Brazilian Journal of Probability and Statistics
2020, Vol. 34, No. 2, 273–290
https://doi.org/10.1214/19-BJPS429

Symmetrical and asymmetrical mixture autoregressive processes

Mohsen Maleki[a], Arezo Hajrajabi[b] and Reinaldo B. Arellano-Valle[c]

[a]*Shiraz University*
[b]*Imam Khomeini International University*
[c]*Universidad Católica de Chile*

Abstract. In this paper, we study the finite mixtures of autoregressive processes assuming that the distribution of innovations (errors) belongs to the class of scale mixture of skew-normal (SMSN) distributions. The SMSN distributions allow a simultaneous modeling of the existence of outliers, heavy tails and asymmetries in the distribution of innovations. Therefore, a statistical methodology based on the SMSN family allows us to use a robust modeling on some non-linear time series with great flexibility, to accommodate skewness, heavy tails and heterogeneity simultaneously. The existence of convenient hierarchical representations of the SMSN distributions facilitates also the implementation of an ECME-type of algorithm to perform the likelihood inference in the considered model. Simulation studies and the application to a real data set are finally presented to illustrate the usefulness of the proposed model.

1 Introduction

Finite and infinite mixtures of distributions are very important for developing flexible statistical inferences. Applying such mixture models to the analysis of real data sets covers some important properties such as multimodality, skewness, kurtosis and unobserved heterogeneity. Many authors, for example, Lindsay (1995), Böhning (2000), McLachlan and Peel (2000), Frühwirth-Schnatter (2006) and Mengersen, Robert and Titterington (2011) have pointed to the importance of mixture distributions. Statistical methodology based on finite mixture modeling has thus been rapidly extended and employed in various fields (see, e.g., Böhning et al., 2007, 2014).

On the other hand, exogenous events such as financial crises alter the behavior of many time series over longer durations. In such situations, the use of time series models with constant parameters may not be adequate. There are also situations in which the data can be regarded as panel time series data. Non-linear time series models, especially those specified as finite mixture of autoregressive (MAR) models, which Wong and Li (2000) suggested to catch multimodal phenomena, are flexible enough to be employed in many fields. Examples include the modeling of financial time series such as market returns and the stock index (Wong and Chan, 2009), interest rates and bond pricing (Lanne and Saikkonen, 2003), the Forex rate (Ni and Yin, 2008), as well as in other fields like telecommunications, hydrology, biology, sociology, medical sciences and many more.

In order to accommodate convenient theoretical and practical implications, the mixing weights of the MAR components are defined in specific ways, therefore MAR models are also Markov-Switching autoregressive (MS-AR) models, and MAR models with two-components are closely related to the threshold autoregressive (TAR) models, and thus to self-exciting threshold autoregressive (SETAR) models as well. Also, MAR models are a special case of random coefficients autoregressive (RCA) models, and so in many cases, like GARCH

Key words and phrases. ECME algorithm, non-linear time series, finite mixtures of autoregressive models, scale mixtures of skew-normal distributions.

Received March 2017; accepted January 2019.

models, can be cast into the framework of RCA models as well as generalized hidden Markov models (HMM) (see Akinyemi, 2013). MAR models can finally be GARCH models with possible structural breaks in the model parameters or slowly evolving parameters (Wood, Rosen and Kohn, 2011).

Most researchers have proposed MAR models based on the Gaussian distribution which are called Gaussian mixture of autoregressive (GMAR) models, and covered by our proposed model. McCulloch and Tsay (1994) obtained inferences on macroeconomic time series via Markov-Switching models. Wong and Li (2001a) introduced a mixture of autoregressive conditional heteroscedastic models. Glasbey (2001) discussed on a first order MAR models based on the multivariate Gaussian distribution and applied it to solar radiation data. Lanne and Saikkonen (2003) determined non-linear time series modeling on the U.S. short-term interest rate by mixing autoregressive processes. Jin and Li (2006) studied panel time series under mixture of autoregressive models. Cervone et al. (2014) suggested location-mixture autoregressive (LMAR) models for online forecasting of lung tumor motions. The GMAR models are, however, sensitive to modeling outliers, heavy tailed and/or asymmetric time series data sets. To overcome this weakness, some authors considered MAR models based on non-Gaussian, heavy tailed and/or asymmetric distributions. Wong and Li (2001b) introduced logistic mixture autoregressive models. Wong and Chan (2009) applied a Student-t mixture autoregressive model for modeling heavy tailed financial data.

Maleki and Arellano-Valle (2017) considered, studied and shown the performance of an autoregressive process for which innovations belong to the finite mixtures of symmetric/asymmetric and light/heavy tailed scale mixtures of skew-normal (FM-SMSN) family of distributions (see also Maleki and Nematollahi, 2017b). The SMSN family is flexible class of distributions which used in celebrated works on the linear mixed models and performances of this class had shown by Arellano-Valle, Bolfarine and Lachos (2005, 2007). In this paper, we propose a more general time series model due to Maleki and Arellano-Valle (2017) and Maleki et al. (2017), that proposes a robust mixture of autoregressive models based on the SMSN class of distributions (hereafter SMSN-MAR models). The SMSN family generalizes the famous scale-mixtures of normal (SMN) class, introduced by Andrews and Mallows (1974), and contains well-known asymmetric distributions such as the skew-t, skew-slash and skew-contaminated-normal. These asymmetric and heavy-tailed distributions have been used in many types of statistical models to obtain flexible and robust inferences, for example, Arellano-Valle et al. (2008), Maleki and Mahmoudi (2017), Maleki and Nematollahi (2017a), Moravveji, Khodadai and Maleki (2018), Ghasami, Khodadadi and Maleki (2019), Hajrajabi and Maleki (2019), Maleki and Wraith (2019) and recently Maleki, Wraith and Arellano-Valle (2018a, 2018b) studied Bayesian inferences of shape mixtures of skewed distributions with application to linear mixed models and mixture models. Moreover, each of these members assigns a specific weight to the contribution of each observation in the likelihood equations, and therefore to the effect of each single observation on the estimation of the model parameters.

The rest of this paper is organized as follows. Section 2 reviews some main properties of the scale mixtures of the skew normal (SMSN) family, although more details and properties of this family can be found in the works of the above authors. The one-dimensional MAR models with scale mixtures of skew normal noise are also introduced in Section 2. Section 3 is devoted to the application of the ECME algorithm for estimating the MAR model's parameters based on SMSN distributions. Section 4 provides numerical studies to evaluate the performance of the ML estimates of the proposed model parameters. An illustration of how the proposed methods can be applied in an analysis of real data is also given in Section 4. Section 5 concludes; and some supplementary results are given in the Appendix.

2 Preliminaries on scale mixture of skew-normal distributions

The SMSN family was first studied by Branco and Dey (2001) and later revisited by Basso et al. (2010) and by Azzalini and Capitanio (2014). Let $X \sim \text{SN}(0, \sigma^2, \lambda)$ denote a skew-normal random variable (Azzalini, 1985) with location parameter $\mu = 0$, scale parameter $\sigma^2 > 0$, skewness parameter $\lambda \in \mathbb{R}$, and density function $f_{\text{SN}}(x; 0, \sigma^2, \lambda) = 2\sigma^{-1}\phi(\sigma^{-1}x)\Phi(\lambda\sigma^{-1}x)$, $x \in \mathbb{R}$ where $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal pdf and cdf, respectively. In this work, it is said that a random variable Y has an SMSN distribution if

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

where U and X are independent, $\mu \in \mathbb{R}$ is a location parameter, $k(U)$ is a positive random function of U, which is a random variable with distribution function $H(\cdot; \boldsymbol{\nu})$ indexed by a scalar or vector parameter $\boldsymbol{\nu}$. This study considers the simple variant $k(u) = u^{-1}$, because of its nice mathematical properties. It thus follows from (1) that the conditional distribution of Y given $U = u$ is given by $Y|U = u \sim \text{SN}(\mu, u^{-1}\sigma^2, \lambda)$. This means that the marginal density of Y can be represented as an infinite scale mixture of the skew-normal density as

$$f(y; \mu, \sigma^2, \lambda, \boldsymbol{\nu}) = \int_0^\infty f_{\text{SN}}(y; \mu, u^{-1}\sigma^2, \lambda)\, dH(u; \boldsymbol{\nu}), \quad y \in \mathbb{R}, \tag{2}$$

in which $H(\cdot\boldsymbol{\nu})$ represents the mixing distribution. In this case, it is said that Y has an SMSN distribution with parameters $(\mu, \sigma^2, \lambda, \boldsymbol{\nu})$, which is denoted by $Y \sim \text{SMSN}(\mu, \sigma^2, \lambda; H)$ or by $Y \sim \text{SMSN}(\mu, \sigma^2, \lambda, \boldsymbol{\nu})$. In the special case of $\lambda = 0$, the skew-normal density becomes the symmetric normal density with mean μ and variance σ^2, namely $f_{\text{SN}}(y; \mu, u^{-1}\sigma^2, 0) = \phi(y; \mu, u^{-1}\sigma^2)$. This way, the SMSN family in (2) reduces to the symmetric SMN class studied by Andrews and Mallows (1974) which has been widely regarded in the literature. Some famous asymmetric distributions in the SMSN class are the skew-normal, skew-t, contaminated skew-normal and skew-slash; the last three are also heavy-tailed distributions. Distributions like this are summarized in Table A.1 in the Appendix.

A very useful property of the SMSN random variable $Y \sim \text{SMSN}(\mu, \sigma^2, \lambda, H)$ is the stochastic representation given by

$$Y = \mu + \Delta W + \gamma U^{-1/2} W_1, \tag{3}$$

where, $\Delta = \sigma\delta$, $\gamma^2 = \sigma^2 - \Delta^2$, $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$, and $W = U^{-1/2}|W_0|$, with W_0 and W_1 being independent standard normal random variables, and $|\cdot|$ denoting the absolute value. Most of the main properties of the SMSN family can be derived from (3). For instance, it is clear from (3) that if $k_1 = E[U^{-1/2}] < \infty$, then $\mu_Y = E[Y] = \mu - b\Delta$, and if $k_2 = E[U^{-1}] < \infty$, then $\sigma_Y^2 = \text{Var}[Y] = \sigma^2 k_2 - b^2\Delta^2$, where $b = -\sqrt{2/\pi}k_1$.

