





Article

A Bayesian Approach to Heavy-Tailed Finite Mixture Autoregressive Models

Mohammad Reza Mahmoudi ^{1,2} , Mohsen Maleki ³ , Dumitru Baleanu ^{4,5} ,
Vu-Thanh Nguyen ⁶ and Kim-Hung Pho ^{7,*} 

¹ Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam; mohammadrezamahmoudi@duytan.edu.vn

² Department of Statistics, Faculty of Science, Fasa University, Fasa 74616-86131, Iran; mahmoudi.m.r@fasau.ac.ir

³ Department of Statistics, University of Isfahan, Isfahan 8174673441, Iran; m.maleki.stat@gmail.com

⁴ Department of Mathematics, Faculty of Art and Sciences, Cankaya University, Balgat, 06530 Ankara, Turkey; dumitru@cankaya.edu.tr

⁵ Institute of Space Sciences, 077125 Magurele-Bucharest, Romania

⁶ University of Economics and Law, Ho Chi Minh City 700000, Vietnam; thanhvn@uel.edu.vn

⁷ Fractional Calculus, Optimization and Algebra Research Group, Faculty of Mathematics and Statistics, Ton Duc Thang University, Ho Chi Minh City 758307, Vietnam

* Correspondence: phokimhung@tdtu.edu.vn

Received: 10 March 2020; Accepted: 18 May 2020; Published: 2 June 2020



Abstract: In this paper, a Bayesian analysis of finite mixture autoregressive (MAR) models based on the assumption of scale mixtures of skew-normal (SMSN) innovations (called SMSN–MAR) is considered. This model is not simultaneously sensitive to outliers, as the celebrated SMSN distributions, because the proposed MAR model covers the lightly/heavily-tailed symmetric and asymmetric innovations. This model allows us to have robust inferences on some non-linear time series with skewness and heavy tails. Classical inferences about the mixture models have some problematic issues that can be solved using Bayesian approaches. The stochastic representation of the SMSN family allows us to develop a Bayesian analysis considering the informative prior distributions in the proposed model. Some simulations and real data are also presented to illustrate the usefulness of the proposed models.

Keywords: Gibbs sampling; MCMC method; non-linear time series; finite mixture autoregressive models; SMSN distributions

1. Introduction

Data analysts apply computational models to describe and infer statements about complex datasets. Mixture models are a valuable class of these models. Finite mixture models are of great importance in statistical inferences. The multimodality, skewness, kurtosis, and unobserved heterogeneity are usually observed in many datasets, for example time series datasets. Importance of mixture distributions, which are the main tools for statistical mixture models, has been noted by many references [1–8]. Various statistical fields containing time series modeling and regression analysis frequently use the mixture modeling. In fact, in analyzing time series data, some events may affect and change the behavior of data over time, for example, finance crises in many financial or panel time series datasets. The finite mixture autoregressive (MAR) models were suggested by Wong and Li [7], to catch the multimodal phenomena, are flexible and applicable in many fields, such as electroencephalogram modeling in medicine [8], interest rates and bond pricing [9], Forex rate [10], and other various fields such as telecommunications, hydrology, biology, sociology, and medical sciences.

Most researchers considered the MAR models based on Gaussian distribution, which are called Gaussian mixture of autoregressive (GMAR) models [7–11]. The GMAR models are sensitive to the existence of outliers, heavy tailed distribution and asymmetry in datasets, and thus some authors considered MAR models based on non-Gaussian, heavy tailed and/or asymmetric distributions. Wong and Li [12] introduced logistic mixture autoregressive model. Wong and Chan [13] also introduced a student t -mixture autoregressive model and discussed its inferences and then applied this model to the heavy tailed financial data.

The SMSN family, introduced by Branco [1], is a rich family containing famous symmetric scale mixtures of normal (SMN; [14]) distributions and asymmetric distributions such as the skew- t , skew-slash and skew-contaminated-normal distribution. The MAR model based on the SMSN innovations, hereafter called SMSN–MAR models, has reasonable performance to model real time series data with outliers and skewness. The various MAR model members can determine weights to each observation, and therefore each single observation affects the estimation of the model's parameters, and it leads to robust inferences. Bayesian inferences of the MAR models have some advantages compared to classical methods, that some of them are discussed below.

For maximum likelihood computations and classical inference of a MAR model, although the numerical optimization algorithms such as the EM-algorithm can be applied, converging to the major mode may fail, while this is not an issue in Bayesian inferences. Furthermore, in the MAR models there exist many situations with indirect information about the parameters or large changes in the results by small changes in the data. When we consider a sample from a MAR model, the sample includes no information about the parameters of one or more components with a positive probability. Therefore, the likelihood function can be unbounded and consequently the resultant likelihood estimates are only local maximum. Using suitable priors and some Bayesian techniques can solve such problems. Another problematic issue on the classical likelihood-based inference is that, in MAR model, the component parameters are not identifiable marginally, so distinguishing any component from other components in the likelihood is hard. The identifiability is not an important problem in Bayesian inferences. Some of the other advantages of the Bayesian inferences of the mixture models are given in [15,16].

In this paper, we consider a Bayesian technique to find estimates of the SMSN–MAR models by the Gibb sampling scheme and application of the Markov Chain Monte Carlo (MCMC) methods. The main characteristics of the SMSN distributions are reviewed in Section 2. In Section 3, we consider the MAR models and apply the Bayesian methodology for estimating the proposed MAR model's parameters. To evaluate the Bayesian estimates of parameters and performance of the proposed model, numerical studies and a real practical example are reported in Section 4. The conclusion is provided in the last section.

2. The SMSN Distributions

A random variable Y is belonged to the *scale mixtures of skew normal* (SMSN) family, if it has the following representation:

$$Y = \mu + k^{1/2}(U)Z, \quad (1)$$

where the random variable Z follows a skew-normal distribution in [17], where σ^2 and λ are respectively scale and skewness parameters, the function $k(u)$ is positive respect to u , U is the scale random variable with density function or pmf $h(\cdot; \boldsymbol{\nu})$ and distribution function $H(\cdot; \boldsymbol{\nu})$, and the vector or scalar $\boldsymbol{\nu}$ indexes the distribution of U . The random variable Y is denoted by $Y \sim \text{SMSN}(\mu, \sigma^2, \lambda, \boldsymbol{\nu})$, and in this work, because of its suitable mathematical properties, we let $k(u) = 1/u$, so $Y|U = u \sim \text{SN}(\mu, u^{-1}\sigma^2, \lambda)$. Therefore, the density of Y is given by

$$f(y; \boldsymbol{\theta}) = 2 \int_0^\infty \phi(y; \mu, u^{-1}\sigma^2) \Phi\left(\frac{u^{1/2}\lambda(y - \mu)}{\sigma}\right) dH(u|\boldsymbol{\nu}), \quad (2)$$

where $\theta = (\mu, \sigma^2, \lambda, \nu^\top)^\top$, and $\phi(\cdot)$ and $\Phi(\cdot)$ are density and distribution functions of the standard normal distribution, respectively.

