Gibbs Sampling for Double Seasonal Moving Average Models

Mohamed Ismail*
Statistics Department, Cairo University, Egypt - mismail@feps.edu.eg

Ayman Amin
Statistics Department, Munofiya University, Egypt - aymanaamin@gmail.com

Abstract

In this paper we develop a Bayesian inference for multiplicative double seasonal moving average (DSMA) model by implementing a fast, easy and accurate Gibbs sampling algorithm. We first show that the conditional posterior distribution of the model parameters and the variance are multivariate normal and inverse gamma respectively, and then we apply the Gibbs sampling to approximate empirically the marginal posterior distributions. The proposed Bayesian methodology is illustrated using a simulated example.

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

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, that is, any periodic pattern of fixed length. For example, hourly electricity load data can exhibit intraday and intraweek seasonal patterns. 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 at least 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) 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 & 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 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 may include Mohamed et al. (2011) and Kim (2013) and references therein, 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 (2003) 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 multiplicative SARIMA model. With respect to the Bayesian analysis of double SARIMA model, as far as the authors know, it has not been tackled in the literature,

with the exception that recently Ismail and Zahran (2014) 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 DSMA model, which has the advantage that is unconditional on the initial values.

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

2. Multiplicative Double Seasonal Moving Average Model (DSMA)

A time series fy_tg is said to be generated by a multiplicative seasonal moving average model of orders q, Q_1, and Q_2, denoted by DSMA(q)$(Q_1)_{s_1}(Q_2)_{s_2}$, if it satis es where $\hat{\theta}_i \in R^q$, $\hat{\Theta}_j \in R^{Q_1}$, and $\hat{\Psi}_\tau \in R^{Q_2}$ are the parameters' estimates. Substituting the residuals in the likelihood function (3) results in an approximate likelihood function

$$l^\star \propto (\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left(\mathbf{y} - \hat{\Sigma}\beta\right)^T\left(\mathbf{y} - \hat{\Sigma}\beta\right)\right\}. \quad (5)$$

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

$$\begin{aligned}\hat{\Sigma}_t = \; & (e_{t-1}, \dots, e_{t-q}, e_{t-s_1}, e_{t-s_1-1}, \dots, e_{t-s_1-q}, \dots\dots, e_{t-Q_1s_1}, e_{t-Q_1s_1-1}, \dots, e_{t-Q_1s_1-q}, e_{t-s_2},\\ & e_{t-s_2-1}, \dots, e_{t-s_2-q}, e_{t-s_1-s_2}, e_{t-s_1-s_2-1}, \dots, e_{t-s_1-s_2-q}, \dots\dots, e_{t-Q_1s_1-s_2}, e_{t-Q_1s_1-s_2-1},\\ & \dots, e_{t-Q_1s_1-s_2-q}, \dots\dots\dots, e_{t-Q_2s_2}, e_{t-Q_2s_2-1}, \dots, e_{t-Q_2s_2-q}, e_{t-s_1-Q_2s_2}, e_{t-s_1-Q_2s_2-1}, \dots,\\ & e_{t-s_1-Q_2s_2-q}, \dots\dots, e_{t-Q_1s_1-Q_2s_2}, e_{t-Q_1s_1-Q_2s_2-1}, \dots, e_{t-Q_1s_1-Q_2s_2-q}).\end{aligned} \quad (6)$$

