Gibbs Sampling for Bayesian Prediction of SARMA Processes

Ayman A Amin
Department of Statistics, Mathematics, and Insurance,
Faculty of Commerce, Menoufia University, Egypt
ayman.a.amin@gmail.com

Abstract

In this article we present a Bayesian prediction of multiplicative seasonal autoregressive moving average (SARMA) processes using the Gibbs sampling algorithm. First, we estimate the unobserved errors using the nonlinear least squares (NLS) method to approximate the likelihood function. Second, we employ conjugate priors on the model parameters and initial values and assume the model errors are normally distributed to derive the conditional posterior and predictive distributions. In particular, we show that the conditional posterior distribution of the model parameters and the variance are multivariate normal and inverse gamma respectively, and the conditional predictive distribution of the future observations is a multivariate normal. Finally, we use these closed-form conditional posterior and predictive distributions to apply the Gibbs sampling algorithm to approximate empirically the marginal posterior and predictive distributions, enabling us easily to carry out multiple-step ahead predictions. We evaluate our proposed Bayesian method using simulation study and real-world time series datasets.

Keywords: Multiplicative SARMA models; Posterior analysis; Predictive analysis; MCMC methods; Gibbs sampler.

1. Introduction

Bayesian analysis of seasonal autoregressive moving average (SARMA) models is difficult since the likelihood function is analytically intractable, which causes problems in the posterior and predictive analysis. Different approaches have been proposed in literature for the Bayesian analysis of SARMA models including analytical approximations and Markov Chain Monte Carlo (MCMC) methods-based approximations.

By the analytical approximations, we mean the proposed Bayesian methods that approximate the posterior and predictive densities to be standard closed-form distributions that are analytically tractable, see for example Newbold (1973), Zellner and Reynolds (1978), Broemeling and Shaarawy (1984), and Amin (2018a, 2018b). However, these methods are conditioning on the initial values leading to waste observations, and treat SARMA model as an additive not a multiplicative model which can increase the number of unnecessary parameters. In order to address these limitations, MCMC methods, especially Gibbs sampling algorithm, have been proposed to ease and advance the Bayesian time series analysis.

Chib and Greenberg (1994) and Marriott et al. (1996) assumed prior distributions on the initial observations and errors and used MCMC methods to develop a Bayesian analysis for ARMA models without dealing with the seasonality feature. Barnett et al. (1996, 1997) estimated the multiplicative seasonal autoregressive (SAR) and ARMA models using MCMC methods. Their algorithm is based on sampling functions of the partial autocorrelations, which makes it relatively complicated compared to other

approaches, and it does not consider the prediction problem. These approaches restrict the coefficients space to satisfy stationarity and invertibility conditions. Vermaak et al. (1998) proposed a Bayesian estimation of SAR models, aiming to model speech production for voiced sounds. Their estimation is based on the state space formulation and using the Metropolis within Gibbs sampling algorithm, without considering the prediction problem.

Amin (2009) and Amin and Ismail (2014,2015) used the Gibbs sampling algorithm to present a Bayesian estimation of multiplicative SARMA and double SAR models, without considering the prediction problem. In addition, Amin (2017a, 2017b) used the Gibbs sampling algorithm to develop a Bayesian estimation of multiplicative double SMA and double SARMA models respectively. Moreover, Amin (2019a) used the Gibbs sampling algorithm to analyze pure multiplicative double (SAR) models. Based on our review for the existing Bayesian analysis of SARMA models, there is no work that considers the prediction problem along with the estimation of SARMA models and deals with the problems of initial values and multiplicativity.

In order to fill this gap in literature, we use the Gibbs sampling algorithm to present a Bayesian prediction of multiplicative SARMA processes. The main idea of our proposed Bayesian prediction is that we first estimate the unobserved errors in the multiplicative SARMA processes using the nonlinear least squares (NLS) method to approximate the likelihood function. Assuming a prior distribution on the initial observations and errors, we then use the approximate likelihood function along with the conjugate priors to derive the conditional posterior and predictive distributions that we require to implement the Gibbs sampling algorithm. The main advantages of this work are: (1) the proposed Bayesian analysis is unconditional on the initial values and treats the SARMA model as a multiplicative not an additive model to hold the parsimonious property, (2) the conditional posteriors of the model parameters are standard distributions, i.e. normal and inverse gamma, for which sampling techniques exist, and (3) the conditional predictive distribution of the future observations is a multivariate normal, which is expressed in a form that enables us to easily carry out multiple-step ahead predictions.

The paper is organized as follows. Section 2 describes the multiplicative SARMA models and their Bayesian concepts. Sections 3 and 4 introduce the posterior and predictive analyses of the multiplicative SARMA models respectively. Section 5 presents the implementation details of the proposed Bayesian prediction based on Gibbs sampling algorithm, including the convergence monitoring. The proposed Bayesian analysis is evaluated in Section 6 using simulation study and two real-world time series datasets. Finally, the paper is concluded in Section 7.

2. Multiplicative SARMA Models and Bayesian Concepts

A mean deleted time series $\{y_t\}$ is said to be generated by a multiplicative SARMA model of orders p, q, P and Q, denoted by SARMA(p,q)(P,Q)$_s$, if it satisfies

$$\phi_p(B)\Phi_P(B^s)y_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t \tag{1}$$

where $\{\varepsilon_t\}$ is a sequence of independent normal variates with zero mean and variance σ^2. The backshift operator B is defined as $By_t = y_{t-1}$, and s is the seasonal period. The nonseasonal autoregressive polynomial is $\phi_p(B) = \left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right)$ with order p, and the nonseasonal moving average polynomial is $\theta_q(B) = \left(1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q\right)$ with order q. The seasonal autoregressive polynomial is $\Phi_P(B^s) = (1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_P B^{Ps})$ with order P, and $\Theta_Q(B^s) = \left(1 + \Theta_1 B^s + \Theta_2 B^{2s} + \cdots + \Theta_Q B^{Qs}\right)$ is the seasonal moving average polynomial with order Q. The nonseasonal and seasonal autoregressive coefficients are $\phi = \left(\phi_1, \phi_2, \cdots, \phi_p\right)^T$ and $\Phi = (\Phi_1, \Phi_2, \cdots, \Phi_P)^T$, and the nonseasonal and seasonal moving average coefficients are $\theta = \left(\theta_1, \theta_2, \cdots, \theta_q\right)^T$ and $\Theta = \left(\Theta_1, \Theta_2, \cdots, \Theta_Q\right)^T$. The time series $\{y_t\}$ is assumed to start at time t=1 with unknown starting observations $y_0 = \left(y_0, y_{-1}, \cdots, y_{1-p-Ps}\right)$ and unknown starting errors $\varepsilon_0 = \left(\varepsilon_0, \varepsilon_{-1}, \cdots, \varepsilon_{1-q-Qs}\right)$.

The model (1) can be expanded and written as

$$y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{P} \Phi_j y_{t-js} - \sum_{i=1}^{p}\sum_{j=1}^{P} \phi_i \Phi_j y_{t-i-js} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \sum_{j=1}^{Q} \Theta_j \varepsilon_{t-js} + \sum_{i=1}^{q}\sum_{j=1}^{Q} \theta_i \Theta_j \varepsilon_{t-i-js} + \varepsilon_t$$
$$= X_t \beta + \Lambda_t \alpha + \varepsilon_t \qquad (2)$$

where,

$X_t = \left(y_{t-1}, \cdots, y_{t-p}; 0_1; y_{t-s}, y_{t-s-1}, \cdots, y_{t-s-p}; 0_1; \cdots; y_{t-Ps}, y_{t-Ps-1}, \cdots, y_{t-Ps-p}\right)$,

$\beta = \left(\phi_1, \cdots, \phi_p; 0_1; \Phi_1, -\phi_1\Phi_1, \cdots, -\phi_p\Phi_1; 0_1; \cdots; \Phi_P, -\phi_1\Phi_P, \cdots, -\phi_p\Phi_P\right)^T$,

$\Lambda_t = \left(\varepsilon_{t-1}, \cdots, \varepsilon_{t-q}; 0_2; \varepsilon_{t-s}, \varepsilon_{t-s-1}, \cdots, \varepsilon_{t-s-q}; 0_2; \cdots; \varepsilon_{t-Qs}, \varepsilon_{t-Qs-1}, \cdots, \varepsilon_{t-Qs-q}\right)$,

$\alpha = \left(\theta_1, \cdots, \theta_q; 0_2; \Theta_1, \theta_1\Theta_1, \cdots, \theta_q\Theta_1; 0_2; \cdots; \Theta_Q, \theta_1\Theta_Q, \cdots, \theta_q\Theta_Q\right)^T$,

and 0_1 and 0_2 are (s-p-1) and (s-q-1) row vectors of zeros respectively. From the model (2), we can observe that the multiplicative SARMA model can be written as ARMA model of order $p + Ps$ and $q + Qs$ with some coefficients are zeros and some are products of nonseasonal and seasonal coefficients. Thus, the SARMA model is nonlinear in the coefficients ϕ, Φ, θ and Θ, which complicates its Bayesian analysis. For more details about the properties of SARMA model, see Box and Jenkins (1976).