Proposition 1. Let $Y \sim \text{SMSN}(\mu, \sigma^2, \lambda; H)$,

- (a) If $E[U^{-1/2}] < \infty$, then $\mu_Y = E[Y] = \mu - b\Delta$,
- (b) If $E[U^{-1}] < \infty$, then $\sigma_Y^2 = \text{Var}[Y] = \sigma^2 k_2 - b^2 \Delta^2$,
- (c) Y has a stochastic representation given by $Y = \mu + \Delta W + \gamma U^{-1/2} W_1$,

where $b = -\sqrt{2/\pi} E[U^{-1/2}]$, $\delta = \lambda / \sqrt{1 + \lambda^2}$, $\Delta = \sigma \delta$, $\gamma^2 = \sigma^2 - \Delta^2$, $W = U^{-1/2} |w_0|$, and W_0 and W_1 are independent standard normal random variables.

The famous members of the SMSN family are the scale mixtures of normal (SMN) family [8,13], as the symmetric members of the SMSN family, and skew-normal (SN), skew- t (ST), skew-slash (SSL) and skew-contaminated-normal (SCN) distributions. More details are given in [18,19].

3. The SMSN–MAR Model and Bayesian Estimates

3.1. The SMSN–MAR Model

In this paper, we study the MAR model, which has been studied by Wong and Li [7], and Wong and Chan [20], due to its flexibility in use for non-linear time series analysis. This model is represented by

$$X_t = \varphi_{i,0} + \sum_{j=1}^{p_i} \varphi_{i,j} X_{t-j} + \varepsilon_{i,t}, \text{ with probability } \pi_i; i = 1, 2, \dots, g; t = 0, \pm 1, \pm 2, \dots, \quad (3)$$

or equivalently,

$$X_t = \begin{cases} \varphi_{1,0} + \sum_{j=1}^{p_1} \varphi_{1,j} X_{t-j} + \varepsilon_{1,t}; & \text{with probability } \pi_1 \\ \varphi_{2,0} + \sum_{j=1}^{p_2} \varphi_{2,j} X_{t-j} + \varepsilon_{2,t}; & \text{with probability } \pi_2, \quad t = 0, \pm 1, \pm 2, \dots, \\ \vdots \\ \varphi_{g,0} + \sum_{j=1}^{p_g} \varphi_{g,j} X_{t-j} + \varepsilon_{g,t}; & \text{with probability } \pi_g \end{cases}$$

where

- g is a known positive integer which indicates the number of components in the model;
- each component occurs with probabilities $\pi_i > 0$, $i = 1, \dots, g$, $\sum_{i=1}^g \pi_i = 1$, which obeys a discrete distribution π ;
- for each $i = 1, \dots, g$,
 - the i th autoregressive component is of order $p_i \geq 1$;
 - $\varphi_{i,j}$, $j = 0, \dots, p_i$, are the autoregressive coefficients of the i th components;
 - each i th innovation's component $\varepsilon_{i,t}$ distributed as following:

$$\varepsilon_{i,t} \stackrel{i.i.d.}{\sim} \text{SMSN}(b_i \Delta_i, \sigma_i^2, \lambda_i, \nu_i); i = 1, \dots, g; t = 0, \pm 1, \pm 2, \dots \quad (4)$$

where $b_i = -\sqrt{2/\pi} E[U_i^{-1/2}]$, $\delta_i = \lambda_i / \sqrt{1 + \lambda_i^2}$, $\Delta_i = \sigma_i \delta_i$ and $\gamma_i^2 = \sigma_i^2 - \Delta_i^2$. Also in each i th component, $\{\varepsilon_{i,t}\}$ are jointly independent and are independent of past X_t s and also independent of other component's innovations.

Hereafter, this model will be called SMSN–MAR(p, g). Note that in each component, when b_i exists, $E(\varepsilon_{i,t}) = 0$, and $\text{Var}(\varepsilon_{i,t}) = \sigma_i^2 k_{i2} - b_i^2 \Delta_i^2$; $i = 1, \dots, g$. Note that we can assume that the vector of coefficients in MAR models, distributed as π , therefore, this model can be also interpreted as random coefficients autoregressive (RCA) model.

Without loss of generality, it is convenient to set $p = \max_{i=1, \dots, g} p_i$ and $\varphi_{i,j} = \dots = \varphi_{i,p} = 0$ for $j > p_i$. Also, let $\{Z_t\}$ be an *i.i.d.* sequence of random variables which are distributed as π (Multinomial), such that $P(Z_t = i) = \pi_i$, $i = 1, \dots, g$, that determines the components, and be independent of innovations and random history of model $\mathbf{X} = (X_1, \dots, X_n)^\top$. Therefore, X_t obeys from i th component of MAR model, if $Z_t = i$. Thus, Equation (3) can be written as

$$X_t = \varphi_{Z_t,0} + \sum_{j=1}^p \varphi_{Z_t,j} X_{t-j} + \varepsilon_{Z_t,t}, \quad t = 0, \pm 1, \pm 2, \dots \quad (5)$$

where $Z_t \sim \text{Multinomial}(1, \pi_1, \dots, \pi_g)$.

It is convenient to set vector and matrix notations of sample $\mathbf{X} = (X_1, \dots, X_n)^\top$ in the form of $\mathbf{X}_{t-1} = (1, X_{t-1}, \dots, X_{t-p})^\top$; $t = 1, \dots, n$, $\boldsymbol{\phi}_{Z_t} = (\varphi_{0,Z_t}, \dots, \varphi_{p,Z_t})^\top$, $\boldsymbol{\varepsilon}_{Z_t} = (\varepsilon_{Z_t,1}, \dots, \varepsilon_{Z_t,n})^\top$ and \mathbf{A} be the $n \times p$ matrix whose t th row is \mathbf{X}_{t-1} , in the rest of the paper. Thus, the following matrix representation

$$\mathbf{X} = \mathbf{A}\boldsymbol{\phi}_{Z_t} + \boldsymbol{\varepsilon}_{Z_t}, \quad (6)$$

Note that in Equation (7), $X_t = \boldsymbol{\phi}_{Z_t}^\top \mathbf{X}_{t-1} + \varepsilon_{Z_t,t}$ where Z_t , $t = 1, \dots, g$ and distributed as π , so by Equation (5) we have

$$X_t | \mathbf{X}_{t-1}, Z_t \sim \text{SMSN}(\boldsymbol{\phi}_{Z_t}^\top \mathbf{X}_{t-1} + b_{Z_t} \Delta_{Z_t}, \sigma_{Z_t}^2, \lambda_{Z_t}, \boldsymbol{\nu}_{Z_t}); Z_t = 1, \dots, g, \quad t = 1, \dots, n. \quad (7)$$

Therefore, the density of N_t is given by

$$f_{X_t | \mathbf{X}_{t-1}, \dots, \mathbf{X}_{t-p}}(x_t; \boldsymbol{\theta}) = \sum_{i=1}^g \pi_i f_i(x_t | \mathbf{X}_{t-1}, \boldsymbol{\theta}_i), \quad (8)$$

where $f_i(\cdot | \mathbf{X}_{t-1}, \boldsymbol{\theta}_i)$ is the pdf of $\text{SMSN}(\boldsymbol{\phi}_i^\top \mathbf{X}_{t-1} + b_i \Delta_i, \sigma_i^2, \lambda_i, \boldsymbol{\nu}_i)$ defined in Equation (2), hereafter note that $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \dots, \boldsymbol{\theta}_g^\top)^\top$, where $\boldsymbol{\theta}_i = (\pi_i, \boldsymbol{\phi}_i^\top, \sigma_i^2, \lambda_i, \boldsymbol{\nu}_i^\top)^\top$, $i = 1, \dots, g$ be the vector of parameters.

