2.22e^{19}$	$2.40e^{19}$	$2.62e^{19}$	-

obtained and displayed in Figure 6. Moreover, a 95% credible interval for η is $-0.711 \leq \eta \geq -0.478$, which does not contain zero and therefore supports the alternative hypothesis of the interaction in the model.

6. Conclusions

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

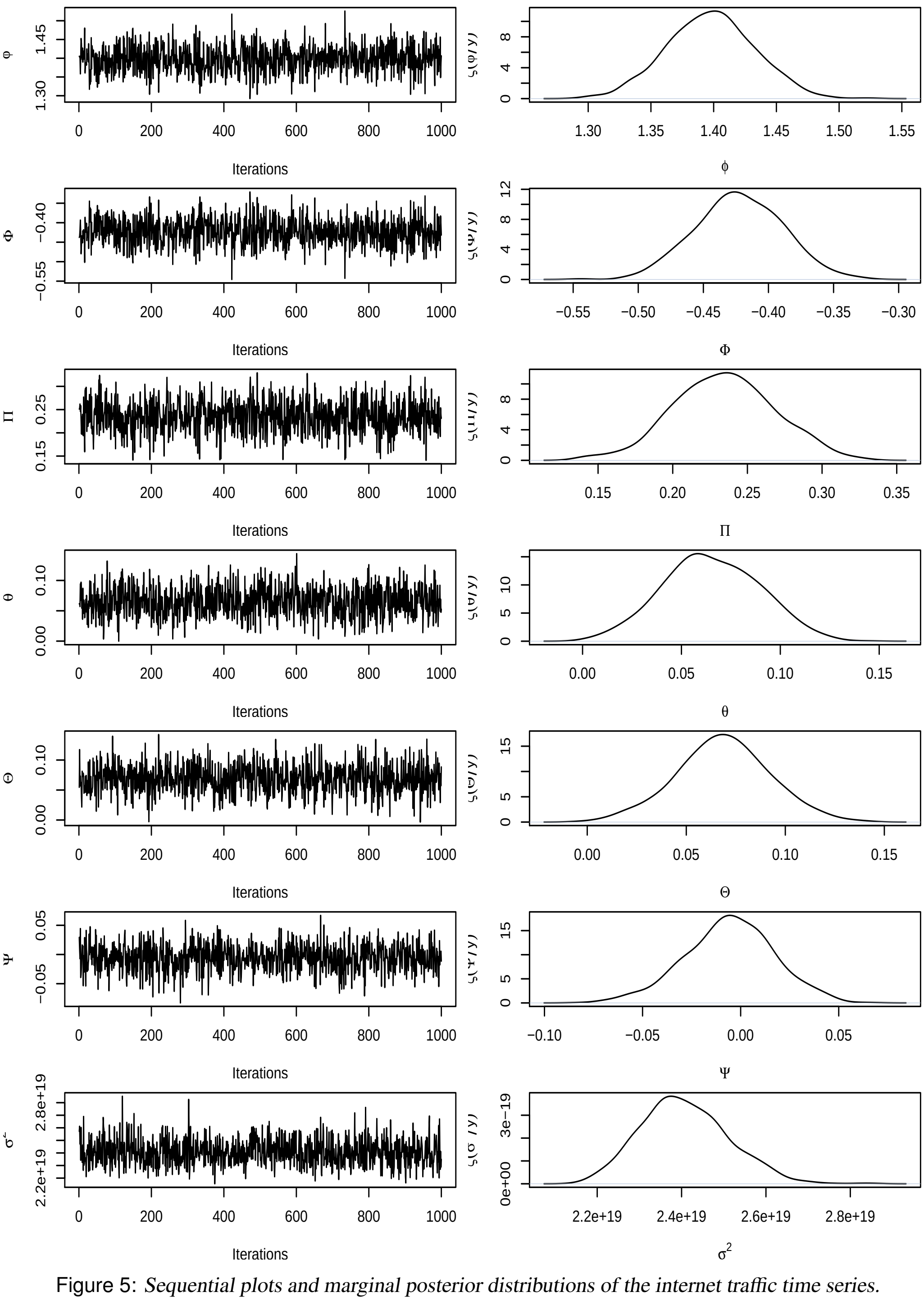

Figure 5: *Sequential plots and marginal posterior distributions of the internet traffic time series.*