Further details and properties of distributions in the SMSN family can be found in Branco and Dey (2001), Azzalini and Capitanio (2014) and Arellano-Valle et al. (2008), among others.

3 Mixture of autoregressive SMSN models

3.1 Model specification

Due to the flexibility in the modeling of non-linear time series analysis, in this section mixture of autoregressive (MAR) models as those studied by Wong and Li (2000) and Wong (1998) are considered. They are described by

$$X_t = \begin{cases} \varphi_{1,0} + \sum_{j=1}^{p_1} \varphi_{1,j} X_{t-j} + \varepsilon_{1,t}; & \text{with probability } \pi_1, \\ \varphi_{2,0} + \sum_{j=1}^{p_2} \varphi_{2,j} X_{t-j} + \varepsilon_{2,t}; & \text{with probability } \pi_2, \\ \vdots & \\ \varphi_{g,0} + \sum_{j=1}^{p_g} \varphi_{g,j} X_{t-j} + \varepsilon_{g,t}; & \text{with probability } \pi_g, \end{cases}$$
$$t = 0, \pm 1, \pm 2, \dots, \tag{4}$$

where $p_i \geq 1$ is the order of the ith AR component, $\boldsymbol{\phi}_i = (\varphi_{i,0}, \dots, \varphi_{i,p_i})^\top$ are the coefficients of the ith AR component, π_i is the weight for the ith AR component, so that $\pi_i > 0$, $i = 1, \dots, g$, and $\sum_{i=1}^{g} \pi_i = 1$, and $\varepsilon_{i,t}$ is the innovation of the ith AR component. In this work, assuming that $k_{i1} = E[U_i^{-1/2}]$ exits, it is considered that the innovations are SMSN distributed as

$$\varepsilon_{i,t} \sim \text{SMSN}(b_i \Delta_i, \sigma_i^2, \lambda_i, \boldsymbol{\nu}_i), \tag{5}$$

where $b_i = -\sqrt{2/\pi} k_1$, $\Delta_i = \sigma_i \delta_i$ and $\delta_i = \lambda_i / \sqrt{1 + \lambda_i^2}$, $i = 1, \dots, g$. Also, for each of the g AR components, the $\{\varepsilon_{i,t}\}$ are assumed mutually independent and independent of the past $\{X_s; s < t\}$. As a consequence of (5), it follows that $E(\varepsilon_{i,t}) = 0$ and, when $k_{i2} = E[U_i^{-1}]$ exist, that $\text{Var}(\varepsilon_{i,t}) = \sigma_i^2 k_{i2} - b_i^2 \Delta_i^2$, $i = 1, \dots, g$. Under the assumption in (5), this process is hereafter called the SMSN-MAR(p, g) model.

Model (4) is related to threshold autoregressive (TAR and SETAR) models, but in TAR models delay parameters, threshold parameter and other parameters must be estimated, while in MAR models the estimation of the weight vector $\boldsymbol{\pi} = (\pi_1, \dots, \pi_g)^\top$ and other parameters must be considered. As indicated below, we can assume vectors of random coefficients in MAR models, distributed as $P_{\boldsymbol{\pi}}$, therefore this model can also be 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 with distribution $P_{\boldsymbol{\pi}}$ such that $P(Z_t = i) = \pi_i$, $i = 1, \dots, g$, which determine the AR component mixtures, and are independent of the innovations $\{\varepsilon_{i,t}\}$ and random history $\{X_s; s < t\}$. Therefore, if $Z_t = i$, then X_t comes from the ith component of the MAR model. Equations in (4) can thus be rewritten as

$$X_t = \varphi_{Z_t,0} + \sum_{j=1}^{p} \varphi_{Z_t,j} X_{t-j} + \varepsilon_{Z_t,t}, \quad \text{and} \quad Z_t \sim P_{\boldsymbol{\pi}}. \tag{6}$$

It can be seen that model (6) is also a Markov switching autoregressive model with hidden process $\{Z_t\}$.

Moreover, by considering a sample vector $\boldsymbol{X} = (X_1, \dots, X_n)^\top$ and the fact that $X_t = \boldsymbol{A}_t^\top \boldsymbol{\phi}_{Z_t} + \varepsilon_{Z_t,t}$, where $\boldsymbol{\phi}_{Z_t} = (\varphi_{Z_t,0}, \dots, \varphi_{Z_t,p})^\top$, $\boldsymbol{A}_t = (1, X_{t-1}, \dots, X_{t-p})^\top$ and $\boldsymbol{a}_t = (1, x_{t-1}, \dots, x_{t-p})^\top$ which is the observed sample of $\boldsymbol{A}_t$, the following convenient matrix representation of the MAR(p, g) model in (6) is obtained in the form of

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{\phi}_{Z_t} + \boldsymbol{\varepsilon}_{Z_t}, \quad \text{and} \quad Z_t \sim P_{\boldsymbol{\pi}}, \tag{7}$$

where $\boldsymbol{A}$ is a $n \times p$ matrix with rows $\boldsymbol{A}_t^\top$, $t = 1, \dots, n$, and $\boldsymbol{\varepsilon}_{Z_t} = (\varepsilon_{Z_t,1}, \dots, \varepsilon_{Z_t,n})^\top$.

In addition, by noting that $X_t \mid X_{t-1} = x_{t-1}, \dots, X_{t-p} = x_{t-p} = X_t|\boldsymbol{A}_t = \boldsymbol{a}_t$, the assumption (5) implies that

$$X_t|\boldsymbol{A}_t = \boldsymbol{a}_t, \qquad Z_t = i \overset{\text{ind.}}{\sim} \text{SMSN}(\boldsymbol{\phi}_i^\top \boldsymbol{a}_t + b_i\Delta_i, \sigma_i^2, \lambda_i, \boldsymbol{\nu}_i), \tag{8}$$

with $P(Z_t = i) = \pi_i, i = 1, \dots, g[(,)]t = 1, \dots, n$. Therefore, the conditional distribution of $X_t|\boldsymbol{A}_t = \boldsymbol{a}_t$ corresponds to a finite mixture of SMSN-AR components, with density given by

$$f_{X_t|\boldsymbol{A}_t=\boldsymbol{a}_t}(x_t;\boldsymbol{\theta}) = \sum_{i=1}^{g} \pi_i h(x_t \mid \boldsymbol{a}_t;\boldsymbol{\theta}_i), \tag{9}$$

where $h(x_t \mid \boldsymbol{a}_t;\boldsymbol{\theta}_i) = \int_0^\infty f_{\text{SN}}(x_t;\boldsymbol{\phi}_i^\top \boldsymbol{a}_t + b_i\Delta_i, u_i^{-1}\sigma_i^2, \lambda_i)\,\mathrm{d}H(u_i;\boldsymbol{\nu}_i)$, $x_t \in \mathbb{R}$, is the density of the $\text{SMSN}(\boldsymbol{\phi}_i^\top \boldsymbol{a}_t + b_i\Delta_i, \sigma_i^2, \lambda_i, \boldsymbol{\nu}_i)$ distribution as defined in (2). Here, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \dots, \boldsymbol{\theta}_g^\top)^\top$, with $\boldsymbol{\theta}_i = (\pi_i, \boldsymbol{\phi}_i^\top, \sigma_i^2, \lambda_i, \boldsymbol{\nu}_i^\top)^\top$, denotes the whole vector of unknown parameters that determines the proposed SMSN-MAR(p, g) model.

By using the Markovian property of this MAR(p, g) model, the (conditional) likelihood function for $\boldsymbol{\theta}$ given an observed sample $\boldsymbol{X} = \boldsymbol{x}$ is $L(\boldsymbol{\theta} \mid \boldsymbol{x}) = \prod_{t=1}^n f_{X_t|\boldsymbol{A}_t=\boldsymbol{a}_t}(x_t;\boldsymbol{\theta})$. The corresponding log-likelihood function is thus given by

$$\begin{aligned}\ell(\boldsymbol{\theta} \mid \boldsymbol{x}) = \sum_{t=1}^{n} \log\Biggl(2\sum_{i=1}^{g} \pi_i \int_0^\infty \phi\bigl(x_t;\boldsymbol{\phi}_i^\top \boldsymbol{a}_t + b_i\Delta_i, u_i^{-1}\sigma_i^2\bigr) \\ \times \Phi\biggl(\frac{\sqrt{u_i}\lambda_i(x_t - \boldsymbol{\phi}_i^\top \boldsymbol{a}_t - b_i\Delta_i)}{\sigma_i}\biggr)\,\mathrm{d}H(u_i|\boldsymbol{\nu}_i)\Biggr).\end{aligned} \tag{10}$$

The maximization of log-likelihood function in (10) with respect to $\boldsymbol{\theta}$ for obtaining the ML estimates requires a high dimensional nonlinear optimization procedure. As described in the next section, the hierarchical formulation of the SMSN distributions given in (3) facilitates the implementation of an ECME algorithm to find the ML estimates.

3.2 ECME algorithm

This section describes the ECME algorithm extension of the EM algorithm (Dempster, Laird and Rubin, 1977) to obtain the ML estimates of the unknown parameters of an SMSN-MAR(p, g) model as that described by equations (4) and (5). The algorithm was proposed by Liu and Rubin (1994) as a more efficient method for finding the ML estimates.

