




Periodic autoregressive models with closed skew-normal innovations

T. Manouchehri¹ · A. R. Nematollahi¹ 

Received: 16 July 2018 / Accepted: 19 April 2019 / Published online: 7 May 2019

© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

This paper is concerned with the estimation problem of a periodic autoregressive model with closed skew-normal innovations. The closed skew-normal (CSN) distribution has some useful properties similar to those of the Gaussian distribution. Maximum likelihood (ML), Maximum a posteriori (MAP) and Bayesian approaches are proposed and compared in order to estimate the model parameters. For the Bayesian approach, the Gibbs sampling algorithm and for computing the ML and MAP estimations, the expectation–maximization algorithms are performed. The simulation studies are then conducted to compare the frequentist average losses of competing estimators and to study the asymptotic properties of the given estimators. The proposed model and methods developed in this paper are also applied to a real time series. The accuracy of the CSN and Gaussian models is compared by cross validation criterion.

Keywords PAR models · VAR models · Closed skew-normal · ECM algorithms · MAP estimate · Bayesian approach · Noninformative priors · Hit-and-run sampler · MCMC algorithms

1 Introduction and motivation

The main aim of this article is to make inference on the parameters of the periodically correlated (PC) time series with a flexible class of skewed innovations. The PC time series has potential applications in describing many phenomena in different area of sciences and technology (e.g. climatology, hydrology, economics, electrical engineering and signal processing). To indicate a part of many relevant works done on the theory and application of PC time, we cite Gladyshev (1961), Noakes et al. (1985), Osborn and Smith (1989), McLeod (1993), Gardner (1994), Hipel and McLeod (1994), McLeod (1994), Franses and Paap (1994), Franses (1996), Novales and de Frutto

✉ A. R. Nematollahi
ar.nematollahi@shirazu.ac.ir

¹ Department of Statistics, Shiraz University, Shiraz, Iran

(1997), Nematollahi and Soltani (2000), Serpedin et al. (2005), Lund et al. (2006), Hurd and Miamee (2007), Ursu and Turkman (2012), Chaari et al. (2014, 2015 and 2017) and Nematollahi et al. (2017).

Several estimation techniques are available for PAR models, namely the least squares method used by Franses and Paap (2004) for the univariate case and by Lutkepohl (2005) for multivariate case, the method of moments based on Yule–Walker equations and asymptotic properties provided by Pagano (1978), Troutman (1979) and Hipel and McLeod (1994) and the maximum likelihood estimation given by Vecchia (1985a, b). Ursu and Duchesne (2009) have studied the asymptotic distributions of the least squares estimators of the model parameters in periodic Vector-AR models. Lund and Basawa (2000) considered the recursive prediction and likelihood evaluation techniques for periodic autoregressive moving average (PARMA) time series models. The asymptotic properties of parameter estimates for causal and invertible PARMA models are studied by Basawa and Lund (2001).

In recent years, consideration has been given to the non-Gaussian time series model in analysis real data sets. Time series models with non-Gaussian innovations are well-studied in the literature. Li and McLeod (1988) considered the ARMA model with non-Gaussian innovations. Ni and Sun (2003) and Sun and Ni (2004, 2005) used the Bayesian inferences on the estimation parameters of vector autoregressive (VAR) model with multivariate normal and multivariate- t innovations and compared them with ML estimators via frequentist risk. Shao (2006, 2007) applied the ML estimation to mixture periodic autoregressive models with asymmetric or multimodal distributions. Bondon (2009) considered the estimation problem of the autoregressive (AR) model with epsilon-skew-normal innovations by using the method of moments and maximum likelihood estimations. Sharafi and Nematollahi (2016) studied the general AR model with skew normal (SN) innovation introduced by Azzalini (1985). The Maximum a posteriori (MAP) estimation of AR processes based on finite mixtures of scale-mixtures of skew-normal distributions is proposed by Maleki and Arellano-Valle (2017). Maleki et al. (2018) used a Bayesian analysis to AR models with scale mixtures of skew-normal (SMSN) innovations.

The main motivation of this paper is to provide a modification and improvement of the results derived by Manouchehri and Nematollahi (2019), where the estimation problems of the PAR(1) time series with symmetric and asymmetric innovations are discussed. In the asymmetric case, they proposed the multivariate skew-normal as the distribution of the T -dimensional innovations. In this paper, the multivariate closed skew normal (CSN) distribution is proposed, as an alternative innovation to the symmetric ones, to provide improved estimates in the PAR(p) time series.

The stationary ARMA models with multivariate skew-normal distributions introduced by Azzalini and Dalla Valle (1996) and Azzalini and Capitanio (1999) are well studied by Pourahmadi (2007), where the innovations are assumed to be correlated and the predictors are assumed to be nonlinear and heteroscedastic. In this case a limitation for modelling real time series will be occurred; the autocorrelations of the ARMA model do not converge to zero for large lags, unlike their Gaussian ARMA counterparts, as pointed out by Pourahmadi (2007). Interestingly, when the multivariate closed skew-normal distributions introduced by González-Farías et al. (2004) (and re-parametrized and re-generalized by Arellano-Valle and Azzalini (2006)) are

used in the ARMA models, the autocorrelations of the ARMA model decay to zero exponentially and the predictors are linear and homoscedastic as in the Gaussian case, when the innovations are considered as a sequence of iid random variables with a univariate distribution in this family, see Pourahmadi (2007) and Bondon (2009) for more details.

The multivariate CSN distribution was first introduced by González-Farías et al. (2004). A p -dimensional vector \mathbf{U} is said to have a CSN-distribution, in symbol $\mathbf{U} \sim CSN_{p,q}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Gamma}, \mathbf{v}, \boldsymbol{\Delta})$, if its density function is given by

$$f_{p,q}(\mathbf{u}) = \Phi_q^{-1}(\mathbf{0}; \mathbf{v}, \boldsymbol{\Delta} + \boldsymbol{\Gamma} \boldsymbol{\Sigma} \boldsymbol{\Gamma}') \varphi_p(\mathbf{u}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \Phi_q(\boldsymbol{\Gamma}(\mathbf{u} - \boldsymbol{\mu}); \mathbf{v}, \boldsymbol{\Delta}), \tag{1.1}$$

where $\boldsymbol{\mu} \in \mathbb{R}^p$, $\mathbf{v} \in \mathbb{R}^q$, $\boldsymbol{\Sigma} \in \mathbb{R}^{p \times p}$ and $\boldsymbol{\Delta} \in \mathbb{R}^{q \times q}$ are both covariance matrices, $\boldsymbol{\Gamma} \in \mathbb{R}^{q \times p}$, $\varphi_p(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\Phi_q(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ are the density and distribution functions of a p -dimensional normal with the indicated mean vector and covariance matrix. Comprehensive listing of the existing references is presented by Genton (2004) and Azzalini (2005). The multivariate CSN distribution has some useful properties similar to those of the Gaussian distribution. Linear combinations of components CSN random variables are also CSN random variables, thus, the CSN distribution is closed under linear transformations. The sum of two independent CSN variables will also be CSN. The CSN random variables conditional of the components are also CSN. The composition of two independent CSN variables will be also CSN. For more details, see the work of González-Farías et al. (2004). These favorable characteristics of the CSN distribution make them analytically tractable, relatively simple to fit observed data with lack of symmetry but with shape of the empirical distribution like normal distribution.

In this paper, it is shown that application of the multivariate CSN for the distribution of T -dimensional innovation associated with a second order PAR(p) time series of period T , is more appropriate and the obtained results are more accurate as compared to those reported by Manouchehri and Nematollahi (2019). The Maximum likelihood (ML), Maximum a posteriori (MAP) and Bayesian estimates of the model parameters were studied and the technical difficulties which are usually encountered in handling these methods were reported. The MAP estimate can be interpreted as a Bayes estimate when the loss function is not specified. It provides a way to incorporate prior information on the estimation process, and can be regarded as an extension of the ML estimation. The MAP estimation procedures are well proposed in the literature, see e.g. Gauvain and Lee (1994), Tolpin and Wood (2015), White et al. (2015) and Maleki and Arellano-Valle (2017) and references therein.

The outline of the paper is as follows: In Sect. 2, the PAR(p) model with the CSN innovations are introduced and the relation between the PC time series and the stationary vector series is recalled here for completeness. The ML and MAP estimation are computed by the EM algorithms in Sect. 3. In this section, the Bayesian estimates of the parameters are also obtained by the MCMC algorithms. In Sect. 4, the simulation studies are performed to check the validity of the estimation method. The consistency and asymptotic normality of estimators are also discussed. The model is then applied to a real data set in Sect. 5. Finally, a brief discussion is given in the last section.

2 Periodic autoregressive models with closed skew-normal innovations

The zero mean and real second order time series $X = \{X_t\}_{-\infty}^{\infty}$ is called a periodic autoregressive of order p (PAR(p)) if

$$X_t = \sum_{j=1}^p \phi_{jt} X_{t-j} + \varepsilon_t, \tag{2.1}$$

where the innovations $\{\varepsilon_t\}_{-\infty}^{\infty}$ are to be assumed independent and closed skew-normally distributed, denoted by $\varepsilon_t \sim CSN_{1,1}(0, \sigma_t^2, \alpha_t, 0, \Delta_t)$, and T is the smallest integer for which $\phi_{jt} = \phi_{j(t+T)}$, $\sigma_t^2 = \sigma_{t+T}^2$, $\alpha_t = \alpha_{t+T}$, $j = 1, \dots, p$. In our applications, it is supposed that the parameter Δ_t is a function of the unknown parameter, σ_t^2 .

Gladyshev (1961) showed that X is periodically correlated time series with period T if and only if the T -dimensional vector $(X_{tT}, X_{tT+1}, \dots, X_{tT+T-1})'$ is stationary in the wide sense. It can be shown that each periodic autoregressive time series X with period T is related to a stationary T dimensional vector autoregressive time series $Y = \{Y_t = (X_{tT}, X_{tT+1}, \dots, X_{tT+T-1})', t \in \mathbb{Z}\}$, (Pagano (1978)). For example, for $p = 2$ and $T = 4$, the model (2.1) can be written as a vector autoregressive model [VAR(1)] given by

$$Y_t = \Phi_0^{-1} \Phi_1^* Y_{t-1} + \Phi_0^{-1} E_t, \tag{2.2}$$

where

$$\Phi_0 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -\phi_{12} & 1 & 0 & 0 \\ -\phi_{23} & -\phi_{13} & 1 & 0 \\ 0 & -\phi_{24} & -\phi_{14} & 1 \end{pmatrix}, \quad \Phi_1^* = \begin{pmatrix} 0 & 0 & \phi_{21} & \phi_{11} \\ 0 & 0 & 0 & \phi_{22} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$Y_t = \begin{pmatrix} X_{4t-3} \\ X_{4t-2} \\ X_{4t-1} \\ X_{4t} \end{pmatrix}, \quad E_t = \begin{pmatrix} \varepsilon_{4t-3} \\ \varepsilon_{4t-2} \\ \varepsilon_{4t-1} \\ \varepsilon_{4t} \end{pmatrix}.$$

In general, each PAR(p) process with period time T has a VAR(P) model representation, where, $P = \lceil \frac{p+T-1}{T} \rceil$ and $\lceil \cdot \rceil$ denotes the integer part (Pagano 1978).

Therefore, assuming the period T is known, one can consider the following general VAR(P) model

$$Y_t = \sum_{j=1}^P \Phi_j Y_{t-j} + U_t, \tag{2.3}$$

where $\Phi_j = \Phi_0^{-1} \Phi_j^*$, $j = 1, \dots, P$ are unknown $T \times T$ coefficients matrices, $\{U_t\}_{-\infty}^{\infty}$ are independent and closed skew normally distributed from $CSN_{T,T}$

$(\mathbf{0}, \Sigma^*, \Gamma^*, \mathbf{0}, I_T)$. Here, $\Sigma^* = \Phi_0^{-1} \Sigma (\Phi_0^{-1})'$, $\Gamma^* = \Delta^{-1/2} \alpha \Phi_0$, $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_T^2)$ and $\alpha = \text{diag}(\alpha_1, \dots, \alpha_T)$ are $T \times T$ diagonal matrices. It can be shown that $Y_t | Y_{t-1}, \dots, Y_{t-P} \sim \text{CSN}_{T,T}(\sum_{j=1}^P \Phi_j y_{t-j}, \Sigma^*, \Gamma^*, \mathbf{0}, I_T)$, (González-Farías et al. 2004). In Sect. 3, the ML, MAP and Bayesian approaches were examined to estimate the parameters of the VAR model (2.3) which in turn can be applied to estimate the parameters of PAR model (2.1).

3 Inference method for estimation of parameter

In this section, three technical methods were applied to estimate parameters in the proposed model (2.1).

