JIRSS (2019)
Vol. 18, No. 01, pp 157-175
DOI: 10.29252/jirss.18.1.157

Classical and Bayesian Estimation of the AR(1) Model with Skew-Symmetric Innovations

Arezo Hajrajabi and Afshin Fallah

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

Received: 06/09/2018, Revision received: 08/11/2018, Published online: 03/28/2019

Abstract. This paper considers a first-order autoregressive model with skew-normal innovations from a parametric point of view. We develop an essential theory for computing the maximum likelihood estimation of model parameters via an Expectation-Maximization (EM) algorithm. Also, a Bayesian method is proposed to estimate the unknown parameters of the model. The efficiency and applicability of the proposed model are assessed via a simulation study and a real-world example.

Keywords. Autoregressive model, Bayesian inference, EM algorithm, Maximum likelihood estimator, Skew-normal innovations.

MSC: 62F10; 62-07; 62F15; 62M10.

1 Introduction

Autoregressive (AR) models are the most widely used class of time series models. Traditionally, an AR model is analyzed based on the normality assumption of innovations by considering them as a sequence of uncorrelated zero mean normal random

Corresponding Author: Arezo Hajrajabi (hajrajabi@sci.ikiu.ac.ir)
Afshin Fallah (a.fallah@sci.ikiu.ac.ir)

variables (Pourahmadi , 2001) and (Shumway and Stofer, 2006). Whereas, in many applications the data violate normality and the conventional normal based theory could not provide an adequate description for data (Tarami and Pourahmadi, 2003). Bondon (2009) introduced a non-Gaussian AR model with epsilon skew-normal innovations and provided the method of moments and maximum likelihood (ML) estimators of the model parameters and their corresponding limiting distributions. Following Bondon (2009), Sharafi and Nematollahi (2016) considered the skew-normal distribution introduced by Azzalini (1985) for the innovations. Also, the semiparametric analyzes of the nonlinear AR(1) model with skew-symmetric innovations has been investigated by Hajrajabi and Fallah (2017).

In this paper, we allow the AR model innovations to have the skew-normal distribution. This is different from the work of Sharafi and Nematollahi (2016) in the way that we are interested in the parametric estimation of the parameters through the ML and Bayes approaches, instead of nonparametric estimation via least squares (LS) method. We provide closed iterative forms for the ML estimators of the parameter using an EM type optimization methodology. We also develop a Gibbs algorithm for estimating the parameters in the Bayesian paradigm.

The rest of the paper unfolds as follows. Section 2, briefly outlines some theoretical and preliminaries of the skew-normal distribution. Conditional maximum likelihood (CML) estimation in the linear AR model using the EM algorithm are investigated in Section 3. Also, Bayesian estimation of the unknown parameters through the Metropolis-Hastings and Gibbs schemes are derived in this section. A simulation study is conducted to verify the accuracy of the proposed method in Section 4. An application of the model in the daily return series of Mellat bank of Iran is explained in Section 5. Finally, some conclusions are given in Section 6.

2 The Proposed Model

In this section, we develop an AR(1) model under the assumption of skew-normal distribution for innovations.

A random variable Y follows a univariate skew-normal distribution with location parameter μ, scale parameter w^2 and skewness parameter α, $Y \sim SN(\mu, w^2, \alpha)$, if its density function is given by

$$f_{SN}(y|\mu, w^2, \alpha) = \frac{2}{w}\phi(\frac{y-\mu}{w})\Phi(\alpha\frac{y-\mu}{w}), \qquad \mu, \alpha \in \Re, w \in \Re^+, \tag{2.1}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the density and cumulative distribution function of the standard normal distribution, respectively. Note that if $\alpha = 0$, the density of Y in equation (2.1) reduces to the $N(\mu, w^2)$ density. Some brief theoretical results about the skew-normal distribution that is necessary in the next sections are presented in the Appendix. Consider an autoregressive model of order one (denoted by AR(1)) for a process y_t as

$$y_t = \phi y_{t-1} + a_t, \qquad t = 2, ..., n, \tag{2.2}$$

where $|\phi| < 1$ and a_t, as the innovation process, follows a skew-normal distribution, $SN(\mu, w^2, \alpha)$.

The motivation for considering skew-normal distribution instead of normal arises from the fact that the former distribution is useful for describing asymmetric and heavy-tailed data. In what follows, we try to estimate the parameters of the model (2.2) using both the CML and Bayes approaches.

2.1 Conditional Maximum Likelihood Approach

Using Lemma 6.3 in the Appendix and defining the variables $S_t \sim TN(0, w^2)I_{\{S_t>0\}}$, $U_t \sim N(0, w^2)$ and $Y_t = \phi Y_{t-1} + \mu + \delta(\alpha)S_t + \sqrt{1-\delta^2(\alpha)}U_t$, where $\delta(\alpha) = \frac{\alpha}{\sqrt{1+\alpha^2}}$ and $TN(a, b)$ denotes the truncated $N(a, b)$ distribution at $(0, \infty)$, the conditional distribution of observations in the model (2.2) can be written as

$$Y_t|Y_{t-1} \sim SN(\phi Y_{t-1} + \mu, w^2, \alpha), \qquad t = 2, ..., n. \tag{2.3}$$

Given data $\boldsymbol{y} = (y_2, ..., y_n)$, the conditional likelihood function of the model (2.3) is

$$\begin{aligned} L(\mu, \phi, w^2, \alpha|\boldsymbol{y}) &= \prod_{t=2}^{n} f_{SN}(y_t|y_{t-1}, \boldsymbol{\theta}) \\ &= \prod_{t=2}^{n} \frac{2}{w}\phi(\frac{y_t - \mu - \phi y_{t-1}}{w})\Phi(\alpha(\frac{y_t - \mu - \phi y_{t-1})}{w})\} \\ &= (\frac{2}{w})^{n-1}(2\pi)^{-\frac{n-1}{2}} \exp\{\frac{-1}{2w^2}\sum_{t=2}^{n}(y_t - \mu - \phi y_{t-1})^2\} \\ &\times \prod_{t=2}^{n} \Phi(\alpha(\frac{y_t - \mu - \phi y_{t-1}}{w})), \end{aligned} \tag{2.4}$$

where $\boldsymbol{\theta} = (\phi, \mu, w^2, \alpha)$ is the vector of unknown parameters. Due to the complexity of the likelihood function (2.4), there are no closed-form expressions of the ML parameter estimators.