An auxiliary determiner of components in MAR models like those specified by (5), can be expressed in terms of a multinomial random vector $\boldsymbol{Z}_t = (Z_{t1}, \dots, Z_{tg})^\top \sim \text{Multinomial}(1, \pi_1, \dots, \pi_g)$, where

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

Only one element of $\boldsymbol{Z}_t$ may be one (while the remaining elements must be zero), and its label/position determines the component distribution for the tth innovation; for further details, see, for example, McLachlan and Peel (2000). Also, by assumption $\boldsymbol{X}_t$ is independent of ε_t, and so of $\boldsymbol{Z}_t$, too. Thus, by noting in (8) that $Z_t = i$ if and only if $Z_{ti} = 1$, and considering the stochastic representation of the SMSN family given in (3), the proposed MAR model in (4)–(5) can be equivalently specified as

$$X_t|\boldsymbol{A}_t = \boldsymbol{a}_t, \qquad W_{ti} = w_{ti}, \qquad U_{ti} = u_{ti},$$

$$Z_{ti} = 1 \overset{\text{ind.}}{\sim} N(\boldsymbol{\phi}_i^\top \boldsymbol{a}_t + \Delta_i w_{ti}, u_{ti}^{-1}\gamma_i^2),$$

$$W_t|U_{ti} = u_{ti}, \qquad Z_{ti} = 1 \overset{\text{ind.}}{\sim} \text{TN}(b_i, u_{ti}^{-1}) I_{(b_i,\infty)},$$
$$U_{ti}|Z_{ti} = 1 \overset{\text{ind.}}{\sim} H(u_{ti}; \boldsymbol{\nu}_i),$$
$$\boldsymbol{Z}_t = (Z_{t1}, \dots, Z_{tg})^\top \overset{\text{ind.}}{\sim} \text{Multinomial}(1, \pi_1, \dots, \pi_g),$$

for $i = 1, \dots, g$, $t = 1, \dots, n$, where $\gamma_i^2 = \sigma_i^2 - \Delta_i^2$ and $\text{TN}(\xi, \omega^2) I_{(a,b)}$ denotes the truncated normal distribution with mean ξ and variance ω^2 to the interval (a, b). As described below, this hierarchical representation provides a convenient ECME algorithm for fitting the proposed SMSN-MAR(p, g) model.

Let $\boldsymbol{D} = (\boldsymbol{X}^\top, \boldsymbol{M}^\top)^\top$ be the complete data where, $\boldsymbol{X} = (X_1, \dots, X_n)^\top$ is the observable part, and $\boldsymbol{M}^\top = ((W_{ti}, U_{ti}, Z_{ti})_{t=1}^n)_{i=1}^g$ is the missing part. Hereafter, the vector of unknown parameters is redefined by convenience as $(\boldsymbol{\theta}, \boldsymbol{\nu})$, where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \dots, \boldsymbol{\theta}_g^\top)^\top$, with $\boldsymbol{\theta}_i = (\pi_i, \boldsymbol{\phi}_i^\top, \Delta_i, \gamma_i^2)^\top$, $i = 1, \dots, g$, and $\boldsymbol{\nu} = (\boldsymbol{\nu}_1^\top, \dots, \boldsymbol{\nu}_g^\top)^\top$. Thus, the conditional likelihood function based on the complete data is

$$L_c(\boldsymbol{\theta}, \boldsymbol{\nu}|\boldsymbol{D}) = \prod_{t=1}^n \prod_{i=1}^g \big(\pi_i f_{X_t|\boldsymbol{A}_t, W_{ti}, U_{ti}, Z_{ti}}(X_t; \boldsymbol{\theta}_i) f_{W_{ti}|U_{ti}, Z_{ti}}(W_{ti}; \boldsymbol{\nu}_i) f_{U_{ti}|Z_{ti}}(U_{ti}; \boldsymbol{\nu}_i)\big)^{Z_{ti}},$$

where $f_{X_t|\boldsymbol{A}_t, W_{ti}, U_{ti}, Z_{ti}}(x_t; \boldsymbol{\theta}_i) = \phi(x_t; \boldsymbol{\phi}_i^\top \boldsymbol{a}_t + \Delta_i w_{ti}, u_{ti}^{-1} \gamma_i^2)$, $f_{W_{ti}|U_{ti}, Z_{ti}}(w_{ti}; \boldsymbol{\nu}_i) = 2\phi(w_{ti}; b_i, u_{it}^{-1})$, where $\phi(\cdot)$ is the normal probability density function, and $f_{U_{ti}|Z_{ti}}(u_{ti}; \boldsymbol{\nu}_i)$ is the density or probability mass function induced by the mixing distribution $H(u_{ti}; \boldsymbol{\nu}_i)$. The respective log-likelihood function is then given by $\ell_c(\boldsymbol{\theta}, \boldsymbol{\nu}|\boldsymbol{D}) = \ell_c(\boldsymbol{\theta}|\boldsymbol{D}) + \ell_c(\boldsymbol{\nu}|\boldsymbol{D})$, where (ignoring constants)

$$\ell_c(\boldsymbol{\theta}|\boldsymbol{D}) = \sum_{i=1}^g \sum_{t=1}^n Z_{ti} \log \pi_i - \frac{1}{2} \sum_{i=1}^g \sum_{t=1}^n Z_{ti} \log \gamma_i^2 - \frac{1}{2} \sum_{i=1}^g \sum_{t=1}^n \frac{Z_{ti} U_{ti}}{\gamma_i^2} (X_t - \boldsymbol{\phi}_i^\top \boldsymbol{A}_t - \Delta_i W_{ti})^2, \tag{11}$$

and $\ell_c(\boldsymbol{\nu}|\boldsymbol{D}) \propto -\frac{1}{2} \sum_{i=1}^g \sum_{t=1}^n Z_{ti} U_{ti} \log W_{ti}^2 - \sum_{i=1}^g \sum_{t=1}^n Z_{ti} \log f_{U_{ti}|Z_{ti}}(u_{ti}; \boldsymbol{\nu}_i)$. Note this last part of the complete log-likelihood function is not relevant to the estimation of $\boldsymbol{\theta}$, and therefore it can be ignored in the implementation of the ECME algorithm described below.

In the E-Step of the ECME algorithm, we should determine the function $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) = E_{\boldsymbol{\theta}^{(k)}}[\ell_c(\boldsymbol{\theta}|\boldsymbol{D})|\boldsymbol{X}]$. Conditional expectations used for the construction of this function follow below. By considering standard properties of conditional expectation, it follows that

$$\hat{Z}_{ti}^{(k)} = E_{\boldsymbol{\theta}^{(k)}}[Z_{ti}|\boldsymbol{X}] = \frac{\pi_i^{(k)} h(x_t|\boldsymbol{A}_t; \hat{\boldsymbol{\theta}}_i^{(k)})}{\sum_{i=1}^g \pi_i^{(k)} h(x_t|\boldsymbol{A}_t; \hat{\boldsymbol{\theta}}_i^{(k)})},$$

where $h(\cdot \mid \boldsymbol{a}_t, \boldsymbol{\theta}_i)$ is defined below Eq. (9). Also, from the results (A.1)–(A.4) in the Appendix, it follows that

$$E_{\boldsymbol{\theta}^{(k)}}[U_{ti}|\boldsymbol{X}] = \hat{u}_{1ti}^{(k)},$$
$$E_{\boldsymbol{\theta}^{(k)}}[U_{ti} W_{ti}|\boldsymbol{X}] = \hat{U}_{ti}^{(k)}(\hat{m}_{ti}^{(k)} + b) + \hat{M}_i^{(k)} \hat{\eta}_{ti}^{(k)},$$
$$E_{\boldsymbol{\theta}^{(k)}}[U_{ti} W_{ti}^2|\boldsymbol{X}] = \hat{M}_i^{2(k)} + (\hat{m}_{ti}^{(k)} + b)^2 \hat{U}_{ti}^{(k)} + \hat{\eta}_{ti}^{(k)} \hat{M}_i^{(k)}(\hat{m}_{ti}^{(k)} + 2b),$$

$i = 1, \ldots, g$ and $t = 1, \ldots, n$, where $\hat{u}_{1ti}^{(k)}$ and $\hat{\eta}_{ti}^{(k)} = \hat{\eta}_{1ti}^{(k)}$ are obtained via (A.1) and (A.2), respectively, with $\hat{M}_i^2 = \frac{\hat{\gamma}_i^2}{\hat{\gamma}_i^2 + \hat{\Delta}_i^2}$ and $\hat{m}_{ti} = \frac{\hat{\Delta}_i}{\hat{\gamma}_i^2 + \hat{\Delta}_i^2}(x_t - \hat{\boldsymbol{\phi}}_i^\top \boldsymbol{A}_t - \hat{b}_i \hat{\Delta}_i)$. All these quantities must be evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{(k)}$, where $\hat{\boldsymbol{\theta}}^{(k)}$ is the estimated value of $\boldsymbol{\theta}$ in the kth step of the algorithm. Conditional expectations $\hat{u}_{1ti}^{(k)}$ and $\hat{\eta}_{1ti}^{(k)}$ and therefore all the above conditional expectations have a closed-form expression, when working with the Skew-t (ST) and Skew contaminated-normal (SCN) distributions of the SMSN class; see Zeller, Lachos and Vilca-Labra (2011) and Basso et al. (2010). In the case of the SSL distribution, the Monte Carlo integration can be employed in order to approximate its integrals, and the so-called MC-ECME algorithm must be implemented; see Wei and Tanner (1990), McLachlan and Krishnan (2008) and Zeller, Lachos and Vilca-Labra (2011). Also, since Z_{ti} and (U_{it}, W_{ti}) are independent, it follows that $\hat{s}_{1ti}^{(k)} = \hat{Z}_{ti}^{(k)} E_{\boldsymbol{\theta}^{(k)}}[U_{ti}|\boldsymbol{X}]$, $\hat{s}_{2ti}^{(k)} = \hat{Z}_{ti}^{(k)} E_{\boldsymbol{\theta}^{(k)}}[U_{ti} W_{ti}|\boldsymbol{X}]$ and $\hat{s}_{3ti}^{(k)} = \hat{Z}_{ti}^{(k)} E_{\boldsymbol{\theta}^{(k)}}[U_{ti} W_{ti}^2|\boldsymbol{X}]$. In order to facilitate the presentation of updating ML estimates in the future, we set $\boldsymbol{\Omega}_i^{(k)} = \operatorname{diag}(\hat{s}_{11i}^{(k)}, \ldots, \hat{s}_{1ni}^{(k)})$, $\boldsymbol{\alpha}_i^{(k)} = (\hat{s}_{21i}^{(k)}, \ldots, \hat{s}_{2ni}^{(k)})$ and $\boldsymbol{\beta}_i^{(k)} = (\hat{s}_{31i}^{(k)}, \ldots, \hat{s}_{3ni}^{(k)})$.