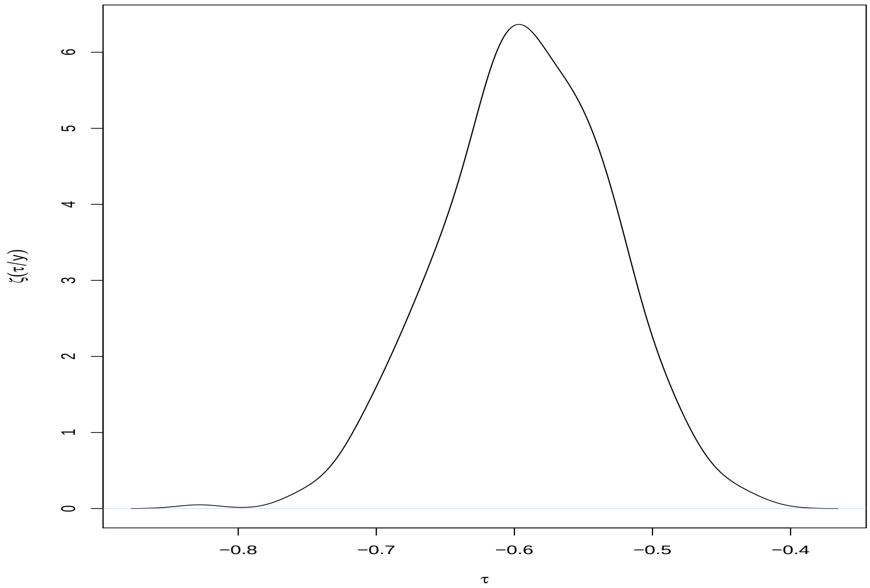

Figure 6: *Marginal posterior distribution of the interaction parameter η.*

achieved. Accordingly, we computed directly the posterior estimates of the parameters via sample averages of the simulation outputs. The empirical results of the simulated and real-world datasets 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

Au, T., Ma, G. Q. and Yeung, S. N. (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.

Baek, M. (2010). Forecasting hourly electricity loads of South Korea: Innovations state space modeling approach, *The Korean Journal of Economics*, **17**, 301–317.

Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control*, 3rd ed, Prentice-Hall, NJ.

Caiado, J. (2008). *Forecasting Water Consumption in Spain Using Univariate Time Series Models*, MPRA Paper No. 6610.

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**, 143–155.

Cruz, A., Munoz, A., Zamora, J. L. and Espinola, R. (2011). The effect of wind generation and weekday on Spanish electricity spot price forecasting, *Electric Power Systems Research*, **81**, 1924–1935.

Feinberg, E. and Genethliou, D. (2005). *Load Forecasting*, In F. W. J. Chow, Applied Mathematics for Restructured Electric Power Systems: Control and Computational Intelligence (269–285). Springer.

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.

Ismail, M. A. (2003a). Bayesian analysis of seasonal autoregressive models, *Journal of Applied Statistical Science*, **12**, 123–136.

Ismail, M. A. (2003b). Bayesian analysis of seasonal moving average model: A Gibbs sampling approach, *Japanese Journal of Applied Statistics*, **32**, 61–75.

Ismail, M. A. and Amin, A. A. (2010). Gibbs Sampling for SARMA Models, Working paper #7-2010, IDSC, Egyptian Cabinet.

Ismail, M. A. and Zahran, A. R. (2014). Bayesian inference on double seasonal autoregressive models, *Journal of Applied Statistical Science*, **12**, 123–136.

Kim, M. S. (2013). Modeling special-day effects for forecasting intraday electricity demand, *European Journal of Operational Research*, **230**, 170–180.

Korisha, S. and Pukkila, T. (1990). Linear methods for estimating ARMA and regression model with serial correlation, *Communications in Statistics: Simulation*, **19**, 71–102.

LeSage, J. P. (1999). *Applied Econometrics using MATLAB*, Department of Economics, University of Toledo, available at http://www.econ.utoledo.edu.

Mohamed, N., Ahmad, M. H., Ismail, Z. and Suhartono (2010). Double seasonal ARIMA model for forecasting load demand, *Matematika*, **26**, 217–231.

Mohamed, N., Ahmad, M. H. and Suhartono (2011). Forecasting short term load demand using double seasonal ARIMA model, *World Applied Sciences Journal*, **13**, 27–35.

Raftrey, A. E. and Lewis, S. (1992). One long run with diagnostics: Implementation strategies for Markov Chain Monte Carlo, *Statistical Science*, **7**, 493–497.

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.

Shaarawy, S. and Ismail, M. A. (1987). Bayesian inference for seasonal ARMA models, *The Egyptian Statistical Journal*, **31**, 323–336.

Taylor, J. W. (2003). Short-term electricity demand forecasting using double seasonal exponential smoothing, *The Journal of the Operational Research Society*, **54**, 799–805.

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**, 1–6.

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**, 645–658.

Taylor, J. W. (2008b). A comparison of univariate time series methods for forecasting intraday arrivals at a call center, *Management Science*, **54**, 253–265.

Thompson, H. E. and Tiao, G. C. (1971). Analysis of telephone data, *The Bell Journal of Economics and Management Science*, **2**, 514–541.

Received January 29, 2015; Revised August 29, 2015; Accepted September 23, 2015