Therefore, we provide an EM algorithm (Dempster *et al.* , 1977) and (Lin *et al.* , 2007) to compute the numerical values of the ML estimates. Hence, it is necessary to formulate the problem in terms of a missing data problem. By considering $\boldsymbol{s} = (s_2, ..., s_n)$ and $\boldsymbol{y}$, respectively, as missing and incomplete data and using Lemma 6.3 in the Appendix, the joint density of the Y_t and S_t is given by

$$\begin{aligned} f_{Y_t,S_t}(y_t, s_t) &= \frac{\sqrt{1+\alpha^2}}{\pi w^2} \exp\{\frac{-1}{2w^2}[(y_t - \phi y_{t-1} - \mu)^2 \\ &+ (1+\alpha^2)(s_t - \frac{\alpha}{\sqrt{1+\alpha^2}}(y_t - \phi y_{t-1} - \mu))^2]\} \\ &= \frac{1}{\pi w^2 \sqrt{1-\delta^2(\alpha)}} \exp\{\frac{-1}{2w^2(1-\delta^2(\alpha))} \\ &\quad [(y_t - \phi y_{t-1} - \mu)^2 - 2s_t\delta(\alpha)(y_t - \phi y_{t-1} - \mu)) + s_t^2]\}. \end{aligned}$$

Hence, the complete data likelihood and log-likelihood functions are obtained as

$$\begin{aligned} L_c(\mu, \phi, w^2, \alpha | \boldsymbol{y}, \boldsymbol{s}) &= \prod_{t=2}^{n} f_{Y_t,S_t}(y_t, s_t) \\ &= (\pi w^2)^{-(n-1)}(1-\delta^2(\alpha))^{-\frac{n-1}{2}} \\ &\times \exp\{\frac{-1}{2w^2(1-\delta^2(\alpha))}[\sum_{t=2}^{n}[(y_t - \phi y_{t-1} - \mu)^2 \\ &- 2s_t\delta(\alpha)(y_t - \phi y_{t-1} - \mu)) + s_t^2]]\}. \end{aligned} \tag{2.5}$$

and

$$\begin{aligned} \ell_c(\mu, \phi, w^2, \alpha | \boldsymbol{y}, \boldsymbol{s}) &= -(n-1)\log(w^2) - \frac{(n-1)}{2}\log(1-\delta^2(\alpha)) \\ &- \frac{1}{2w^2(1-\delta^2(\alpha))}[\sum_{t=2}^{n}(y_t - \phi y_{t-1} - \mu)^2 \\ &- 2\delta(\alpha)\sum_{t=2}^{n} s_t(y_t - \phi y_{t-1} - \mu)) + \sum_{t=2}^{n} s_t^2], \end{aligned}$$

respectively. For executing the EM algorithm, using Lemma 6.4 in the Appendix, we have

$$S_t|Y_t = y_t \sim TN(\mu_{S_t}, \sigma_S^2)I_{s_t>0},$$

where

$$\begin{aligned}\mu_{S_t} &= \delta(\alpha)(y_t - \phi y_{t-1} - \mu),\\ \sigma_S^2 &= w^2(1 - \delta^2(\alpha)).\end{aligned} \tag{2.6}$$

In the E-step of the EM algorithm, the conditional expectation of complete data log-likelihood given the observed data and the current parameters is computed as below

$$\begin{aligned}E[\ell_c(\mu, \phi, w^2, \alpha|\boldsymbol{y}, \boldsymbol{s})|\boldsymbol{y}] &= -(n-1)\log(w^2) - \frac{(n-1)}{2}\log(1 - \delta^2(\alpha))\\ &- \frac{1}{2w^2(1 - \delta^2(\alpha))}[\sum_{t=2}^{n}(y_t - \phi y_{t-1} - \mu)^2\\ &- 2\delta(\alpha)\sum_{t=2}^{n} A_{1t}(y_t - \phi y_{t-1} - \mu)) + \sum_{t=2}^{n} A_{2t}],\end{aligned} \tag{2.7}$$

where A_{1t} and A_{2t} are computed, using Lemma 6.4 in the Appendix, as follow

$$\begin{aligned}A_{1t} &= E[S_t|Y_{t-1} = y_{t-1}] = \mu_{S_t} + \frac{\phi(\frac{\mu_{S_t}}{\sigma_S})}{\Phi(\frac{\hat{\mu}_{S_t}}{\sigma_S})}\sigma_S,\\ A_{2t} &= E[S_t^2|Y_{t-1} = y_{t-1}] = \mu_{S_t}^2 + \sigma_S^2 + \frac{\phi(\frac{\mu_{S_t}}{\sigma_S})}{\Phi(\frac{\mu_{S_t}}{\sigma_S})}\sigma_S\mu_{S_t},\end{aligned}$$

and μ_{S_t} and σ_S^2 are given in equation (2.6). In the M-step, the algorithm finds values in the parameter space that maximizes the conditional expectation in equation (2.7). Given the values of parameters in iteration k, the updated estimates are obtained as follow