3.1 ML estimation

The ML estimation of the parameters is the values for which the exact (full) likelihood $f(Y_1, \dots, Y_n) = f(Y_1)f(Y_2) \cdots f(Y_P)f(Y_{P+1}, \dots, Y_n | Y_1, Y_2, \dots, Y_P)$, is maximized. In this section, we apply the conditional likelihood $f(Y_{P+1}, \dots, Y_n | Y_1, Y_2, \dots, Y_P)$ due to the stationarity condition, see Manouchehri and Nematollahi (2019). When the non-normal innovations are found in a general framework and under the stationarity condition, Li and McLeod (1988) have shown that the conditional MLE and the MLE are consistent and have the same limiting normal distribution.

The conditional likelihood function of $\theta = (\Phi_1, \dots, \Phi_P, \Sigma^*, \Gamma^*)$ provided the observed data matrix $Y = (Y_1, \dots, Y_n)$ is

$$\begin{aligned}
 CL(\theta | Y) &= f(Y_{P+1}, \dots, Y_n | Y_1, \dots, Y_P) \\
 &= |\Sigma^*|^{-n^*/2} \prod_{t=P+1}^n \Phi_T \left(\Gamma^* \left(y_t - \sum_{j=1}^P \Phi_j y_{t-j} \right); 0, I_T \right) \\
 \exp \left[\text{trace} \left\{ -\frac{1}{2} \Sigma^{*-1} \sum_{t=P+1}^n \left(y_t - \sum_{j=1}^P \Phi_j y_{t-j} \right) \left(y_t - \sum_{j=1}^P \Phi_j y_{t-j} \right)' \right\} \right], \tag{3.1}
 \end{aligned}$$

where $n^* = n - P$.

The objective is to find the conditional MLE of θ which requires a high dimensional nonlinear procedure. Instead, the latent structure of the model proposed by Theorem 3.1 was applied for the beneficial EM-based methods.

Theorem 3.1 Let $V_t \sim N_T(\sum_{j=1}^P \Phi_j y_{t-j}, G)$ and $W_t \sim N_T^0(0, \Lambda)$ (N_T^0 denotes the truncated multivariate normal at $\mathbf{0}$) and V_t is independent of W_t and

$$Y_t = V_t + DW_t, t = P, \dots, n$$

where $D = \Sigma^* \Gamma^{*'} \Lambda^{-1}$ is a full rank matrix, $G = \Sigma^* - D \Lambda D'$ and $\Lambda = I_T + \Gamma^* \Sigma^* \Gamma^{*'}$, then

- (a) $Y_t \sim CSN_{T,T} \left(\sum_{j=1}^P \Phi_j y_{t-j}, \Sigma^*, \Gamma^*, \mathbf{0}, I_T \right)$,
- (b) $Y_t | W_t = w_t \sim N_T \left(\sum_{j=1}^P \Phi_j y_{t-j} + D w_t, G \right)$,
- (c) $W_t | Y_t = y_t \sim C(y_t, \theta_1) N_T^0(v^*, \Lambda^*)$,

where $\Lambda^* = \left(\Lambda^{-1} + D' G^{-1} D \right)^{-1}$, $v^* = \Lambda^* D' G^{-1} \left(y_t - \sum_{j=1}^P \Phi_j y_{t-j} \right)$ and C is a function of parameters $\theta_1 = (\Phi_j, \Lambda, G, D)$ and observed data y_t . Also we have

$$E(W_t | Y_t) = C(y_t, \theta_1) (v^* + \Lambda^* \xi_1), \tag{3.2}$$

$$E(W_t' W_t | Y_t) = C(y_t, \theta_1) \left(\Lambda^* \xi_2 + \Lambda^* \xi_1 v^{*'} + \left(\Lambda^* \xi_1 v^{*'} \right)' + v^* v^{*' + \Lambda^*} \right), \tag{3.3}$$

where

$$\xi_1 = \frac{\partial \Phi_T(s; -v^*, \Lambda^*)}{\partial s} \Big|_{s=0},$$

$$\xi_2 = \frac{\partial^2 \Phi_T(s; -v^*, \Lambda^*)}{\partial s \partial s'} \Big|_{s=s'=0}$$

and $\frac{\partial \Phi_T(s; -v^*, \Lambda^*)}{\partial s}$, $\frac{\partial^2 \Phi_T(s; -v^*, \Lambda^*)}{\partial s \partial s'}$ are the first and second derivatives of multivariate normal distribution function, respectively.

The proof is left to “Appendix A”.

It is not possible to compute the high dimensional Eqs. (3.2) and (3.3), analytically, since there is no any analytical solution for $\Phi_T(., ., .)$. Thus the numerical approximations is needed to solve these equations. Alternatively, $E(W_i) = \int_{w_i} w_i p_i(w_i) dw_i$ can be approximated by a numerical method. For example, a sample of size N , say x_1, x_2, \dots, x_N can be generated from $N_T^0(v^*, \Lambda^*)$ and so the traditional estimates are given by

$$E(\widehat{W}_t | Y_t) = C(y_t, \theta_1) \frac{1}{N} \sum_{j=1}^N x_j, \quad E(\widehat{W}_t' \widehat{W}_t | Y_t) = C(y_t, \theta_1) \frac{1}{N} \sum_{j=1}^N x_j' x_j. \tag{3.4}$$

Therefore, the conditional log-likelihood function of $\theta = (\Phi, \Sigma, \alpha)$ based on the complete data $Y = (Y_1, \dots, Y_n)$ (observed data) and $W = (W_1, \dots, W_n)$ (hidden variables or missing data) is given by

$$\begin{aligned}
 Cl(\theta|Y, W) &= \log \left(\prod_{t=P}^n f(W_t) f(Y_t|W_t, Y_{t-1}, \dots, Y_{t-P}) \right) \\
 &= \log \left(\prod_{t=P}^n \Phi_T^{-1}(0; 0, \Lambda) \varphi_T(\mathbf{w}_t; 0, \Lambda) \varphi_T \left(Y_t; \sum_{j=1}^P \Phi_j y_{t-j} + D\mathbf{w}_t, G \right) \right) \\
 &= \log \left(\prod_{t=P}^n 2^T (2\pi)^{-T} |\Lambda G|^{-1/2} \right. \\
 &\quad \times \exp \left(\mathbf{w}'_t \Lambda^{-1} \mathbf{w}_t + \left(y_t - \sum_{j=1}^P \Phi_j y_{t-j} - D\mathbf{w}_t \right)' \right. \\
 &\quad \left. \left. \times G^{-1} \left(y_t - \sum_{j=1}^P \Phi_j y_{t-j} - D\mathbf{w}_t \right) \right) \right), \tag{3.5}
 \end{aligned}$$

And by simplifying the parameters $G = \Phi_0^{-1} \Sigma \Lambda^{-1} \Phi_0^{-1'}$, $D = \Phi_0^{-1} \Sigma \Lambda^{-1/2} \alpha \Lambda^{-1}$ and $\Lambda = I_T + \Delta^{-1} \Sigma \alpha$ in (3.5), we have

$$\begin{aligned}
 Cl(\theta|Y, W) &= n^* \log(|\Phi_0|) - \frac{n^*}{2} \log(|\Sigma|) + \text{tr} \left(-\frac{1}{2} W'_{-P} W_{-P} \right) \\
 &\quad + \text{tr} \left\{ -\frac{1}{2} (\Sigma^{-1} \Lambda) \left[\Phi_0 (Y_{-P} - Z_{-P} \Phi)' (Y_{-P} - Z_{-P} \Phi) \Phi_0' - \Phi_0 \right. \right. \\
 &\quad \left. \left. (Y_{-P} - Z_{-P} \Phi)' W_{-P} \Lambda^{-1} \alpha \Delta^{-\frac{1}{2}} \Sigma - \Sigma \Delta^{-\frac{1}{2}} \alpha \Lambda^{-1} W'_{-P} (Y_{-P} - Z_{-P} \Phi) \Phi_0' \right] \right\}, \tag{3.6}
 \end{aligned}$$

where

$$\begin{aligned}
 W_{-P} &= \begin{pmatrix} W'_{P+1} \\ \vdots \\ W'_n \end{pmatrix}, \quad Y_{-P} = \begin{pmatrix} Y'_{P+1} \\ \vdots \\ Y'_n \end{pmatrix}, \quad Z_{-P} = \begin{pmatrix} Z'_{P+1} \\ \vdots \\ Z'_n \end{pmatrix}, \\
 \Phi &= \begin{pmatrix} \Phi'_1 \\ \vdots \\ \Phi'_P \end{pmatrix} \quad \text{and} \quad Z'_t = (Y'_{t-1}, \dots, Y'_{t-P}).
 \end{aligned}$$

The EM algorithm is a helpful technique for ML estimation in models with hidden variables W and has several good features such as stability of monotone convergence and simplicity of implementation (Liu and Rubin (1994)). However, if the M-step of this algorithm is not in closed form, EM loses some of its attraction. The ECM

algorithm proposed by Meng and Rubin (1993) is a simple modification of EM in which the M(maximization)-step is replaced by a sequence of computationally conditional maximization (CM)-steps.

Here, the ECM algorithm is used for finding ML estimates of parameters. Conditional expectation used in the ECM algorithm is $Q(\theta, \theta^{(k)}) = E_{\theta^{(k)}}[Cl(\theta|Y, W)|Y]$, where, $\theta^{(k)} = (\Phi^{(k)}, \alpha^{(k)}, \Sigma^{(k)}, \Phi_0^{(k)})$ is the estimated value of θ in the k -th step of algorithm as follows:

E-step Calculate $M^{(k)}$ and $R^{(k)}$ obtained from conditional expectations of $N_q^0(v^*, \Lambda^*)$, given by (3.4),

$$M^{(k)} = \left\{ E_{\theta^{(k)}}[W_t|U_t] \right\}_{t=P}^n, \quad R^{(k)} = \sum_{t=P}^n E_{\theta^{(k)}}[W_t' W_t|U_t], \quad U_t = Y_t - \sum_{j=1}^P \Phi_j^{(k)} Y_{t-j}, \tag{3.7}$$

So in the E-step of the algorithm, we have

$$\begin{aligned} Q(\theta, \theta^{(k)}) &= E_{\theta^{(k)}}[Cl(\theta|Y, W)|Y] \\ &= n^* \log(|\Phi_0^{(k)}|) - \frac{n^*}{2} \log(|\Sigma^{(k)}|) \\ &\quad + \text{tr} \left\{ -\frac{1}{2} (\Sigma^{(k)})^{-1} \Lambda^{(k)} \right\} \left[\Phi_0^{(k)} (Y_{-P} - Z_{-P} \Phi^{(k)})' (Y_{-P} \right. \\ &\quad \left. - Z_{-P} \Phi^{(k)}) \Phi_0^{(k)'} - \Phi_0^{(k)} (Y_{-P} - Z_{-P} \Phi^{(k)})' M^{(k)} \Lambda^{(k)-1} \alpha^{(k)} \Delta^{(k)-1/2} \Sigma^{(k)} \right. \\ &\quad \left. - \Sigma^{(k)} \Delta^{(k)-1/2} \alpha^{(k)} \Lambda^{(k)-1} M^{(k)'} (Y_{-P} - Z_{-P} \Phi^{(k)}) \Phi_0^{(k)'} \right]. \end{aligned} \tag{3.8}$$

- *CM-steps*

Updating of parameters in the CM-steps evidently, will be done in the following parts:

- *CM-step 1* Update $\Phi^{(k)}$ by maximizing (3.8) over Φ which gives

$$\Phi^{(k+1)} = (Z'_{-P} Z_{-P})^{-1} \left[Z'_{-P} Y_{-P} - Z'_{-P} M^{(k)} \left(\Phi_0^{(k)-1} \Delta^{(k)-\frac{1}{2}} \Sigma^{(k)} \alpha^{(k)} \Lambda^{(k)-1} \right)' \right]. \tag{3.9}$$

- *CM-step 2* Update $\Sigma^{(k)}$ by maximizing (3.8) over Σ by solve the following equation

$$\begin{aligned} \Sigma^{(k+1)} &= \frac{\text{diag} \left[\Phi_0^{(k+1)} (Y_{-P} - Z_{-P} \Phi^{(k+1)})' (Y_{-P} - Z_{-P} \Phi^{(k+1)}) \Phi_0^{(k+1)'} \right]}{n^*} \\ &\quad + \Sigma^{2(k+1)} \left(\frac{\delta \Delta^{-1}}{\delta \Sigma} \right)^{(k+1)} \\ &\quad \times \frac{\text{diag} \left[\Phi_0^{(k+1)} (Y_{-P} - Z_{-P} \Phi^{(k+1)})' (Y_{-P} - Z_{-P} \Phi^{(k+1)}) \Phi_0^{(k+1)'} \alpha^{(k)2} \right]}{n^*} \end{aligned}$$

$$+ \Sigma^{2(k+1)} \left(\frac{\delta \Delta^{-\frac{1}{2}}}{\delta \Sigma} \right)^{(k+1)} \frac{\text{diag} \left[\Phi_0^{(k+1)} (Y_{-P} - Z_{-P} \Phi^{(k+1)})' M^{(k)} \alpha^{(k)} \right]}{n^*} = 0. \tag{3.10}$$

- *CM-step 3* Update $\alpha^{(k)}$ by maximizing (3.8) over α which gives

$$\alpha^{(k+1)} = \left\{ \frac{\text{diag} \left[\Phi_0^{(k+1)} (Y_{-P} - Z_{-P} \Phi^{(k+1)})' (Y_{-P} - Z_{-P} \Phi^{(k+1)}) \Phi_0^{(k+1)} \Delta^{(k+1)-1} \right]}{\text{diag} \left[\Delta^{(k+1)-\frac{1}{2}} \Phi_0^{(k+1)} (Y_{-P} - Z_{-P} \Phi^{(k+1)})' M^{(k)} \right]} \right\}^{-1} \tag{3.11}$$

The simple starting values of the parameters θ can be the diagonal positive matrix (identity matrix) for Σ and α and also $(Z'_{-P} Z_{-P})^{-1} Z'_{-P} Y_{-P}$ for Φ . The E- and CM-steps are alternated repeatedly until a suitable convergence rule is satisfied, for example $|CI(\theta^{(k+1)}|Y)/CI(\theta^{(k)}|Y) - 1| \leq \text{tolerance}$. The value considered here is 10^{-3} , but the choice of tolerance may vary with different users.