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

E-step:

$$
\begin{aligned}
Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) = {}& \sum_{i=1}^{g}\sum_{t=1}^{n} \hat{Z}_{ti}^{(k)} \log \pi_i - \frac{1}{2}\sum_{i=1}^{g}\sum_{t=1}^{n} \hat{Z}_{ti}^{(k)} \log \gamma_i^2 \\
& - \frac{1}{2}\sum_{i=1}^{g}\sum_{t=1}^{n} \frac{\hat{s}_{1ti}^{(k)}}{\gamma_i^2}(X_t - \boldsymbol{\phi}_i^\top \boldsymbol{A}_t)^2 + \sum_{i=1}^{g}\sum_{t=1}^{n} \frac{\hat{s}_{2ti}^{(k)}}{\gamma_i^2}(X_t - \boldsymbol{\phi}_i^\top \boldsymbol{A}_t)\Delta_i \\
& - \frac{1}{2}\sum_{i=1}^{g}\sum_{t=1}^{n} \frac{\hat{s}_{3ti}^{(k)}}{\gamma_i^2}\Delta_i^2.
\end{aligned}
$$

By considering the matrix representation in (7) and the vector notations of the conditional expectations defined above, we have

$$
\begin{aligned}
Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) = {}& \sum_{i=1}^{g}\sum_{t=1}^{n} \hat{Z}_{ti}^{(k)} \log \pi_i - \frac{1}{2}\sum_{i=1}^{g}\sum_{t=1}^{n} \hat{Z}_{ti}^{(k)} \log \gamma_i^2 \\
& - \Biggl\{ \sum_{i=1}^{g} \frac{1}{2\gamma_i^2}\bigl((\boldsymbol{X} - \boldsymbol{A}\boldsymbol{\phi}_i)^\top \boldsymbol{\Omega}_i^{(k)} (\boldsymbol{X} - \boldsymbol{A}\boldsymbol{\phi}_i) + \Delta_i^2 \mathbf{1}^\top \boldsymbol{\beta}_i^{(k)} \\
& - 2\Delta_i (\boldsymbol{X} - \boldsymbol{A}\boldsymbol{\phi}_i)^\top \boldsymbol{\alpha}_i^{(k)}\bigr) \Biggr\},
\end{aligned}
$$

where $\mathbf{1}$ denotes a vector of ones.

Conditional maximization (CM)-steps:

In all steps for $i = 1, \ldots, g$ we have

$$
\begin{aligned}
\hat{\pi}_i^{(k+1)} &= \frac{\sum_{t=p}^{n} \hat{Z}_{ti}^{(k)}}{n}, \\
\hat{\boldsymbol{\phi}}_i^{(k+1)} &= (\boldsymbol{A}^\top \boldsymbol{\Omega}_i^{(k)} \boldsymbol{A})^{-1} \boldsymbol{A}^\top (\boldsymbol{\Omega}_i^{(k)} \boldsymbol{X} - \Delta_i^{(k)} \boldsymbol{\alpha}_i^{(k)}), \\
\hat{\Delta}_i^{(k+1)} &= \frac{\boldsymbol{\varepsilon}_i^{\top(k+1)} \boldsymbol{\alpha}_i^{(k)}}{\mathbf{1}^\top \boldsymbol{\beta}_i^{(k)}}, \quad \text{where } \boldsymbol{\varepsilon}_i^{(k+1)} = \boldsymbol{X} - \boldsymbol{A}\boldsymbol{\phi}_i^{(k+1)}, \\
\hat{\gamma}_i^{2(k+1)} &= \frac{\boldsymbol{\varepsilon}_i^{\top(k+1)} \boldsymbol{\Omega}_i^{(k)} \boldsymbol{\varepsilon}_i^{(k+1)} - 2\Delta_i^{(k+1)} \boldsymbol{\varepsilon}_i^{\top(k+1)} \boldsymbol{\alpha}_i^{(k)} + (\Delta_i^2)^{(k+1)} \mathbf{1}^\top \boldsymbol{\beta}_i^{(k)}}{\sum_{t=p}^{n} \hat{Z}_{ti}^{(k)}}.
\end{aligned}
$$

Conditional maximization likelihood (CML)-step:

$$\hat{\boldsymbol{\nu}}_i^{(k+1)} = \operatorname*{argmax}_{\boldsymbol{\nu}_i}\left(\sum_{t=1}^{n}\log\left(\sum_{i=1}^{g}\pi_i^{(k+1)}h\big(X_t \mid \boldsymbol{A}_t; \boldsymbol{\phi}_i^{(k+1)}, \Delta_i^{(k+1)}, \hat{\gamma}_i^{2(k+1)}, \boldsymbol{\nu}_i\big)\right)\right), \quad i = 1, \dots, g, \tag{12}$$

where $h(\cdot \mid \boldsymbol{A}_t, \boldsymbol{\theta}_i)$ defined below equation (9).

In order to approximate the asymptotic covariance matrix of the ML estimates obtained from the proposed MAR model, the observed information matrix is considered in the Appendix.

3.3 Initial values and model selection criteria

Determination of initial values on the ECME algorithm depends first on clustering time series data and, second, on initial values of the unknown parameters in each partition of the correspondent components of MAR models. There exist many algorithms for subsequence time series (STS) clustering. For example, Goldin, Mardales and Nagy (2006) introduced an algorithm of STS with a new distance measure, Lau and So (2008) used the so-called WCR (Weighted Chinese Restaurant) process in order to cluster time series data, and Van Wijk and Van Selow (1999) used clustering time series data with different mean levels. Time series which appear to obey threshold models can be clustered with the K-means clustering method or the partitions around medoids (PAM) clustering method; see Xiong and Yeung (2004). These methods apply to partitions of single time series data, panel time series data, clustered by ordinary clustering method of time series data into homogeneous groups.

In order to fit the SMSN-MAR(p, g) model on data set, by clustering to g categories, proportion of data points in each cluster category can be adopted for the weight initial values $\pi_i^{(0)}$, $i = 1, \dots, g$. Initial values for the vector of AR coefficients $\boldsymbol{\phi}_i^{(0)}$ in each cluster can be provided by the ordinary least square method and regression techniques as follows: for each t in each cluster, determine the p-tuples $(X_{t-p}, X_{t-p+1}, \dots, X_t)$ and regress X_t on the $(X_{t-p}, X_{t-p+1}, \dots, X_{t-1})$ and consider the estimate of the regression coefficients as the initial values of AR coefficients. In the next stage, by computing preliminary residuals of the SMSN-MAR model, initial values of $\sigma_i^{2(0)}$, $\lambda_i^{(0)}$ and $\boldsymbol{\nu}_i^{(0)}$ can be chosen via a method similar to that used by Garay, Lachos and Abanto-Valle (2011). Then the ECME algorithm iterates until a sufficient convergence rule is satisfied, e.g. if $|\ell(\hat{\boldsymbol{\theta}}^{(k+1)})/\ell(\hat{\boldsymbol{\theta}}^{(k)}) - 1| \le \varepsilon$ under the determined tolerance ε. Usually users employ $\varepsilon = 10^{-3}$, but choice of tolerance may vary with different users.

Other important issues are the determination of the order (p) of the AR process and selecting the most appropriate model from of all the competing ones (Wong and Li, 2000). Although we can use the partial auto-correlation function (PACF) to determine a tentative order p, it seems more desirable to have a systematic order selection criterion for a general AR model. Two commonly used methods are the Akaike information criterion (AIC) (Akaike, 1974) and Bayesian information criterion (BIC) (Schwarz, 1978). These criteria are defined by $\mathrm{AIC} = 2k - 2\ell(\hat{\boldsymbol{\theta}})$ and $\mathrm{BIC} = k\log n - 2\ell(\hat{\boldsymbol{\theta}})$, where $\updownarrow(\hat{\boldsymbol{\theta}})$ is the maximized log-likelihood function, k is the number of estimated parameters in the proposed AR model and n is the length (sample size) of the AR sample. In the calculation of AIC and BIC, note that $\pi_g = 1 - \pi_1 - \cdots - \pi_{g-1}$, so the number of probability components is $g - 1$.

4 Numerical studies

In this section, to compare the ability of the $SMSN - \mathrm{MAR}(p, g)$ models to fit data, at first, by simulating data from a IG-MAR(2, 2) process (The innovations from Inverse Gaussian distribution), we compute the root of the mean squared error of the predictions (RMSEP) for different members of the SMSN family. Second, by simulating from SN-MAR(2, 2) (skew-normal MAR), ST-MAR(2, 2) (skew-t MAR) and SSL-MAR(2, 2) (skew-slash MAR), we find in each case the ML estimates to ensure that the proposed SMSN-MAR models work satisfactorily. The asymptotic properties of the EM estimates are studied in terms of bias and root mean squared error (RMSE) of the estimated parameters in the third study. For computational convenience, we assume $\boldsymbol{\nu}_1 = \cdots = \boldsymbol{\nu}_g = \boldsymbol{\nu}$. The implementation of the ECME algorithms is based on the R software version 3.2.1 with a core i5 760 processor 2.8 GHz. A relative tolerance of 10^{-3} is used for convergence of the ECME algorithms. The proposed models and methods are then applied on the closing price of the Iran telecommunication company stock.