$$\begin{aligned}\hat{\mu}^{(k+1)} &= \frac{1}{n-1}[\sum_{t=2}^{n}(y_t - \hat{\phi}^{(k)}y_{t-1}) - \delta(\hat{\alpha}^{(k)})\sum_{t=2}^{n} A_{1t}^{(k)}],\\ \hat{\phi}^{(k+1)} &= \frac{\sum_{t=2}^{n} y_{t-1}(y_t - \hat{\mu}^{(k+1)}) - \sum_{t=2}^{n}\delta(\hat{\alpha}^{(k)})A_{1t}^{(k)}y_{t-1}}{\sum_{t=2}^{n} y_{t-1}^2},\end{aligned}$$

$$\hat{w}^{2(k+1)} = \frac{1}{2(n-1)(1-\delta^2(\hat{\alpha}^{(k)}))}[\sum_{t=2}^{n} A_{2t}^{(k)} - 2\delta(\hat{\alpha}^{(k)})\sum_{t=2}^{n} A_{1t}^{(k)}(y_t - \hat{\phi}^{(k+1)}y_{t-1} - \hat{\mu}^{(k+1)})$$
$$+ \sum_{t=2}^{n}(y_t - \hat{\phi}^{(k+1)}y_{t-1} - \hat{\mu}^{(k+1)})^2],$$
$$\hat{\alpha}^{(k+1)} = \arg\max_{\alpha} \sum_{t=2}^{n} \log\left[\Phi(\alpha(\frac{y_t - \hat{\mu}^{(k+1)} - \hat{\phi}^{(k+1)}y_{t-1}}{\hat{w}^{(k+1)}}))\right].$$

As can be seen, computing $\hat{\alpha}^{(k+1)}$ requires a one-dimensional optimization, which can be easily done by using any iterative algorithm such as Newton-Raphson. The values of the moment estimates, given in Lemma 6.2 in the Appendix, can be used as initial values for the EM algorithm.

2.2 Bayesian Approach

It is well known that the ML estimates are less accurate for small sample sizes. In this situation, the Bayesian approach provides an interesting alternative methodology for inference and modeling especially when prior information about the unknown parameters is available (Ibazizen and Fellag , 2003). The Bayesian approach incorporates our prior knowledge about parameters, in terms of the prior distributions, and the information obtained from observations via Bayes rule.

To estimate the unknown parameters $\boldsymbol{\theta} = (\phi, \mu, w^2, \alpha)$ using Bayesian methodology, it is necessary to assume $\boldsymbol{\theta}$ is a random vector with some probability distribution. Therefore, to complete the model setup, a prior distribution for the parameter vector $\boldsymbol{\theta}$ needs to be specified. Following Arellano-Valle and Azzalini (2006), Kastner (2016) and De Oliveira (2012), we set the following prior distributions

$$\begin{aligned}
\mu &\sim N(\lambda_0, \frac{1}{k_0}),\\
\phi &\sim U(-1,1),\\
w^{-2} &\sim Gamma(a_0, b_0),\\
\alpha|\mu_\alpha, \sigma_\alpha^2 &\sim N(\mu_\alpha, \sigma_\alpha^2),\\
\mu_\alpha &\sim TN(-5, 5, m_0, s_0^2),\\
\sigma_\alpha^{-2} &\sim Gamma(c_0, d_0),
\end{aligned} \tag{2.8}$$

where $TN(a, b, \mu, \sigma^2)$ denotes the truncated normal distribution with parameters μ and σ^2 within the range (a, b). As it can be seen, due to the importance of skewness parameter α the proposed model, we considered a two-level hierarchical prior distribution for this parameter to improve the robustness of the resulting Bayes estimators. As it is noted by Robert (2001), it is rarely necessary to consider more than two conditional levels in the hierarchical modeling and usually considering only two levels of hierarchy suffices to achieve robust results, see Lindley and Smith (1972) and Good (1980) for more information about hierarchical Bayes analysis. By considering the complete likelihood of data in equation (2.5), the joint posterior distribution of parameters obtained as follow

$$\begin{aligned}
\pi(\mu, \phi, w^2, \alpha|\boldsymbol{y}, \boldsymbol{s}) &= L_c(\phi, \mu, w^2, \alpha|\boldsymbol{y}, \boldsymbol{s})\pi(\mu, \phi, w^2, \alpha) \\
&\propto L_c(\phi, \mu, w^2, \alpha|\boldsymbol{y}, s)\pi(\mu)\pi(\phi)\pi(w^2)\pi(\alpha|\mu_\alpha, \sigma^2_\alpha)\pi(\mu_\alpha)\pi(\sigma^2_\alpha) \\
&\propto (\pi w^2)^{-(n-1)}(1-\delta^2(\alpha))^{-\frac{n-1}{2}} \\
&\times \exp\{\frac{-1}{2w^2(1-\delta^2(\alpha))}[\sum_{t=2}^{n}[(y_t - \phi y_{t-1} - \mu)^2 \\
&- 2s_t\delta(\alpha)(y_t - \phi y_{t-1} - \mu)) + s_t^2]]\} \\
&\times \exp\{-\frac{k_0}{2}(\mu - \lambda_0)^2\} \times (w^{-2})^{(a_0-1)} \exp\{-b_0 w^{-2}\} \\
&\times \exp\{-\frac{1}{2\sigma^2_\alpha}(\alpha - \mu_\alpha)^2\} \times (\sigma_\alpha^{-2})^{(c_0-1)} \exp\{-d_0\sigma_\alpha^{-2}\} \\
&\times \frac{\phi(\frac{\mu_\alpha - m_0}{s_0})}{s_0(\phi(\frac{5-m_0}{s_0}) - \phi(\frac{-5-m_0}{s_0}))}.
\end{aligned} \tag{2.9}$$