3.2 Bayesian estimation

The Bayesian analysis is implemented here for the proposed model. So, we need to consider prior distribution for all the unknown parameters $\theta = (\Phi, \Sigma, \alpha)$. Since no any prior information is available from historical data or from previous experiment, we choose noninformative prior distributions for the parameters. We also suppose that the prior distributions of parameters are independent. The prior for Φ is assumed to be the constant prior ($\pi_C(\Phi) \propto 1$) and the Jeffreys prior and RATS (Regression Analysis of Time Series) prior (a modified version of the Jeffreys prior, a software package popular among macroeconomists) are chosen for Σ given by $\pi_J(\Sigma) \propto |\Sigma|^{-\frac{T+1}{2}}$ and $\pi_A(\Sigma) \propto |\Sigma|^{-(T+1)}$, respectively, See Sun and Ni (2004, 2005). Similar to the approximate Jeffreys priors for α in the univariate case, which is $t(0, \frac{\pi^2}{4}, \frac{1}{2})$ (Bayes and Branco (2007)), we consider its multivariate version given by

$$\pi_J(\alpha) = \prod_{j=1}^T \sqrt{\frac{\pi}{2}} \left(1 + \frac{2\alpha_j^2}{\pi^2} \right)^{-\frac{3}{2}}.$$

So, the joint prior distributions of θ are

$$\pi_{CJJ}(\Phi, \Sigma, \alpha) \propto |\Sigma|^{-\frac{T+1}{2}} \prod_{j=1}^T \sqrt{\frac{\pi}{2}} \left(1 + \frac{2\alpha_j^2}{\pi^2} \right)^{-\frac{3}{2}}, \tag{3.12}$$

and

$$\pi_{CAJ}(\Phi, \Sigma, \alpha) \propto |\Sigma|^{-(T+1)} \prod_{j=1}^T \sqrt{\frac{\pi}{2}} \left(1 + \frac{2\alpha_j^2}{\pi^2} \right)^{-\frac{3}{2}}, \tag{3.13}$$

where CJJ and CAJ indicate the Constant–Jeffreys–Jeffreys and Constant–RATS–Jeffreys priors.

The full conditional posteriors of $(W, \Phi, \Sigma, \alpha)$ are as follows

- (a) $W_t | \Phi, \Sigma, \alpha, Y_{-P} \sim C(y_t, \theta_1) N_q^0(v^*, \Lambda^*), t = P, \dots, n$, where, v^* and Λ^* are given by part (c) in Theorem 3.1.
- (b) The conditional distribution of Φ given $(\Sigma, \alpha, Y_{-P}, W_{-P})$ is given by

$$\pi(\Phi | \Sigma, \alpha, Y_{-P}, W_{-P}) \propto \text{etr} \left\{ -\frac{1}{2} \left[\Lambda \Sigma^{-1} (\Phi_0(Y_{-P} - Z_{-P}\Phi))' (Y_{-P} - Z_{-P}\Phi) \Phi_0' - (\Delta^{-\frac{1}{2}} (\Phi_0(Y_{-P} - Z_{-P}\Phi))' W_{-P} \alpha + \alpha' W_{-P}' (Y_{-P} - Z_{-P}\Phi) \Phi_0') \right] \right\}. \tag{3.14}$$

- (c) The conditional density of Σ given $(\Phi, \alpha, Y_{-P}, W_{-P})$ for CJJ priors is given by

$$\pi(\Sigma | \Phi, \alpha, Y_{-P}, W_{-P}) \propto |\Sigma|^{-\frac{n^*+p+1}{2}} \text{etr} \left\{ -\frac{1}{2} \left[\Sigma^{-1} (\Phi_0(Y_{-P} - Z_{-P}\Phi))' (Y_{-P} - Z_{-P}\Phi) \Phi_0' + \Delta^{-1} \alpha^2 (\Phi_0(Y_{-P} - Z_{-P}\Phi))' (Y_{-P} - Z_{-P}\Phi) \Phi_0' - \Delta^{-\frac{1}{2}} (\Phi_0(Y_{-P} - Z_{-P}\Phi))' W_{-P} \alpha + \alpha' W_{-P}' (Y_{-P} - Z_{-P}\Phi) \Phi_0' \right] \right\}. \tag{3.15}$$

- (d) The conditional density of α given $(\Phi, \Sigma, Y_{-P}, W_{-P})$ is given by

$$\pi(\alpha | \Phi, \Sigma, Y_{-P}, W_{-P}) \propto \text{etr} \left\{ -\frac{1}{2} \left[\alpha^2 \Phi_0(Y_{-P} - Z_{-P}\Phi)' (Y_{-P} - Z_{-P}\Phi) \Phi_0' \Delta^{-1} - \Delta^{-\frac{1}{2}} (\Phi_0(Y_{-P} - Z_{-P}\Phi))' W_{-P} + W_{-P}' (Y_{-P} - Z_{-P}\Phi) \Phi_0' \alpha \right] \right\} \times \prod_{j=1}^T \sqrt{\frac{\pi}{2}} \left(1 + \frac{2\alpha_j^2}{\pi^2} \right)^{-\frac{3}{2}}. \tag{3.16}$$

In this study, the Gibbs sampling Markov chain Monte Carlo (MCMC) methods were applied for sample from the posteriors. The conditional posterior density of (Φ, Σ, α) are not available in closed form. Thus, a MC algorithm was developed for sample W_{-P} directly from the posterior distribution and sample from the conditional distribution of (Φ, Σ, α) , adopting a hit-and-run algorithm. For this, the one-to-one transformation $\Sigma^* = \log(\Sigma)$, or $\Sigma = \exp(\Sigma^*)$ was considered. It can be shown that

the conditional posterior density of Σ^* for CJJ prior given $(\alpha, \Phi, \Sigma, Y_{-P}, W_{-P})$ is then

$$\begin{aligned} \pi(\Sigma^* | \Phi, \alpha, Y_{-P}, W_{-P}) \propto & \text{etr} \left\{ - (n^* + p - 1/2) D^* \right. \\ & + \frac{1}{2} \left[\exp(\Sigma^*)^{-1} \left(\Phi_0(Y_{-P} - Z_{-P}\Phi)'(Y_{-P} - Z_{-P}\Phi)\Phi_0' \right) \right. \\ & + \Delta_{\exp(\Sigma^*)}^{-1} \alpha^2 \left(\Phi_0(Y_{-P} - Z_{-P}\Phi)'(Y_{-P} - Z_{-P}\Phi)\Phi_0' \right) \\ & \left. \left. - \Delta_{\exp(\Sigma^*)}^{-\frac{1}{2}} \left(\Phi_0(Y_{-P} - Z_{-P}\Phi)' W_{-P} \alpha + \alpha W_{-P}'(Y_{-P} - Z_{-P}\Phi)\Phi_0' \right) \right] \right\}, \end{aligned} \tag{3.17}$$

where $\Sigma^* = \mathbf{O} D^* \mathbf{O}'$, \mathbf{O} is $T \times T$ orthogonal matrix, and $D^* = \text{diag}(d_1^*, \dots, d_T^*)$, satisfies $d_1^* \geq d_2^* \geq \dots \geq d_T^*$. Assume we have a Gibbs sample $(W_k, \Phi_k, \Sigma_k, \alpha_k)$ from the previous cycle. At cycle $k + 1$, we then use our proposed algorithm (CJJ) for sample $(W_{k+1}, \Phi_{k+1}, \Sigma_{k+1}, \alpha_{k+1})$. Details and steps of the CJJ algorithm can be found in “Appendix B”.

Finally, the Bayes estimators depend on the loss function that is characterized as follows:

1. The following three loss functions for Σ were considered

$$L_{\Sigma}^{(1)}(\hat{\Sigma}, \Sigma) = \text{tr}(\hat{\Sigma}^{-1} \Sigma) - \log |\hat{\Sigma}^{-1} \Sigma| - T, \tag{3.18}$$

$$L_{\Sigma}^{(2)}(\hat{\Sigma}, \Sigma) = \text{tr}(\hat{\Sigma}^{-1} \Sigma - \mathbf{I})^2, \tag{3.19}$$

and

$$L_{\Sigma}^{(3)}(\hat{\Sigma}, \Sigma) = \text{tr}(\hat{\Sigma} \Sigma^{-1}) - \log |\hat{\Sigma} \Sigma^{-1}| - T. \tag{3.20}$$

2. The following two well-known loss function for Φ were also considered

$$L_{\Phi}^{(1)}(\hat{\Phi}, \Phi) = \text{tr} \left\{ (\hat{\Phi} - \Phi)' W (\hat{\Phi} - \Phi) \right\}, \tag{3.21}$$

where, W is a constant weighting matrix, and

$$L_{\Phi}^{(2)}(\hat{\Phi}, \Phi) = \sum_{i=1}^{PT} \sum_{j=1}^T \left[\exp \{ a_{ij} (\hat{\phi}_{ij} - \phi_{ij}) \} - a_{ij} (\hat{\phi}_{ij} - \phi_{ij}) - 1 \right], \tag{3.22}$$

where, a_{ij} is a given constant.

3. The most common loss for α are the quadratic loss

$$L_{\alpha}^{(1)}(\hat{\alpha}, \alpha) = \text{trace} \left\{ (\hat{\alpha} - \alpha)' A (\hat{\alpha} - \alpha) \right\}, \tag{3.23}$$

where, A is a constant weighting matrix. If the weighting matrix A is the identity matrix, then the loss of $L_{\alpha}^{(1)}$ is simply the sum of squared errors of all elements of α , $\sum_{i=1}^T (\hat{\alpha}_i - \alpha_i)^2$.

The Bayesian estimates of Σ , α and Φ under the above loss functions can be derived separately from minimizing expected posterior loss functions regarding (Φ, Σ, α) provided the minimum is finite. In the following well-known facts, the results are summarized.

- (a) The generalized Bayesian estimators of Σ under loss functions $L_{\Sigma}^{(1)}$, $L_{\Sigma}^{(2)}$ and $L_{\Sigma}^{(3)}$ are given by

$$\hat{\Sigma}_1 = E(\Sigma|Y_{-P}), \quad (3.24)$$

$$\text{vec}(\hat{\Sigma}_2) = \left[E\left\{ \left(\Sigma^{-1} \otimes \Sigma^{-1} \right) | Y_{-P} \right\} \right]^{-1} \text{vec} \left\{ E\left(\Sigma^{-1} | Y_{-P} \right) \right\}, \quad (3.25)$$

and

$$\hat{\Sigma}_3 = \{E(\Sigma^{-1}|Y_{-P})\}^{-1}, \quad (3.26)$$

respectively.

- (b) The generalized Bayesian estimators of Φ under loss functions $L_{\Phi}^{(1)}$ and $L_{\Phi}^{(2)}$ are given by

$$\hat{\Phi}_1 = E(\Phi|Y_{-P}), \quad (3.27)$$

and

$$\hat{\phi}_{ij} = -\frac{1}{a_{ij}} \log \left[E\left\{ \exp(-a_{ij}\phi_{ij}) | Y_{-P} \right\} \right], \quad (3.28)$$

For $i = 1, \dots, PT$, $j = 1, \dots, T$, where $\hat{\phi}_{ij}$ is the (i, j) -th element of the Bayesian estimator $\hat{\Phi}_2$, respectively.