Bayesian analysis of time series models is based on Bayes' theorem that combines the prior distribution of the model parameters with the likelihood function of observed sample to get the posterior distribution (Amin, 2019b). Regarding the specification of prior distribution, assuming the parameters ϕ, Φ, θ, Θ, y_0 and ε_0 are independent apriori given the model variance σ^2, the joint prior distribution is given as

$$\zeta(\phi, \Phi, \theta, \Theta, \sigma^2, y_0, \varepsilon_0) = \zeta(\phi|\sigma^2) \times \zeta(\Phi|\sigma^2) \times \zeta(\theta|\sigma^2) \times \zeta(\Theta|\sigma^2) \times \zeta(y_0|\sigma^2) \times \zeta(\varepsilon_0|\sigma^2) \times \zeta(\sigma^2)$$

$$= N_p(\mu_\phi, \sigma^2\Sigma_\phi) \times N_P(\mu_\Phi, \sigma^2\Sigma_\Phi) \times N_q(\mu_\theta, \sigma^2\Sigma_\theta) \times$$
$$N_Q(\mu_\Theta, \sigma^2\Sigma_\Theta) \times$$
$$N_{p+Ps}(\mu_{y_0}, \sigma^2\Sigma_{y_0}) \times N_{q+Qs}(\mu_{\varepsilon_0}, \sigma^2\Sigma_{\varepsilon_0}) \times IG(\tfrac{\nu}{2}, \tfrac{\lambda}{2}), \quad (3)$$

where $N_r(\mu, \Delta)$ 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. Accordingly, the joint prior distribution can be simplified and written as

$$\zeta(\phi, \Phi, \theta, \Theta, \sigma^2, y_0, \varepsilon_0) \propto (\sigma^2)^{-\left(\frac{\nu+2p+P+Ps+2q+Q+Qs}{2}+1\right)} \exp\left\{-\frac{1}{2\sigma^2}\left[\lambda + (\phi-\mu_\phi)^T\Sigma_\phi^{-1}(\phi-\mu_\phi) + (\Phi-\mu_\Phi)^T\Sigma_\Phi^{-1}(\Phi-\mu_\Phi) + (\theta-\mu_\theta)^T\Sigma_\theta^{-1}(\theta-\mu_\theta) + (\Theta-\mu_\Theta)^T\Sigma_\Theta^{-1} \times (\Theta-\mu_\Theta) + (y_0-\mu_{y_0})^T\Sigma_{y_0}^{-1}(y_0-\mu_{y_0}) + (\varepsilon_0-\mu_{\varepsilon_0})^T\Sigma_{\varepsilon_0}^{-1}(\varepsilon_0-\mu_{\varepsilon_0})\right]\right\} . \quad (4)$$

The prior distribution (4) is chosen to be conditionally a conjugate prior in order to facilitate the mathematical calculations. Employing a straightforward random variable transformation from ε_t to y_t, the likelihood function is given by

$$L(\phi, \Phi, \theta, \Theta, \sigma^2, \varepsilon_0|y) \propto (\sigma^2)^{-\frac{n}{2}}\exp\left\{-\tfrac{1}{2\sigma^2}\textstyle\sum_{t=1}^{n}\varepsilon_t^2\right\} = (\sigma^2)^{-\frac{n}{2}}\exp\left\{-\tfrac{1}{2\sigma^2}(y - X\beta - \Lambda\alpha)^T(y - X\beta - \Lambda\alpha)\right\} \quad (5)$$

Multiplying the joint prior distribution (4) by the likelihood function (5) results in the joint posterior density of $\phi, \Phi, \theta, \Theta, \sigma^2, y_0, \varepsilon_0$ that can be written as

$$\zeta(\phi, \Phi, \theta, \Theta, \sigma^2, y_0, \varepsilon_0|y) \propto (\sigma^2)^{-\left(\frac{n+\nu+2p+P+Ps+2q+Q+Qs}{2}+1\right)}\exp\left\{-\frac{1}{2\sigma^2}\left[\lambda + (\phi-\mu_\phi)^T\Sigma_\phi^{-1} \times (\phi-\mu_\phi) + (\Phi-\mu_\Phi)^T\Sigma_\Phi^{-1}(\Phi-\mu_\Phi) + (\theta-\mu_\theta)^T\Sigma_\theta^{-1}(\theta-\mu_\theta) + (\Theta-\mu_\Theta)^T\Sigma_\Theta^{-1}(\Theta-\mu_\Theta) + (y_0-\mu_{y_0})^T\Sigma_{y_0}^{-1}(y_0-\mu_{y_0}) + (\varepsilon_0-\mu_{\varepsilon_0})^T \times \Sigma_{\varepsilon_0}^{-1}(\varepsilon_0-\mu_{\varepsilon_0}) + (y - X\beta - \Lambda\alpha)^T(y - X\beta - \Lambda\alpha)]\right\} \quad (6)$$

3. Posterior Analysis for SARMA Models

The joint posterior density (6) is analytically intractable because of two reasons. First, the elements of the matrix Λ are unknown lagged errors. Second, the likelihood function (5)

is nonlinear function in the coefficients ϕ, Φ, θ, and Θ. In order to address the first problem, the unknown errors can be estimated recursively as:

$$\hat{\varepsilon}_t = y_t - \sum_{i=1}^{p} \hat{\phi}_i y_{t-i} - \sum_{j=1}^{P} \hat{\Phi}_j y_{t-js} + \sum_{i=1}^{p}\sum_{j=1}^{P} \hat{\phi}_i \hat{\Phi}_j y_{t-i-js} - \sum_{i=1}^{q} \hat{\theta}_i \hat{\varepsilon}_{t-i} - \sum_{j=1}^{Q} \hat{\Theta}_j \hat{\varepsilon}_{t-js} - \sum_{i=1}^{q}\sum_{j=1}^{Q} \hat{\theta}_i \hat{\Theta}_j \hat{\varepsilon}_{t-i-js}, \quad (7)$$

where, $\hat{\phi}_i \in R^p, \hat{\Phi}_j \in R^P, \hat{\theta}_i \in R^q$, and $\hat{\Theta}_j \in R^Q$ are nonlinear least squares (NLS) estimates obtained by minimizing $\sum_{t=1}^{n} \varepsilon_t^2 = SS(\phi, \Phi, \theta, \Theta)$ with respect to ϕ, Φ, θ, and Θ over the staionarity and invertibility regions. Accordingly, the joint posterior density (6) can be approximated as

$$\zeta^{å}(\phi, \Phi, \theta, \Theta, \sigma^2, y_0, \varepsilon_0 | y) \propto (\sigma^2)^{-\left(\frac{n+\nu+2p+P+Ps+2q+Q+Qs}{2}+1\right)} \exp\left\{-\frac{1}{2\sigma^2}\left[\lambda + \left(\phi-\mu_\phi\right)^T \Sigma_\phi^{-1} \times \left(\phi-\mu_\phi\right) + (\Phi-\mu_\Phi)^T \Sigma_\Phi^{-1}(\Phi-\mu_\Phi) + (\theta-\mu_\theta)^T \Sigma_\theta^{-1}(\theta-\mu_\theta) + (\Theta-\mu_\Theta)^T \Sigma_\Theta^{-1}(\Theta-\mu_\Theta) + \left(y_0-\mu_{y_0}\right)^T \Sigma_{y_0}^{-1}\left(y_0-\mu_{y_0}\right) + \left(\varepsilon_0-\mu_{\varepsilon_0}\right)^T \Sigma_{\varepsilon_0}^{-1}\left(\varepsilon_0-\mu_{\varepsilon_0}\right) + \left(y - X\beta - \hat{\Lambda}\alpha\right)^T \left(y - X\beta - \hat{\Lambda}\alpha\right)\right]\right\}, \quad (8)$$

where $\hat{\Lambda}$ is an $n \times (q + Qs)$ matrix with the t^{th} row:

$$\hat{\Lambda}_t = \left(\hat{\varepsilon}_{t-1}, \cdots, \hat{\varepsilon}_{t-q}; 0_2; \hat{\varepsilon}_{t-s}, \hat{\varepsilon}_{t-s-1}, \cdots, \hat{\varepsilon}_{t-s-q}; 0_2; \cdots; \hat{\varepsilon}_{t-Qs}, \hat{\varepsilon}_{t-Qs-1}, \cdots, \hat{\varepsilon}_{t-Qs-q}\right)$$

Since the likelihood function (5) is nonlinear in the coefficients ϕ, Φ, θ, and Θ, the marginal approximate posteriors of the model coefficients cannot be obtained. In order to solve this problem, we introduce the Gibbs sampling method, as one of the MCMC methods, to approximate the marginal posteriors of the model coefficients. However, to be able to achieve this work the full conditional posteriors of the model parameters have to be standard distributions or at least in closed form as one of the main requirements of implementing the Gibbs sampling algorithm. In general, we can derive the conditional posterior distribution of each of the model parameters by grouping together terms in the approximate joint posterior (8) that depend on this parameter, and then finding the appropriate normalizing constant to form a proper density. In our previous work (Amin, 2009), we showed in detail that all the conditional posteriors are standard distributions, in particular the conditional posteriors of $\phi, \Phi, \theta, \Theta, y_0$, and ε_0 are the multivariate normal and the conditional posterior of σ^2 is the inverse gamma, for which sampling techniques exist. In the following, we present these conditional posteriors of the model parameters without proofs, and for more details about the proofs see Amin (2009).