As clearly be seen, the complexity of the posterior distribution in the equation (2.9) precludes analytical treating of the Bayes estimators of the parameters. Therefore, in what follows, we provide a Gibbs algorithm to sample from joint posterior distribution and provide desirable posterior inference about parameters. Using the joint posterior distribution of the parameters given in equation (2.9) and considering the missing data as the unknown parameters, we have

$$\begin{aligned}
\pi(s_t|others) &\propto \exp\{\frac{-1}{2w^2(1-\delta^2(\alpha))}[\sum_{t=2}^{n}[s_t^2 - 2s_t\delta(\alpha)(y_t - \phi y_{t-1} - \mu)]]\}, \\
\pi(\mu|others) &\propto \exp\{\frac{-1}{2w^2(1-\delta^2(\alpha))}[\mu^2[(n-1) + kw^2(1-\delta^2(\alpha))] \\
&- 2\mu[\sum_{t=2}^{n}(y_t - \phi y_{t-1}) - \delta_(\alpha)\sum_{t=2}^{n} s_t + k\lambda w^2(1-\delta^2(\alpha))]]\},
\end{aligned}$$

$$
\begin{aligned}
\pi(\phi|others) &\propto \exp\{\frac{-1}{2w^2(1-\delta^2(\alpha))}[\sum_{t=2}^{n}\phi^2 y_{t-1}^2 - 2\phi\sum_{t=2}^{n}[y_{t-1}(y_t-\mu)-\delta(\alpha)s_t y_{t-1}]]\},\\
\pi(w^{-2}|others) &\propto (w^{-2})^{a_0+n-2}\exp\{-w^{-2}[\frac{-1}{2w^2(1-\delta^2(\alpha))}\sum_{t=2}^{n}[(y_t-\phi y_{t-1}-\mu)^2\\
&- 2s_t\delta(\alpha)(y_t-\phi y_{t-1}-\mu))+s_t^2]+b]\},\\
\pi(\mu_\alpha|others) &\propto \exp\{-\frac{1}{2\sigma_\alpha^2}(\alpha-\mu_\alpha)^2\}\frac{\phi(\frac{\mu_\alpha-m_0}{s_0})}{s_0(\phi(\frac{5-m_0}{s_0})-\phi(\frac{-5-m_0}{s_0}))},\\
\pi(\sigma_\alpha^{-2}|others) &\propto \exp\{-\frac{1}{2\sigma_\alpha^2}(\alpha-\mu_\alpha)^2\}(\sigma_\alpha^{-2})^{(c_0-1)}\exp\{-d_0\sigma_\alpha^{-2}\},\\
\pi(\alpha|others) &\propto (1-\delta^2(\alpha))^{-\frac{n-1}{2}}\exp\{\frac{-1}{2w^2(1-\delta^2(\alpha))}[\sum_{t=2}^{n}[(y_t-\phi y_{t-1}-\mu)^2\\
&- 2s_t\delta(\alpha)(y_t-\phi y_{t-1}-\mu))+s_t^2]]\}\exp\{-\frac{1}{2\sigma_\alpha^2}(\alpha-\mu_\alpha)^2\},
\end{aligned}
$$

Therefore, the necessary full conditional distributions at the (k+1)-th iteration of the Gibbs algorithm are given by

$$
\begin{aligned}
s_t^{(k+1)}|others &\sim TN(\delta(\alpha^{(k)})(y_t-\phi^{(k)}y_{t-1}-\mu^{(k)}), w^{2(k)}(1-\delta^2(\alpha^{(k)})),\\
\mu^{(k+1)}|others &\sim N(M_\mu^{(k+1)}, S_\mu^{2(k+1)}),\\
\phi^{(k+1)}|others &\sim N(M_\phi^{(k+1)}, S_\phi^{2(k+1)}),\\
w^{-2(k+1)}|others &\sim Gamma((n-1)+a_0, \gamma^{(k+1)}+b),\\
\sigma_\alpha^{-2(k+1)}|others &\sim Gamma(c_0+\frac{1}{2}, d_0+\frac{(\alpha^{(k)}-\mu_\alpha^{(k+1)})^2}{2}).
\end{aligned}
$$

where

$$
\begin{aligned}
M_\mu^{(k+1)} &= \frac{\sum_{t=2}^{n}(y_t-\phi^{(k)}y_{t-1})-\delta(\alpha^{(k)})\sum_{t=2}^{n}s_t^{(k+1)}+k\lambda w^{2(k)}(1-\delta^2(\alpha^{(k)})}{(n-1)+kw^{2(k)}(1-\delta^2(\alpha^{(k)}))},\\
S_\mu^{2(k+1)} &= [\frac{(n-1)}{w^{2(k)}(1-\delta^2(\alpha^{(k)})}+k]^{-1},\\
\gamma^{(k+1)} &= \frac{-1}{2w^{2(k)}(1-\delta^2(\alpha^{(k)}))}[\sum_{t=2}^{n}[(y_t-\phi^{(k)}y_{t-1}-\mu^{(k+1)})^2\\
&- 2s_t^{(k)}\delta(\alpha^{(k)})(y_t-\phi^{(k)}y_{t-1}-\mu^{(k+1)}))+s_t^{2(k+1)}]],
\end{aligned}
$$

$$M_{\phi}^{(k+1)} = \frac{\sum_{t=2}^{n} y_{t-1}(y_t - \mu^{(k+1)}) - \delta(\alpha^{(k)})s_t^{(k+1)}y_{t-1}}{\sum_{t=2}^{n} y_{t-1}^2},$$
$$S_{\phi}^{2(k+1)} = [\frac{\sum_{t=2}^{n} y_{t-1}^2}{w^{2(k)}(1-\delta^2(\alpha^{(k)}))}]^{-1}.$$

Because of to the lack of distributional construction in full conditional posterior distribution of parameters μ_α and α, the Metropolis-Hastings algorithm is used for sampling of these distributions.