- (c) Under the loss $L_{\alpha}^{(1)}$, the generalized Bayesian estimator of α is given by

$$\hat{\alpha}_1 = E(\alpha|Y_{-P}). \quad (3.29)$$

3.3 MAP estimation

In this part, the ECM algorithm was applied to obtain the MAP estimates of the parameters (Φ, Σ, α) . Constant–Jeffreys–Jeffreys and Constant–RATS–Jeffreys priors were used for $\theta = (\Phi, \Sigma, \alpha)$ given by (3.12) and (3.13), respectively. According to the results obtained in the previous sections, the posterior function based on data to be maximized is $\pi(\Phi, \Sigma, \alpha|Y) \propto CL(\Phi, \Sigma, \alpha|Y)\pi(\Phi, \Sigma, \alpha)$, where $CL(\Phi, \Sigma, \alpha|Y)$ and $\pi(\Phi, \Sigma, \alpha)$ are given by (3.6) and (3.12 and 3.13), respectively.

E-step Conditional expectation used in the ECM algorithm for finding MAP estimates of parameters θ is

$$Q(\theta, \theta^{(k)}) = E_{\theta^{(k)}}[\pi(\theta|Y, W)|Y] = E_{\theta^{(k)}}[Cl(\theta|Y, W)|Y] + \log(\pi(\theta)). \tag{3.30}$$

where $E_{\theta^{(k)}}[l(\theta|Y, W)|Y]$ is given by (3.8).

CM-steps Updating of parameters in the CM-steps obviously, will be done in the following parts:

- CM-step 1 Update Φ by (3.9)
- CM-step 2 Update Σ by solving the following equation

$$\begin{aligned} \Sigma^{(k+1)} - & \frac{\text{diag}\left[\Phi_0^{(k+1)}(Y_{-P} - Z_{-P}\Phi^{(k+1)})'(Y_{-P} - Z_{-P}\Phi^{(k+1)})\Phi_0^{(k+1)'}\right]}{n^* + T + 1} \\ & - \Sigma^{2(k+1)}\left(\frac{\delta\Delta^{-1}}{\delta\Sigma}\right)^{(k+1)} \\ & \times \frac{\text{diag}\left[\Phi_0^{(k+1)}(Y_{-P} - Z_{-P}\Phi^{(k+1)})'(Y_{-P} - Z_{-P}\Phi^{(k+1)})\Phi_0^{(k+1)'}\alpha^{(k)2}\right]}{n^* + T + 1} \\ & + \Sigma^{2(k+1)}\left(\frac{\delta\Delta^{-\frac{1}{2}}}{\delta\Sigma}\right)^{(k+1)} \frac{\text{diag}\left[\Phi_0^{(k+1)}(Y_{-P} - Z_{-P}\Phi^{(k+1)})'M^{(k)}\alpha^{(k)}\right]}{n^* + T + 1} = 0. \end{aligned} \tag{3.31}$$

- CM-step 3 Update α by solving the following equation

$$\begin{aligned} \alpha_j^3 - & \frac{\left[\Delta^{(k+1)-\frac{1}{2}}\Phi_0^{(k+1)}(Y_{-P} - Z_{-P}\Phi^{(k+1)})'M^{(k)}\right]_{j,j}}{\left[\Phi_0^{(k+1)}(Y_{-P} - Z_{-P}\Phi^{(k+1)})'(Y_{-P} - Z_{-P}\Phi^{(k+1)})\Phi_0^{(k+1)'}\Delta^{(k+1)-1}\right]_{j,j}}\alpha_j^2 \\ & + \left\{ \frac{\pi^2}{8} + \frac{3}{\left[\Phi_0^{(k+1)}(Y_{-P} - Z_{-P}\Phi^{(k+1)})'(Y_{-P} - Z_{-P}\Phi^{(k+1)})\Phi_0^{(k+1)'}\Delta^{(k+1)-1}\right]_{j,j}} \right\} \alpha_j \\ & - \frac{\pi^2\left[\Delta^{(k+1)-\frac{1}{2}}\Phi_0^{(k+1)}(Y_{-P} - Z_{-P}\Phi^{(k+1)})'M^{(k)}\right]_{j,j}}{8\left[\Phi_0^{(k+1)}(Y_{-P} - Z_{-P}\Phi^{(k+1)})'(Y_{-P} - Z_{-P}\Phi^{(k+1)})\Phi_0^{(k+1)'}\Delta^{(k+1)-1}\right]_{j,j}} = 0, \end{aligned} \tag{3.32}$$

$j = 1, \dots, T.$

4 Simulation studies

This section includes the numerical results of two simulation studies for different models. In the first part, performance of the ML, MAP and Bayes estimators in the PAR models with closed skew normal innovation are compared by using the frequentist risks under a loss function L , that is, $E_{Y|\theta}L(\theta, \hat{\theta})$ which is estimated by $\frac{1}{n} \sum_{i=1}^n L$

Table 1 True parameters model

	Coefficient model	Scale innovation	Skewness parameter
Model 1	(1.5,0.4,- 2,0.9)	(0.7,0.06,1, 2)	(0.2,- 1.5,- 0.5,1)
Model 2	(- 0.3,- 1.4,0.8,- 2)	(0.5,1,0.04,2)	(- 0.1,2,- 1.2,0.6)
Model 3	(1.5,0.2,2,1)	(0.5,1,0.01,2)	(- 1.0,5,1.5,- 0.5)

$(\theta, \hat{\theta}_i)$, where $\hat{\theta}_i$ is the estimate θ in the i -th sample ($i = 1, \dots, n$). In the second part, some numerical results are provided to show the consistency and asymptotic distribution of the estimators.

4.1 Simulation 1

In this part, a PAR(1) model with $T = 4$ and closed skew normal innovations for a sample of $N = 200$ observations was considered. In the study, it is supposed that $\Delta_t = \sigma_t^2$. 200 samples were generated from three PAR(1) models with true different parameters shown in Table 1. To compare the ML, MAP and Bayesian estimates under priors and loss functions proposed in Sect. 3, 10,000 MCMC iterations were run after 500 burn-in cycles. The weighting matrix in the first loss function for Φ (3.21) and α (3.23) is the identity matrix. The parameter a in the second loss function for Φ (3.22) for all elements is also assumed to be -3 .

Tables 2, 3 and 4 show the frequentist average losses and standard deviations (in parentheses) of Σ , Φ and α , respectively, under different loss functions and Constant–Jeffreys–Jeffreys (CJJ) and Constant–RATS–Jeffreys (CAJ) priors for three different considered models. For example, in these tables $\hat{\Sigma}_{1CJJ}$ represents the estimator of Σ under loss $L_{\Sigma}^{(1)}$ and Constant–Jeffreys–Jeffreys prior.

Tables 5, 6 and 7 provide the means and standard deviations (in parentheses) of the estimated coefficient, scale and skewness parameters of three different PAR models with skew-normal innovation, respectively, under different loss functions and CJJ or CAJ priors. In these tables, the best results are highlighted in bold and also, the MAP1 and MAP2 denote the maximum a posteriori estimates form CJJ and CAJ priors, respectively.

In general, it is observed that the Bayes estimates provide the best performance with the lowest frequentist average losses as compared to the ML and MAP estimates in all parameters. The estimates of Σ , Φ and α under Constant–Jeffreys–Jeffreys (CJJ) prior also performed better than Constant–RATE–Jeffreys (CAJ) prior in almost all the cases.

According to Table 2, the estimates of Σ under the loss functions $L_{\Sigma}^{(1)}$ and $L_{\Sigma}^{(3)}$ and CJJ prior provided the best result. Table 6 also shows that, in most cases, the estimates of the scale parameters under loss functions $L_{\Sigma}^{(1)}$ and CJJ prior are very close to the true parameters and has the lowest biases.

The frequentist average losses of Φ under CJJ and CAJ priors with loss function $L_{\Phi}^{(1)}$ have almost similar performances; however, it seems that the CAJ prior performs better than CJJ prior with LINEX loss function (Table 3). In all the cases and methods,

Table 2 Frequentist average losses (with standard deviations in parentheses) of $\hat{\Sigma}$ with closed skew-normal innovation

	Model 1			Model 2			Model 3		
	$L_{\Sigma}^{(1)}$	$L_{\Sigma}^{(2)}$	$L_{\Sigma}^{(3)}$	$L_{\Sigma}^{(1)}$	$L_{\Sigma}^{(2)}$	$L_{\Sigma}^{(3)}$	$L_{\Sigma}^{(1)}$	$L_{\Sigma}^{(2)}$	$L_{\Sigma}^{(3)}$
$\hat{\Sigma}_{ML}$.262(.277)	3.392(8.086)	.451(.692)	.142(.139)	.728(1.917)	.181(.252)	.135(.138)	.750(2.781)	.177(.337)
$\hat{\Sigma}_{MAP1}$.240(.240)	2.264(5.713)	.371(.565)	.147(.143)	.478(1.440)	.163(.220)	.145(.141)	.533(2.069)	.163(.297)
$\hat{\Sigma}_{MAP2}$.247(.214)	1.502(3.863)	.323(.457)	.179(.158)	.435(.887)	.166(.183)	.184(.158)	.521(1.456)	.173(.254)
$\hat{\Sigma}_{1CJJ}$.106 (.086)	.202 (.281)	.091 (.063)	.090 (.082)	.183 (.266)	.084 (.066)	.102 (.078)	.230 (.454)	.102 (.097)
$\hat{\Sigma}_{2CJJ}$.146(.118)	.356(.392)	.114(.078)	.114(.108)	.234(.288)	.093(.072)	.124(.095)	.249(.293)	.103(.073)
$\hat{\Sigma}_{3CJJ}$.123(.101)	.261(.328)	.099 (.069)	.098 (.093)	.181 (.239)	.085 (.066)	.108 (.083)	.202 (.308)	.097 (.076)
$\hat{\Sigma}_{1CAJ}$.158(.122)	.405(.413)	.122(.080)	.123(.117)	.268(.321)	.099(.077)	.133(.100)	.265(.298)	.109(.075)
$\hat{\Sigma}_{2CAJ}$.227(.159)	.692(.535)	.167(.101)	.177(.149)	.500(.451)	.134(.094)	.189(.132)	.507(.431)	.143(.087)
$\hat{\Sigma}_{3CAJ}$.190(.141)	.540(.477)	.142(.090)	.147(.133)	.368(.388)	.114(.085)	.158(.115)	.371(.364)	.123(.079)

Table 3 Frequentist average losses (with standard deviations in parentheses) of Φ with closed skew-normal innovation

	Model 1		Model 2		Model 3	
	$L_{\Phi}^{(1)}$	$L_{\Phi}^{(2)}$	$L_{\Phi}^{(1)}$	$L_{\Phi}^{(2)}$	$L_{\Phi}^{(1)}$	$L_{\Phi}^{(2)}$
$\hat{\Phi}_{ML}$.0596(.055)	.261(.242)	.261(.315)	1.810(4.753)	.065(.084)	.367(.638)
$\hat{\Phi}_{MAP1}$.0594(.055)	.261(.245)	.261(.318)	1.806(4.754)	.064(.084)	.365(.633)
$\hat{\Phi}_{MAP2}$.0581(.053)	.262(.250)	.260(.308)	1.699(3.987)	.064(.083)	.351(.605)
$\hat{\Phi}_{1CJJ}$.0103(.015)	.047(.070)	.069(.101)	.393(1.173)	.031(.089)	.242(1.304)
$\hat{\Phi}_{2CJJ}$.0107(.016)	.048(.072)	.078(.112)	.356(.832)	.029(.063)	.171(.646)
$\hat{\Phi}_{1CAJ}$.0095 (.012)	.043 (.058)	.061 (.071)	.312(.573)	.027 (.052)	.169(.611)
$\hat{\Phi}_{2CAJ}$.0098 (.013)	.044 (.059)	.068 (.084)	.293 (.440)	.028 (.058)	.143 (.468)

Table 4 Frequentist average losses (with standard deviations in parentheses) of α with closed skew-normal innovation

	Model 1	Model 2	Model 3
	$L_{\alpha}^{(1)}$	$L_{\alpha}^{(1)}$	$L_{\alpha}^{(1)}$
$\hat{\alpha}_{ML}$.530(.489)	.729(.729)	.725(.852)
$\hat{\alpha}_{MAP1}$.644(.785)	1.075(1.179)	.963(1.161)
$\hat{\alpha}_{MAP2}$.592(.600)	.929(.997)	.851(1.032)
$\hat{\alpha}_{1CJJ}$.388(.473)	.752(2.264)	.522 (.655)
$\hat{\alpha}_{1CAJ}$.369 (.509)	.710 (1.498)	.548(.773)

the estimate of the coefficient parameters showed the same bias but the results for the estimate of CAJ prior with $L_{\Phi}^{(1)}$ perform were better than that of CJJ prior with LINEX loss (Table 5).

The results shown in Tables 4 and 7, also suggest that the estimates of skewness parameter α under CAJ prior has lower bias and lower frequentist average losses than CJJ prior.