4.1 First simulation study

The first simulation study is performed from the following IG-MAR(2, 2) model with 1000 samples of different sizes $n = 100, 200, 500, 1000$:

$$X_t = \begin{cases} \boldsymbol{\phi}_1^\top \boldsymbol{A}_t + \varepsilon_{t1}; & \text{w.p. } \pi_1 = 0.4, \\ \boldsymbol{\phi}_2^\top \boldsymbol{A}_t + \varepsilon_{t2}; & \text{w.p. } \pi_2 = 0.6 \end{cases}$$

where $\boldsymbol{\phi}_1 = (\varphi_{1,0}, \varphi_{1,1}, \varphi_{1,2})^\top = (0.1, 0.5, 0.3)^\top$, $\boldsymbol{\phi}_2 = (\varphi_{2,0}, \varphi_{2,1}, \varphi_{2,2})^\top = (0.1, 0.6, 0.2)^\top$, $\{\varepsilon_{t1}\}$ are i.i.d. from IG(3, 1) distribution and $\{\varepsilon_{t2}\}$ are i.i.d from IG(2, .5) distribution, where IG(μ, λ) denotes the inverse Gaussian with mean μ and the shape parameter λ. The RMSEP of models that is calculated by

$$\mathrm{RMSEP} = \sqrt{\frac{1}{n-1}\sum_{t=2}^{n}(Y_t - \hat{Y}_t)^2},$$

under Gaussian, Student's t, skew-normal and skew-t and skew slash that are given in Table 1. As we see from Table 1, the results show that for different sample sizes, the SMSN families (SN, ST, SSL distributions) have a better fit based on the RMSE when the data is generated from the IG-MAR(2, 2) distribution.

4.2 Second simulation study

The second simulation study is devoted to an SMSN-MAR(1, 2) model with 1000 samples of length $n = 200$ obtained from the SN, ST, SSL and SCN distributions for the innovations, using the following parameters: $\boldsymbol{\phi}_1 = (\varphi_{1,0}, \varphi_{1,1})^\top = (0.1, 0.4)^\top$, $\boldsymbol{\phi}_2 = (\varphi_{2,0}, \varphi_{2,1})^\top = (0.1, 0.7)^\top$, $\pi_1 = 0.4$, $\nu_1 = \nu_2 = \nu = 3$, $\alpha_1 = \alpha_2 = \alpha = 0.2$, $\sigma_1^2 = 1$, $\sigma_2^2 = 2$, $\lambda_1 = 2$ and $\lambda_2 = 4$.

Table 1 *The RMSEP of the different models with data set from IG-MAR*(2, 2)

Sample size	Normal	Student's t	SN	ST	SSL
100	0.9658	0.9381	0.6829	0.6290	0.4901
200	0.9501	0.7881	0.5669	0.5391	0.4721
500	0.8152	0.7206	0.6281	0.5502	0.4706
1000	0.8136	0.7001	0.6291	0.4490	0.4272

Table 2 *The average Mean and SD of proposed estimates for SMSN-MAR*(2, 2) *distributions with ECME algorithm*

Parameters (values)	SN		ST		SSL		SCN	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
$\varphi_{1,0}(0.1)$	0.1012	0.02131	0.1002	0.02273	0.1021	0.02281	0.1006	0.02647
$\varphi_{2,0}(0.1)$	0.1009	0.02625	0.1006	0.02477	0.1007	0.02152	0.1001	0.02746
$\varphi_{1,1}(0.4)$	0.4101	0.02536	0.4027	0.11248	0.4301	0.20183	0.4270	0.10383
$\varphi_{2,1}(0.7)$	0.7101	0.07623	0.7007	0.02901	0.6901	0.06251	0.6903	0.02438
$\sigma_1^2(1)$	1.1023	0.02064	0.9690	0.31741	1.0210	0.27493	0.8972	0.20374
$\sigma_2^2(2)$	1.9782	0.03027	2.0190	0.28950	1.9920	0.31171	2.3847	0.36248
$\lambda_1(2)$	2.0184	0.81651	2.0025	0.94561	1.9367	0.82846	1.9026	0.47463
$\lambda_2(4)$	4.0038	1.09817	3.9014	0.95928	3.9930	0.90563	4.0436	0.64762
$\nu(3)$	–	–	3.8957	0.56842	3.6473	1.14587	3.4524	1.03821
$\alpha(0.2)$	–	–	–	–	–	–	0.1983	0.02631
$\pi_1(0.4)$	0.4011	0.04113	0.4008	0.02795	0.3957	0.01758	0.4287	0.12843

Table 2, summarizes the average means and the standard deviations (SD) of the ML estimates obtained via the ECME algorithm. In each case the proposed SMSN-MAR models perform very well.

4.3 Third simulation study

Some asymptotic properties of the estimates obtained by using the suggested ECME algorithm are investigated. Only the ST-MAR(3, 2) model and the following sets of true parameter values $\boldsymbol{\phi}_1 = (\varphi_{1,0}, \varphi_{1,1}, \varphi_{1,2})^\top = (0.2, 0.4, 0.3)^\top$, $\boldsymbol{\phi}_2 = (\varphi_{2,0}, \varphi_{2,1}, \varphi_{2,2})^\top = (0.1, 0.6, 0.2)^\top$, $\pi_1 = 0.4$, $\nu = 3$, $\sigma_1^2 = 0.1$, $\sigma_2^2 = 0.2$, $\lambda_1 = 2$ and $\lambda_2 = 4$ are considered. The bias and MSE of the ECME estimates with considering 1000 samples of different sizes $n = 100, 200, 500$ and 1000 are obtained. For example, for $\varphi_{2,0}$ these criteria are defined as

$$\text{Bias} = \frac{1}{1000}\sum_{j=1}^{1000}(\hat{\varphi}_{2,0}^j - \varphi_{2,0}), \qquad \text{MSE} = \frac{1}{1000}\sum_{j=1}^{1000}(\hat{\varphi}_{2,0}^j - \varphi_{2,0})^2$$

respectively, where $\hat{\varphi}_{2,0}^j$ is the ECME estimate of $\varphi_{2,0}$ when the data is sample j. Definitions for the other parameters are obtained by analogy. The result are shown in Table 3 that we can see when the sample size increases the ECME estimates will be precisely based on the bias and MSE.

4.4 Real data

To illustrate the potential of SMSN-MAR models, we consider the closing price of Iran telecommunication company stock (I.T.C) from July 2th, 2011, to July 2th, 2013 for 446 observations. This data set is available at Tehran securities exchange technology management company site (www.tsetmc.com). Figure 1(a) shows the time series plot of the closing price (Iranian Rial) for the I.T.C stock. From this figure, we can see that this series is not stationary, therefore the return series of the closing price is computed as $r_t = \log(p_t/p_{t-1})$ where p_t is the closing price at time t. Time plot of the stationary return series is shown in Figure 1(b). Also, the histogram of the closing price of the I.T.C. stock is shown in Figure 1(c). The histogram shows the bimodal marginal distribution of the closing price series. We consider one, two and three component of SMSN-MAR models. The order of the AR

Table 3 *Bias and MSE for ECME estimates of ST-MAR*(3, 2)

Measure	Parameters	Sample size			
		100	200	500	1000
Bias	$\varphi_{1,0}$	0.9029	0.8073	0.6072	0.4759
	$\varphi_{2,0}$	−0.8921	−0.8119	−0.9742	−0.3810
	$\varphi_{1,1}$	0.9842	0.8293	0.5703	0.5103
	$\varphi_{2,1}$	0.8091	0.8031	−0.6679	0.4089
	$\varphi_{1,2}$	−0.8341	−0.8220	−0.6011	−0.3220
	$\varphi_{2,2}$	0.7379	0.7834	0.7007	0.4193
	σ_1^2	−0.5146	−0.5011	−0.4833	−0.4023
	σ_2^2	0.6167	0.5003	0.5281	0.4461
	λ_1	−5.2981	−3.1193	−3.0081	−0.7828
	λ_2	1.0800	0.9959	0.7183	0.5030
	ν	7.1730	0.6207	0.6081	0.1172
	π_1	−0.0841	0.0681	0.0558	0.020
MSE	$\varphi_{1,0}$	0.2102	0.2048	0.2001	0.1420
	$\varphi_{2,0}$	0.1078	0.1173	0.1054	0.0991
	$\varphi_{1,1}$	0.3391	0.1841	0.1082	0.1004
	$\varphi_{2,1}$	0.0481	0.1073	0.0982	0.0429
	$\varphi_{1,2}$	0.1082	0.0892	0.0693	0.0294
	$\varphi_{2,2}$	0.6901	0.7082	0.5829	0.1073
	σ_1^2	1.2847	1.0011	1.0383	0.0403
	σ_2^2	1.4681	1.3003	0.8204	0.4307
	λ_1	9.3081	9.0029	0.9381	0.4927
	λ_2	10.0800	9.1619	0.6061	0.3954
	ν	13.1730	10.6207	1.1673	0.7379
	π_1	0.0673	0.0595	0.0393	0.0073

components is chosen by minimum the AIC. For example for the two components, Figure 2 shows the AIC plots of each component versus the different orders and we see that for the first component, minimum of the AIC is in order 3 ($p_1 = 3$) and for the second component, minimum of the AIC is in order 1 ($p_2 = 1$). By considering this data, we fit the models N-MAR (mixture AR based on the Gaussian model), SN-MAR, ST-MAR, SSL-MAR and SCN-MAR models with $g = 1, 2, 3$ components. The corresponding log-likelihood, AIC and BIC criteria are shown in Table 4. These criteria show that among the models considered, the ST-MAR(3, 2) model presents the best fit. The ML estimates with standard errors of the parameters under an ST-MAR(3, 2) model for the return series of the I.T.C. stock are given in Table 5. This two component of ST-MAR for the return series of the I.T.C. stock, is transformed back to two component of ST-MAR for the closing price of the I.T.C. stock.