The conditional posterior of $\boldsymbol{\phi}$

The conditional posterior of ϕ given $y, \sigma^2, \Phi, \theta, \Theta, y_0$, and ε_0 is a multivariate normal with location vector $\mu_{\phi}^{\text{å}}$, and dispersion matrix $v_{\phi}^{\text{å}}$, where,

$$\mu_{\phi}^{\text{å}} = \left[\left(H^T H + \Sigma_{\phi}^{-1}\right)^{-1}\left(\Sigma_{\phi}^{-1}\mu_{\phi} + H^T\left(y - L\Phi - \widehat{\Lambda}\alpha\right)\right)\right], v_{\phi}^{\text{å}} = \sigma^2\left(H^T H + \Sigma_{\phi}^{-1}\right)^{-1},$$

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

The conditional posterior of $\boldsymbol{\Phi}$

The conditional posterior of Φ given $y, \sigma^2, \phi, \theta, \Theta, y_0$, and ε_0 is a multivariate normal with location vector $\mu_{\Phi}^{\text{å}}$, and dispersion matrix $v_{\Phi}^{\text{å}}$, where,

$$\mu_{\Phi}^{\text{å}} = \left[(G^T G + \Sigma_{\Phi}^{-1})^{-1}\left(\Sigma_{\Phi}^{-1}\mu_{\Phi} + G^T\left(y - R\phi - \widehat{\Lambda}\alpha\right)\right)\right], v_{\Phi}^{\text{å}} = \sigma^2(G^T G + \Sigma_{\Phi}^{-1})^{-1},$$

G is an $n \times P$ matrix with tj-th element $G_{tj} = \left(y_{t-js} - \sum_{i=1}^{p} \phi_i y_{t-js-i}\right)$ and R is an $n \times p$ matrix with ti-th element $R_{ti} = (y_{t-i})$.

The conditional posterior of $\boldsymbol{\theta}$

The conditional posterior of θ given $y, \sigma^2, \phi, \Phi, \Theta, y_0$, and ε_0 is a multivariate normal with location vector $\mu_{\theta}^{\text{å}}$, and dispersion matrix $v_{\theta}^{\text{å}}$, where,

$$\mu_{\theta}^{\text{å}} = \left[\left(A^T A + \Sigma_{\theta}^{-1}\right)^{-1}\left(\Sigma_{\theta}^{-1}\mu_{\theta} + A^T(y - K\Theta - \ X\beta)\right)\right], v_{\theta}^{\text{å}} = \sigma^2\left(A^T A + \Sigma_{\theta}^{-1}\right)^{-1},$$

A is an $n \times q$ matrix with ti-th element $A_{ti} = \left(\hat{\varepsilon}_{t-i} + \sum_{i=1}^{Q} \Theta_j \hat{\varepsilon}_{t-js-i}\right)$ and K is an $n \times Q$ matrix with tj-th element $K_{tj} = \left(\hat{\varepsilon}_{t-js}\right)$.

The conditional posterior of $\boldsymbol{\Theta}$

The conditional posterior of Θ given $y, \sigma^2, \phi, \Phi, \theta, y_0$, and ε_0 is a multivariate normal with location vector $\mu_{\Theta}^{\text{å}}$, and dispersion matrix $v_{\Theta}^{\text{å}}$, where,

$$\mu_{\Theta}^{\text{å}} = \left[(W^T W + \Sigma_{\Theta}^{-1})^{-1}\left(\Sigma_{\Theta}^{-1}\mu_{\Theta} + W^T(y - Z\theta - \ X\beta)\right)\right], v_{\Theta}^{\text{å}} = \sigma^2(W^T W + \Sigma_{\Theta}^{-1})^{-1},$$

W is an $n \times Q$ matrix with tj-th element $W_{tj} = \left(\hat{\varepsilon}_{t-js} + \sum_{i=1}^{q} \theta_i \hat{\varepsilon}_{t-js-i}\right)$ and Z is an $n \times q$ matrix with ti-th element $Z_{ti} = (\hat{\varepsilon}_{t-i})$.

The conditional posterior of σ^2

The conditional posterior of σ^2 given $y, \phi, \Phi, \theta, \Theta, y_0$, and ε_0 is an inverse gamma with parameters $a = \frac{\nu^{å}}{2}$, and $b = \frac{\lambda + nS^2}{2}$, where,

$\nu^{å} = n + \nu + 2p + P + Ps + 2q + Q + Qs$, and

$$nS^2 = \Big[(\phi - \mu_\phi)^T \Sigma_\phi^{-1}(\phi - \mu_\phi) + (\Phi - \mu_\Phi)^T \Sigma_\Phi^{-1}(\Phi - \mu_\Phi) + (\theta - \mu_\theta)^T \Sigma_\theta^{-1}(\theta - \mu_\theta) + (\Theta - \mu_\Theta)^T \Sigma_\Theta^{-1}(\Theta - \mu_\Theta) + (y_0 - \mu_{y_0})^T \Sigma_{y_0}^{-1}(y_0 - \mu_{y_0}) + (\varepsilon_0 - \mu_{\varepsilon_0})^T \Sigma_{\varepsilon_0}^{-1}(\varepsilon_0 - \mu_{\varepsilon_0}) + (y - X\beta - \widehat{\Lambda}\alpha)^T (y - X\beta - \widehat{\Lambda}\alpha)\Big].$$

Accordingly, we can sample the parameter σ^2 from the Chi square distribution using the transformation $\frac{\lambda + nS^2}{\sigma^2} \sim \chi^2_{\nu^{å}}$.

The conditional posterior of $\mathbf{y_0}$ and $\boldsymbol{\varepsilon_0}$

Using the model (2), the equations for the elements of y_0 and ε_0 can be expanded and written as

$$Fy_{p+Ps} = Dy_0 + M\varepsilon_0 + N\varepsilon_{p+Ps}$$

where,

$$F = \begin{pmatrix} 1 & 0 & 0 & \cdots & \cdots & 0 & 0 \\ -\beta_1 & 1 & 0 & \cdots & \cdots & 0 & 0 \\ -\beta_2 & -\beta_1 & 1 & \cdots & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \cdots & \vdots & \vdots \\ -\beta_{p+Ps-1} & -\beta_{p+Ps-2} & \cdots & \cdots & -\beta_2 & -\beta_1 & 1 \end{pmatrix}_{(p+Ps)\times(p+Ps)},$$

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

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

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

$\alpha_r=0\forall r=q+Qs+1,\cdots,p+Ps$ if $p+Ps>q+Qs$, $y_{p+Ps}=\left(y_1,y_2,\cdots,y_{p+Ps}\right)^T$ and $\varepsilon_{p+Ps}=\left(\varepsilon_1,\varepsilon_2,\cdots\varepsilon_{p+Ps}\right)^T$ which has the (p+Ps) normal distribution with zero mean and variance $(\sigma^2 I_{p+Ps})$ where I_{p+Ps} is the unit matrix of order (p+Ps). With some manipulations, we derived the conditional posterior of y_0 given $y,\sigma^2,\phi,\Phi,\theta,\Theta$, and ε_0 as a multivariate normal with location vector $\mu_{y_0}^{å}$, and dispersion matrix $v_{y_0}^{å}$, where,

$$\mu_{y_0}^{å}=\left[D^T(NN^T)^{-1}D+\Sigma_{y_0}^{-1}\right]^{-1}\left[\Sigma_{y_0}^{-1}\mu_{y_0}+D^T(NN^T)^{-1}\left(Fy_{p+Ps}-M\varepsilon_0\right)\right],$$

$$\text{and } v_{y_0}^{å}=\sigma^2\left(D^T(NN^T)^{-1}D+\Sigma_{y_0}^{-1}\right)^{-1}$$

Similarly, the conditional posterior of ε_0 given $y,\sigma^2,\phi,\Phi,\theta,\Theta$, and y_0 is a multivariate normal with location vector $\mu_{\varepsilon_0}^{å}$, and dispersion matrix $v_{\varepsilon_0}^{å}$, where,

$$\mu_{\varepsilon_0}^{å}=\left[M^T(NN^T)^{-1}M+\Sigma_{\varepsilon_0}^{-1}\right]^{-1}\left[\Sigma_{\varepsilon_0}^{-1}\mu_{\varepsilon_0}+M^T(NN^T)^{-1}\left(Fy_{p+Ps}-Dy_0\right)\right],$$

$$\text{and } v_{\varepsilon_0}^{å}=\sigma^2\left[M^T(NN^T)^{-1}M+\Sigma_{\varepsilon_0}^{-1}\right]^{-1}$$

4. Predictive Analysis for SARMA Models

The predictive density of the next k future observations is the conditional density of future observations $y_f=(y_{n+1},y_{n+2},\ldots y_{n+k})$ given the time series data y. This predictive density can be obtained by averaging the distribution of future values with respect to the joint posterior density of the model parameters as

$$\zeta(y_f|y)=\int_\phi\int_\Phi\int_\theta\int_\Theta\int_{\sigma^2}\int_{y_0}\int_{\varepsilon_0}\zeta\left(y_f,\phi,\Phi,\theta,\Theta,\sigma^2,y_0,\varepsilon_0|y\right)d\phi d\Phi d\theta d\Theta d\sigma^2 dy_0 d\varepsilon_0 \quad (9)$$