For μ_α we use the following schema: we draw $\mu_\alpha^{(k+1)} = \mu_\alpha^{(k)} + N(0, \tau_1)$ where τ_1 is a tuning parameter used to adjust the acceptance probability by an opening of the window technique, i.e. by trying several tuning parameters and deciding for a good compromise. Then, $\mu^{(k+1)}$ is accepted as an observation from the corresponding posterior distribution if $u < \rho$, where

$$\rho = \min\{1, \exp(\log(\pi(\mu_\alpha'|others)) - \log(\pi(\mu_\alpha^{(k)}|others))\}.$$

denotes the Metropolis ratio and u is an observation from the uniform distribution on the interval $(0, 1)$. The parameter α is updated in exactly the same manner.

3 Simulation Study

In this section, we perform a simulation study to assess the efficiency of the ML and Bayesian estimators in the model (2.2). For the first purpose, the values of the model parameters are set to be $\mu = 1, w^2 = 1$ and $\phi = 0.2$. To evaluate the effect of sample size on the efficiency of the proposed methods, we consider different sample sizes $n = 50, 100, 500, 1000$. Also, to assess the ability of the proposed model for modeling, observations with both symmetric and asymmetric structures, we consider different values for the skewness parameter as $\alpha \in \{-3, -1, 0, 1, 3\}$.

For the Bayesian approach, we adopt the vague and non-informative priors with considering hyperparameters ($\lambda_0 = 0, k_0 = 100$) for μ, ($a_0 = 0.01, b_0 = 0.01$) for w^{-2}, ($m_0 = 0, s_0^2 = 100$) for μ_α and ($c_0 = 0.01, d_0 = 0.01$) for σ_α^{-2}, for more details, Ntzoufras (2009). Taking these considerations into account, the values of the root mean square error (RMSE) for the ML and Bayesian estimates of the model parameters are presented in Table 1. The number of repetitions is fixed at be $R = 1000$ in order to take into account the uncertainty in a random number generating procedure. For example, the RMSE of ϕ is calculated as $RMSE = \sqrt{\frac{1}{R-1}\sum_{t=2}^{R}(\hat{\phi}_i - \phi)^2}$, where $\hat{\phi}_i$ is an estimation of ϕ in each repetition. Also, for checking the sensitivity of the method to different values of $\phi = -0.8, -0.5, 0.5, 0.8$, the values of $\mu = 2, w^2 = .5$ and $\alpha = 2$ are fixed and the RMSE for the ML and Bayes estimates of the model parameters are computed that are shown in Table 2. The results from Tables 1 and 2, show that if the sample size is small, the Bayesian estimates will be more precise than the ML estimates based on the RMSE values.

Table 1: The values of the RMSE for the ML and Bayes estimators of the model parameters by considering different values of n and α.

			Parameter			
α	n	Method	ϕ	μ	w^2	α
3	50	ML	0.0140	0.2286	0.4305	1.9012
		Bayes	0.0027	0.1217	0.4153	1.9002
	100	ML	0.0231	0.1073	0.2797	1.8827
		Bayes	0.0031	0.1003	0.2682	1.8081
	500	ML	0.0138	0.1085	0.2150	1.8070
		Bayes	0.0139	0.1049	0.2183	1.8072
	1000	ML	0.0127	0.1092	0.1395	1.8034
		Bayes	0.0125	0.1061	0.1330	1.8031
0	50	ML	0.0734	0.5371	0.4405	1.8846
		Bayes	0.0521	0.2990	0.3221	1.8707
	100	ML	0.0631	0.1135	0.2372	1.8745
		Bayes	0.0617	0.1117	0.2280	1.8633
	500	ML	0.0583	0.1084	0.1840	1.8074
		Bayes	0.0562	0.1093	0.1802	1.8021
	1000	ML	0.0539	0.1062	0.1306	1.8053
		Bayes	0.0542	0.1022	0.1302	1.8095
-3	50	ML	0.0471	0.2282	0.8331	1.7741
		Bayes	0.0311	0.2021	0.3838	1.6103
	100	ML	0.0172	0.1201	0.2172	1.6490
		Bayes	0.0160	0.1027	0.2081	1.5652
	500	ML	0.0146	0.1293	0.1183	1.0314
		Bayes	0.0129	0.1213	0.1140	1.0283
	1000	ML	0.0150	0.1078	0.1064	0.9058
		Bayes	0.0157	0.1044	0.1029	1.0018

Table 2: The values of the RMSE for the ML and Bayes estimators of the model parameters by considering different values of n and ϕ.