5 Conclusion

In this paper, we have explored the idea of using robust MAR models in which the distribution of the innovations belong to a flexible family called scale mixtures of skew normal distributions. The proposed models serve for modeling non-linear time series data and include the family of N-MAR models as special case. In addition, they offer greater flexibility for modeling skewness and heavy tails, keeping a good degree of tractability in computing. Numerical studies show the suitability of the ML estimates under the proposed models, which

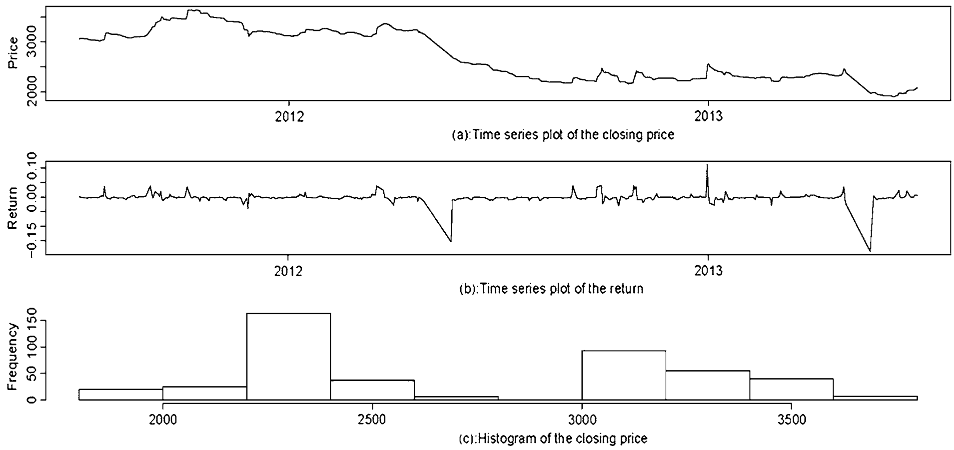

Figure 1 *Time series plot of (a): the closing price, (b) the return and (c) histogram of the I.T.C stock from* 2011–2013.

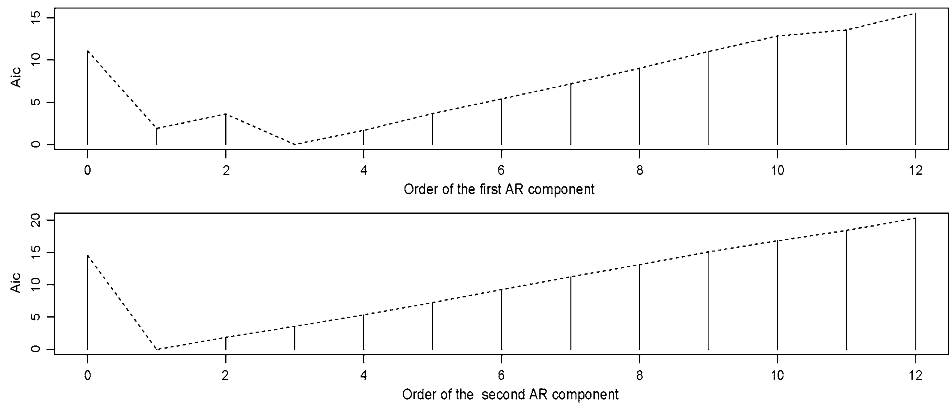

Figure 2 *AIC plot for different orders (a): the first component, (b) the second component preliminary partitioned annual Canadian Lynx Trappings data.*

Table 4 *Model selection criteria for SMSN-MAR with* $p = 3$ *and* $g = 1, 2, 3$ *under Gaussian, skew-normal and skew-t. assumption*

Model	Number of component (g)	Log-likelihood	AIC	BIC
N-MAR	1	−1638.7081	3281.4162	3288.6024
	2	−1589.6039	3079.3961	3304.0078
	3	−1606.8208	3182.0986	3217.6104
SN-MAR	1	−1629.9906	3261.8760	3270.8840
	2	−1607.8500	3040.0568	3166.3749
	3	−1610.8062	3178.6792	3170.6933
ST-MAR	1	−1457.8280	2978.9483	2985.8682
	2	**−1407.9038**	**2819.2360**	**2889.7582**
	3	−1414.0427	2913.7505	2980.7920
SSL-MAR	1	−1549.7902	3174.7835	3110.8493
	2	−1538.8931	3001.8462	3104.6580
	3	−1538.1921	3149.4836	3160.3711
SCN-MAR	1	−1560.6643	3167.3489	3193.9370
	2	−1508.2897	3012.9738	3100.9377
	3	−1557.4358	3166.8036	3183.6481

Table 5 *ML estimates with standard errors of the parameters under an ST-MAR(3, 2) model for the return series of the I.T.C. stock*

Component	Parameters	Estimates	S.E.
First Comp.	π_1	0.5617	0.0145
	$\varphi_{1,0}$	0.0043	0.0017
	$\varphi_{1,1}$	0.1498	0.0352
	$\varphi_{1,2}$	−0.0442	0.0016
	$\varphi_{1,3}$	−0.0225	0.0008
	σ_1^2	0.0031	0.0021
	λ_1	1.0290	0.2209
	ν	2.0000	0.4763
Second Comp.	π_2	0.4383	0.0145
	$\varphi_{2,0}$	0.0033	0.0009
	$\varphi_{2,1}$	0.2229	0.0521
	σ_2^2	1.0720	0.0103
	λ_2	0.9474	0.0875
	ν	2.0000	0.4763

are found using an EM-type algorithm that can be coded via free open source R software. Finally, these studies demonstrate that the SMSN-MAR models may be useful tools for modeling non-linear and non-stationary time series. For future works, researchers can apply the proposed methodology of Mahmoudi, Maleki and Pak (2017) to SMSN-MAR models, or extend the new flexible symmetric/asymmetric skew reflected Gompertz (SRG) distribution (Hoseinzadeh et al., 2018) with desirable properties to the MAR models and finally also extend them to the asymmetrical Autoregressive Moving average (ARMA) model analogy of Zarrin et al. (2018).

Appendix

A.1 Some members and properties of SMSN distributions

Also, from the hierarchical representation induced by (3) it is easy to compute the following necessary conditional moments for the application of the ECME algorithm:

$$\begin{aligned} u_r &= E\big(U^r|Y=y\big) \\ &= 2\frac{f_0(y;\mu,\sigma^2,\nu)}{f(y;\mu,\sigma^2,\lambda,\nu)}E\big(U_y^r\Phi\big(U_y^{1/2}a\big)\big), \end{aligned} \tag{A.1}$$

$$\begin{aligned} \eta_r &= E\big(U^{r/2}W_\Phi\big(U^{-1/2}a\big)|Y=y\big) \\ &= 2\frac{f_0(y;\mu,\sigma^2,\nu)}{f(y;\mu,\sigma^2,\lambda,\nu)}E\big(U_y^{r/2}\Phi\big(U_y^{1/2}a\big)\big), \end{aligned} \tag{A.2}$$

$$E(UW|Y=y) = mu_1 + M\eta_1, \tag{A.3}$$

$$E\big(UW^2|Y=y\big) = M^2 + m^2u_1 + \eta_1 Mm, \tag{A.4}$$

where $W_\Phi(\cdot) = \frac{\phi(\cdot)}{\Phi(\cdot)}$, $a = \lambda z$, $z = \frac{y-\mu}{\sigma}$, f_0 is the pdf of $Y_0 \sim \text{SMN}(\mu,\sigma^2;H)$, $U_y =^d U|(Y = y)$, $d = (y-\mu)^2/\sigma^2$, $M^2 = \frac{\gamma^2}{\gamma^2+\Delta^2}$, $m = \frac{\Delta}{\gamma^2+\Delta^2}(y-\mu)$.

Table A.1 *Some members of the SMSN distributions*

Distribution	Density ($z=\sigma^{-1}(y-\mu)$; $y \in \mathbb{R}$)	Distribution of U
Skew-normal (SN)	$2\sigma^{-1}\phi(z)\Phi(\lambda z)$	$U=1$ w.p.1
Skew-t (ST)	$2\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}\sigma}(1+\frac{z^2}{\nu})^{-\frac{\nu+1}{2}}$ $\times T(\sqrt{\frac{\nu+1}{z^2+\nu}}\lambda z; \nu+1)$	Gamma$(\frac{\nu}{2},\frac{\nu}{2})$
Skew-slash (SSL)	$2\sigma^{-1}\nu\int_0^1 u^{\nu-1}\phi(u^{-1/2}z)\Phi(u^{1/2}\lambda z)\,\mathrm{d}u$	Beta$(\nu,1)$, $\nu>0$
Skew contaminated-normal (SCN)	$2\sigma^{-1}\{\nu\alpha^{-1/2}\phi(\alpha^{1/2}z)\Phi(\alpha^{1/2}\lambda z)$ $+(1-\nu)\phi(z)\Phi(\lambda z)\}$	$\nu I_{(u_t=\alpha)}+(1-\nu)I_{(u_t=1)}$, $0<\nu<1, 0<\gamma\le 1$