This predictive density does not have an analytical closed form, which complicates the predictive analysis of SARMA models. Therefore, we try to obtain the conditional

predictive distribution of y_f given $y, \phi, \Phi, \theta, \Theta, \sigma^2, y_0$, and ε_0, and then employ the Gibbs sampling method to approximate the predictive density. In order to derive the conditional predictive distribution of y_f, we use the model (2) to obtain the equations of the k future observations as

$$\Psi y_f = \Omega y_L + \Pi \varepsilon_L + \Gamma \varepsilon_f$$

where,

$$\Psi = \begin{pmatrix} 1 & 0 & 0 & \cdots & \cdots & 0 & 0 \\ -\psi_1 & 1 & 0 & \cdots & \cdots & 0 & 0 \\ -\psi_2 & -\psi_1 & 1 & \cdots & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \cdots & \vdots & \vdots \\ -\psi_{k-1} & -\psi_{k-2} & \cdots & \cdots & -\psi_2 & -\psi_1 & 1 \end{pmatrix}_{k\times k},$$

$$\Omega = \begin{pmatrix} \omega_1 & \omega_2 & \omega_3 & \cdots & \cdots & \omega_{p+Ps-1} & \omega_{p+Ps} \\ \omega_2 & \omega_3 & \omega_4 & \cdots & \cdots & \omega_{p+Ps} & 0 \\ \omega_3 & \omega_4 & & \cdots & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \cdots & \vdots & \vdots \\ \omega_k & \omega_{k+1} & \cdots & \omega_{p+Ps} & 0 & \cdots & 0 \end{pmatrix}_{k\times(p+Ps)},$$

$$\Pi = \begin{pmatrix} \pi_1 & \pi_2 & \pi_3 & \cdots & \cdots & \pi_{q+Qs-1} & \pi_{q+Qs} \\ \pi_2 & \pi_3 & \pi_4 & \cdots & \cdots & \pi_{q+Qs} & 0 \\ \pi_3 & \pi_4 & & \cdots & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \cdots & \vdots & \vdots \\ \pi_k & \pi_{k+1} & \cdots & \pi_{q+Qs} & 0 & \cdots & 0 \end{pmatrix}_{k\times(q+Qs)},$$

$$\Gamma = \begin{pmatrix} 1 & 0 & 0 & \cdots & \cdots & 0 & 0 \\ \gamma_1 & 1 & 0 & \cdots & \cdots & 0 & 0 \\ \gamma_2 & \gamma_1 & 1 & \cdots & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \cdots & \vdots & \vdots \\ \gamma_{k-1} & \gamma_{k-2} & \cdots & \cdots & \gamma_2 & \gamma_1 & 1 \end{pmatrix}_{k\times k},$$

$\psi_r = \beta_r$ and $\omega_r = \beta_r \forall r \le p + Ps$ and zeros otherwise, $\pi_r = \alpha_r$ and $\gamma_r = \alpha_r \forall r \le q + Qs$ and zeros otherwise, $y_L = \left(y_n, y_{n-1}, \cdots, y_{n-p-Ps+1}\right)^T$, $\varepsilon_L = \left(\varepsilon_n, \varepsilon_{n-1}, \cdots, \varepsilon_{n-q-Qs+1}\right)^T$, and $\varepsilon_f = (\varepsilon_{n+1}, \varepsilon_{n+2}, \cdots, \varepsilon_{n+k})^T$. It is worth noting that we assume ε_f has a multivariate normal distribution with zero mean and variance $(\sigma^2 I_k)$, and ε_f and ε_L are independent. Using the above defined matrices and with some manipulations, we derive the conditional predictive distribution of y_f given $y, \phi, \Phi, \theta, \Theta, \sigma^2$, and y_0 to be a multivariate normal with location vector $\mu_{y_f}^{\text{å}}$, and dispersion matrix $v_{y_f}^{\text{å}}$, where,

$$\mu_{y_f}^{å} = \Psi^{-1}[\Omega y_L + \Pi \varepsilon_L], \text{ and } v_{y_f}^{å} = \sigma^2[(\Gamma^{-1}\Psi)^T(\Gamma^{-1}\Psi)]^{-1}$$

It should be noted that the elements in ε_L are unknown and can be estimated as discussed in Section 3. Accordingly, using this derived conditional predictive distribution of y_f along with the conditional posteriors of the model parameters presented in Section 3, we apply the Gibbs sampling method not only to approximate the marginal posteriors but also to approximate the predictive density (9). In the following section we introduce our proposed Gibbs sampling algorithm for Bayesian analysis of SARMA models.

5. Proposed Gibbs Sampling Algorithm for SARMA Models

The proposed Gibbs sampling algorithm for the Bayesian analysis of SARMA models can be implemented in the following steps:

1. Specify starting values $\{\phi^0, \Phi^0, \theta^0, \Theta^0, (\sigma^2)^0, y_0^0, \varepsilon_0^0, y_f^0\}$ and set j=0. These starting values can be obtained from the initial estimates of the model parameters obtained using the NLS method, as we discussed in Section 3.

2. Generate one value for each model parameter from its conditional posterior distribution and for future observations y_f from the conditional predictive distribution as
   - $\phi^j \sim \zeta(\phi^j | y, (\sigma^2)^{j-1}, \Phi^{j-1}, \theta^{j-1}, \Theta^{j-1}, {y_0}^{j-1}, {\varepsilon_0}^{j-1}, {y_f}^{j-1})$,
   - $\Phi^j \sim \zeta(\Phi^j | y, (\sigma^2)^{j-1}, \phi^j, \theta^{j-1}, \Theta^{j-1}, {y_0}^{j-1}, {\varepsilon_0}^{j-1}, {y_f}^{j-1})$,
   - $\theta^j \sim \zeta(\theta^j | y, (\sigma^2)^{j-1}, \phi^j, \Phi^j, \Theta^{j-1}, {y_0}^{j-1}, {\varepsilon_0}^{j-1}, {y_f}^{j-1})$,
   - $\Theta^j \sim \zeta(\Theta^j | y, (\sigma^2)^{j-1}, \phi^j, \Phi^j, \theta^j, {y_0}^{j-1}, {\varepsilon_0}^{j-1}, {y_f}^{j-1})$,
   - $(\sigma^2)^j \sim \zeta((\sigma^2)^j | y, \phi^j, \Phi^j, \theta^j, \Theta^j, {y_0}^{j-1}, {\varepsilon_0}^{j-1}, {y_f}^{j-1})$,
   - ${y_0}^j \sim \zeta({y_0}^j | y, (\sigma^2)^j, \phi^j, \Phi^j, \theta^j, \Theta^j, {\varepsilon_0}^{j-1}, {y_f}^{j-1})$,
   - ${\varepsilon_0}^j \sim \zeta({\varepsilon_0}^j | y, (\sigma^2)^j, \phi^j, \Phi^j, \theta^j, \Theta^j, {y_0}^j, {y_f}^{j-1})$,
   - ${y_f}^j \sim \zeta({y_f}^j | y, (\sigma^2)^j, \phi^j, \Phi^j, \theta^j, \Theta^j, {y_0}^j, {\varepsilon_0}^j)$.

   Now, the algorithm gives the first value of the Markov chain $\{\phi^j, \Phi^j, \theta^j, \Theta^j, (\sigma^2)^j, y_0^j, \varepsilon_0^j, y_f^j\}$.

3. Set j = j+1 and go to 2. The algorithm gives the next value of the Markov chain, say $\{\phi^{j+1}, \Phi^{j+1}, \theta^{j+1}, \Theta^{j+1}, (\sigma^2)^{j+1}, y_0^{j+1}, \varepsilon_0^{j+1}, y_f^{j+1}\}$ by simulating each of the full conditionals where the conditioning elements are revised each iteration.

This iterative process is repeated for a large number of iterations and its convergence is monitored. Once the chain has converged, the simulated values from the conditional posterior and predictive distributions are used as samples from the joint posterior and from the predictive distribution respectively. Accordingly, posterior estimates of the model parameters and Bayesian forecasts of the future observations can be computed directly by sample averages of the simulation outputs. Regarding the convergence of the obtained Markov chain, three groups of diagnostics can be used to monitor including 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 be generated to get accurate posterior estimates. Second, diagnostics

proposed by Raftery and Lewis (1992,1995) include (1) Burn: the number of draws used as initial draws or "burn-in" before starting to sample draws for the purpose of posterior approximation, (2) Total: the total number of draws needed to achieve the desired level of accuracy, (3) Nmin: the 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 a convergence problem may be indicated when the I-stat values exceed 5. Third, diagnostics proposed by Geweke (1992) include two groups:

1. The first group includes the numerical standard errors (NSE) and the relative numerical efficiency (RNE). The NSE estimates reflect the variation that can be expected if the simulation is to be repeated. The RNE estimates indicate the number of draws that is required 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.

These convergence diagnostics are used in the following section to monitor the convergence of the proposed Bayesian method.

6. Simulation Study and Real-World Application

We have two parts in this section. First, we present a simulation study to evaluate the accuracy of our proposed Bayesian analysis for SARMA models. Second, we apply the proposed Bayesian analysis to real-world time series datasets.

6.1 Simulation Study

In this subsection we present four simulations of SARMA models, and Table 1 shows the design of these simulations including true parameters values, sample size, seasonality period, model variance, and next k future values. By these simulations we try to represent different seasonality patterns with different data types.

Table 1: Simulation design.