By using Markovian property of this model, we have that $L(\boldsymbol{\theta} | \mathbf{X}) = f_X(X_1, \dots, X_n; \boldsymbol{\theta}) = \prod_{t=1}^n f_{X_t | \mathbf{X}_{t-p}, \dots, \mathbf{X}_{t-1}}(X_t; \boldsymbol{\theta})$, where $f_X(\cdot)$ be the density of \mathbf{X} . An auxiliary determiner of components in MAR models (6), can be expressed in the random vector $\mathbf{Z}_t = (Z_{t1}, \dots, Z_{tg})^\top$, where

$$Z_{ti} = \begin{cases} 1, & \text{if } t\text{th observation of MAR process obeys the } i\text{th component} \\ 0, & \text{otherwise} \end{cases}$$

Therefore, under the above approach, the latent random vector \mathbf{Z}_t , $t = 1, \dots, n$ has the following multinomial $\mathbf{Z}_t \sim \text{Multinomial}(1, \pi_1, \dots, \pi_g)$ with the following probability mass function:

$$P(Z_i = z_i) = \pi_1^{z_{i1}} \pi_2^{z_{i2}} \dots \pi_g^{z_{ig}}; \quad t = 1, \dots, n,$$

such that $\sum_{i=1}^g \pi_i = 1$, $\pi_i > 0$, $i = 1, \dots, g$. Therefore,

$$X_t | \mathbf{X}_{t-1}, Z_{ti} = 1 \sim \text{SMSN}(\boldsymbol{\phi}_i^\top \mathbf{X}_{t-1} + b_i \Delta_i, \sigma_i^2, \lambda_i, \boldsymbol{\nu}_i), \quad i = 1, \dots, g; \quad t = 1, \dots, n.$$

Also, to provide a MCMC method, we have:

$$X_t | \mathbf{X}_{t-1}, W_{ti} = w_{ti}, U_{ti} = u_{ti}, Z_{ti} = 1 \sim N(\boldsymbol{\phi}_i^\top \mathbf{X}_{t-1} + \Delta_i w_{ti}, u_{ti}^{-1} \gamma_i^2),$$

$$\begin{aligned}
W_{ti}|U_{ti} = u_{ti}, Z_{ti} = 1 &\sim TN_1(b_i, u_{ti}^{-1})I_{(b_i, \infty)}, \\
U_{ti}|Z_{ti} = 1 &\sim H(u_{ti}; \mathbf{v}_i), \\
\mathbf{Z}_t &\sim \text{Multinomial}(1, \pi_1, \dots, \pi_g),
\end{aligned}$$

for $t = 1, \dots, n$. Let $\mathbf{D} = (\mathbf{X}^\top, \mathbf{M}^\top)^\top$ as the complete data, where \mathbf{X} is the observed part of data, $\mathbf{M}^\top = ((W_{ti}, U_{ti}, Z_{ti})_{t=1}^n)_{i=1}^g$ is the missing parts of data, and $\boldsymbol{\theta}$ is the vector of unknown parameters. The conditional likelihood function based on the complete data is provided by

$$L(\boldsymbol{\theta}|\mathbf{D}) = \prod_{t=1}^n \prod_{i=1}^g (\pi_i f_{X_t|X_{t-1}, W_{ti}, U_{ti}, Z_{ti}}(X_t; \boldsymbol{\theta}_i) f_{W_{ti}|U_{ti}, Z_{ti}}(W_{ti}; \boldsymbol{\theta}_i) f_{U_{ti}|Z_{ti}}(U_{ti}; \boldsymbol{\theta}_i))^{Z_{ti}}.$$

3.2. Bayesian Approach

In this part, we use the Bayesian approach by applying MCMC algorithms for the model in Equation (3). We consider informative priors under squared error loss function (SEL). To consider the prior distributions for the parameters of the considered model, we assign prior distributions to the parameters as follow:

$$\begin{aligned}
\boldsymbol{\phi}_i &\sim N_{p+1}(\boldsymbol{\phi}_i, \boldsymbol{\Sigma}_i), \quad \Delta_i \sim N(\mu_{\Delta_i}, \sigma_{\Delta_i}^2), \quad \gamma_i^2 \sim \text{IG}(\rho_i/2, \varrho_i/2), \\
\boldsymbol{\pi} &= (\pi_1, \dots, \pi_g) \sim \text{Dir}(\delta_1, \dots, \delta_g),
\end{aligned} \tag{9}$$

for $i = 1, \dots, g$, where Dir and IG denote the Dirichlet distribution and the Inverse-Gamma distribution, respectively.

The prior distribution assigned to \mathbf{v}_i , $i = 1, \dots, g$, varies with the particular SMSN distributions. For skew- t model, we can consider $v_i \sim \exp(\varsigma_i/2)I_{(2, \infty)}$, exponential distribution with mean $2/\varsigma_i$ before truncation, for skew-slash model, we can consider $v_i \sim \text{Gamma}(a_i, b_i)$ with small positive values a_i and b_i such that $b_i \ll a_i$, and for skew-contaminated-normal model, the independent and non-informative $U(0, 1)$ prior distributions are considered for each component of $\mathbf{v}_i = (v_i, \gamma_i)^\top$. We also assume that priors are independent and all hyper-parameters (parameters of priors) are known, which guarantees that all posteriors are proper. The joint posterior of $\boldsymbol{\theta}$ and missing variables is in the form of $\pi(\boldsymbol{\theta}|\mathbf{D}) \propto L(\boldsymbol{\theta}|\mathbf{D})\pi(\boldsymbol{\theta})$, which is not analytically tractable, but the MCMC methods such as the Gibbs sampling, will be applied to draw samples from the marginal of conditional posteriors which are shown as follows, when $\boldsymbol{\theta}_{(-m)}$, means the m th element of $\boldsymbol{\theta}$ has been removed (Algorithm 1).