A.2 The observed information matrix

To approximate the asymptotic covariance matrix of ML estimates from the proposed AR model, we use the observed information matrix, $\boldsymbol{I}(\boldsymbol{\theta} \mid \boldsymbol{X}) = -\partial^2\ell(\boldsymbol{\theta}|\boldsymbol{X})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$. Under some regularity conditions, ML estimates of the inverse of the observed information matrix, $\hat{\mathbf{I}}(\boldsymbol{\theta} \mid \boldsymbol{X})$, can be employed as an estimate of the covariance matrix of ML estimates. The elements of $\boldsymbol{I}(\boldsymbol{\theta} \mid \boldsymbol{X})$ in blocks can be calculated as $I_{\boldsymbol{\alpha},\boldsymbol{\beta}} = -\partial^2\ell_t(\boldsymbol{\theta}|\boldsymbol{X})/\partial\boldsymbol{\alpha}\partial\boldsymbol{\beta}^\top$ for $\boldsymbol{\alpha}, \boldsymbol{\beta} = \boldsymbol{\phi}_r, \pi_r, \mu_r, \sigma_r^2, \lambda_r$ and $\boldsymbol{\nu}_r$, $r=1,\ldots,g$. Hence, by considering the log-likelihood (12), we have

$$I_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \sum_{i=1}^{g}\frac{\partial^2\log\pi_i}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\beta}^\top} - \frac{1}{2}\sum_{i=1}^{g}\frac{\partial^2\log\sigma_i^2}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\beta}^\top} - \sum_{i=1}^{g}\frac{1}{K_{ti}^2}\frac{\partial K_{ti}}{\partial\boldsymbol{\alpha}}\frac{\partial K_{ti}}{\partial\boldsymbol{\beta}^\top} + \sum_{i=1}^{g}\frac{1}{K_{ti}}\frac{\partial^2 K_{ti}}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\beta}^\top},$$

where $\frac{\partial K_{ti}}{\partial\boldsymbol{\alpha}} = l_{ti}^{\phi}(1)\frac{\partial A_{ti}}{\partial\boldsymbol{\alpha}} - \frac{1}{2}l_{ti}^{\Phi}(\frac{5}{2})\frac{\partial d_{ti}}{\partial\boldsymbol{\alpha}}$, and $\frac{\partial^2 K_{ti}}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\beta}^\top} = \frac{1}{4}l_{ti}^{\Phi}(\frac{5}{2})\frac{\partial d_{ti}}{\partial\boldsymbol{\alpha}}\frac{\partial d_{ti}}{\partial\boldsymbol{\beta}^\top} - \frac{1}{2}l_{ti}^{\Phi}(\frac{3}{2})\frac{\partial^2 d_{ti}}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\beta}^\top} - \frac{1}{2}l_{ti}^{\phi}(2)(\frac{\partial A_{ti}}{\partial\boldsymbol{\alpha}}\frac{\partial d_{ti}}{\partial\boldsymbol{\beta}^\top} + \frac{\partial d_{ti}}{\partial\boldsymbol{\alpha}}\frac{\partial A_{ti}}{\partial\boldsymbol{\beta}^\top}) - l_{ti}^{\phi}(2)A_{ti}\frac{\partial A_{ti}}{\partial\boldsymbol{\alpha}}\frac{\partial d_{ti}}{\partial\boldsymbol{\beta}^\top} + l_{ti}^{\phi}(1)\frac{\partial^2 A_{ti}}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\beta}^\top}$, with $l_{ti}^{\Phi}(\omega) = E_U(U^\omega \times \exp(-Ud_{ti}/2)\Phi(\sqrt{U}A_{ti}))$, $l_{ti}^{\phi}(\omega) = E_U(U^\omega\exp(-U(d_{ti}+A_{ti}^2)/2))$, $K_{ti} = E_U(\sqrt{U} \times \exp(-Ud_{ti}/2)\Phi(\sqrt{U}A_{ti}))$, $d_{ti} = (X_t - \boldsymbol{\phi}_i^\top\boldsymbol{A}_t - b_i\Delta_i)^2/\sigma_i^2$ and $A_{ti} = \lambda_i(X_t - \boldsymbol{\phi}_i^\top\boldsymbol{A}_t - b_i\Delta_i)/\sigma_i$.

The expressions $l_{ti}^{\Phi}(\cdot)$ and $l_{ti}^{\phi}(\cdot)$ for some members of the SMSN family are as follows (for more details see Basso et al., 2010).

- *Skew-t:*

$$l_{ti}^{\Phi}(\omega) = \frac{2^\omega\nu^{\frac{\nu}{2}}\Gamma(\omega+\frac{\nu}{2})}{\sqrt{2\pi}\Gamma(\frac{\nu}{2})(\nu+d_{ti})^{\omega+\frac{\nu}{2}}}T\left(\frac{A_{ti}\sqrt{\nu+2\omega}}{(\nu+d_{ti})^{\frac{1}{2}}};\nu+2\omega\right),$$

$$l_{ti}^{\phi}(\omega) = \{\partial\nu\}\frac{2^\omega\nu^{\frac{\nu}{2}}\Gamma(\omega+\frac{\nu}{2})}{\sqrt{2\pi}\Gamma(\frac{\nu}{2})}\left(\frac{1}{\nu+A_{ti}^2+d_{ti}}\right)^{\omega+\frac{\nu}{2}},$$

and

$$\frac{\partial\ell_t(\boldsymbol{\theta}\mid\boldsymbol{X})}{\partial\nu} = \frac{1}{\sqrt{2\pi}\sigma_i}\Big(1+\log(\nu/2)+DG(\nu/2)l_{ti}^{\Phi}(1/2) - l_{ti}^{\Phi}(3/2) + \int_0^\infty u^{1/2}\log(u)\exp(-ud_{ti}/2)\Phi(\sqrt{u}A_{ti})\,\mathrm{d}H(u\mid\boldsymbol{\nu}_i)\Big),$$

where DG denotes the digamma function.

- *Skew-slash:*

$$l_{ti}^{\Phi}(\omega) = \frac{2^{2+\nu}\Gamma(\omega+\nu)}{d_{ti}^{\omega+\nu}} P(\omega+\nu, d_{ti}/2) E\big(\Phi(\sqrt{G_{ti}})A_{ti}\big),$$

$$l_{ti}^{\phi}(\omega) = \frac{\nu 2^{\omega+\nu}\Gamma(\omega+\nu)}{\sqrt{2\pi}(A_{ti}^2+d_{ti})^{\omega+\nu}} P\big(\omega+\nu, (A_{ti}^2+d_{ti})/2\big),$$

where $G_{ti} \sim \text{Gamma}(\omega+\nu, d_{ti}/2) I_{(0,1)}$ and $P(a,b)$ denotes the Gamma(a,b)-distribution function.

$$\frac{\partial \ell_t(\boldsymbol{\theta}|\boldsymbol{X})}{\partial \nu} = 2\int_0^1 u^{\nu-1}(1+\nu\log u)\phi(X_t - \boldsymbol{\phi}_i^\top \boldsymbol{A}_t; 0, u^{-1}\sigma_i^2)\Phi(\sqrt{u}A_{ti})\,du.$$

- *Skew-contaminated-normal:*

$$\begin{aligned} l_{ti}^{\Phi}(\omega) &= \sqrt{2\pi}\big(\nu\gamma^{\omega-1/2}\phi(\sqrt{d_{ti}}; 0, \gamma^{-1})\Phi(\sqrt{\gamma}A_{ti}) \\ &\quad + (1-\nu)\phi(\sqrt{d_{ti}}; 0, 1)\Phi(A_{ti})\big), \\ l_{ti}^{\phi}(\omega) &= \nu\gamma^{\omega-1/2}\phi\Big(\sqrt{A_{ti}^2+d_{ti}}; 0, \gamma^{-1}\Big) + (1-\nu)\phi\Big(\sqrt{A_{ti}^2+d_{ti}}; 0, 1\Big), \\ \frac{\partial \ell_t(\boldsymbol{\theta}|\boldsymbol{X})}{\partial \nu} &= 2(\phi(X_t - \boldsymbol{\phi}_i^\top \boldsymbol{A}_t; 0, \gamma^{-1}\sigma_i^2)\Phi(\sqrt{\gamma}A_{ti}) \\ &\quad - \phi(X_t - \boldsymbol{\phi}_i^\top \boldsymbol{A}_t; 0, \sigma_i^2)\Phi(A_{ti})\big), \end{aligned}$$

and

$$\begin{aligned} \frac{\partial \ell_t(\boldsymbol{\theta}|\boldsymbol{X})}{\partial \gamma} &= \frac{\nu}{\sqrt{2\pi}\sigma_i}\gamma^{1/2}\exp(-\gamma d_{ti}/2) \\ &\quad \times \big(\gamma^{-1}\Phi(\sqrt{\gamma}A_{ti}) + \phi(A_{ti}/\sqrt{\gamma})A_{ti}/\sqrt{\gamma} - \Phi(\sqrt{\gamma}A_{ti})d_{ti}\big). \end{aligned}$$

Acknowledgment

The authors would like thank the associate editor and two anonymous referees for their suggestions, corrections and encouragements, which helped us to improve earlier versions of the manuscript. Also to Arellano-Valle's research was partially supported by a grant from FONDECYT 1120121-Chile.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**, 716–723. MR0423716 https://doi.org/10.1109/tac.1974.1100705

Akinyemi, M. I. (2013). Mixture autoregressive models: Asymptotic properties and application to financial risk. PhD thesis, University of Manchester. MR3052080

Andrews, D. F. and Mallows, C. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society, Series B* **36**, 99–102. MR0359122

Arellano-Valle, R. B., Bolfarine, H. and Lachos, V. H. (2005). Skew-normal linear mixed models. *Journal of Data Science* **3**, 415–438. MR2410041 https://doi.org/10.1080/02664760701236905

Arellano-Valle, R. B., Bolfarine, H. and Lachos, V. H. (2007). Bayesian inference for skew-normal linear mixed models. *Journal of Applied Statistics* **34**, 663–682. MR2410041 https://doi.org/10.1080/02664760701236905

Arellano-Valle, R. B., Castro, L. M., Genton, M. G. and Gómez, H. W. (2008). Bayesian inference for shape mixtures of skewed distributions, with application to regression analysis. *Bayesian Analysis* **3**, 513–539. MR2434401 https://doi.org/10.1214/08-BA320

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* **12**, 171–178. MR0808153

Azzalini, A. and Capitanio, A. (2014). *The Skew-Normal and Related Families*. *IMS Monographs Series*. Cambridge: Cambridge University Press. MR3468021

Basso, R. M., Lachos, V. H., Cabral, C. R. B. and Ghosh, P. (2010). Robust mixture modeling based on the scale mixtures of skew-normal distributions. *Computational Statistics & Data Analysis* **54**, 2926–2941. MR2727724 https://doi.org/10.1016/j.csda.2009.09.031

Böhning, D. (2000). *Computer-Assisted Analysis of Mixtures and Applications. Meta-Analysis, Disease Mapping and Others*. Boca Raton: Chapman & Hall/CRC. MR1684363

Böhning, D., Hennig, C., McLachlan, G. J. and McNicholas, P. D. (2014). Editorial: The 2nd special issue on advances in mixture models. *Computational Statistics & Data Analysis* **71**, 1–2. MR3131949 https://doi.org/10.1016/j.csda.2013.10.010