Model	ϕ	Φ	θ	Θ	n	s	σ^2	k
I	0.8	0.3	0.5	0.8	100	4	1.0	4
II	0.3	0.4	0.4	0.7	100	4	1.0	4
III	0.6	0.5	0.3	0.5	300	12	1.0	12
IV	-0.5	0.6	-0.5	0.8	300	12	1.0	12

Once the time series datasets are generated from these SARMA models, the Bayesian analysis is performed by assuming a non informative prior for the parameters ϕ, Φ, θ , Θ, and σ^2 and a normal prior with zero mean for both initial observations y_0 with variance $\sigma^2 I_{p+Ps}$ and initial errors ε_0 with variance $\sigma^2 I_{q+Qs}$. To apply the proposed Gibbs sampler, the starting values for the parameters ϕ, Φ, θ, Θ, and σ^2 are obtained using the NLS method, and the starting values for y_0, ε_0, and y_f are assumed to be zeros.

For each dataset, the Gibbs sampler was run 6,000 iterations where the first 1,000 draws are ignored and every fifth value in the sequence of the last 5,000 draws is recorded to have an approximately independent sample. All the posterior estimates of the model parameters and the Bayesian forecasts are computed directly as sample averages of the Gibbs sampler draws. In the following, we discuss the results and investigate the convergence diagnostics.

Table 2 presents the Bayesian estimates of the model parameters and the Bayesian forecasts of the next k future observations with the corresponding true values for Model I. The 95% credible intervals using the 0.025 and 0.975 percentiles of the simulated draws is computed and presented in Table 2. From Table 2, it is clear that the Bayesian estimates and forecasts are close to the true values and the 95% credible interval includes the true value for each parameter and future observation. Sequential plots of generated sequences of the model parameters together with their marginal posterior densities are displayed in Figure 1, while marginal predictive densities of the next four future observations are displayed in Figure 2. Figure 1 shows that the posterior draws of the proposed algorithm are stable and fluctuate in the neighborhood of the true values. On the other hand, the marginal posteriors show that the true value of each parameter falls in the constructed 95% credible interval, as indicated by the vertical line. Similar conclusions can be obtained from Figure 2 about the predictive performance of the proposed Gibbs sampling algorithm.

Table 2: Bayesian results for Model I.

Parameter	True values	Mean	Std. Dev.	Lower 95 % limit	Median	Upper 95 % limit
ϕ	0.80	0.82	0.03	0.76	0.82	0.88
Φ	0.30	0.20	0.08	0.04	0.20	0.34
θ	0.50	0.36	0.09	0.20	0.36	0.54
Θ	0.80	0.83	0.11	0.61	0.82	1.04
σ^2	1.00	0.97	0.13	0.74	0.96	1.27
y_{n+1}	0.53	0.43	0.99	-1.54	0.44	2.29
y_{n+2}	-0.24	-0.42	1.52	-3.49	-0.34	2.43
y_{n+3}	0.69	-0.41	1.87	-4.04	-0.36	3.17
y_{n+4}	-0.66	-1.53	2.06	-5.63	-1.59	2.54

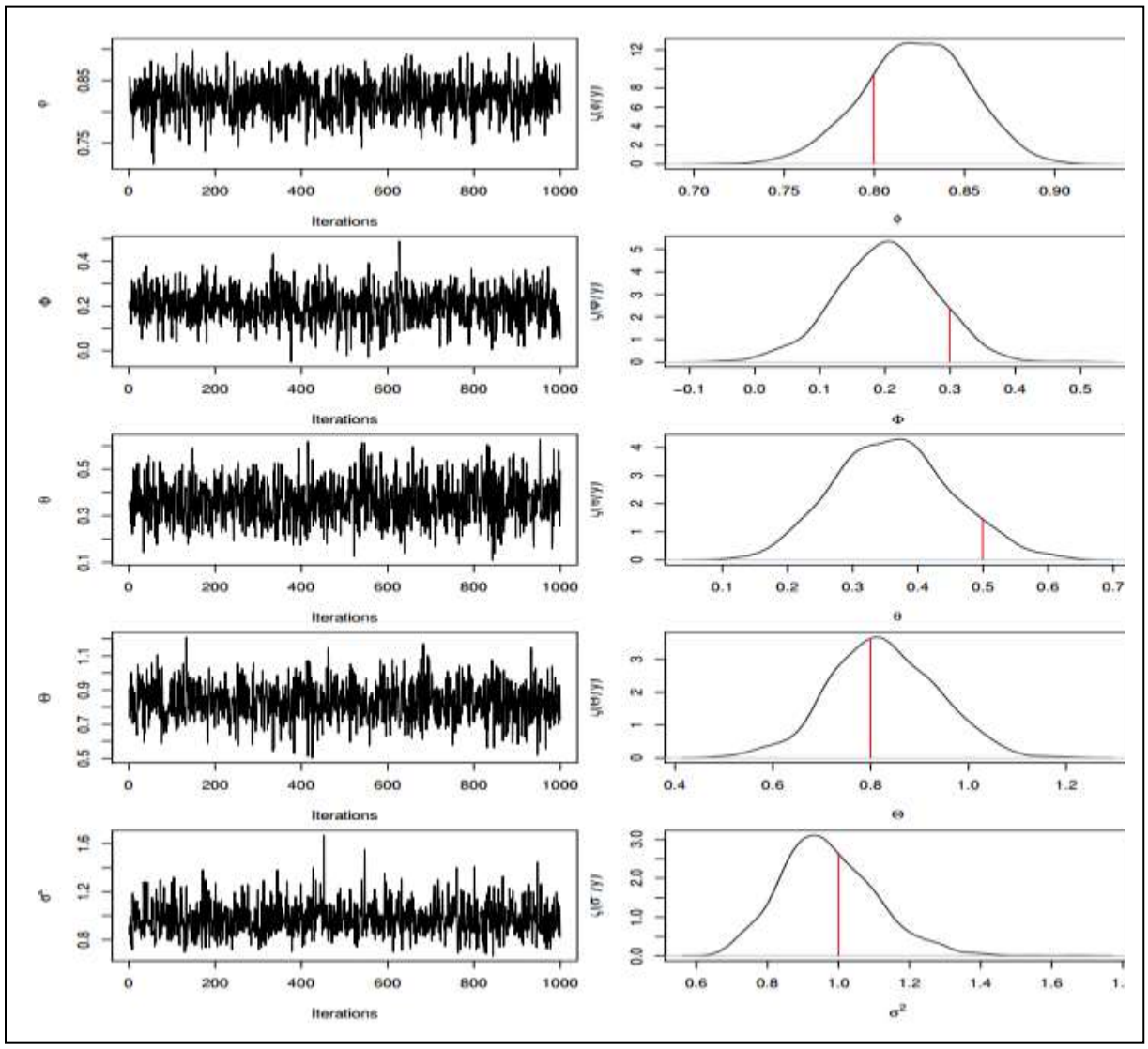


Figure 1: Sequential plots and marginal posterior distributions of Model I

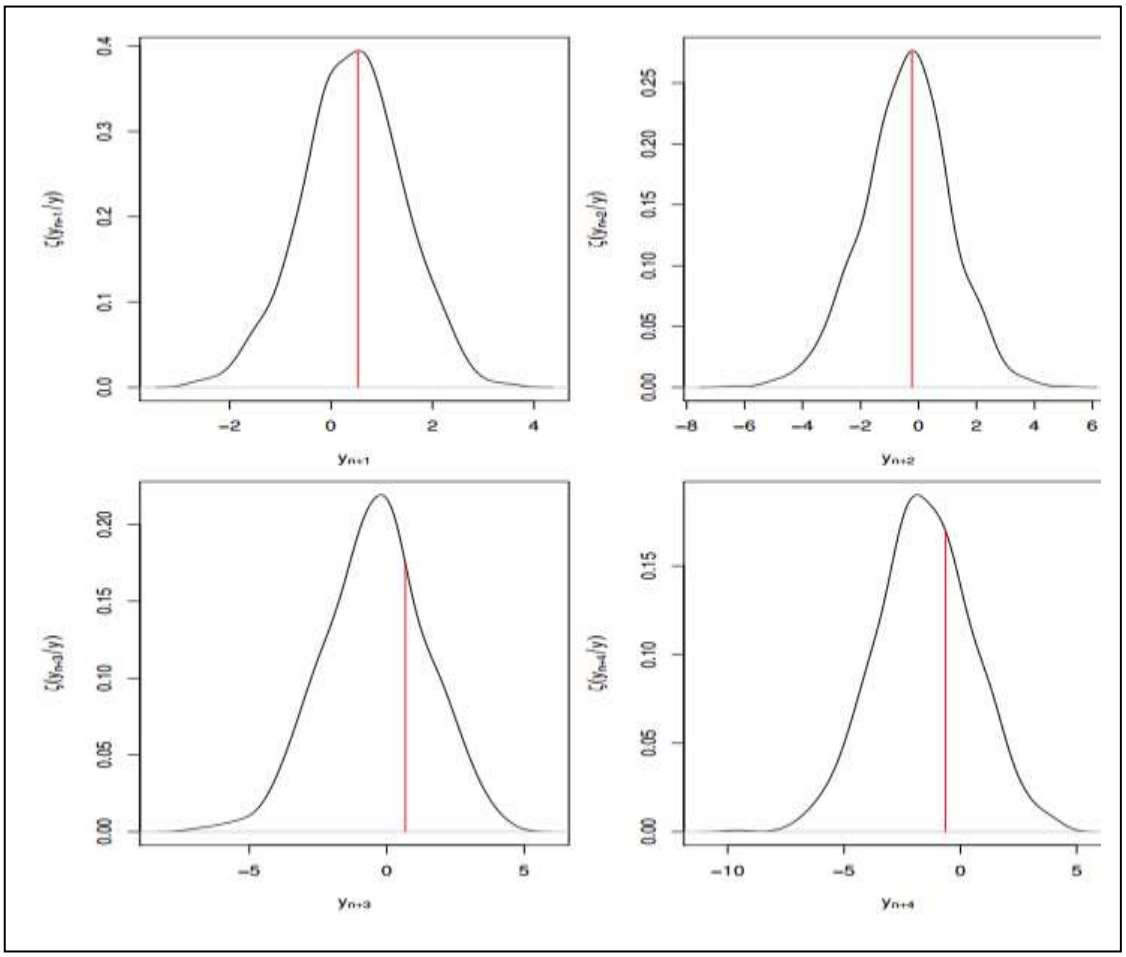


Figure 2: Marginal predictive distributions of Model I

Using the convergence diagnostics summarized in Section 5, the convergence of the proposed Gibbs sampling algorithm is monitored. In particular, the autocorrelations and Raftery and Lewis diagnostics are displayed in Table 3 and Geweke diagnostics are displayed in Table 4. Table 3 shows that the draws for each parameter have small autocorrelations, indicating that there is no convergence problem. This conclusion was confirmed by the diagnostic measures of Raftery and Lewis where the reported (Nmin) is 994 which 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 close to 1 which indicates the iid nature of the output sample.