Algorithm 1: MCMC

For $i = 1, \dots, g$, and $t = 1, \dots, n$,

1. $\mathbf{Z}_t|\boldsymbol{\theta}, \mathbf{U}, \mathbf{W}; \mathbf{X} \sim \text{Multinomial}(1, p_1, \dots, p_g)$,
2. $\boldsymbol{\pi}|\boldsymbol{\theta}_{(-\pi)}, \mathbf{D} \sim \text{Dir}(\delta_1 + \sum_{t=1}^n Z_{t1}, \dots, \delta_g + \sum_{t=1}^n Z_{tg})$,
3. $W_{ti}|\boldsymbol{\theta}, \mathbf{U}, Z_{ti} = 1; \mathbf{X} \sim TN_1(\mu_{W_{ti}} + b_i, u_{ti}^{-1}M_i^2)I_{(b_i, \infty)}$,
4. $\boldsymbol{\phi}_i|\boldsymbol{\theta}_{(-\phi_i)}, Z_{ti} = 1, \mathbf{D} \sim N_{p+1}(\boldsymbol{\mu}_{\phi_i}, \boldsymbol{\Sigma}_{\phi_i})$,
5. $\gamma_i^2|\boldsymbol{\theta}_{(-\gamma_i)}, Z_{ti} = 1, \mathbf{D} \sim \text{IG}(\rho_i, \tau_i)$,
6. $\Delta_i|\boldsymbol{\theta}_{(-\Delta_i)}, Z_{ti} = 1, \mathbf{D} \sim N\left(\frac{\beta_i\sigma_{\Delta_i}^2 + \gamma_i^2\mu_{\Delta_i}}{\alpha_i\sigma_{\Delta_i}^2 + \gamma_i^2}, \frac{\alpha_i\sigma_{\Delta_i}^2 + \gamma_i^2}{\gamma_i^2\sigma_{\Delta_i}^2}\right)$,
7. $U_{ti}|\boldsymbol{\theta}, Z_{ti} = 1, \mathbf{D} \sim H_{\text{post}}(u; \boldsymbol{\theta}, \mathbf{v})$,
8. $\mathbf{v}_i|\boldsymbol{\theta}_{(-v_i)}, \mathbf{U}, Z_{ti} = 1, \mathbf{D} \sim f_{\text{post}}(\mathbf{v}; \boldsymbol{\theta})$,

where $\mathbf{Z}_t = (Z_{t1}, \dots, Z_{tg})^\top$, $p_i = \frac{\pi_i f(X_t|\mathbf{X}_{t-1}; \boldsymbol{\theta}_i)}{\sum_{i=1}^g \pi_i f(X_t|\mathbf{X}_{t-1}; \boldsymbol{\theta}_i)}$, $M_i^2 = \frac{\gamma_i^2}{\gamma_i^2 + \Delta_i^2}$, $\mu_{W_{ti}} = M_i^2 \frac{\Delta_i}{r_i^2} (X_t - \boldsymbol{\phi}_i^\top \mathbf{X}_{t-1} - b_i \Delta_i)$,

$\boldsymbol{\mu}_{\phi_i} = (\gamma_i^{-2}(\mathbf{Z}^\top \boldsymbol{\Lambda}_i \mathbf{Z}) + \mathbf{M}_{i0}^{-1})^{-1} (\gamma_i^{-2}(\mathbf{Z}^\top \boldsymbol{\Lambda}_i \mathbf{Z}) \mathbf{M}_i + \mathbf{M}_{i0}^{-1} \boldsymbol{\phi}_{i0})$, $\boldsymbol{\Sigma}_{\phi_i} = (\gamma_i^{-2}(\mathbf{Z}^\top \boldsymbol{\Lambda}_i \mathbf{Z}) + \mathbf{M}_{i0}^{-1})^{-1}$, $\rho_i = \frac{n-p+p_i+i}{2}$, $\tau_i = \frac{1}{2}(\varrho_i + (\boldsymbol{\phi}_i - \mathbf{M}_i)^\top (\mathbf{Z}^\top \boldsymbol{\Lambda}_i \mathbf{Z}) (\boldsymbol{\phi}_i - \mathbf{M}_i))$, $\alpha_i = \mathbf{W}_i^\top \boldsymbol{\Lambda}_i \mathbf{W}$ and $\beta_i = \mathbf{W}_i^\top \boldsymbol{\Lambda}_i (\mathbf{X}_t - \boldsymbol{\phi}_i^\top \mathbf{X}_{t-1} - b_i \Delta_i)$.

Finally, in order to complete the specifications of the sampling via MCMC methods, the posterior distribution of the latent variable $U_{ti}, t = 1, \dots, n, i = 1, \dots, g$, $H_{post}(u; \theta, \nu)$, and its indexed parameter ν_i , $f_{post}(\nu; \theta)$, should be determined. Set $c_{ti} = (X_t - \phi_i^T X_{t-1} - \Delta_i W_{ti})^2 / \sigma_i^2 + (W_{ti} - b_i)^2$, we then have, for $t = -1, \dots, n$ and the several models given as:

■ ST-MAR:

$$U_{ti} | \theta, Z_{ti} = 1, D \sim \text{Gamma}\left(\frac{\nu_i + 2}{2}, \frac{\nu_i + c_{ti}}{2}\right), i = 1, \dots, g, t = 1, \dots, n, \quad (10)$$

and

$$\begin{aligned} \pi(\nu_i | \theta_{(-\nu_i)}, Z_{ti} = 1, D) &\propto \left(2^{\nu_i/2} \Gamma(\nu_i/2)\right)^{-n} \\ &\times \text{Gamma}\left(\frac{n\nu_i}{2} - 1, \frac{1}{2} \left[\sum_{t=1}^n (U_{ti} - \log U_{ti}) + \varsigma_i \right] \right) I_{(2, \infty)}(\nu_i). \end{aligned} \quad (11)$$

■ SSL-MAR:

$$U_{ti} | \theta, Z_{ti} = 1, D \sim \text{Gamma}\left(\nu_i + 1, \frac{c_{ti}}{2}\right) I_{(0,1)}(U_{ti}), i = 1, \dots, g, \quad (12)$$

$$\nu_i | \theta_{(-\nu_i)}, \mathbf{U}, Z_{ti} = 1, D \sim \text{Gamma}\left(n + a_i, b_i - \sum_{t=1}^n \log U_{ti}\right), i = 1, \dots, g. \quad (13)$$

■ SCN-MAR:

$$\pi(u_{ti} | \theta, Z_{ti} = 1, D) = \frac{A_{ti}}{A_{ti} + B_{ti}} I(U_{ti} = \gamma_i) + \frac{B_{ti}}{A_{ti} + B_{ti}} I(U_{ti} = 1), i = 1, \dots, g, \quad (14)$$

where $A_{ti} = \nu_i \gamma_i \exp(-\gamma_i c_{ti}/2)$, $B_{ti} = (1 - \nu_i) \exp(-c_{ti}/2)$, and

$$\nu_i | \theta_{(-\nu_i)}, Z_{ti} = 1, D \sim \text{Beta}\left(\frac{n - \sum_{t=1}^n U_{ti}}{1 - \gamma_i} + 1, \frac{\sum_{t=1}^n U_{ti} - n\gamma_i}{1 - \gamma_i} + 1\right), i = 1, \dots, g, \quad (15)$$

$$\pi(\gamma_i | \theta_{(-\gamma_i)}, Z_{ti} = 1, D) \propto \nu_i^{\frac{n - \sum_{t=1}^n U_{ti}}{1 - \gamma_i}} (1 - \nu_i)^{\frac{\sum_{t=1}^n U_{ti} - n\gamma_i}{1 - \gamma_i}}, i = 1, \dots, g. \quad (16)$$

Note that Equations for ν_i and γ_i do not have closed forms, but a Metropolis–Hasting algorithm which has used in [21] can be employed to obtain draws of them.