In summary, the Bayes approach showed the best results for estimation of unknown parameters of the PAR models with closed skew normal innovation in almost all the cases.

4.2 Simulation 2

In this part, we conduct Monte Carlo simulation to evaluate the accuracy and asymptotic properties of the various estimation methods. We generate samples from the CSN PAR(1) model with $T = 2$ and true parameters shown in Table 8 for small ($n = 60$), moderate ($n = 100$) and large ($n = 140, 200$) sample sizes. In this study, in addition of ML estimate, we also consider the posterior mean for Bayes estimate and MAP estimate under CJJ and CAJ prior distributions.

Table 5 The means and standard deviations (in parentheses) of the estimated coefficient parameter of the PAR model with closed skew-normal innovation

	Model 1				Model 2				Model 3			
	1.5	.4	-2	.9	-3	-1.4	.8	-2	1.5	.2	2	1
$\hat{\phi}_{ML}$	1.490 (.097)	.398 (.015)	-1.931 (.177)	.873 (.114)	-.307 (.076)	-1.405 (.263)	.800 (.048)	-1.981 (.429)	1.505 (.065)	.212 (.065)	2.001 (.038)	.952 (.229)
$\hat{\phi}_{MAP1}$	1.490 (.097)	.398 (.016)	-1.932 (.177)	.873 (.115)	-.307 (.076)	-1.403 (.266)	.801 (.048)	-1.979 (.428)	1.504 (.065)	.212 (.066)	2.001 (.038)	.953 (.229)
$\hat{\phi}_{MAP2}$	1.493 (.096)	.397 (.015)	-1.944 (.177)	.862 (.114)	-.308 (.077)	-1.415 (.265)	.799 (.048)	-1.964 (.425)	1.508 (.065)	.209 (.065)	2.000 (.038)	.962 (.229)
$\hat{\phi}_{1CJJ}$	1.484 (.035)	.398 (.004)	-1.961 (.056)	.856 (.046)	-.303 (.038)	-1.402 (.118)	.803 (.021)	- 2.006 (.231)	1.506 (.035)	.204 (.044)	2.001 (.015)	.972 (.165)
$\hat{\phi}_{2CJJ}$	1.485 (.035)	.398 (.004)	-1.956 (.060)	.858 (.045)	-. 302 (.038)	-1.384 (.122)	.803 (.021)	-1.947 (.243)	1.507 (.035)	.205 (.044)	2.002 (.016)	1.000 (.159)
$\hat{\phi}_{1CAJ}$	1.485 (.034)	.398 (.004)	-1.962 (.052)	.856 (.044)	-.303 (.039)	- 1.399 (.124)	.802 (.021)	-2.012 (.209)	1.503 (.034)	.202 (.046)	2.001 (.016)	.973 (.158)
$\hat{\phi}_{2CAJ}$	1.486 (.034)	.398 (.004)	-1.958 (.055)	.857 (.043)	-. 302 (.040)	-1.383 (.127)	.802 (.021)	-1.957 (.218)	1.504 (.034)	.203 (.046)	2.002 (.016)	.993 (.154)

Table 6 The means and standard deviations (in parentheses) of the estimated scale parameter of the PAR model with closed skew-normal innovation

	Model 1			Model 2			Model 3					
	.7	.06	1	2	.5	1	.04	2	.5	1	.01	2
	$\hat{\sigma}_{ML}^2$.921 (.466)	.077 (.046)	1.332 (.700)	2.32 (.717)	.527 (.123)	1.067 (.284)	.047 (.019)	2.326 (.707)	.519 (.125)	1.052 (.285)	.0116 (.0058)
$\hat{\sigma}_{MAP1}^2$.839 (.424)	.072 (.041)	1.218 (.630)	2.153(.657)	.480 (.112)	1.004 (.268)	.044 (.018)	2.143 (.658)	.480 (.119)	.967 (.270)	.0109 (.0054)	2.079 (.532)
$\hat{\sigma}_{MAP2}^2$.766 (.382)	.066 (.039)	1.103 (.552)	1.939(.560)	.441 (.103)	.917 (.247)	.040 (.016)	1.951 (.579)	.438 (.108)	.883 (.237)	.0100 (.0050)	1.902 (.478)
$\hat{\sigma}_{1CJJ}^2$.690 (.152)	.057 (.011)	.969 (.206)	1.874(.395)	.491 (.097)	.977 (.189)	.041 (.009)	2.075 (.444)	.483 (.101)	1.004 (.215)	.0100 (.0021)	2.174 (.586)
$\hat{\sigma}_{2CJJ}^2$.635 (.137)	.053 (.010)	.893 (.184)	1.732(.364)	.453 (.089)	.905 (.175)	.038 (.008)	1.892 (.388)	.446 (.094)	.927 (.194)	.0092 (.0019)	1.938 (.443)
$\hat{\sigma}_{3CJJ}^2$.661 (.143)	.055 (.010)	.928 (.192)	1.800(.378)	.471 (.093)	.939 (.182)	.039 (.008)	1.979 (.413)	.464 (.097)	.964 (.204)	.0096 (.0020)	2.043 (.491)
$\hat{\sigma}_{1CAJ}^2$.626 (.133)	.052 (.010)	.882 (.192)	1.704(.349)	.446 (.086)	.891 (.174)	.037 (.008)	1.881 (.383)	.438 (.093)	.913 (.192)	.0092 (.0021)	1.923 (.413)
$\hat{\sigma}_{2CAJ}^2$.581 (.124)	.048 (.009)	.817 (.171)	1.588(.327)	.414 (.081)	.831 (.162)	.034 (.007)	1.726 (.346)	.408 (.087)	.849 (.176)	.0084 (.0019)	1.752 (.361)
$\hat{\sigma}_{3CAJ}^2$.602 (.128)	.050 (.010)	.847 (.178)	1.643(.337)	.430 (.084)	.860 (.168)	.036 (.007)	1.799 (.362)	.423 (.090)	.880 (.184)	.0088 (.0020)	1.832 (.383)

Table 7 The means and standard deviations (in parentheses) of the estimated skewness parameter of the PAR model with closed skew-normal innovation

	Model 1					Model 2					Model 3				
	.2	-1.5	-.5	1	-1.229	-.1	2	-1.229	.6	-1	.5	1.5	-.5		
$\hat{\alpha}_{ML}$.189 (.197)	-1.360 (.573)	-.467 (.198)	.845 (.285)	1.923 (.618)	-.088 (.194)	1.923 (.618)	-.1.229 (.523)	.601 (.312)	-1.070 (.431)	.527 (.264)	1.474 (.604)	-.460 (.315)		
$\hat{\alpha}_{MAP1}$.195 (.206)	-1.472 (.670)	-.485 (.210)	.892 (.314)	2.124 (.729)	-.091 (.201)	2.124 (.729)	-1.332 (.601)	.630 (.338)	-1.150 (.491)	.550 (.284)	1.614 (.706)	-.481 (.332)		
$\hat{\alpha}_{MAP2}$.184 (.195)	-1.393 (.625)	-.462 (.201)	.844 (.298)	2.004 (.671)	-.087 (.192)	2.004 (.671)	-1.259 (.584)	.594 (.321)	-1.096 (.469)	.524 (.275)	1.545 (.667)	-.463 (.319)		
$\hat{\alpha}_{1CJJ}$.204 (.203)	-1.532 (.489)	-.513 (.198)	.899 (.243)	1.988 (.704)	-.086 (.189)	1.988 (.704)	-1.269 (.395)	.618 (.253)	-1.042 (.389)	.506 (.231)	1.467 (.490)	-.458 (.275)		
$\hat{\alpha}_{1CAJ}$.194 (.191)	-1.471 (.472)	-.491 (.190)	.860 (.234)	1.908 (.683)	-.082 (.184)	1.908 (.683)	-1.213 (.389)	.593 (.229)	-0.999 (.372)	.488 (.235)	1.422 (.534)	-.438 (.250)		

Table 8 True model parameters

Coefficient model	Scale innovation	Skewness parameter
(1.5, - 0.2)	(0.5,2)	(0.5,- 1.5)

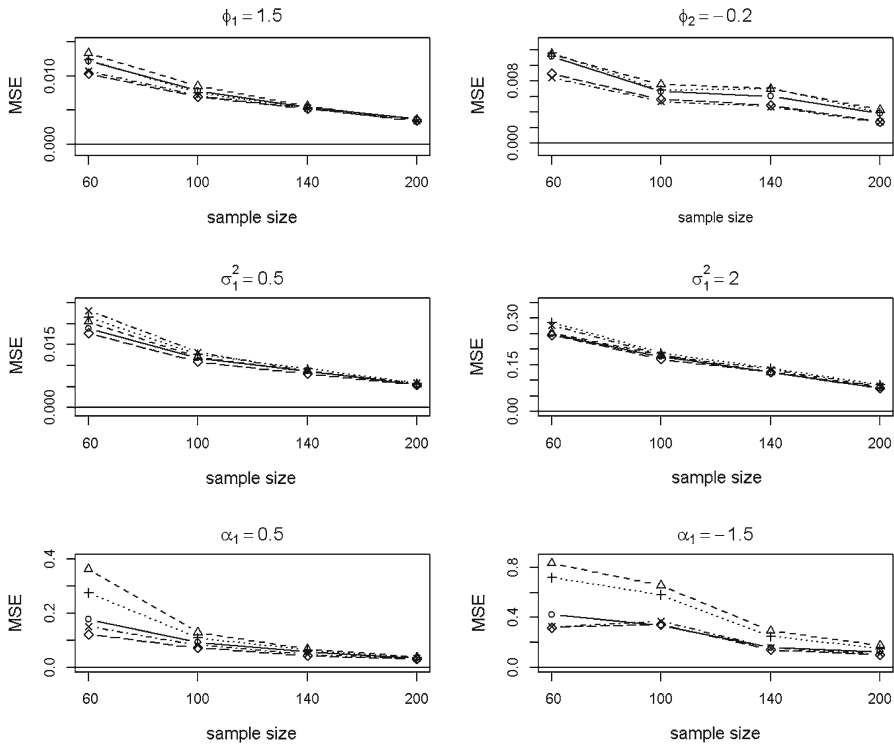


Fig. 1 The MSE of the proposed estimators for different sample size, where the piecewise linear functions with nodes indicated by \circ , Δ , $+$, \times and \diamond illustrate the MSE of the ML, MAP-CJJ, MAP-CAJ, Bayesian CJJ and Bayesian CAJ estimators, respectively

We compute the mean square error (MSE) and deficiency (Def) criteria for different sample sizes. The Def criterion is an essential measure for comparing the joint efficiencies of the different methods used for estimating a set of parameters (here, (Φ, Σ, α)). It is defined as the sum of the MSE values of the estimators of the unknown parameters (Gebizlioglu et al. (2011)) and is given by

$$Def(\Phi, \Sigma, \alpha) = \sum_{k=1}^T MSE(\hat{\phi}_k) + MSE(\hat{\sigma}_k^2) + MSE(\hat{\alpha}_k),$$

where $MSE(\hat{\theta}) = Var(\hat{\theta}) + (Bias(\hat{\theta}))^2$.

The result based on 200 Monte Carlo runs are provided in Figs. 1, 2, 3, 4 and 5. Figure 1 and 2 show the MSE and Def criteria of the proposed estimates. Figures 3,

Fig. 2 The Def criteria of the proposed estimators for different sample size

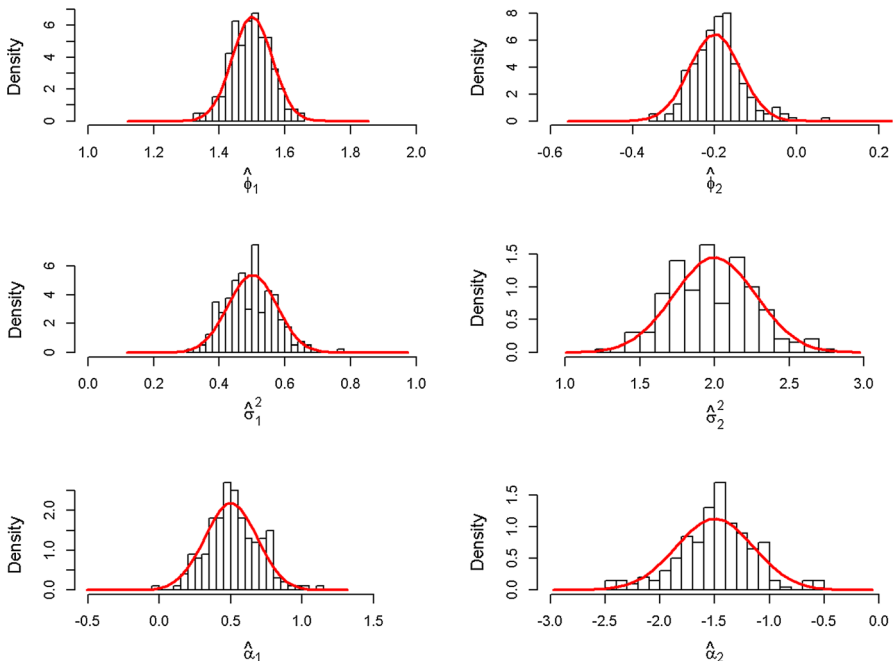
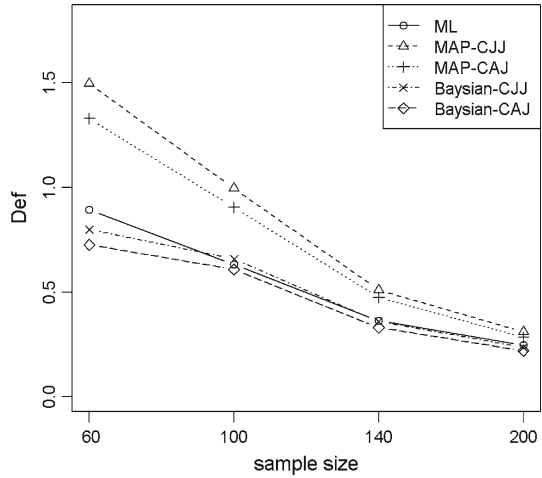


Fig. 3 Histograms of the ML estimated of parameters for $N = 200$

4, 5 show the theoretical Gaussian density super-imposed over the histograms of the proposed estimates when $N = 200$.