A similar analysis, to that applied to Model I, is applied to Models II, III, and IV, and all the Bayesian results are presented in Tables 5, 6, and 7. From all these results we can claim that similar conclusions to those of Model I are obtained for the other three models, which confirm the accuracy of the proposed Gibbs sampling algorithm.

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

parameter	Autocorrelations				Raftery-Lewis Diagnostics			
	Lag 1	Lag 5	Lag 10	Lag 50	Burn	Total(N)	(Nmin)	I-stat
ϕ	-0.00	0.00	-0.02	0.02	3	1117	994	1.12
Φ	-0.02	0.01	-0.05	0.01	2	948	994	0.95
θ	0.01	-0.01	0.03	0.07	2	1028	994	1.03
Θ	0.04	-0.05	0.02	0.03	2	1028	994	1.03
σ^2	0.01	-0.02	0.04	0.02	2	1028	994	1.03
y_{n+1}	-0.01	-0.05	0.01	0.02	2	1028	994	1.03
y_{n+2}	0.03	-0.04	0.03	0.02	2	948	994	0.95
y_{n+3}	0.06	-0.07	0.03	0.00	2	948	994	0.95
y_{n+4}	0.06	-0.01	0.07	0.02	2	948	994	0.95

Table 4: Geweke diagnostics for Model I.

Param.	NSE iid	RNE iid	NSE 4%	RNE 4%	NSE 8%	RNE 8%	NSE 15%	RNE 15%	χ^2
ϕ	0.0009	1	0.0009	1.0232	0.0008	1.3247	0.0007	1.8775	0.1923
Φ	0.0024	1	0.0028	0.7065	0.0025	0.9109	0.0021	1.2951	0.2428
θ	0.0028	1	0.0026	1.1402	0.0026	1.2052	0.0026	1.1569	0.2212
Θ	0.0034	1	0.0030	1.2620	0.0023	2.1333	0.0022	2.3668	0.3100
σ^2	0.0042	1	0.0043	0.9884	0.0036	1.3734	0.0033	1.6717	0.6940
y_{n+1}	0.0313	1	0.0313	1.0035	0.0311	1.0130	0.0272	1.3280	0.8612
y_{n+2}	0.0480	1	0.0478	1.0049	0.0459	1.0895	0.0321	2.2326	0.9631
y_{n+3}	0.0592	1	0.0574	1.0632	0.0529	1.2506	0.0421	1.9789	0.9746
y_{n+4}	0.0653	1	0.0717	0.8283	0.0674	0.9366	0.0666	0.9594	0.9384

Table 5: Bayesian results for example II.

Parameter	True values	Mean	Std. Dev.	Lower 95 % limit	Median	Upper 95 % limit
ϕ	0.30	0.20	0.10	0.02	0.20	0.39
Φ	0.40	0.31	0.07	0.17	0.31	0.46
θ	0.40	0.39	0.14	0.13	0.39	0.68
Θ	0.70	0.71	0.13	0.45	0.71	0.96
σ^2	1.00	0.96	0.14	0.73	0.95	1.26
y_{n+1}	-0.06	-0.17	0.99	-2.14	-0.15	1.69
y_{n+2}	-0.67	-0.63	1.13	-2.91	-0.59	1.49
y_{n+3}	0.83	-0.07	1.19	-2.44	-0.08	2.24
y_{n+4}	-0.98	-1.13	1.23	-3.47	-1.18	1.29

Table 6: Bayesian results for Model III.

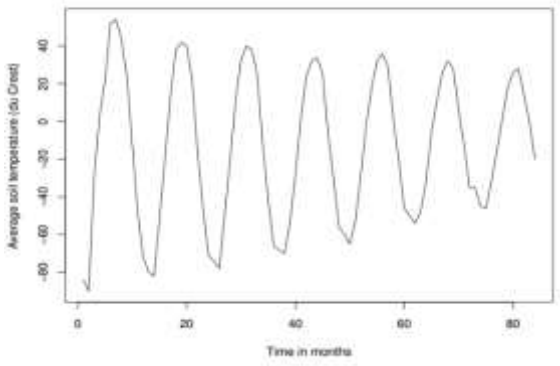
Parameter	True values	Mean	Std. Dev.	Lower 95 % limit	Median	Upper 95 % limit
ϕ	0.60	0.64	0.05	0.55	0.64	0.74
Φ	0.50	0.51	0.05	0.41	0.51	0.61
θ	0.30	0.21	0.07	0.08	0.21	0.35
Θ	0.50	0.41	0.09	0.24	0.41	0.58
σ^2	1.00	0.92	0.07	0.79	0.92	1.08
y_{n+1}	-2.42	-2.62	0.95	-4.53	-2.64	-0.69
y_{n+2}	0.18	-1.50	1.27	-3.90	-1.55	0.97
y_{n+3}	-0.48	-2.00	1.37	-4.64	-2.06	0.66
y_{n+4}	-2.73	-2.98	1.41	-5.65	-3.03	-0.20
y_{n+5}	-3.59	-3.09	1.44	-5.86	-3.13	-0.13
y_{n+6}	-4.77	-2.67	1.43	-5.43	-2.68	-0.02
y_{n+7}	-2.93	-1.65	1.43	-4.36	-1.71	1.14
y_{n+8}	0.29	-0.84	1.49	-3.78	-0.88	2.06
y_{n+9}	-0.52	-1.51	1.49	-4.37	-1.53	1.42
y_{n+10}	0.04	-1.86	1.45	-4.90	-1.88	1.11
y_{n+11}	-0.24	-2.01	1.45	-4.82	-1.99	0.81
y_{n+12}	-2.36	-2.09	1.49	-5.13	-2.08	0.73

Table 7: Bayesian results for Model IV.

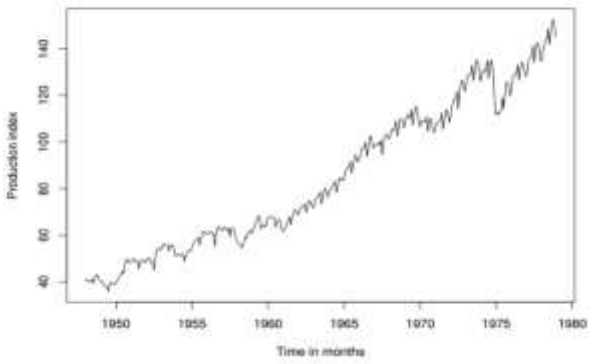
Parameter	True values	Mean	Std. Dev.	Lower 95 % limit	Median	Upper 95 % limit
ϕ	-0.50	-0.57	0.04	-0.65	-0.58	-0.49
Φ	0.60	0.60	0.04	0.51	0.60	0.68
θ	-0.50	-0.51	0.07	-0.64	-0.51	-0.37
Θ	0.40	0.29	0.06	0.17	0.29	0.42
σ^2	1.00	0.93	0.07	0.80	0.93	1.10
y_{n+1}	-2.17	-2.31	0.97	-4.22	-2.33	-0.32
y_{n+2}	4.05	2.71	1.43	-0.10	2.75	5.55
y_{n+3}	-4.16	-2.72	1.53	-5.82	-2.75	0.17
y_{n+4}	0.74	0.77	1.60	-2.27	0.74	4.12
y_{n+5}	-0.66	-0.54	1.63	-3.73	-0.49	2.49
y_{n+6}	-1.11	0.30	1.60	-2.90	0.30	3.46
y_{n+7}	2.00	-0.02	1.59	-3.04	-0.06	3.30
y_{n+8}	0.61	0.11	1.60	-3.05	0.09	3.12
y_{n+9}	-2.56	-1.16	1.62	-4.27	-1.22	1.91
y_{n+10}	2.75	0.59	1.59	-2.54	0.54	3.67
y_{n+11}	-2.24	-0.67	1.62	-3.80	-0.61	2.35
y_{n+12}	-0.36	0.40	1.62	-2.74	0.36	3.75