In the Bayesian technique, some model selection criteria such as the expected Bayesian information criterion (EBIC) and the expected Akaike information criterion (EAIC) [22,23] can be considered. Let $\theta_1, \dots, \theta_m$ as a sample from $\pi(\theta | D)$. The posterior mean of the deviance is estimated by $\bar{D} = \sum_{i=1}^m D(\theta_i) / m$, such that $D(\theta) = -2 \sum_{t=1}^n \log f(X_t | X_{t-1}, \theta)$, and $f(\cdot | X_{t-1}, \theta)$ is the density given in Equation (8). These criteria can be respectively estimated by $\text{E}\hat{\text{AIC}} = \bar{D} + 2k$ and $\text{E}\hat{\text{BIC}} = \bar{D} + k \log n$, where k is the number of parameters, n is the sample size and $\bar{D} = D(\bar{\theta})$, where $\bar{\theta} = \sum_{i=1}^m \theta_i / m$.

4. Numerical Studies

In this section, numerical studies and a real data example are given to study the ability of the considered SMSN–MAR(p, g) model. The statistical *R software version 3.6.1* is used. A sample size of $n = 200$, and the following prior distributions are selected: $\phi_i \sim N_{p+1}(\mathbf{0}, 10^3 \mathbf{I}_p)$, $\Delta_i \sim N_1(0, 10^3)$ and $\gamma_i^2 \sim IG(0.001, 0.001)$, with $\nu_i \sim \exp(0.1) I_{(2, \infty)}$ for the skew- t model, $\nu_i \sim \text{gamma}(1, 0.01)$ for the skew-slash model, and $\nu_i \sim U(0, 1)$ and independent of $\gamma_i \sim U(0, 1)$ for the skew-contaminated normal model. Also, a Gibbs sampling with iterations equal to 60,000 and a burn-in of 10,000 cycles is applied.

In the simulation parts, we considered two schemes. In the first scheme, simulations are based on the SMSN–MAR models to show the performances of the considered model and Bayesian estimates, and in the second scheme, simulations are mixture autoregressive model based on Generalized–Hyperbolic distribution (GH–MAR model). The GH distribution is well known because of its asymmetry and heavy-tailed nature [24,25]. The second scheme will show the performance of the proposed Bayesian estimates of the heavy-tailed SMSN–MAR models to the time series with heavy-tailed and asymmetric innovations. Finally, the SMSN–MAR model is applied on the closing price of the Iran Telecommunication Company stock via the proposed Bayesian methodology.

4.1. First Scheme

The first simulation study is devoted to the following SMSN–MAR(2,2) model with 1000 samples simulations from the SN, ST and SSL distributions for the innovations with parameters $\phi_1 = (\phi_{1,0}, \phi_{1,1}, \phi_{1,2})^\top = (0.1, 0.5, 0.3)^\top$, $\phi_2 = (\phi_{2,0}, \phi_{2,1}, \phi_{2,2})^\top = (0.1, 0.6, 0.2)^\top$, $\pi_1 = 0.4$, $\sigma_1^2 = 1$, $\sigma_2^2 = 2$, $\lambda_1 = 2$ and $\lambda_2 = 4$, and $\nu_1 = \nu_2 = 3$ for ST and SSL, and $(\nu_1, \gamma_1) = (\nu_2, \gamma_2) = (0.5, 0.5)$ for the SCN. Means and standard deviations of the proposed Bayesian estimates are shown in Table 1.

Table 1. Means and standard deviations of the Bayesian estimates of the simulated SMSN–MAR(2,2) models.

Model	SN–MAR		ST–MAR		SSL–MAR		SCN–MAR	
Parameters (Values)	Mean	SD	Mean	SD	Mean	SD	Mean	SD
$\phi_{1,0}(0.1)$	0.1003	0.02131	0.1001	0.02743	0.1004	0.02309	0.1010	0.02301
$\phi_{2,0}(0.1)$	0.1002	0.02242	0.1007	0.02436	0.1005	0.02419	0.1012	0.02502
$\phi_{1,1}(0.5)$	0.5017	0.01483	0.5011	0.01738	0.5010	0.01333	0.5033	0.01501
$\phi_{1,2}(0.3)$	0.3011	0.01746	0.3009	0.01846	0.3024	0.01963	0.3032	0.01758
$\phi_{2,1}(0.6)$	0.6032	0.01592	0.6048	0.01977	0.6024	0.01846	0.6101	0.01610
$\phi_{2,2}(0.2)$	0.2101	0.02032	0.2097	0.02006	0.2008	0.01926	0.2112	0.02094
$\sigma_1^2(1)$	1.1023	0.02064	1.0969	0.31741	1.1021	0.27493	1.1026	0.02080
$\sigma_2^2(2)$	2.1978	0.03027	2.2019	0.28950	2.1992	0.31171	2.2011	0.03311
$\lambda_1(2)$	2.0184	0.81651	2.0025	0.94561	1.9367	0.82846	2.0221	0.82331
$\lambda_2(4)$	4.0038	1.09817	3.9014	0.95928	3.9930	0.90563	4.0112	1.09831
$\nu_1(3)$	–	–	3.8957	0.56842	3.9473	1.14587	0.5211	0.01918
$\nu_2(3)$	–	–	3.8957	0.46877	3.9473	1.24587	0.5192	0.02111
$\gamma_1(0.5)$	–	–	–	–	–	–	0.5103	0.02013
$\gamma_2(0.5)$	–	–	–	–	–	–	0.5201	0.02321
$\pi_1(0.4)$	0.4011	0.04113	0.4008	0.02795	0.3957	0.01758	0.4011	0.04113

4.2. Second Scheme

The second simulation study is devoted to the following GH–MAR(2,2) model with 200 samples as follows:

$$X_t = \begin{cases} \phi_1^\top X_{t-1} + \varepsilon_{t1}; & w.p. \pi_1 = 0.4, \\ \phi_2^\top X_{t-1} + \varepsilon_{t2}; & w.p. \pi_2 = 0.6, \end{cases}$$

with the autoregressive coefficients in the first scheme, for which $\{\varepsilon_{t1}\}$ be an *i.i.d.* sequence of $GH(\mu = 0, \sigma = 1, \gamma = -0.1, \bar{\alpha} = 2, \lambda = 1)$ with weak skewness, and $\{\varepsilon_{t2}\}$ be an *i.i.d.* sequence of $GH(\mu = 0, \sigma = 1, \gamma = +3, \bar{\alpha} = 2, \lambda = 1)$ with strong skewness for innovations. The EAIC criteria of all Bayesian fitted SMSN–MAR models given in Figure 1, satisfies that the asymmetry and heavy-tailed members (i.e., ST–MAR, SSL–MAR and SCN–MAR models), especially ST–MAR, are the best fitted models in comparison with the light-tailed SN–MAR member.