3.2 Prior Specification

Assuming that the parameters θ, Θ, Ψ and $\varepsilon_\mathbf{0}$ are independent apriori, given the error variance parameter σ^2, the joint prior distribution is

$$\begin{aligned}\zeta\left(\theta, \Theta, \Psi, \sigma^2, \varepsilon_0\right) = \; & N_q(\mu_\theta, \sigma^2\Sigma_\theta) \times N_{Q_1}(\mu_\Theta, \sigma^2\Sigma_\Theta) \times N_{Q_2}(\mu_\Psi, \sigma^2\Sigma_\Psi) \times N_{q^\star}(\mu_{\varepsilon_\mathbf{0}}, \sigma^2\Sigma_{\varepsilon_0})\\ & \times IG(\frac{\nu}{2}, \frac{\lambda}{2}),\end{aligned} \quad (7)$$

where $q^\star = q + Q_1s_1 + Q_2s_2$, $N_r(\mu, \Sigma)$ is the r-variate normal distribution with mean μ and variance Σ, and IG(a, b) is the inverse gamma distribution with parameters a and b. The prior distribution (7) is chosen for several reasons, especially it is a conjugate prior and thus facilitates the mathematical calculations. Multiplying the joint prior distribution by the approximate likelihood function (5) results in the joint posterior $\zeta\left(\theta, \Theta, \Psi, \sigma^2, \varepsilon_0 \mid \mathbf{y}\right)$ which may be written as

$$\begin{aligned}\zeta\left(\theta, \Theta, \Psi, \sigma^2, \varepsilon_0 \mid \mathbf{y}\right) \propto \; & (\sigma^2)^{-\left(\frac{n+\nu^\star}{2}+1\right)} \exp\left\{-\frac{1}{2\sigma^2}\Big[\lambda + (\theta - \mu_\theta)^T\Sigma_\theta^{-1}(\theta - \mu_\theta)\right.\\ & + (\Theta - \mu_\Theta)^T\Sigma_\Theta^{-1}(\Theta - \mu_\Theta) + (\Psi - \mu_\Psi)^T\Sigma_\Psi^{-1}(\Psi - \mu_\Psi) +\\ & \left.(\varepsilon_\mathbf{0} - \mu_{\varepsilon_\mathbf{0}})^T\Sigma_{\varepsilon_0}^{-1}(\varepsilon_\mathbf{0} - \mu_{\varepsilon_\mathbf{0}}) + \left(\mathbf{y} - \hat{\Sigma}\beta\right)^T\left(\mathbf{y} - \hat{\Sigma}\beta\right)\Big]\right\}\end{aligned} \quad (8)$$

where $\nu^\star = \nu + 2q + Q_1(1 + s_1) + Q_2(1 + s_2)$.

3.3 Full Conditional Distributions

The conditional posterior distribution for each of the unknown parameters is obtained from the joint posterior distribution (8) by grouping together terms in the joint posterior that depend on this parameter, and finding the appropriate normalizing constant to form a proper density. In this study all the conditional posteriors that are obtained for the unknown parameters are normal and inverse gamma distributions for which sampling techniques exist. These conditional posterior distributions are presented in details in a research report that is available online and can be downloaded [1].

4. The Proposed Gibbs Sampler

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

1. Specify starting values θ^0, $\mathbf{\Theta}^0$, $\mathbf{\Psi}^0$, $(\sigma^2)^0$ and ε_0^0 and set j=0. A set of initial estimates of the model parameters can be obtained using the IS technique of Koreisha and Pukkila (1990).

2. Calculate the residuals recursively using (5) and the IS parameter estimates.

3. Simulate

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

[1] https://sites.google.com/site/aymanaamin/ResearchReports

$$\quad^{j} \quad \zeta \quad {}^{j} \mid \mathbf{y}, (\sigma^2)^{j-1}, \theta^j, \quad {}^{j}, \varepsilon_{\mathbf{0}}{}^{j-1} \quad ,$$
$$(\sigma^2)^j \quad \zeta \quad (\sigma^2)^j \mid \mathbf{y}, \theta^j, \quad {}^{j}, \quad {}^{j}, \varepsilon_{\mathbf{0}}{}^{j-1} \quad ,$$
$$\varepsilon_{\mathbf{0}}{}^{j} \quad \zeta \quad \varepsilon_{\mathbf{0}}{}^{j} \mid \mathbf{y}, (\sigma^2)^j, \theta^j, \quad {}^{j}, \quad {}^{j} \quad ,$$