Figure 1 clearly illustrates the consistency properties of the estimators. When N is small, all estimation methods suffer from the bias problem, especially for the scale and skewness parameters, however, the ML and Bayesian estimates has the lowest biased in all the parameters. By increasing the sample size, the MSE of parameters converge

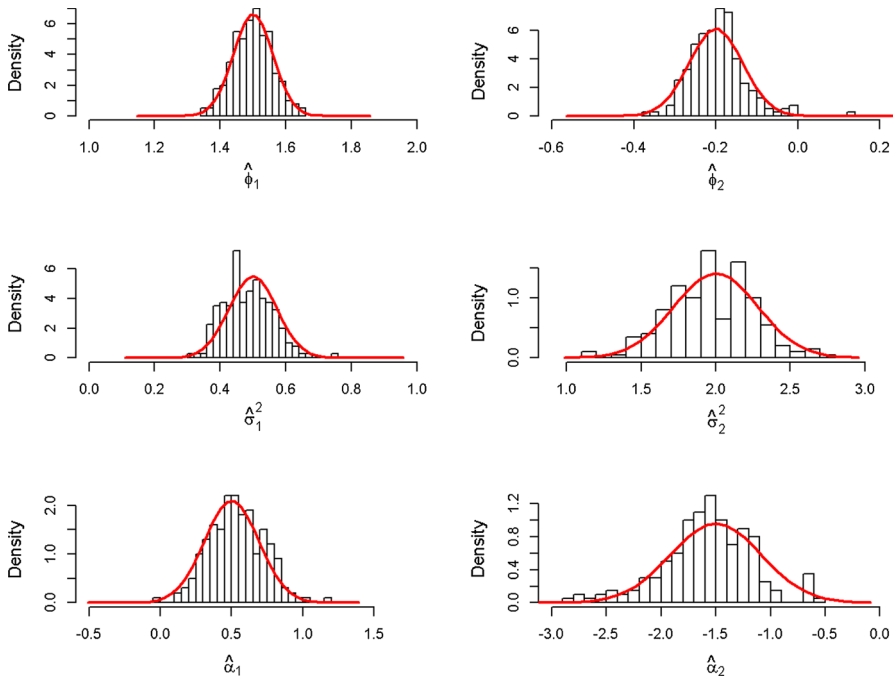


Fig. 4 Histograms of the MAP estimated of parameters for $N = 200$ and CJJ prior distribution

to zero, which indicate the consistency of the parameter estimators. For large N , the Bayesian estimator has the lowest bias in all the parameters.

Figure 2 indicates that as the sample size increases, the Def criteria decreases in all methods and the Bayesian estimators show the best performance among the estimators.

Figures 3, 4 and 5 are demonstrating that the histograms match pretty well to the asymptotic normal distribution, especially for coefficients parameters in all method estimations. The results of the Kolmogorov–Smirnov test (not given here for the reason of space limitation) also show that the asymptotic normality of the proposed estimates. The p value of Kolmogorov–Smirnov normality test for all methods are greater than 0.05.

5 Real data analysis

In this section, as an illustrative case study of the proposed estimation methods, we present an analysis of the quarterly United Kingdom macroeconomic variables. We consider the average quarterly Final Consumption Expenditure in the United Kingdom (FCEUK) from January 1, 1988, to December 31, 2017, from the National Statistical Institute of the UK (<https://www.ons.gov.uk>). This data set consists of 120 data points. The first 104 data points, that is, data from January, 1988 to December, 2013 were used for model building, and the remaining 16 data points, that is, data from January, 2014 to

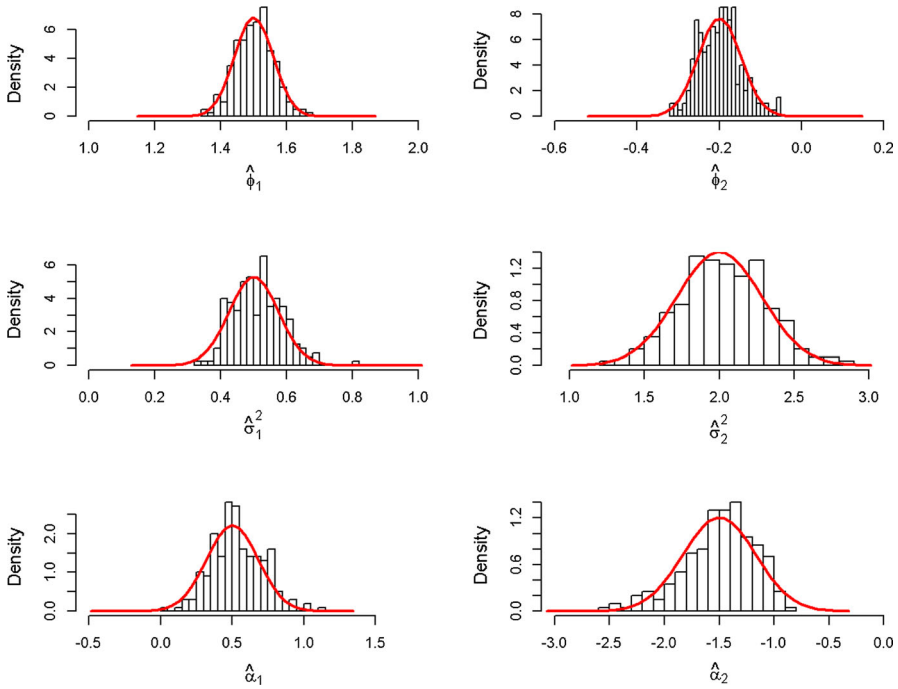


Fig. 5 Histograms of the Bayes estimated of parameters for $N = 200$ and CJJ prior distribution

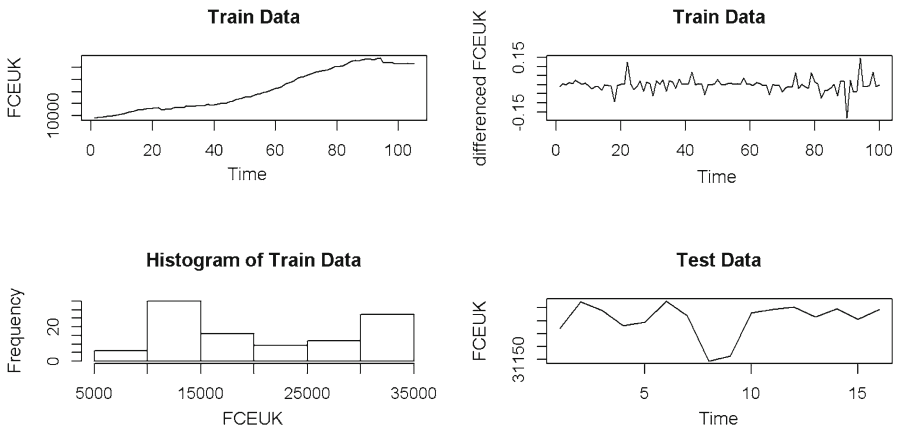


Fig. 6 Quarterly final consumption expenditure in the United kingdom 1988–2017

December, 2017, were used for model validation. After removing the linear trend and seasonality by applying two differencing of order 4 and 1 from the original quarterly data, respectively, some diagnostic methods are applied to determine the presence of periodic correlation. The coherent, incoherent and the measure of fitness (MoF) statistics are good tools for detecting the period of periodically correlated processes, see e.g. Broszkiewicz-Suwaj et al. (2004) and Nematollahi et al. (2017) for more

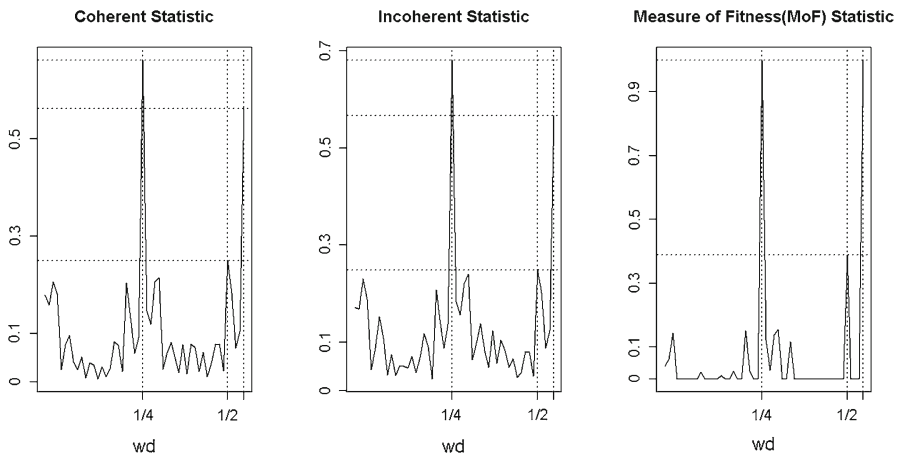


Fig. 7 The test parameters for coherent, incoherent and MoF statistics are $M = 20$, $B = 100$ and $\alpha = 0.01$, $N = 220$

details. These statistics take real values in the interval $[0, 1]$ and due to the symmetry are plotted only in the interval $(0, N/2)$. Peaks at points w_d, w_{2d}, w_{3d} , etc. indicate periodic correlation with the period of length $T = 1/w_d$. Figures 6 and 7 show the descriptive plot of the original time series, Coherent statistic, Incoherent statistic and MoF statistic of differenced data.

The coherent, incoherent and MoF statistics detect a 4-quarter period; peaks appear at frequencies being multiples of $\frac{1}{4}$. Table 9 provide the result of fitting the PAR model with normal and skew normal innovations (Manouchehri and Nematollahi (2019)) and closed skew normal innovations for quarterly FCEUK data. The root mean square prediction error (RMSPE) value, mean absolute prediction error (MAPE) value, relative mean absolute prediction error (RMAPE) value and also, AIC and BIC for validation of the forecasts for fitted PAR model for the present data are summarized in these tables.

The results of modeling and prediction of the PAR model with the normal, skew-normal and closed skew normal innovations based on the ML, MAP and Bayes estimates which are computed in several proposed methods, are also listed.

Figure 7 shows the plots of the individual elements of one step predicted, that is, \hat{X}_{104+j} , for $j = 1, \dots, 12$ in terms of $X_1, \dots, X_{104+1}, \dots, X_{104+j-1}$, assuming that the innovations are distributed according to the normal, skew-normal and closed skew normal laws. The figure also includes the 95% prediction intervals of the one step forecasts which are calculated by $\hat{X}_{104+4t+k} \mp c_{1-\alpha/2} \hat{\sigma}_k$, where, $c_{1-\alpha/2}$ is the $(1 - \alpha/2)$ -th quantile of the normal and skew-normal distributions, and $t = 0, 1, 2$ and $k = 1, 2, 3, 4$ and $\hat{\sigma}_k$ denotes the square root of the variance of one step forecasts.

Results in Table 9 and Fig. 8 suggest that the PAR model with closed skew-normal innovation gives the best results.