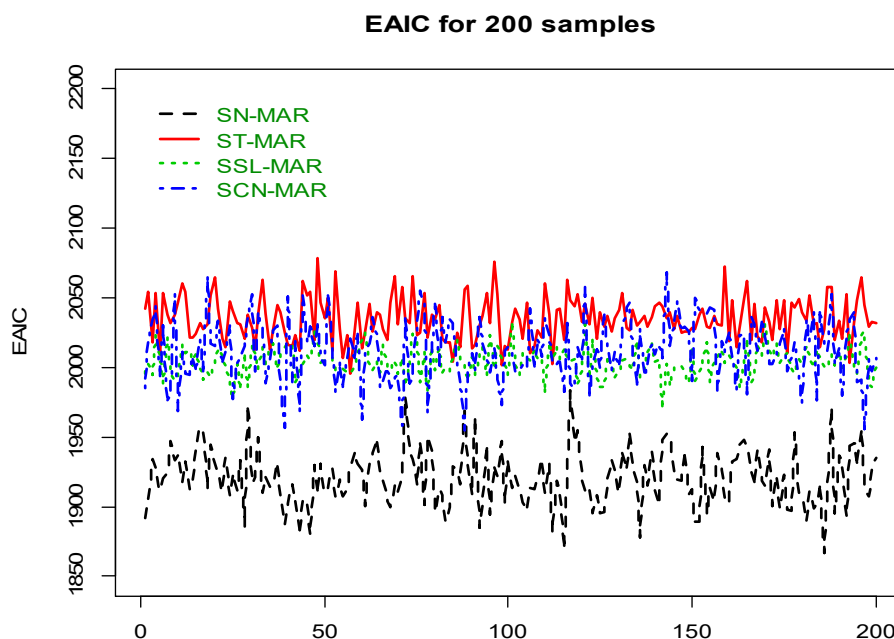


Figure 1. EAIC criterion for 200 MAR samples based on the GH distributions.

4.3. Real Data

In this section, we consider the 446 daily observations of the closing price of Iran Telecommunication Company (I.T.C.) stock based on the Iranian Rial from 2011-07-02 up to 2013-07-02 (see data at Tehran securities exchange technology management company site: www.tsetmc.com). Original time series plot of the I.T.C. stock data $\{x_t\}$, where x_t is the closing price at time t , are given in Figure 2. It can be observed that the time series $\{x_t\}$ is not stationary. A histogram of the data is plotted in Figure 3.

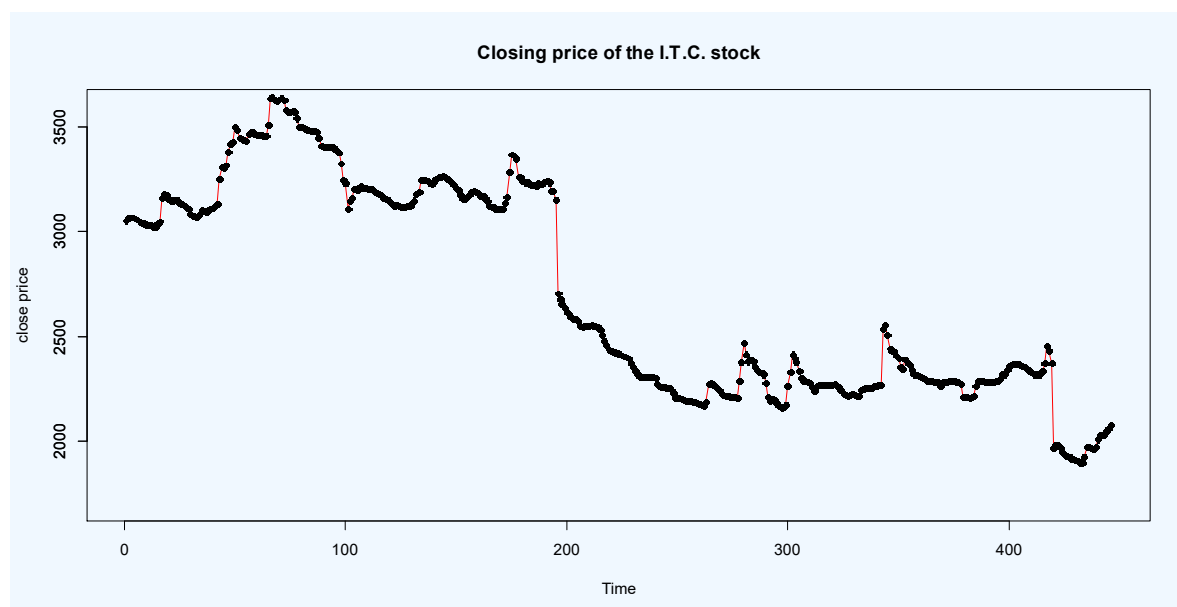


Figure 2. Time series plot of the I.T.C. stock from 2011-07-02 up to 2013-07-02.

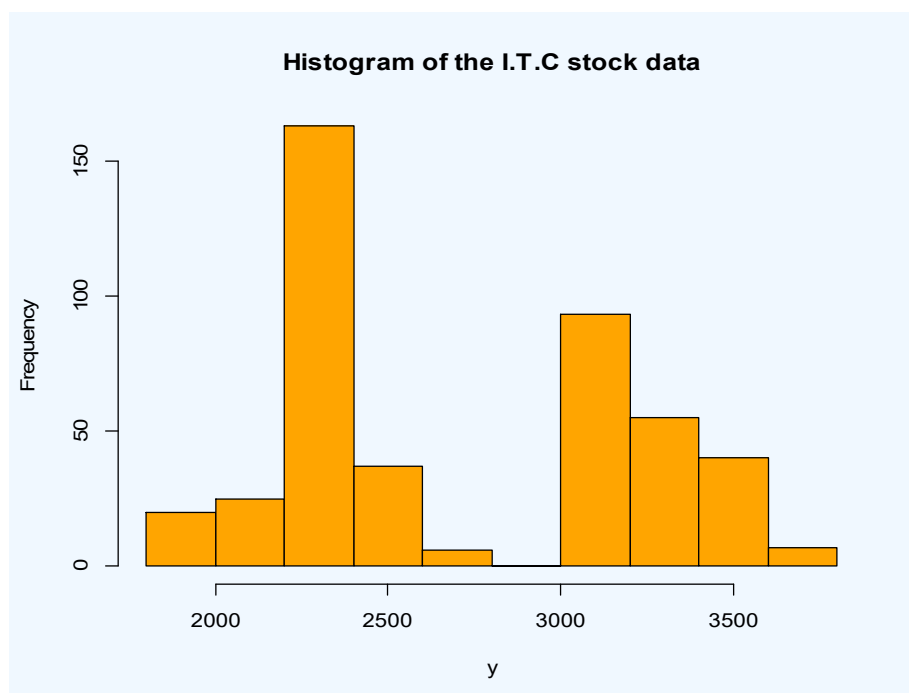


Figure 3. Histogram of the I.T.C. stock from 2011-07-02 up to 2013-07-02.