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

This algorithm gives the next value of the Markov chain $\left\{\theta^{j+1}, \mathbf{\Theta}^{j+1}, \mathbf{\Psi}^{j+1}, (\sigma^2)^{j+1}, \varepsilon_0^{j+1}\right\}$ 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 $\left\{\theta^{j+1}, \mathbf{\Theta}^{j+1}, \mathbf{\Psi}^{j+1}, (\sigma^2)^{j+1}, \varepsilon_0^{j+1}\right\}$ 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 Markov chain can be monitored by three groups of diagnostics, which include 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 get 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 proosed 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.

These diagnostics will be used in section 5 to monitor the convergence of the proposed algorithm.

5. Illustrative Simulated Example

In this section we present one simulated example of DSMA model with true parameters values are θ = 0.6, $\mathbf{\Theta}$ = 0.2, $\mathbf{\Psi}$ = -0.3, and σ^2 = 1.0. Once the time series dataset is generated from this DSMA model, the Bayesian analysis is performed by assuming a non informative prior for the parameters θ, , and σ^2 and a normal prior with zero mean for initial errors $\varepsilon_{\mathbf{0}}$ with variance $\sigma^2 I_{q^\star}$. To apply the proposed Gibbs sampler, the starting values for the parameters θ, , and σ^2 are obtained using IS method, and the starting values for $\varepsilon_{\mathbf{0}}$ are assumed to be zeros. 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.

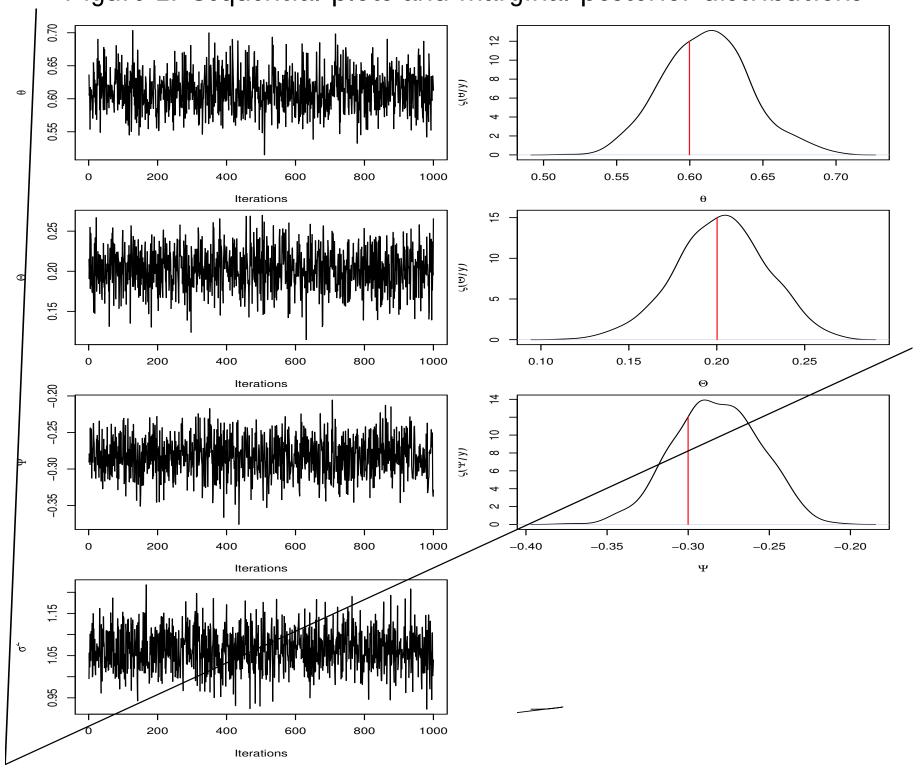
Table 1 presents the true values and Bayesian estimates of the parameters. Moreover, a 95% confidence interval using the 0.025 and 0.975 percentiles of the simulated draws is constructed for every parameter. From Table 1, it is clear that the Bayesian estimates are close to the true values and the 95%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

The convergence of the proposed algorithm is monitored using the diagnostic measures explained in Section 4. The autocorrelations and Raftery and Lewis diagnostics are displayed in Table 2 whereas Table 3 presents

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

Table 1: Bayesian results.