ϕ	Sample size	Method	Parameter ϕ	μ	w^2	α
-0.8	50	ML	0.0583	0.3198	0.7141	1.8246
		Bayes	0.0404	0.2863	0.7064	1.7814
	100	ML	0.0560	0.3146	0.7017	1.8217
		Bayes	0.0488	0.29950	0.6880	1.7721
	500	ML	0.0548	0.3086	0.6912	1.7824
		Bayes	0.0503	0.2910	0.6871	1.7715
	1000	ML	0.0524	0.3068	0.6551	1.7818
		Bayes	0.0599	0.3086	0.6543	1.7062
-0.5	50	ML	0.0471	0.1280	0.3598	1.8472
		Bayes	0.0359	0.1219	0.3419	1.8114
	100	ML	0.0410	0.1271	0.3363	1.8390
		Bayes	0.0405	0.1242	0.3280	1.8021
	500	ML	0.0405	0.1560	0.3301	1.8061
		Bayes	0.0407	0.1543	0.3362	1.8055
	1000	ML	0.0400	0.1541	0.3205	1.7930
		Bayes	0.0406	0.1554	0.3243	1.7933
0.5	50	ML	0.0492	0.1580	0.3998	1.6031
		Bayes	0.0475	0.1419	0.3598	1.5714
	100	ML	0.0417	0.1572	0.3663	1.5823
		Bayes	0.0412	0.1402	0.3628	1.4792
	500	ML	0.0395	0.1560	0.3310	1.5088
		Bayes	0.0392	0.1563	0.3412	1.5038
	1000	ML	0.0346	0.1541	0.3305	1.4739
		Bayes	0.0339	0.1524	0.3343	1.4710
0.8	50	ML	0.0899	0.4580	0.3725	1.0849
		Bayes	0.0698	0.4419	0.3561	1.0489
	100	ML	0.0867	0.4312	0.2963	1.0492
		Bayes	0.0710	0.4172	0.2628	1.0481
	500	ML	0.0860	0.2561	0.2314	1.0094
		Bayes	0.0854	0.2563	0.2302	1.0002
	1000	ML	0.0809	0.241	0.2155	1.0021
		Bayes	0.0900	0.2473	0.2143	1.0063

We also provide the estimator of the model parameters with normal innovations (Shumway and Stofer, 2006) as the usual traditional model to compare and evaluate the robustness of the proposed model versus violation of innovations from normality. The root mean square error of prediction defined by

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

is also computed in order to assess the predictive power of the proposed model and the results are presented in Table 3. To ensure the convergence of the generated Markov chains to their corresponding stationary distributions, some convergence diagnostics have been used.

Table 3: The RMSEP values of the model under the normal and skew-normal innovations.

			Distribution of innovations	
Sample size	α	Method	Skew-normal	Normal
10	-3	ML	2.7482	2.9034
		Bayes	1.2116	1.7436
	-1	ML	1.4046	1.8759
		Bayes	0.9084	1.3282
	0	ML	1.8508	1.8728
		Bayes	1.4285	1.4035
	1	ML	1.4984	2.5550
		Bayes	1.1148	1.9538
	3	ML	0.9181	1.6658
		Bayes	0.9950	1.3715
50	-3	ML	1.4370	2.4772
		Bayes	1.3020	1.8031
	-1	ML	1.2218	1.9230
		Bayes	1.1154	1.6717
	0	ML	1.4790	1.4960
		Bayes	1.3044	1.3180
	1	ML	1.7360	2.5348
		Bayes	1.6331	2.3502
100	-3	ML	1.1762	2.6414
		Bayes	1.1279	2.6051
	-1	ML	2.0075	2.7316
		Bayes	2.1902	2.7050
	0	ML	1.2920	1.2541
		Bayes	1.2001	1.2084
	1	ML	1.1090	1.8540
		Bayes	1.1496	1.8900
	3	ML	0.8475	2.8364
		Bayes	0.9816	2.8210

As can be seen from Table 3, the values of the RMSEP for the normal and the skew-normal model are almost the same for $\alpha = 0$ that indicate for this value, the skew-normal distribution reduces to the normal distribution. For positive and negative values of the skewness parameter, which respectively correspond to the right-skewed and left-skewed data, the skew-normal model provides more efficiency based on the RMSEP than the normal model because it truly takes into account the skewed structure of data.

4 A Real Example

In this section, we report some empirical results based on the analysis of the daily returns of Mellat bank stock of Iran from January 1, 2011 to January 14, 2015 for 894 observations. The data set has been downloaded from Tehran securities exchange technology management company site (www.tsetmc.com). The return series is shown in Figure 1 (a). We see that the daily returns of the Mellat bank stock are weakly stationary.

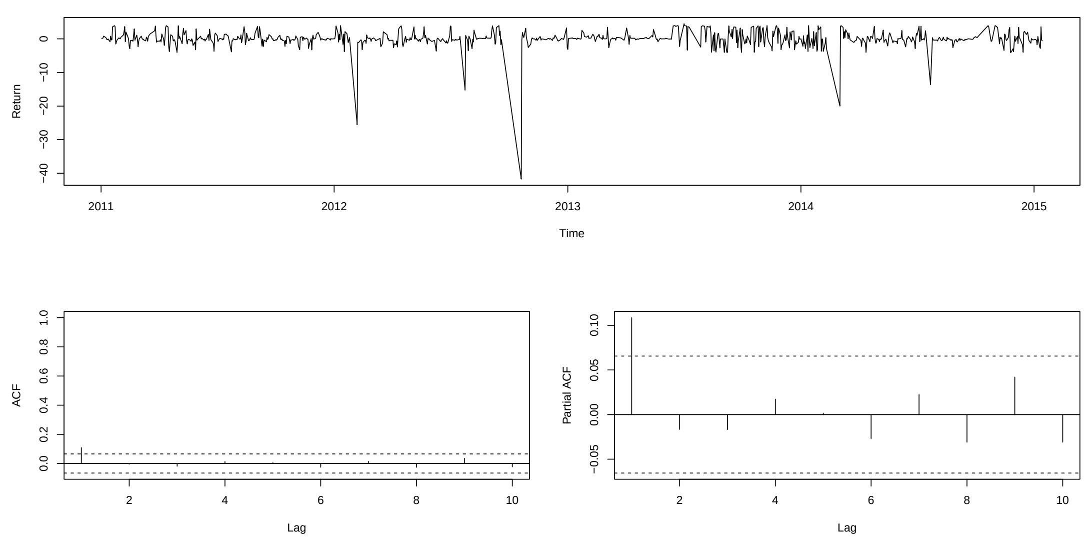

Figure 1: (a) Time plot of the daily returns of Mellat bank stock of Iran from January 1, 2011 to January 14, 2015: (b) The ACF of the return (c) The PACF of the return.