Using the Box-Cox transformation with $\lambda = 0$, and the difference with order of one, we have

$$y_t = x_t^* - x_{t-1}^* = \log(x_t/x_{t-1}), \quad (17)$$

where

$$x_t^* = \log x_t. \quad (18)$$

The time series $\{y_t\}$ has been plotted in in Figure 4. The Dickey–Fuller test verifies that the transformed time series data are stationary (p -value = 0.01).

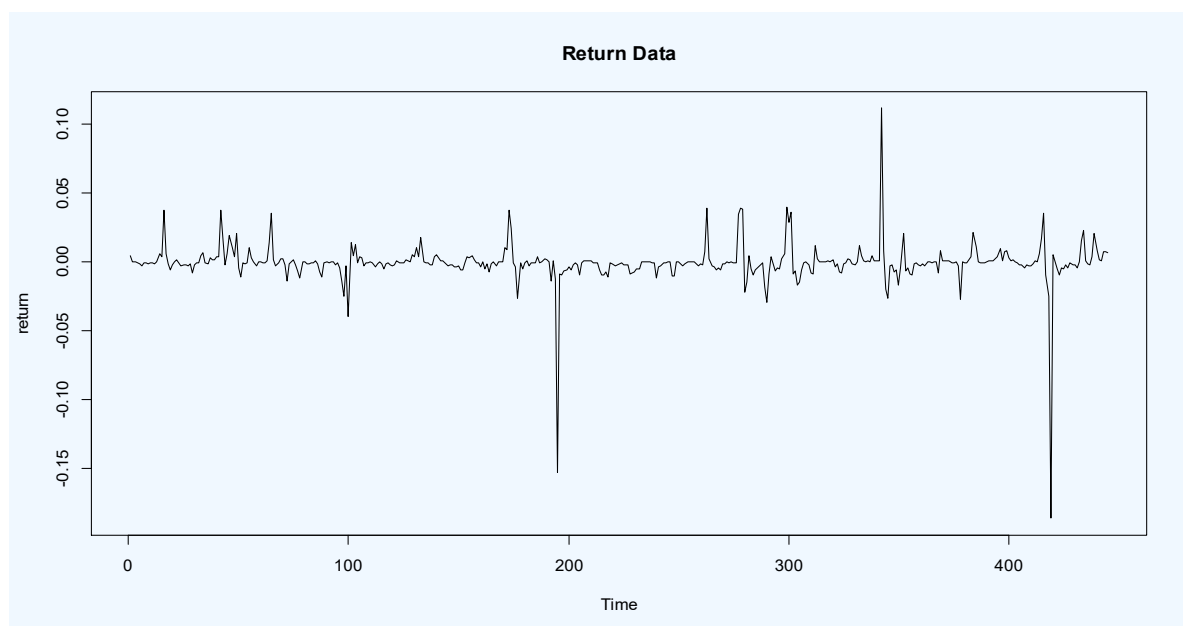


Figure 4. Plot of returned I.C.T stock data.

The EAIC and EBIC criteria demonstrate that the best SMSN–MAR model has two components with order $p_1 = 3$ for the first AR component and $p_2 = 1$ for the second AR component. Using the proposed Bayesian approach, we fitted the N–MAR (GMAR), SN–MAR, ST–MAR, SSL–MAR and SCN–MAR models with $g = 1, 2, 3$ components. The Bayesian model selection criteria for them are given in Table 2. The Bayesian estimates (posterior mean) of the ST–MAR (3,2) (the best fitted) model parameters for the return series of the I.T.C. stock data are given in Table 3.

Table 2. Bayesian model selection criteria for the proposed SMSN–MAR(3,2) with various number of components $g = 1, 2, 3$.

Model	Number of Component (g)	EAIC	EBIC
N-MAR	1	3290.6352	3297.8210
	2	3067.3456	3291.9571
	3	3175.4758	3210.9882
SN-MAR	1	3262.7364	3271.7441
	2	3038.7464	3165.0641
	3	3175.7387	3183.7251
ST-MAR	1	2979.3748	2986.2951
	2	2816.8374	2887.3601
	3	2910.6364	2977.6781
SSL-MAR	1	3176.3647	3240.2990
	2	2997.0694	3099.8811
	3	3146.7564	3157.6441
SCN-MAR	1	3164.7564	3191.3452
	2	3009.3745	3097.3383
	3	3165.7564	3182.6013

Note: The best values (between all possible models) are indicated in bold.

Table 3. Bayesian estimates with standard deviations of the parameters of fitted ST–MAR (3,2) model for the return series of the I.T.C. stock data.

Component	Parameters	Bayesian Estimates	S.E.
First AR Component	π_1	0.5802	0.0156
	$\varphi_{1,0}$	0.0046	0.0100
	$\varphi_{1,1}$	0.1478	0.0304
	$\varphi_{1,2}$	−0.0487	0.0102
	$\varphi_{1,3}$	−0.0238	0.0011
	σ_1^2	0.0036	0.0032
	λ_1	1.0312	0.0193
	ν	2.0034	0.0722
Second AR Component	π_2	0.4198	0.0121
	$\varphi_{2,0}$	0.0030	0.0013
	$\varphi_{2,1}$	0.2256	0.0345
	σ_2^2	1.0679	0.0094
	λ_2	0.9017	0.0435
	ν	2.0042	0.0847

5. Conclusions

In this paper, we have considered the robust MAR models based on the scale mixtures of skew normal distributions, called SMSN–MAR, and a Bayesian approach to estimate the models' parameters. These models are used for modeling non-linear time series data, which involve GMAR models, and offer greater flexibility than normal distribution. Numerical studies verified the good performance of the considered models and suitability of Bayesian estimates. A real data example demonstrated that SMSN–MAR models can be useful tools for non-linear and non-stationary time series

modeling. In future studies, the researchers can consider cyclostationary or almost cyclostationary processes [26–33] with SMSN errors or mixture models based on these processes.

Author Contributions: M.R.M.: Conceptualization, Methodology, Software, Supervision; M.M.: Data curation, Validation, Writing—Original draft preparation; D.B.: Visualization, Writing—Reviewing and Editing; V.-T.N.: Visualization, Writing—Reviewing and Editing; K.-H.P.: Visualization, Investigation, Writing—Reviewing and Editing. All authors have read and agreed to the published version of the manuscript

Funding: This research received no external funding.

Acknowledgments: We would like to thank the anonymous referees for their careful reading of this manuscript and also for their constructive suggestions which considerably improved the article.

Conflicts of Interest: The authors declare that they have no conflict of interest.