6.2 Real-World Application

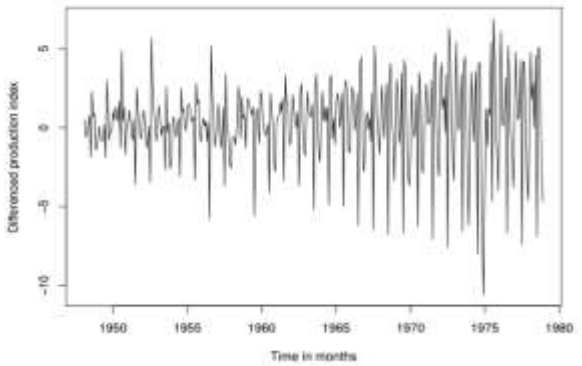
In this subsection we apply the proposed Bayesian analysis of SARMA models to two real-world time series that characterized by exhibiting seasonal pattern. The first time series dataset is 84 observations of average monthly soil temperature near Zurich at seven different depths, averaged over four years beginning in 1762, and the temperature measurements are related to the 'du Crest' scale (see Lambert, 1779, Page 358). This time series dataset is one of the earliest time series in scientific literature. The second time series dataset is the Federal Reserve Board Production Index that consists of 372 monthly values from January 1948 to December 1978 (Shumway and Stoffer, 2006). These two time series datasets are shown in Figure 3. Figure 3 shows that the average soil temperature time series is stationary, however, Figure 3 shows that the production index time series is nonstationary. We used the first (nonseasonal) difference to stationarize the production index time series but still it seems nonstationary in the seasonal component as shown in Figure 3, and accordingly we used the seasonal difference and now it seems stationary as displayed in Figure 3. This implies that we apply our proposed Bayesian analysis for the stationary differenced production index, not for its nonstationary raw data.



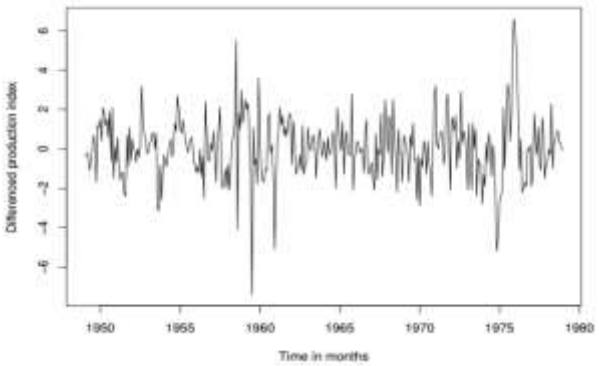
(a) Average monthly soil temperature near Zurich



(b) Federal reserve board production index



(c) Nonseasonal difference of production index



(d) Nonseasonal and seasonal differences of production index

Figure 3: Time-plot of the two real-world time series datasets

We use the first 78 observations of the average soil temperature data and the first 360 observations of the production index data to identify the suitable SARMA models and apply our proposed algorithm to conduct the posterior and predictive analysis to these two time series datasets, and the remaining observations (i.e. $k = 6$ for the average soil temperature data and $k = 12$ for the production index data) we use to evaluate out of sample prediction performance. In order to identify the best suitable order of the SARMA models for the two time series, we use the Akaike information criterion (AIC) with assuming the maximum order of each of the nonseasonal and seasonal polynomials is two in the SARMA model (1) according to the recommendation of Laing and Smith (1987). We compute the AIC for all the combinations of SARMA models and select the best model with smallest value of AIC. In particular, the identified models are SARMA(2,2)(1,1) $_{12}$ for the average soil temperature data and SARMA(1,1)(2,2) $_{12}$ for the differenced production index data.

Using the identified models, we apply the proposed Bayesian analysis to conduct the posterior and predictive analysis to the two real-world time series datasets with choosing the hyperparameters as in the simulation study. Table 8 summarizes the Bayesian results for the average soil temperature data, wheretheir marginal predictive densities are displayed in Figure 4. On the other hand, the Bayesian results for the differenced production index are presented in Table 9, and their marginal predictive densities are displayed in Figure 5. From the Bayesian forecasts of the next six future values of the average soil temperature data (displayed in Figure 4) and those of the next twelve future values of the production index data (displayed in Figure 5), we can confirm the applicability of our proposed Gibbs sampling algorithm to real-world time series.

Table 8: Bayesian results for the average soil temperature.

Parameter	True values	Mean	Std. Dev.	Lower 95 % limit	Median	Upper 95 % limit
ϕ_1	-	1.58	0.08	1.39	1.58	1.71
ϕ_2	-	-0.75	0.09	-0.88	-0.76	-0.55
Φ	-	0.62	0.06	0.51	0.62	0.74
θ_1	-	-0.39	0.10	-0.58	-0.39	-0.19
θ_2	-	-0.19	0.11	-0.41	-0.19	0.03
Θ	-	-0.12	0.13	-0.36	-0.12	0.09
σ^2	-	14.13	2.75	10.03	13.65	20.51
y_{n+1}	18.00	17.35	3.98	10.00	17.26	25.45
y_{n+2}	26.00	26.33	6.23	13.90	26.50	38.62
y_{n+3}	28.00	26.44	7.33	12.49	26.39	41.13
y_{n+4}	14.00	12.60	8.23	-2.45	12.15	30.77
y_{n+5}	0.00	-1.03	8.53	-16.96	-1.57	16.68
y_{n+6}	-20.00	-17.66	8.48	-33.59	-17.78	0.30

Table 9: Bayesian results for the differenced production index.

Parameter	True values	Mean	Std. Dev.	Lower 95 % limit	Median	Upper 95 % limit
ϕ	-	0.63	0.09	0.44	0.64	0.81
Φ_1	-	-0.23	0.10	-0.41	-0.24	-0.03
Φ_2	-	-0.30	0.06	-0.41	-0.30	-0.19
θ	-	-0.31	0.11	-0.52	-0.31	-0.10
Θ_1	-	-0.46	0.12	-0.70	-0.45	-0.25
Θ_2	-	0.05	0.08	-0.11	0.05	0.21
σ^2	-	1.19	0.09	1.03	1.18	1.37
y_{n+1}	-0.60	-0.30	1.13	-2.49	-0.29	1.91
y_{n+2}	0.00	-0.71	1.10	-2.96	-0.70	1.25
y_{n+3}	-0.30	-1.33	1.14	-3.65	-1.33	0.84
y_{n+4}	2.30	0.32	1.19	-2.05	0.32	2.57
y_{n+5}	-1.00	0.10	1.20	-2.35	0.08	2.44
y_{n+6}	0.30	0.13	1.21	-2.07	0.09	2.69
y_{n+7}	0.50	0.98	1.16	-1.21	0.97	3.25
y_{n+8}	0.90	1.15	1.21	-1.10	1.14	3.57
y_{n+9}	0.90	0.11	1.23	-2.26	0.10	2.44
y_{n+10}	0.30	-0.51	1.19	-2.88	-0.49	1.76
y_{n+11}	0.30	0.25	1.20	-2.05	0.24	2.60
y_{n+12}	-0.10	-0.73	1.23	-3.07	-0.79	1.75