Table 9 The MAPE, RMSPE, RMAPE, AIC and BIC values of fitting PAR model with normal, skew-normal and closed skew normal innovation for FCEUK data

	AIC	BIC	MAPE	RMSPE	RMAPE
PAR model (normal)					
ML	- 377.909	- 295.85	369.144	442.346	1.179
CJ11	- 370.518	- 288.459	485.698	572.619	1.551
CJ12	- 377.894	- 295.835	366.102	434.556	1.169
CJ13	- 377.664	- 295.605	413.682	489.147	1.321
CA11	- 377.572	- 295.512	417.622	498.02	1.334
CA12	- 373.897	- 291.838	334.926	403.13	1.069
CA13	- 377.882	- 295.823	369.717	442.807	1.181
CR11	- 372.695	- 290.635	400.304	451.665	1.279
CR12	- 376.752	- 294.692	315.578	355.016	1.008
CR13	- 377.054	- 294.995	351.286	395.183	1.122
PAR model (skew normal)					
ML	- 417.888	- 387.115	78.466	89.639	0.251
CJJ111	- 414.003	- 383.231	75.084	90.252	0.24
CJJ121	- 409.39	- 378.618	79.595	91.68	0.254
CJJ131	- 413.493	- 382.721	76.789	90.162	0.245
PAR model (closed skew normal)					
ML	- 418.792	- 388.02	60.831	74.872	0.194
MAP1	- 416.958	- 386.186	63.979	71.878	0.204
MAP2	- 412.195	- 381.423	63.048	73.086	0.201
CJJ211	- 417.042	- 386.269	45.932	58.23	0.147
CJJ221	- 423.342	- 392.57	57.3	64.305	0.183
CJJ231	- 423.764	- 392.992	50.603	60.125	0.162
CAJ211	- 421.485	- 390.713	50.654	60.503	0.162
CAJ221	- 417.261	- 386.489	62.287	68.7	0.199
CJJ231	- 421.785	- 391.013	57.452	64.499	0.184

6 Concluding remarks

In this paper, it is shown that application of the multivariate CSN for the distribution of T -dimensional innovation associated with a second order PAR(p) time series of period T , is more appropriate with respect to the multivariate SN. The multivariate CSN has some useful properties similar to those of the Gaussian distribution, which make them analytically tractable, relatively simple to fit observed data with lack of symmetry but with shape of the empirical distribution like to normal distribution. The Maximum likelihood, Maximum a posteriori and Bayesian estimates of the model parameters are examined and some technical recommendations are recommended to the users.

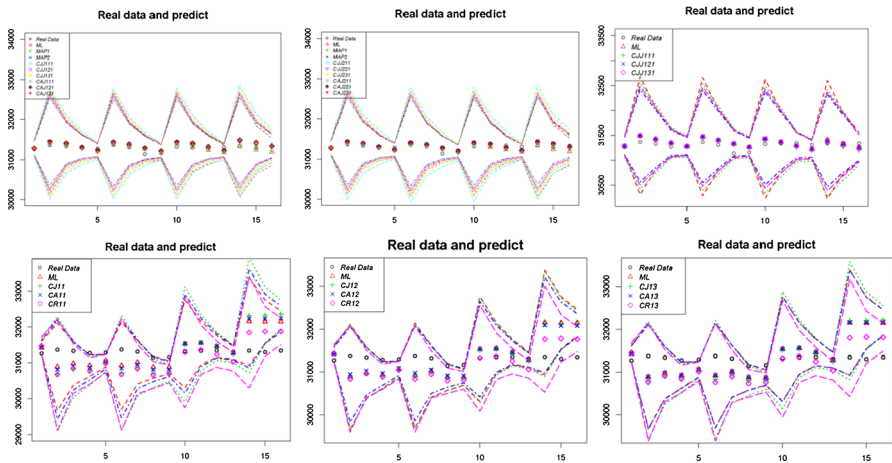


Fig. 8 The one step prediction and the 95% prediction intervals of fitting PAR model with normal, skew-normal and closed skew normal innovation for FCEUK data

Appendix A: Proof of Theorem 3.1

In order to prove of Theorem 3.1, we need some preliminary definitions and properties.

Definition 1 (Truncated multivariate normal). If $W \sim N_q(\mu, \Sigma)$ and $U = \begin{cases} W & \text{if } W \geq c \\ \mathbf{0} & \text{if } W < c \end{cases}$ where $W \geq c$ means $W_j \geq c_j, j = 1, \dots, q$, then the density function of U is:

$$f(u; \mu, \Sigma, c) = \Phi_q^{-1}(0; c - \mu, \Sigma) \varphi_q(u; \mu, \Sigma), u \geq c.$$

U is truncated multivariate normal denote by $U \sim N_q^c(\mu, \Sigma)$.

Property 1 If $U \sim N_q^c(\mu, \Sigma)$ then the moment generating function of U is given by

$$M_U(t) = \Phi_q^{-1}(0; c - \mu, \Sigma) e^{t' \mu + \frac{1}{2} t' \Sigma t} \Phi_q(\Sigma t; c - \mu, \Sigma), t \in \mathbb{R}^q.$$

Property 2 If $Z \sim CSN_{p,q}(\mu, \Sigma, \Gamma, v, \Delta)$, then the moment generative function of Z is given in González-Farías et al. (2004) as

$$M_Z(s) = \frac{\Phi_q(\Gamma \Sigma s; v, \Delta + \Gamma \Sigma \Gamma')}{\Phi_q(0; v, \Delta + \Gamma \Sigma \Gamma')} \exp\left(s' \mu + \frac{1}{2} s' \Sigma s\right), s \in \mathbb{R}^p.$$

Proof of Theorem 3.1 The result of part (a) is proved by using the uniqueness property of the moment generating functions. Note that

$$\begin{aligned}
 M_{Y_t}(s) &= M_{V_t}(s)M_{W_t}(D's) = e^{s' \left(\sum_{j=1}^P \Phi_j y_{t-j} \right) + \frac{1}{2} s' G s} \\
 &\quad \times \Phi_T^{-1}(\mathbf{0}; \mathbf{0}, \Lambda) e^{\frac{1}{2} s' D \Lambda D' s} \Phi_T(\Lambda D' s; \mathbf{0}, \Lambda) \\
 &= \frac{\Phi_T(\Lambda D' s; \mathbf{0}, \Lambda)}{\Phi_T(\mathbf{0}; \mathbf{0}, \Lambda)} e^{s' \left(\sum_{j=1}^P \Phi_j y_{t-j} \right) + \frac{1}{2} s' (G + D \Lambda D') s} \\
 &= \frac{\Phi_T(\Gamma^* \Sigma^* s; \mathbf{0}, I_T + \Gamma^* \Sigma^* \Gamma^{*'})}{\Phi_T(\mathbf{0}; \mathbf{0}, I_T + \Gamma^* \Sigma^* \Gamma^{*'})} e^{s' \left(\sum_{j=1}^P \Phi_j y_{t-j} \right) + \frac{1}{2} s' \Sigma^* s} = M_Z(s)
 \end{aligned}$$

where $Z \sim CSN_{T,T} \left(\sum_{j=1}^P \Phi_j y_{t-j}, \Sigma^*, \Gamma^*, \mathbf{0}, I_T \right)$.

(b) It is proved by using the linearity property of the multivariate normal distributions.

(c) It can be proved by the following arguments:

$$\begin{aligned}
 f(W_t|Y_t) &= \frac{f(Y_t, W_t)}{f(Y_t)} = \frac{f(W_t)f(Y_t|W_t)}{f(Y_t)} \\
 &= \frac{\Phi_T^{-1}(\mathbf{0}; \mathbf{0}, \Lambda) \varphi_T(\mathbf{w}_t; \mathbf{0}, \Lambda) \varphi_T\left(y_t; \sum_{j=1}^P \Phi_j y_{t-j} + D W_t, G\right)}{\varphi_T\left(y_t; \sum_{j=1}^P \Phi_j y_{t-j}, \Sigma^*\right) \Phi_T\left(\Gamma^*\left(y_t - \sum_{j=1}^P \Phi_j y_{t-j}\right); \mathbf{0}, I_T\right) \Phi_T^{-1}(\mathbf{0}; \mathbf{0}, I_T)} \\
 &= \frac{\varphi_T(\mathbf{w}_t; \mathbf{0}, \Lambda) \varphi_T\left(y_t; \sum_{j=1}^P \Phi_j y_{t-j} + D \mathbf{w}_t, G\right)}{\varphi_T\left(y_t; \sum_{j=1}^P \Phi_j y_{t-j}, \Sigma^*\right) \Phi_T\left(\Gamma^*\left(y_t - \sum_{j=1}^P \Phi_j y_{t-j}\right); \mathbf{0}, I_T\right)} \\
 &= C_1(y_t, \theta_1) \varphi_q(\mathbf{w}_t; \mathbf{v}^*, \Lambda^*) = C_2(y_t, \theta_1) \Phi_T^{-1}(\mathbf{0}; -\mathbf{v}^*, \Lambda^*) \varphi_q(\mathbf{w}_t; \mathbf{v}^*, \Lambda^*) \\
 &= C(y_t, \theta_1) N_T^0(\mathbf{w}_t; \mathbf{v}^*, \Lambda^*), \mathbf{w}_t \geq 0,
 \end{aligned}$$

where $\Lambda^* = (\Lambda^{-1} + D'G^{-1}D)^{-1}$, $\mathbf{v}^* = \Lambda^* D'G^{-1} \left(y_t - \sum_{j=1}^P \Phi_j y_{t-j} \right)$, and C is function of parameters $\theta_1 = (\Phi_j, \Lambda, G, D)$ and observed data y_t .

The moment generating function of $W_t|Y_t$ is given by

$$M_{W_t|Y_t}(s) = C(Y_t, \theta_1) \frac{\Phi_T(\Lambda^* s; -\mathbf{v}^*, \Lambda^*)}{\Phi_T(\mathbf{0}; -\mathbf{v}^*, \Lambda^*)} e^{s' \mathbf{v}^* + \frac{1}{2} s' \Lambda^* s}, s \in \mathcal{R}^T,$$

and so

$$E(W_t|Y_t) = \frac{\partial M_{W_t|Y_t}(s)}{\partial s} \Big|_{s=0},$$

where

$$\frac{\partial M_{W_t|Y_t}(s)}{\partial s} = C(Y_t, \theta_1) \left\{ \Lambda^* \frac{\partial \Phi_T(s; -\mathbf{v}^*, \Lambda^*)}{\partial s} e^{s' \mathbf{v}^* + \frac{1}{2} s' \Lambda^* s} + \frac{\Phi_T(\Lambda^* s; -\mathbf{v}^*, \Lambda^*)}{\Phi_T(\mathbf{0}; -\mathbf{v}^*, \Lambda^*)} (\mathbf{v}^* + \Lambda^* s) e^{s' \mathbf{v}^* + \frac{1}{2} s' \Lambda^* s} \right\}.$$

Therefore

$$E(W_t|Y_t) = \frac{\partial M_{W_t|Y_t}(s)}{\partial s} \Big|_{s=0} = C(Y_t, \theta_1) (\mathbf{v}^* + \Lambda^* \xi_1),$$

where $\xi_1 = \frac{\partial \Phi_T(s; -\mathbf{v}^*, \Lambda^*)}{\partial s} \Big|_{s=0}$. Also,

$$E(W_t' W_t | Y_t) = \frac{\partial^2 M_{W_t|Y_t}(s)}{\partial s \partial s'} \Big|_{s=s'=0},$$

Where

$$\begin{aligned} \frac{\partial^2 M_{W_t|Y_t}(s)}{\partial s \partial s'} &= C(Y_t, \theta_1) \left\{ \Lambda^* \frac{\partial^2 \Phi_T(s; -\mathbf{v}^*, \Lambda^*)}{\partial s \partial s'} e^{s' \mathbf{v}^* + \frac{1}{2} s' \Lambda^* s} \right. \\ &+ \Lambda^* \frac{\partial \Phi_T(s; -\mathbf{v}^*, \Lambda^*)}{\partial s} (\mathbf{v}^* + \Lambda^* s)' e^{s' \mathbf{v}^* + \frac{1}{2} s' \Lambda^* s} \\ &+ \left(\Lambda^* \frac{\partial \Phi_T(s; -\mathbf{v}^*, \Lambda^*)}{\partial s} (\mathbf{v}^* + \Lambda^* s)' e^{s' \mathbf{v}^* + \frac{1}{2} s' \Lambda^* s} \right)' \\ &\left. + \frac{\Phi_T(\Lambda^* s; -\mathbf{v}^*, \Lambda^*)}{\Phi_T(\mathbf{0}; -\mathbf{v}^*, \Lambda^*)} \left((\mathbf{v}^* + \Lambda^* s)(\mathbf{v}^* + \Lambda^* s)' e^{s' \mathbf{v}^* + \frac{1}{2} s' \Lambda^* s} + \Lambda^* e^{s' \mathbf{v}^* + \frac{1}{2} s' \Lambda^* s} \right) \right\} \end{aligned}$$

Therefore

$$\begin{aligned} E(W_t' W_t | Y_t) &= \frac{\partial^2 M_{W_t|Y_t}(s)}{\partial s \partial s'} \Big|_{s=s'=0} \\ &= C(Y_t, \theta_1) \left(\Lambda^* \xi_2 + \Lambda^* \xi_1 \mathbf{v}^{*'} + (\Lambda^* \xi_1 \mathbf{v}^{*'})' + \mathbf{v}^* \mathbf{v}^{*'} + \Lambda^* \right), \end{aligned}$$

where $\xi_2 = \frac{\partial^2 \Phi_T(s; -\mathbf{v}^*, \Lambda^*)}{\partial s \partial s'} \Big|_{s=s'=0}$.