Figure 1: Sequential plots and marginal posterior distributions



the diagnostic measures of Geweke (1992). Table 2 shows that the draws for each of the parameter have small autocorrelations at lags 1, 5, 10 and 50 which indicates that there is no convergence problem. This conclusion was confirmed by the the diagnostic measures of Raftery and Lewis where the reported (Nmin) is 994 which is close to the 1000 draws we used and I-stat value is about 1 which is less than 5. Scanning the entries of Table 3, confirms the convergence of the proposed algorithm where Chi squared probabilities show that the equal means hypothesis can not be rejected and no dramatic differences between the NSE estimates are found. In addition, the RNE estimates are close to 1 which indicates the iid nature of the output sample.

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

Table 2: Autocorrelations and Raftery-Lewis diagnostics.

5. Conclusion

In this paper we showed that the conditional posterior distribution of the DSMA model parameters and the variance are multivariate normal and inverse gamma, respectively. Exploiting that the conditional posterior

	NSE iid	RNE iid	NSE 4%	RNE 4%	NSE 8%	RNE 8%	NSE 15%	RNE 15%	χ^2
θ	0.0010	1	0.0010	1.0149	0.0009	1.2064	0.0009	1.4460	0.7383
	0.0010	1	0.0008	1.7299	0.0007	2.2224	0.0006	2.7406	0.8864
	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

Table 3: Geweke diagnostics.

densities are standard distributions, we used the simple MCMC Gibbs sampling algorithm to develop a Bayesian method for estimating the parameters of the multiplicative DSMA model. Simply, we applied the Gibbs sampling algorithm to approximate empirically the marginal posterior distributions along with using several diagnostics that showed the convergence of the proposed algorithm was achieved. Accordingly, we computed directly the posterior estimates of the parameters via sample averages of the simulation outputs. The empirical results of the simulated dataset con rmed the accuracy of the proposed methodology.

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

References

Baek, M. (2010). Forecasting Hourly Electricity Loads of South Korea: Innovations State Space Modeling Approach. The Korean Journal of Economics 17(2), 301-317.

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

Cruz, A., Munoz,A., Zamora, J.L., Espinola, R. (2011). The e ect of Wind Generation and Weekday on Spanish Electricity Spot price Forecasting. Electric Power Systems Research 81(10), 19241935.

Feinberg, E. and Genethliou, D. (2005). Load forecasting. In F. W. J. Chow, Applied Mathematics for Restructured Electirc Power Systems: Control and Computational Intelligence (pp. 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. (2003). Bayesian Analysis of Seasonal Moving Average Model: A Gibbs Sampling Approach. Japanese Journal of Applied Statistics, **32(2)**, 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(2)**, 123-136.

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

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.

Kim, M.S. (2013). Modeling Special-day E ects for Forecasting Intraday Electricity demand, European Journal of Operational Research 230(1), 170-180.

Mohamed, N, Ahmad, MH and Suhartono (2011). Forecasting Short Term Demand Using Double Seasonal ARIMA Model. World Applied Sciences Journal 13, 27-35.

Taylor, J.W. (2003). Short-Term Electricity Demand Forecasting Using Double Seasonal Exponential Smoothing. The Journal of the Operational Research Society 54(8), 799-805.

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: 116.

Thompson, H.E, and Tiao, G.C. (1971). Analysis of Telephone Data. The Bell Journal of Economics and Management Science 2, 514-541.