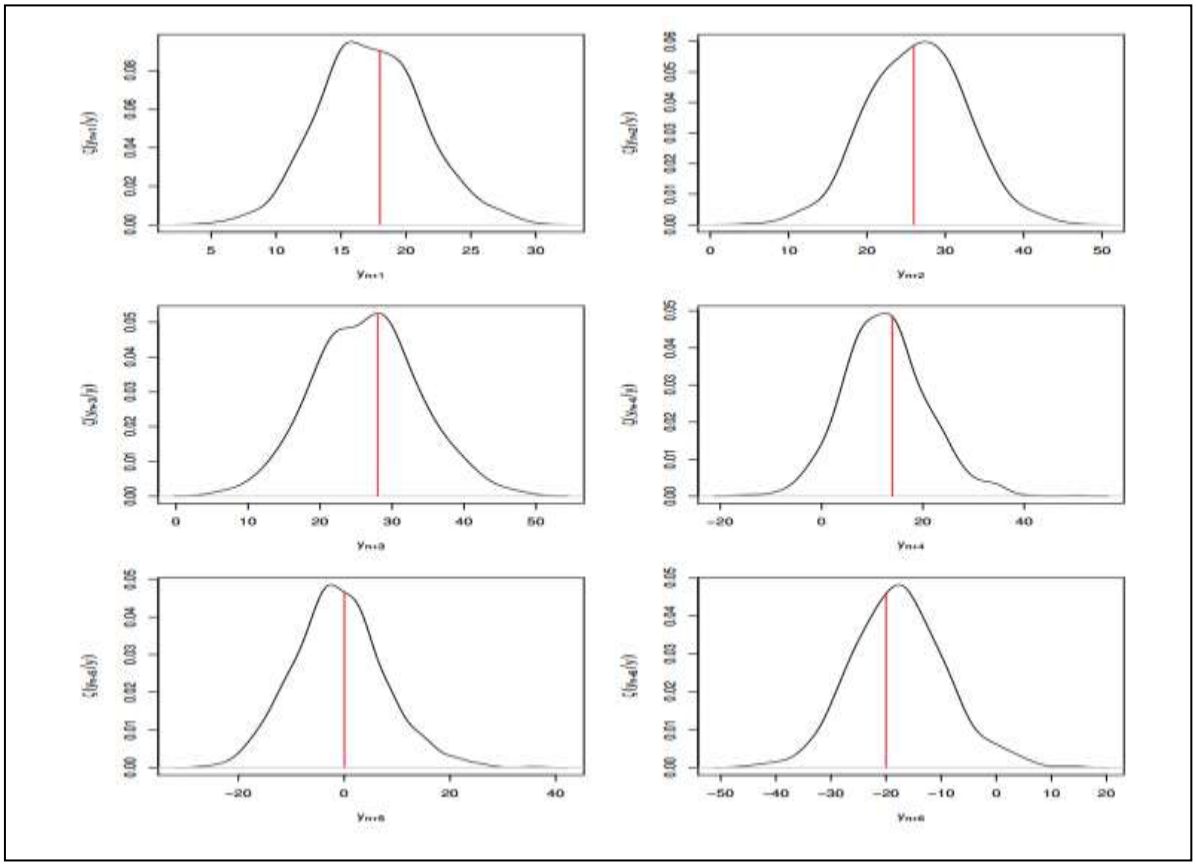


Figure 4: Marginal predictive distributions of the average soil temperature

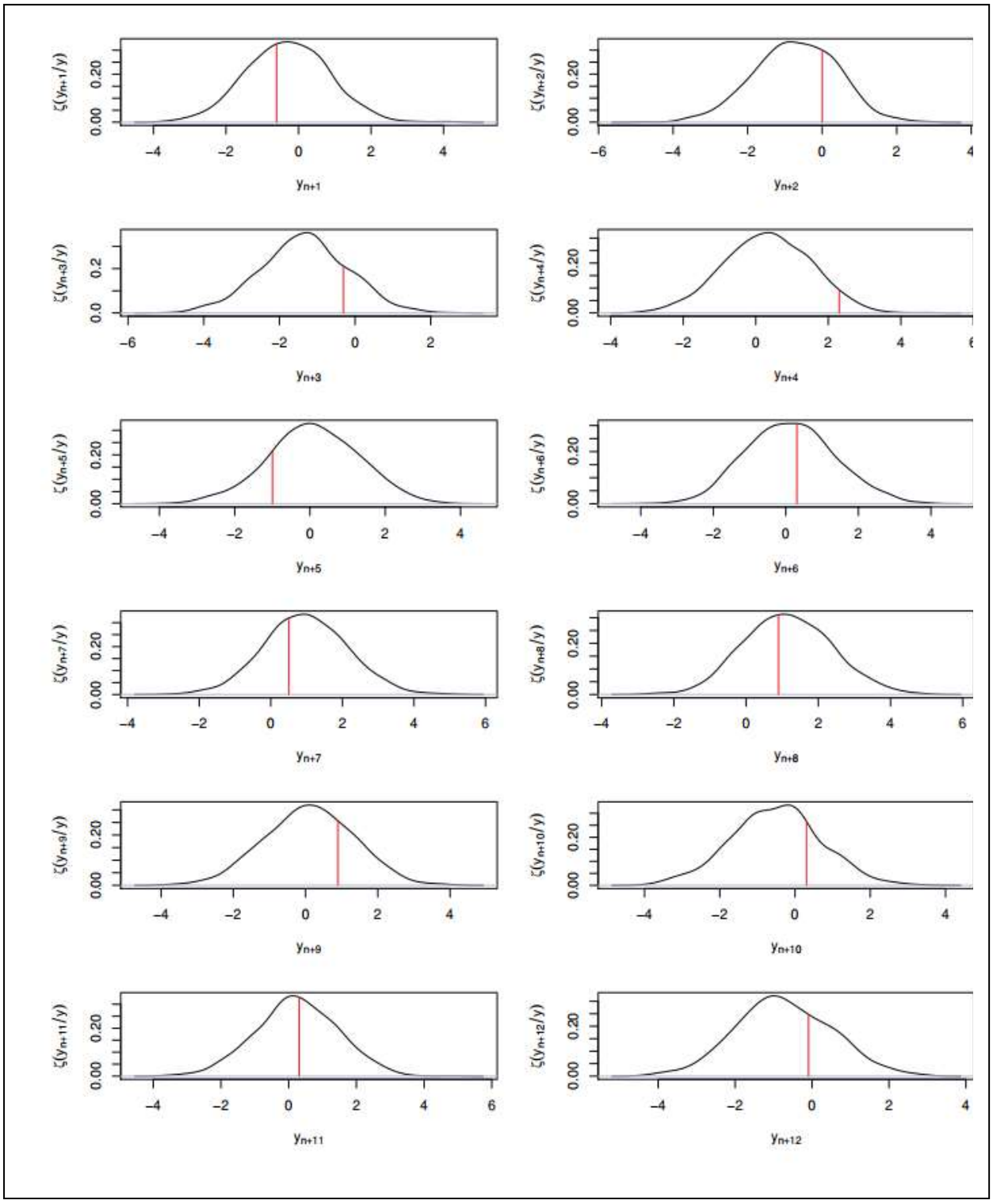


Figure 5: Marginal predictive distributions of the differenced production index

7. Conclusions

In this paper we first used the nonlinear least squares (NLS) method to estimate the unobserved errors in the multiplicative SARMA models and accordingly approximate its likelihood function. We employed conjugate priors on the model parameters and initial values to show that the conditional posterior distributions of the model parameters and variance are multivariate normal and inverse gamma respectively, and the conditional predictive distribution of the future observations is a multivariate normal. Exploiting that the conditional posterior and predictive densities are standard distributions, we used the Gibbs sampling algorithm to present a Bayesian method for estimating the SARMA model parameters and obtaining multiple-step ahead predictions. Simply, we applied the Gibbs sampling algorithm to approximate empirically the marginal posterior and predictive distributions along with using several convergence diagnostics. Accordingly, we computed directly the posterior estimates of the model parameters and the Bayesian forecasts of the future values as sample averages of the Gibbs sampling chains. The empirical results of the simulated and real-world time series datasets confirmed the accuracy of the proposed Bayesian method. Future work may include model identification using stochastic search variable selection, outliers detection, and extension to multivariate seasonal time series models.

References

1. Amin, A. (2009). Bayesian inference for seasonal ARMA models: A Gibbs sampling approach. Master's thesis, Statistics Department, Faculty of Economics and Political Science, Cairo University, Egypt.
2. Amin, A. (2017a). Bayesian inference for double seasonal moving average models: A Gibbs sampling approach. *Pakistan Journal of Statistics and Operation Research*, 13(3), 483–499.
3. Amin, A. (2017b). Gibbs sampling for double seasonal ARMA models. Proceedings of the 29th Annual International Conference on Statistics and Computer Modeling in Human and Social Sciences.
4. Amin, A. (2018a). Bayesian inference for double SARMA models. *Communications in Statistics: Theory and Methods*, doi: 10.1080/03610926.2017.1390132.
5. Amin, A. (2018b). Bayesian identification of double seasonal autoregressive time series models. *Communications in Statistics: Simulation and Computation*, doi: 10.1080/03610918.2018.1458130.
6. Amin, A. (2019a). Bayesian analysis of double seasonal autoregressive models. *Sankhya B: The Indian Journal of Statistics*, doi: 10.1007/s13571-019-00192-z.
7. Amin, A. (2019b). Kullback-Leibler divergence to evaluate posterior sensitivity to different priors for autoregressive time series models. *Communications in Statistics: Simulation and Computation*, doi: 10.1080/03610918.2017.1410709.
8. 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.
9. 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.
10. Barnett, G., Kohn, R., and Sheather, S. (1997). Robust Bayesian estimation of an autoregressive moving-average models. *Journal of Time Series Analysis*, 18, 11-28.
11. Box, G.E.P and Jenkins, G. M. (1976). *Time Series Analysis, Forecasting and Control*. Holden-Day, San Francisco.

12. Broemeling, L. D. and Shaarawy, S. (1984). Bayesian inferences and forecasts with moving average processes. *Communications in Statistics: Theory and Methods*, 13, 1871-1888.
13. Chib, S. and Greenberg, E. (1994). Bayes inference in regression models with ARMA(p,q) errors. *Journal of Econometrics*, 64, 183-206.
14. 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.
15. Ismail, M.A. and Amin, A.A. (2014). Gibbs Sampling for SARMA Models. Pakistan Journal of Statistics, 30(2), 153-168.
16. Laing, W.D. and Smith, D.G. (1987). A comparison of time series forecasting methods for predicting the CEGB demand. Proceedings of the Ninth Power Systems Computation Conference.
17. Lambert, J.H. (1779), Johann Heinrich Lambert... Pyrometrie oder vom Maasze des Feuers und der Warme. bey Haude und Spender.
18. 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.
19. Newbold, P. (1973). Bayesian estimation of Box Jenkins transfer function noise models.JRSSB, 35, 323-336.
20. Raftrey, A. E. and Lewis, S. (1992). One long run with diagnostics: Implementation strategies for Markov Chain Monte Carlo. *Statistical Science*, 7, 493-497.
21. 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.
22. Shumway, R.H. and Stoffer, D.S. (2006). Time series analysis and its applications with R examples. New York: Springer.
23. Vermaak, J., Niranjan, M. and Godstil, S. (1998). Markov chain Monte Carlo estimation for the seasonal autoregressive process with application to pitch modelling. *Technical report*. Engineering department, Cambridge University, Cambridge, England.
24. Zellner, A. and Reynolds, R. (1978). Bayesian analysis of ARMA models. Presented at the Sixteenth Seminar on Bayesian Inference in Econometrics, 23, 1-5.