References

1. Branco, M.D.; Dey, D.K. A general class of multivariate skew-elliptical distributions. *J. Multivar. Anal.* **2001**, *79*, 99–113. [\[CrossRef\]](#)
2. Lindsay, B.G. Mixture models: Theory geometry and applications. In *NSF-CBMS Regional Conference Series in Probability and Statistics*; Institute of Mathematical Statistics: Hayward, CA, USA, 1995; Volume 51.
3. Böhning, D. Computer-assisted analysis of mixtures and applications. In *Meta-Analysis, Disease Mapping and Others*; Chapman&Hall/CRC: Boca Raton, FL, USA, 2000.
4. McLachlan, G.J.; Peel, D. *Finite Mixture Models*; Wiley: New York, NY, USA, 2000.
5. Frühwirth-Schnatter, S. *Finite Mixture and Markov Switching Models*; Springer: New York, NY, USA, 2006.
6. Mengersen, K.; Robert, C.P.; Titterton, D.M. *Mixtures: Estimation and Applications*; John Wiley and Sons: New York, NY, USA, 2011.
7. Wong, C.S.; Li, W.K. On a mixture autoregressive model. *J. R. Stat. Soc. Ser. B* **2000**, *62*, 95–115. [\[CrossRef\]](#)
8. Maleki, M.; Nematollahi, A.R. Autoregressive Models with Mixture of Scale Mixtures of Gaussian Innovations. *Iran. J. Sci. Technol. Trans. Sci.* **2017**, *41*, 1099–1107. [\[CrossRef\]](#)
9. Ni, H.; Yin, H. A self-organizing mixture autoregressive network for time series modeling and prediction. *Neurocomputing* **2008**, *72*, 3529–3537. [\[CrossRef\]](#)
10. McCulloch, R.E.; Tsay, R.S. Statistical inference of macroeconomic time series via markov switching models. *J. Time Ser. Anal.* **1994**, *15*, 523–539. [\[CrossRef\]](#)
11. Glasbey, C.A. Non-linear autoregressive time series with multivariate Gaussian mixtures as marginal distributions. *J. R. Stat. Soc. Ser. C* **2001**, *50*, 143–154. [\[CrossRef\]](#)
12. Wong, C.S.; Li, W.K. On a logistic mixture autoregressive model. *Biometrika* **2001**, *88*, 833–846. [\[CrossRef\]](#)
13. Wong, C.S.; Chan, W.S. A student t-mixture autoregressive model with applications to heavy tailed financial data. *Biometrika* **2009**, *96*, 751–760. [\[CrossRef\]](#)
14. Andrews, D.F.; Mallows, C.L. Scale mixtures of normal distributions. *J. R. Stat. Soc. Ser. B* **1974**, *36*, 99–102. [\[CrossRef\]](#)
15. Maleki, M.; Wraith, D.; Arellano-Valle, R.B. Robust finite mixture modeling of multivariate unrestricted skew-normal generalized hyperbolic distributions. *Stat. Comput.* **2019**, *29*, 415–428. [\[CrossRef\]](#)
16. Maleki, M.; Wraith, D. Mixtures of multivariate restricted skew-normal factor analyzer models in a Bayesian framework. *Comput. Stat.* **2019**, *34*, 1039–1053. [\[CrossRef\]](#)
17. Zarrin, P.; Maleki, M.; Khodadadi, Z.; Arellano-Valle, R.B. Time series models based on the unrestricted skew-normal process. *J. Stat. Comput. Sim.* **2019**, *89*, 38–51. [\[CrossRef\]](#)
18. Zeller, C.B.; Cabral CR, B.; Lachos, V.H. Robust mixture regression modeling based on scale mixtures of skew-normal distributions. *Comput. Stat. Data Anal.* **2010**, *54*, 2926–2941. [\[CrossRef\]](#)
19. Maleki, M.; Wraith, D.; Arellano-Valle, R.B. A flexible class of parametric distributions for Bayesian linear mixed models. *TEST* **2019**, *28*, 543–564. [\[CrossRef\]](#)
20. Wong, C.S.; Chan, W.S. Statistical Inference for Non-Linear Time Series Models. Ph.D. Thesis, University of Hong Kong, Hong Kong, China, 1998.
21. Moravveji, B.; Khodadadi, Z.; Maleki, M. A Bayesian Analysis of Two-Piece Distributions Based on the Scale Mixtures of Normal Family. *Iran. J. Sci. Technol. Trans. Sci.* **2019**, *43*, 991–1001. [\[CrossRef\]](#)

22. Brooks, S.P. Discussion on the paper by Spiegelhalter, Best, Carlin, and van der Linde. *J. R. Stat. Soc. B* **2002**, *64*, 616–618.
23. Carlin, B.P.; Louis, T.A. *Bayes and Empirical Bayes Methods for Data Analysis*, 2nd ed.; Chapman & Hall/CRC: Boca Raton, FL, USA, 2001.
24. Barndorff-Nielsen, O. Normal inverse Gaussian distributions and stochastic volatility modelling. *Scand. J. Stat.* **1997**, *24*, 1–13. [[CrossRef](#)]
25. McNeil, A.J.; Frey, R.; Embrechts, P. *Quantitative Risk Management: Concepts, Techniques and Tools*; Princeton University Press: Princeton, NJ, USA, 2005.
26. Mahmoudi, M.R.; Nematollahi, A.R.; Soltani, A.R. On the detection and estimation of the simple harmonizable processes. *Iran. J. Sci. Technol. A (Sci.)* **2015**, *39*, 239–242.
27. Mahmoudi, M.R.; Maleki, M. A new method to detect periodically correlated structure. *Comput. Stat.* **2017**, *32*, 1569–1581. [[CrossRef](#)]
28. Nematollahi, A.R.; Soltani, A.R.; Mahmoudi, M.R. Periodically correlated modeling by means of the periodograms asymptotic distributions. *Stat. Pap.* **2017**, *58*, 1267–1278. [[CrossRef](#)]
29. Mahmoudi, M.R.; Heydari, M.H.; Avazzadeh, Z.; Pho, K.H. Goodness of fit test for almost cyclostationary processes. *Digit. Signal Process.* **2020**, *96*, 102597. [[CrossRef](#)]
30. Mahmoudi, M.R.; Heydari, M.H.; Roohi, R. A new method to compare the spectral densities of two independent periodically correlated time series. *Math. Comput. Simulat.* **2019**, *160*, 103–110. [[CrossRef](#)]
31. Mahmoudi, M.R.; Heydari, M.H.; Avazzadeh, Z. Testing the difference between spectral densities of two independent periodically correlated (cyclostationary) time series models. *Commun. Stat. Theory Methods* **2019**, *48*, 2320–2328. [[CrossRef](#)]
32. Mahmoudi, M.R.; Maleki, M.; Pak, A. Testing the Difference between Two Independent Time Series Models. *Iran. J. Sci. Technol. A (Sci.)* **2017**, *41*, 665–669. [[CrossRef](#)]
33. Mahmoudi, M.R.; Heydari, M.H.; Avazzadeh, Z. On the asymptotic distribution for the periodograms of almost periodically correlated (cyclostationary) processes. *Digit. Signal Process.* **2018**, *81*, 186–197. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).