The sample autocorrelation function (ACF) and partial autocorrelation function (PACF) of the returns series, presented in panels (b) and (c) of Figure 1, indicate that an AR(1) model can be suggested for data.

The ML and Bayesian estimates of parameters and the RMSE and Akaike information criterion (AIC) for the proposed model (Akaike , 1973) are presented in Table 4. Although the descriptive statistics of return, given in Table 3, show that the data has a negative skewness.

We also fit the model with normal innovations in order to assess the effect of ignoring the skewness in modeling the process. It is seen from the results that the skew-normal model has a better fit to the data than the normal models based on the RMSEP and AIC criteria.

Table 4: The descriptive statistics of the return series.

n	Min	Max	Mean	Variance	Skewness
894	-41.7300	4.4760	0.0117	6.3361	-7.1341

Table 5: The ML and Bayesian estimation of the model parameters, the RMSEP and AIC of the model with normal and skew-normal innovations.

		Parameter				Criterion	
method	Innovations	ϕ	μ	w^2	α	RMSEP	AIC
ML	Normal	0.1085	0.0103	6.2615	-	3.6297	4178.3666
	Skew-normal	0.1046	0.0076	8.4595	-1.8140	3.3139	3965.0531
Bayes	Normal	0.2210	0.0098	6.9017	-	3.5020	4088.2619
	Skew-normal	0.2019	0.0015	9.0074	-2.0080	3.4044	3704.5025

Time plot of the daily return series (dotted line) and estimation of it (dashed line) are presented in the upper panel of Figure 2. We analyze the residual of the model for

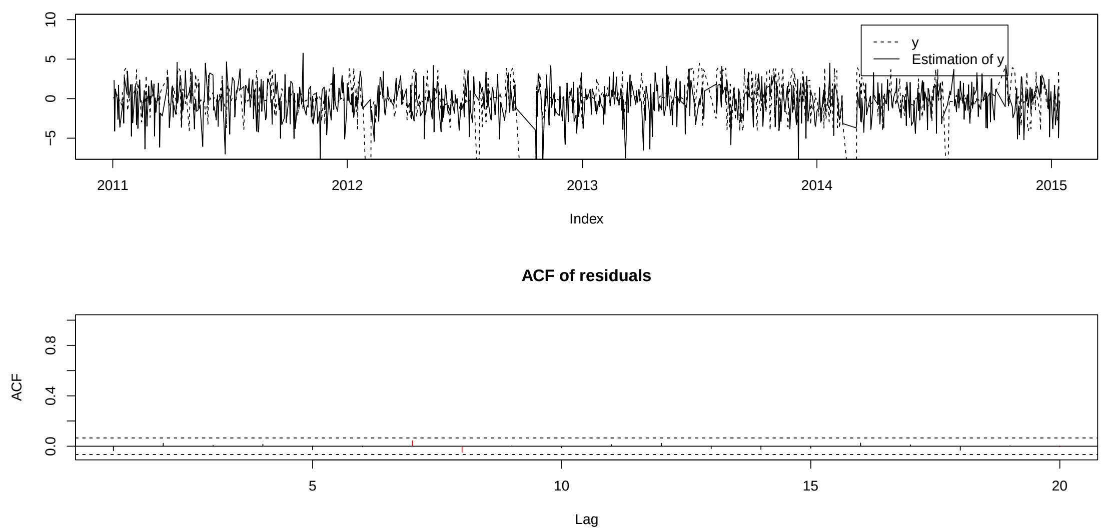

Figure 2: (a) Time plot of the daily return series (dotted line) and its estimation of it (dashed line) (b) The ACF of residuals.

checking goodness of fit by using the ACF plot of them, that are presented in the bottom panel of Figure 2. Based on this figure, there are no significant serial correlations, thus, the model appears to be adequate in describing the daily return series of Mellat bank.

5 Conclusion

In this paper, we considered an autoregressive model with the skew-normal innovations and developed a parametric schema for fitting the model both from the frequentist and Bayesian point of view. The proposed model allows a flexible treatment of asymmetry in the conditional distribution of the observations. The ML and Bayes estimates of the unknown parameters obtained via EM type optimization and MCMC methods, respectively. The result of a simulation study and the empirical application are indications of a good performance of the proposed methodology.

References

Akaike, H. (1973), Information theory and an extension of the maximum likelihood principle. *2nd International Symposium Information Theory (C. B. Petrov and F. Csaki, Eds)*, 267–281.

Arellano-Valle, R. B and Azzalini, A. (2006), On the unification of families of skew-normal distributions. *Scandinavian Journal of Statistics*, **33**(3), 561–574.

Arnold, B. C., Beaver, R. J., Groeneveld, R. A. and Meeker, W. Q. (1993), The non-truncated marginal of a truncated bivariate normal distribution. *Psychometrika*, **58**(3), 471–488.

Azzalini, A. (1985), A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, **12**(2), 171–178.

Azzalini, A. (1986), Further results on a class of distributions which includes the normal ones. *Statistica*, **46**(2), 199–208.

Bondon, P. (2009), Estimation of autoregressive models with epsilon-skew-normal innovations. *Journal of Multivariate Analysis*, **100**(8), 1761–1776.

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*, **39**(1), 1-22.