Appendix B. Algorithm CJJ

Step 1 Compute $\mathbf{v}_{(k+1)} = \mathbf{I}_T$ and $\mathbf{m}_{(k+1)} = \alpha_k \Delta_k^{-1/2} \Phi'_{0k} (\mathbf{Y}_{-P} - \mathbf{Z}_{-P} \Phi_k)$. Simulate \mathbf{W}_{k+1} from a multivariate truncated normal with mean $\mathbf{m}_{(k+1)}$ and $T \times T$ variance—covariance matrix \mathbf{v}_{k+1} .

Step 2 Select a T -dimension random vector \mathbf{V}_1 with elements $v_{1i} = z_{1i} / (\sum_j z_{1j}^2)^{1/2}$, where, $z_{1i}, 1 \leq i \leq T$ are $iid \sim N(0, 1)$. Generate $\lambda_1 \sim N(0, 1)$ and set $\Upsilon_1 = \Phi_k + \lambda_1 \mathbf{V}_1$. Compute

$$\tau_{k+1} = \log\{\pi(\Upsilon_1 | \mathbf{W}_{k+1})\} - \log\{\pi(\Phi_k | \mathbf{W}_{k+1})\}$$

Simulate $u_1 \sim \text{Unif}(0, 1)$. If $u_1 \leq \min(1, \exp(\tau_{k+1}))$, let $\Phi_{k+1} = \Upsilon_1$. Otherwise, let $\Phi_{k+1} = \Phi_k$.

Step 3 Decompose $\Sigma_k = \mathbf{O} \mathbf{D} \mathbf{O}'$, where, $\mathbf{D} = \text{diag}(d_1, \dots, d_T), d_1 \geq d_2 \geq \dots \geq d_T$, and $\mathbf{O} \mathbf{O}' = \mathbf{I}$. Let $d_i^* = \log(d_i), \mathbf{D}^* = \text{diag}(d_1^*, \dots, d_T^*)$ and $\Sigma_k^* = \mathbf{O} \mathbf{D}^* \mathbf{O}'$.

Select a random symmetric $T \times T$ matrix \mathbf{V}_2 with elements $v_{2ij} = z_{2ij} / (\sum_{l \leq m} z_{2lm}^2)^{1/2}$, where, $z_{2ij}, 1 \leq i \leq j \leq T \times (T + 1)/2$, are $iid \sim N(0, 1)$. (the other elements of \mathbf{V}_2 are defined by symmetry).

Generate $\lambda_2 \sim N(0, 1)$ and set $\Upsilon_2 = \Sigma_k^* + \lambda_2 \mathbf{V}_2$. Decompose $\Upsilon_2 = \mathbf{Q} \mathbf{C}^* \mathbf{Q}'$, where, $\mathbf{C}^* = \text{diag}(c_1^*, \dots, c_T^*), c_1^* \geq c_2^* \geq \dots \geq c_T^*$, and $\mathbf{Q} \mathbf{Q}' = \mathbf{I}$. Compute

$$\tau_{k+1} = \log\{\pi(\Upsilon_2 | \mathbf{W}_{k+1}, \Phi_{k+1})\} - \log\{\pi(\Sigma_k^* | \mathbf{W}_{k+1}, \Phi_{k+1})\}$$

Simulate $u_2 \sim \text{Unif}(0, 1)$. If $u_2 \leq \min(1, \exp(\tau_{k+1}))$, let $\Sigma_{k+1}^* = \Upsilon_2, \mathbf{C} = \text{diag}(e^{c_1^*}, \dots, e^{c_T^*})$ and $\Sigma_{k+1} = \mathbf{Q} \mathbf{C} \mathbf{Q}'$. Otherwise, let $\Sigma_{k+1}^* = \Sigma_k^*$ and $\Sigma_{k+1} = \Sigma_k$.

Step 4 Select a T -dimension random vector \mathbf{V}_3 with elements $v_{3i} = z_{3i} / (\sum_j z_{3j}^2)^{1/2}$, where, $z_{3i}, 1 \leq i \leq T$ are $iid \sim N(0, 1)$. Generate $\lambda_3 \sim N(0, 1)$ and set $\Upsilon_3 = \alpha_k + \lambda_3 \mathbf{V}_3$. Compute

$$\tau_{k+1} = \log\{\pi(\Upsilon_3 | (\mathbf{W}_{k+1}, \Sigma_{k+1}, \Phi_{k+1}))\} - \log\{\pi(\alpha_k | (\mathbf{W}_{k+1}, \Sigma_{k+1}, \Phi_{k+1}))\}.$$

Simulate $u_3 \sim \text{Unif}(0, 1)$. If $u_3 \leq \min(1, \exp(\tau_{k+1}))$, let $\alpha_{k+1} = \Upsilon_3$. Otherwise, let $\alpha_{k+1} = \alpha_k$.

References

Arellano-Valle RB, Azzalini A (2006) On the unification of families of skew-normal distributions. *J Stat Theory Appl* 33(3):561–574
 Azzalini A (1985) A class of distribution which includes the normal ones. *Scand J Stat* 12(2):171–178
 Azzalini A (2005) The skew normal distribution and related multivariate families. *Scand J Stat* 32(2):159–200 (with discussion)
 Azzalini A, Capitanio A (1999) Statistical applications of the multivariate skew-normal distribution. *J R Stat Soc B* 61:579–602
 Azzalini A, Dalla Valle A (1996) The multivariate skew-normal distribution. *Biometrika* 83:715–726

- Basawa IV, Lund RB (2001) Large sample properties of parameter estimates for periodic arma models. *J Time Ser Anal* 22:651–663
- Bayes CL, Branco MD (2007) Bayesian inference for the skewness parameter of the scalar skew-normal distribution. *Braz J Probab Stat* 21:141–163
- Bondon P (2009) Estimation of autoregressive models with epsilon-skew-normal innovations. *J Multivar Anal* 100(8):1761–1776
- Broszkiewicz-Suwaj E, Makagon A, Weron R, Wylomanska A (2004) On detecting and modeling periodic correlation in financial data. *Physica A* 336(1–2):196–205
- Chaari F, Leskow J, Napolitano A, Sanchez-Ramirez A (eds) (2014) Cyclostationarity: theory and methods. Lecture notes in mechanical engineering. Springer, Cham
- Chaari F, Leskow J, Napolitano A, Zimroz R, Wylomanska A, Dudek A (eds) (2015) Cyclostationarity: theory and methods II. Applied condition monitoring. Springer, Cham
- Chaari F, Leskow J, Napolitano A, Zimroz R, Wylomanska A (eds) (2017) Cyclostationarity: theory and methods III. Applied condition monitoring. Springer, Cham
- Franses PH (1996) Periodicity and stochastic trends in economic time series. Oxford University Press, Oxford
- Franses PH, Paap R (1994) Model selection in periodic autoregressive. *Oxf Bull Econ Stat* 56(4):421–439
- Franses PH, Paap R (2004) Periodic time series models. Oxford University Press, Oxford
- Gardner WA (1994) Cyclostationarity in communications and signal processing. IEEE Press, New York
- Gauvain J, Lee C (1994) Maximum a posteriori estimation for multivariate Gaussian mixture observations of markov Chains. *IEEE Trans Speech Audio Process* 2(2):291–298
- Gebizlioglu OL, Senoglu B, Kantar YM (2011) Comparison of certain value-at-risk estimation methods for the two-parameter Weibull loss distribution. *J Comput Appl Math* 235(11):3304–3314
- Genton ME (2004) Skew elliptical distributions and their applications: a journey beyond normality. CRC, London
- Gladyshev EG (1961) Periodically correlated random sequences. *Sov Math* 2:385–388
- González-Farías G, Domínguez-Molina J, Gupta A (2004) Additive properties of skew normal random vectors. *J Stat Plan Inference* 126:521–534
- Hipel KW, McLeod AI (1994) Time series modelling of water resources and environmental systems. Elsevier, Amsterdam
- Hurd HL, Miamee A (2007) Periodically correlated random sequences: spectral theory and practice. Wiley, Hoboken
- Li WK, McLeod AI (1988) ARMA modelling with non-Gaussian innovations. *J Time Ser Anal* 9(2):155–168
- Liu C, Rubin DB (1994) The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence. *Biometrika* 81:633–648
- Lund RB, Basawa IV (2000) Recursive prediction and likelihood evaluation for periodic arma models. *J Time Ser Anal* 21:75–93
- Lund R, Shao Q, Basawa I (2006) Parsimonious periodic time series models. *Aust N Z J Stat* 48(1):33–47
- Lutkepohl H (2005) New introduction to multiple time series analysis. Springer, Berlin
- Maleki M, Arellano-Valle RB (2017) Maximum a posteriori estimation of autoregressive processes based on finite mixtures of scale-mixtures of skew-normal distributions. *J Stat Comput Simul* 87(2):1061–1083
- Maleki M, Arellano-Valle RB, Dey DK, Mahmoudi MR, Jalili SMJ (2018) A Bayesian approach to robust skewed autoregressive processes. *Calcutta Stat Assoc Bull* 69(2):165–182
- Manouchehri T, Nematollahi AR (2019) On the estimation problem of periodic autoregressive time series: symmetric and asymmetric innovations. *J Stat Comput Simul* 89(1):71–97
- McLeod AI (1993) Parsimony, model adequacy, and periodic autocorrelation in time series forecasting. *Int Stat Rev* 61:387–393
- McLeod AI (1994) Diagnostic checking of periodic autoregression models with application. *J Time Ser Anal* 15(2):221–233
- Meng XL, Rubin DB (1993) Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika* 80:267–278
- Nematollahi AR, Soltani AR (2000) Discrete time periodically correlated Markov processes. *Math Stat* 20:127–140
- Nematollahi AR, Soltani AR, Mahmoudi MR (2017) Periodically correlated modeling by means of the periodograms asymptotic distributions. *Stat Pap* 1(1):1–12
- Ni S, Sun D (2003) Noninformative priors and frequentist risks of bayesian estimators of vector-autoregressive models. *J Econom* 115:159–197

- Noakes DJ, McLeod AI, Hipel KW (1985) Forecasting monthly river-flow time series. *Int J Forecast* 1:179–190
- Noales A, de Frutto RF (1997) Forecasting with periodic models: a comparison with time invariant coefficient models. *Int J Forecast* 13:393–405
- Osborn D, Smith J (1989) The performance of periodic autoregressive models in forecasting seasonal U.K. consumption. *J Bus Econ Stat* 7:117–127
- Pagano M (1978) On periodic and multiple autoregressions. *Ann Stat* 6:1310–1317
- Pourahmadi M (2007) Skew-normal ARMA models with nonlinear heteroscedastic predictors. *Commun Stat Theory Methods* 36:1803–1819
- Serpedin E, Panduru F, Sari I, Giannakis GB (2005) Bibliography on cyclostationarity. *Signal Process* 85(12):2233–2303
- Shao Q (2006) Mixture periodic autoregressive time series models. *Stat Probab Lett* 76(6):609–618
- Shao Q (2007) Robust estimation for periodic autoregressive time series. *J Time Ser Anal* 29:251–263
- Sharafi M, Nematollahi AR (2016) AR(1) model with skew-normal innovations. *Metrika* 79(8):1011–1029
- Sun D, Ni S (2004) Bayesian analysis of vector-autoregressive models with noninformative priors. *J Stat Plan Inference* 121(2):291–309
- Sun D, Ni S (2005) Bayesian estimates for vector autoregressive models. *J Bus Econ Stat* 23(1):105–117
- Tolpin D, Wood F (2015) Maximum a posteriori estimation by search in probabilistic programs. In: The 8th annual symposium on combinatorial search will take place in EinGedi, the Dead Sea, Israel, from June 11–13, 2015. The proceedings and workshop technical reports will be published by AAAI Press
- Troutman BM (1979) Some results in periodic autoregression. *Biometrika* 66:219–228
- Ursu E, Duchesne P (2009) On modelling and diagnostic checking of vector periodic autoregressive time series models. *J Time Ser Anal* 30:70–96
- Ursu E, Turkman KF (2012) Periodic autoregressive model identification using genetic algorithm. *J Time Ser Anal* 33:398–405
- Vecchia AV (1985a) Periodic autoregressive-moving average (PARMA) modeling with applications to water resources. *Water Resour Bull* 21:721–730
- Vecchia AV (1985b) Maximum likelihood estimation for periodic autoregressive moving average models. *Technometrics* 27:375–384
- White M, Wen J, Bowling M, Schuurmans D (2015) Optimal estimation of multivariate ARMA models. In: Proceedings of the twenty-ninth AAAI conference on artificial intelligence

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.