Böhning, D., Seidel, W., Alfó, M., Garel, B., Patilea, V. and Walther, G. (2007). Editorial: Advances in mixture models. *Computational Statistics & Data Analysis* **51**, 5205–5210. MR2370865 https://doi.org/10.1016/j.csda.2006.10.025

Branco, M. D. and Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis* **79**, 99–113. MR1867257 https://doi.org/10.1006/jmva.2000.1960

Cervone, D., Pillai, N. S., Pati, D., Berbeco, R. and Lewis, J. H. (2014). A location-mixture autoregressive model for online. Forecasting of lung tumor motion. *Annals of Applied Statistics* **8**, 1341–1371. MR3271335 https://doi.org/10.1214/14-AOAS744

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B, Methodological* **39**, 1–22. MR0501537

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. New York: Springer. MR2265601

Garay, A. M., Lachos, V. H. and Abanto-Valle, C. A. (2011). Nonlinear regression models based on scale mixtures of skew-normal distributions. *Journal of the Korean Statistical Society* **40**, 115–124. MR2752409 https://doi.org/10.1016/j.jkss.2010.08.003

Ghasami, S., Khodadadi, Z. and Maleki, M. (2019). Autoregressive processes with generalized hyperbolic innovations. *Communications in Statistics Simulation and Computation*. https://doi.org/10.1080/03610918.2018.1535066

Glasbey, C. A. (2001). Non-linear autoregressive time series with multivariate Gaussian mixtures as marginal distributions. *Journal of the Royal Statistical Society Series C* **50**, 143–154. MR1833269 https://doi.org/10.1111/1467-9876.00225

Goldin, D. Q., Mardales, R. and Nagy, G. (2006). In search of meaning for time series subsequence clustering: Matching algorithms based on a new distance measure. In *Proc. of the ACM International Conference on Information and Knowledge Management*, 347–356.

Hajrajabi, A. and Maleki, M. (2019). Nonlinear semiparametric autoregressive model with finite mixtures of scale mixtures of skew normal innovations. *Journal of Applied Statistics* **46**(11), 2010–2029. https://doi.org/10.1080/02664763.2019.1575953

Hoseinzadeh, A., Maleki, M., Khodadadi, Z. and Contreras-Reyes, J. E. (2018). The Skew-Reflected-Gompertz distribution for analyzing the symmetric and asymmetric data. *Journal of Computational and Applied Mathematics* **349**, 132–141. MR3886701 https://doi.org/10.1016/j.cam.2018.09.011

Jin, S. and Li, W. K. (2006). Modeling panel time series with mixture autoregressive model. *Journal of Data Science* **4**, 425–446.

Lanne, M. and Saikkonen, P. (2003). Modeling the U.S. short-term interest rate by mixture autoregressive processes. *Journal of Financial Econometrics* **1**, 96–125.

Lau, J. and So, M. K. P. (2008). Bayesian mixture of autoregressive models. *Computational Statistics & Data Analysis* **53**, 38–60. MR2528591 https://doi.org/10.1016/j.csda.2008.06.001

Lindsay, B. G. (1995). *Mixture Models: Theory Geometry and Applications*. *NSF-CBMS Regional Conference Series in Probability and Statistics* **51**. Harward: Institute of Mathematical Statistics.

Liu, C. and Rubin, D. B. (1994). The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence. *Biometrika* **80**, 267–278. MR1326414 https://doi.org/10.1093/biomet/81.4.633

Liu, M. and Lin, T. I. (2014). A skew-normal mixture regression model. *Educational and Psychological Measurement* **74**, 139–162.

Mahmoudi, M. R., Maleki, M. and Pak, A. (2017). Testing the difference between two independent time series models. *Iranian Journal of Science and Technology, Transaction A, Science* **41**, 665–669. MR3714808 https://doi.org/10.1007/s40995-017-0288-8

Maleki, M. and Arellano-Valle, R. B. (2017). Maximum a-posteriori estimation of autoregressive processes based on finite mixtures of scale-mixtures of skew-normal distributions. *Journal of Statistical Computation and Simulation* **87**, 1061–1083. MR3606839 https://doi.org/10.1080/00949655.2016.1245305

Maleki, M., Arellano-Valle, R. B., Dey, D. K., Mahmoudi, M. R. and Jalali, S. M. J. (2017). A Bayesian approach to robust skewed autoregressive processes. *Calcutta Statistical Association Bulletin* **69**, 165–182. MR3746983 https://doi.org/10.1177/0008068317732196

Maleki, M. and Mahmoudi, M. R. (2017). Two-piece location-scale distributions based on scale mixtures of normal family. *Communications in Statistics Theory and Methods* **46**, 12356–12369. MR3740871 https://doi.org/10.1080/03610926.2017.1295160

Maleki, M. and Nematollahi, A. R. (2017a). Bayesian approach to epsilon-skew-normal family. *Communications in Statistics Theory and Methods* **46**, 7546–7561. MR3648583 https://doi.org/10.1080/03610926.2016.1157186

Maleki, M. and Nematollahi, A. R. (2017b). Autoregressive models with mixture of scale mixtures of Gaussian innovations. *Iranian Journal of Science and Technology Transaction A, Science* **41**, 1099–1107. MR3736837 https://doi.org/10.1007/s40995-017-0237-6

Maleki, M. and Wraith, D. (2019). Mixtures of multivariate restricted skew-normal factor analyzer models in a Bayesian framework. *Computational Statistics* **34**, 1039–1053. https://doi.org/10.1007/s00180-019-00870-6

Maleki, M., Wraith, D. and Arellano-Valle, R. B. (2018a). Robust finite mixture modeling of multivariate unrestricted skew-normal generalized hyperbolic distributions. *Statistics and Computing* Available at https://doi.org/10.1007/s11222-018-9815-5. https://doi.org/10.1007/s11222-018-9815-5

Maleki, M., Wraith, D. and Arellano-Valle, R. B. (2018b). A flexible class of parametric distributions for Bayesian linear mixed models. *Test* Available at https://doi.org/10.1007/s11749-018-0590-6. https://doi.org/10.1007/s11749-018-0590-6

McCulloch, R. E. and Tsay, R. S. (1994). Statistical inference of macroeconomic time series via Markov switching models. *Journal of Time Series Analysis* **15**, 523–539. MR1263893 https://doi.org/10.1111/j.1467-9892.1994.tb00188.x

McLachlan, G. J. and Krishnan, T. (2008). *The eM Algorithm and Extensions*. New Jersey: Wiley. MR2392878 https://doi.org/10.1002/9780470191613

McLachlan, G. J. and Peel, D. (2000). *Finite Mixture Models*. New York: Wiley. MR1789474 https://doi.org/10.1002/0471721182

Mengersen, K., Robert, C. P. and Titterington, D. M. (2011). *Mixtures: Estimation and Applications*. Chichester: Wiley. MR2867716 https://doi.org/10.1002/9781119995678

Moravveji, B., Khodadai, Z. and Maleki, M. (2018). A Bayesian analysis of two-piece distributions based on the scale mixtures of normal family. *Iranian Journal of Science and Technology Transaction A, Science* Available at https://doi.org/10.1007/s40995-018-0541-9. https://doi.org/10.1007/s40995-018-0541-9

Ni, H. and Yin, H. (2008). A self-organizing mixture autoregressive network for time series modeling and prediction. *Neurocomputing* **72**, 3529–3537.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics* **6**, 461–464. MR0468014

Van Wijk, J. J. and Van Selow, E. R. (1999). Cluster and calendar-based visualization of time series data. In *Proceedings of the IEEE Symposium on Information Visualization*, 4–9.

Wei, G. C. G. and Tanner, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association* **85**, 699–704.

Wong, C. S. (1998). Statistical inference for some non-linear time series models. PhD thesis, University of Hong.Kong. Hong Kong.

Wong, C. S. and Chan, W. S. (2009). A Student-*t* mixture autoregressive model with applications to heavy tailed financial data. *Biometrika* **96**, 751–760. MR2538770 https://doi.org/10.1093/biomet/asp031

Wong, C. S. and Li, W. K. (2000). On a mixture autoregressive model. *Journal of the Royal Statistical Society, Series B* **62**, 95–115. MR1747398 https://doi.org/10.1111/1467-9868.00222

Wong, C. S. and Li, W. K. (2001a). On a mixture autoregressive conditional heteroscedastic model. *Journal of American Statistical Association, Series B* **96**, 982–995. MR1946366 https://doi.org/10.1198/016214501753208645

Wong, C. S. and Li, W. K. (2001b). On a logistic mixture autoregressive model. *Biometrika* **88**, 833–846. MR1859413 https://doi.org/10.1093/biomet/88.3.833

Wood, S., Rosen, O. and Kohn, R. (2011). Bayesian mixtures of autoregressive models. *Journal of Computational and Graphical Statistics* **20**, 174–195. MR2816544 https://doi.org/10.1198/jcgs.2010.09174

Xiong, Y. and Yeung, D. (2004). Time series clustering with ARMA mixtures. *Pattern Recognition* **37**, 1675–1689.

Zarrin, P., Maleki, M., Khodadadi, Z. and Arellano-Valle, R. B. (2018). Time series process based on the unrestricted skew normal process. *Journal of Statistical Computation and Simulation* **89**, 38–51. MR3871128 https://doi.org/10.1080/00949655.2018.1533962

Zeller, C. B., Lachos, V. H. and Vilca-Labra, F. E. (2011). Local influence analysis for regression models with scale mixtures of skew-normal distributions. *Journal of Applied Statistics* **38**, 348–363. MR2749591 https://doi.org/10.1080/02664760903406504

M. Maleki
Department of Statistics
Shiraz University
Shiraz
Iran
E-mail: m.maleki.stat@gmail.com

A. Hajrajabi
Department of Statistics
Faculty of Basic Sciences
Imam Khomeini International University
Qazvin
Iran

R. B. Arellano-Valle
Departamento de Estadística Pontificia
Universidad Católica de Chile
Santiago
Chile