De Oliveira, V. (2012), Bayesian analysis of conditional autoregressive models. *Annals of the Institute of Statistical Mathematics*, **64**(1), 107–133.

Good, I. J. (1980), Some history of the hierarchical Bayesian methodology. *Trabajos de estadistica y de investigacion operativa*, **31**(1), 489.

Hajrajabi, A. and Fallah, A. (2017), Nonlinear semiparametric AR(1) model with skew-symmetric innovations. *Communications in Statistics - Simulation and Computation*, **47**(5), 1453–1462.

Henze, N. (1986), A probabilistic representation of the skew-normal. *Scandinavian Journal of Statistics*, **13**(4), 271–275.

Ibazizen, M. and Fellag, H. (2003), Bayesian estimation of an AR(1) process with exponential white noise. *Statistics: A Journal of Theoretical and Applied Statistics*, **37**(5), 365-372.

Kastner, G. (2016), Dealing with stochastic volatility in time Series using the R package stochvol. *Journal of Statistical Software*, **69**(5), 1–30.

Lin, T. I., Lee, J. C. and Yen, Shu Y. (2007), Finite mixture modeling using the skew-normal distribution. *Statistica Sinica*, **17**(3), 909–927.

Lindley, D. V. and Smith, A. F. M. (1972), Bayes estimates for the linear model (with discussion). *Journal of the Royal Statistical Society, Series B*, **34**(1), 1–44.

Ntzoufras, I. (2009), *Bayesian modeling using winbugs*. New York: wiley.

Pourahmadi, M. (2001), *Foundation of time series analysis and prediction theory* New York: Wiley.

Robert, C. (2001), *Bayesian choice*. New York: Springer.

Sharafi, M. and Nematollahi, A. R. (2016), AR(1) model with skew-normal innovations. *Metrika*, **79**(8), 1011–1029.

Shumway, R. H. and Stofer, D. S. (2006), *Time series analysis and its applications: with R examples*. New York: Springer.

Tarami, B. and Pourahmadi, M. (2003), Multi-variate t autoregressions: innovations, prediction variances and exact likelihood equations. *Journal of Time Series Analysis*, **24**(6), 739–754.

6 Appendix

In this appendix some brief theoretical results about the skew-normal distribution are presented.

Lemma 6.1. *If* $Y \sim SN(\mu, w^2, \alpha)$*, then*
(i) $E(Y) = \mu + \sqrt{\frac{2}{w}}\delta(\alpha)$,
(ii) $Var(Y) = \{1 - \frac{2}{\pi}\delta^2(\alpha)\}w^2$,

where $\delta(\alpha) = \frac{\alpha}{\sqrt{1+\alpha^2}}$.

Lemma 6.2. *If* $Y \sim SN(\mu, w^2, \alpha)$*; then the following moment estimators can be obtained from the work of Arnold et al. (1993)*

(i) $\hat{\mu} = m_1 - c_1(\frac{m_3}{d_1})^{\frac{1}{3}}$,
(ii) $\hat{w}^2 = m_2 + c_1^2(\frac{m_3}{d_1})^{\frac{2}{3}}$,
(iii) $\hat{\delta}(\alpha) = \{c_1^2 + m_2(\frac{d_1}{m_3})^{\frac{2}{3}}\}^{-\frac{1}{2}}$,

where $c_1 = \sqrt{\frac{2}{\pi}}$, $d_1 = (\frac{4}{\pi} - 1)c_1$, $m_1 = \frac{\sum_{i=1}^n Y_i}{n}$, $m_2 = \frac{\sum_{i=1}^n (Y_i - m_1)^2}{n-1}$ and $m_3 = \frac{\sum_{i=1}^n (Y_i - m_1)^3}{n-1}$.

The following lemma provides an stochastic representation for the skew-normal distribution as a mixture of a truncated normal (TN) and a normal distribution.

Lemma 6.3. *According to Azzalini (1986) and Henze (1986), if* $S \sim TN(0, w^2)I_{\{T>0\}}$, $U \sim N(0, w^2)$ *such that* T *and* U *be independent, then*

$$Y = \mu + \delta(\alpha)S + \sqrt{1 - \delta^2(\alpha)}U,$$

distributed as $SN(\mu, w^2, \alpha)$*. Also, the joint density of* Y *and* S *is given by*

$$\begin{aligned} f_{Y,S}(y,s) &= \frac{\sqrt{1+\alpha^2}}{\pi w^2} \\ &\times \exp\{\frac{-1}{2w^2}[(y-\mu)^2 + (1+\alpha^2)(s - \frac{\alpha}{\sqrt{1+\alpha^2}}(y-\mu))^2]\}. \end{aligned}$$

Lemma 6.4. *If* $S \sim TN(0, w^2)I_{\{S>0\}}$, $U \sim N(0, w^2)$ *and* Y *is defined as* $Y = \mu + \delta(\alpha)S + \sqrt{1-\delta^2(\alpha)}U$, *then*

$$S|Y = y \sim TN(\mu_S, \sigma_S^2)I_{a_1=0<s<a_2=\infty},$$

Also, we have

$$\begin{aligned} E(S|Y = y) &= \mu_S + \frac{\phi(\frac{\mu_S}{\sigma_S})}{\Phi(\frac{\mu_S}{\sigma_S})}\sigma_S, \\ E(S^2|Y = y) &= \mu_S^2 + \sigma_S^2 + \frac{\phi(\frac{\mu_S}{\sigma_S})}{\Phi(\frac{\mu_S}{\sigma_S})}\sigma_S\mu_S, \end{aligned}$$

where

$$\mu_S = \delta(\alpha)(y - \mu),$$
$$\sigma_S^2 = w^2(1 - \delta^2(\alpha)).$$