Journal of Mathematical Extension
Vol. 18, No. 6, (2024) (1)1-22
URL: https://doi.org/10.30495/JME.2024.2918
ISSN: 1735-8299
Original Research Paper

Time Series Based on Random Process of Unbounded Asymmetric Normal Distribution

P. Zarin
Islamic Azad University, Marvdasht Branch

Z. Khodadadi*
Islamic Azad University, Marvdasht Branch

M. Maleki
Isfahan University

H. Jafarpour
Islamic Azad University, Shiraz Branch

Abstract. One of the generalizations of normal distribution unlimited skew normal distribution, which is more flexible than classical normal distribution. In contrast to normal distribution, unlimited skewed normal distribution is asymmetric with various types of skewness which makes it applicable in fitting different types of real data. Therefore, this study intended to investigate moving average autoregressive time series process based on asymmetric normal coefficients of unbounded skew, or SUN-ARMA process for short. Providing a hierarchical representation of unbounded skew normal distribution facilitated the simulation of this distribution in practice. The parameters of the asymmetric SUN-ARMA process were estimated using maximum likelihood method with EM algorithm approach. The performance and accuracy of the maximum likelihood method in estimating the parameters of the

Received: August 2024; Accepted: September 2024
*Corresponding Author

SUN-ARMA process were investigated based on simulated data under different sample sizes. Also, using two real data series, the efficiency of SUN-ARMA process was studied in comparison to classical autoregressive process of moving average with normal coefficients, and the results confirmed the superiority of SUN-ARMA process in fitting asymmetric real data.

Keywords and Phrases: EM algorithm, skewness, auto regression, moving average, unbounded Skew normal.

1 Introduction

Time series Statistical modeling is used in many sciences and natural phenomena. Using statistical models, the relationships between the considered observations are mathematically equated. Time series statistical models are used in modeling dependent data collected in a time frame. $\{x_t, t\in T\}$ time series is an ordered sequence of observations that are usually presented in terms of time, especially in equal time intervals. The inherent nature of time series is the correlation of its observations and its analysis which are used in various sciences such as examining the number of crops and annual grain prices in agricultural sciences, the price of goods at the end of the day in commercial sciences and economics, the voltage required in a device and car production in engineering sciences as well as air temperature and annual precipitation in meteorology. Normal distribution, despite its unique and useful properties, is inefficient in fitting asymmetric data. Therefore, the researchers introduced generalizations of the normal distribution and presented asymmetric versions of the normal distribution. The family of elliptic distributions introduced by Kelker (1970) includes a wide set of symmetric distributions such as normal, t-Student and Pearson type II distributions. A simple but powerful method to generate a family of multivariate Chole elliptic distributions is to use the conditioning approach. The elliptic distribution of Chula was investigated by Genton (2004) and the elliptic distribution of integrated multivariate Chula was investigated and studied by Arellano- called normal family of mixed scalar Chule Valle and Genton (2010). Also, Malaki and Arellano-Valle (2017) considered an autoregressive model called normal family of mixed scalar Chule with innovators that are a mixture of the family of heavy-tailed/light-tailed and symmetric/asymmetric distributions. For further reading on time series topics such as Kalman-filter, spatial prediction and signal detection based on normal Chule process, refer to Gualtierotti (2005). Camassa et al. (2021) in a study persisting asymmetry in probability distribution function for a random advection-diffusion equation in impermeable channels they paid.The analytical result reported in this study verifies the conclusion of the linear shear flow obtained from numerical simulations in [10]. It is also shown that limiting distribution is negatively skewed for any shear flow at sufficiently low Péclet number. Additionally, we demonstrate the convergence of the Ornstein-Uhlenbeck case to the white noise case in the limit $\gamma \rightarrow infity$ of the OU damping parameter, which generalizes to the channel domain problem the results for free space in [11]. We specify that the long-time limit of the first three moments depends explicitly on the value of γ , which is in contrast to the conclusion in [12] for the limiting PDF in free space.To find a benchmark for theoretical analysis, we derive the exact formula of the N-point correlator for a flow with no spatial dependence and Gaussian temporal fluctuation, generalizing the results of [13]. The long-time analysis of this formula is consistent with our theory for a general shear flow. All results are verified by Monte-Carlo simulations.

the efficiency of SUN distribution in modeling time series observations. In the introduction of the time series model, we consider the non-normal autoregressive moving average (ARMA) linear process with elements of the SUN distribution, and we examine the structure of the unlimited Skew normal ARMA (SUN-ARMA) models. We obtain the unknown parameters of the process using the maximum likelihood (ML) method with the mathematical expectation-maximization (EM) approach. Using the approach of simulated studies, we examine the consistency of ML estimators under different sample sizes and measure the efficiency of the new process in fitting real data related to the import of goods and services in Australia and the estimated Australian resident population compared to the ARMA model with normal coefficients. R statistical software was used to perform the calculations and analyzes required in this article. In this article, considering the first method, we focus on the unbounded multivariate Skew normal distribution and examine its application in presenting the stochastic process and modeling asymmetric time series observations. Therefore, he introduced a new ARMA process based on SUN features, which is used in fitting different types of symmetric or asymmetric observations. As expected, the presented asymmetric time series process provides better performance than its symmetric counterpart. According to the hierarchical display of SUN distribution, the parameters of the process can be estimated and it has a better performance than the classical ARMA process based on normal symmetric distribution

2 Introduction and Investigation of Normal Distributions and Multivariate Normal Distributions

In recent years, the study of parametric families of probability distributions for modeling continuous multivariate random variables has received much attention. The main motivation of this approach is to introduce skewness in the family of normal distributions, which has been studied by several authors in different fields.

In the following sections, the Skew Normal (SN), Multivariate Skew Normal (MSN) distribution and some of their practical and basic features are introduced and reviewed. The density function of the standard Skew normal distribution is presented as follows:

$$f\left(z,\lambda\right)=2\varphi\left(z|0,1\right)\Phi\left(\lambda z|0,1\right),\lambda\in R,$$

which λ is the skewness parameter, $\varphi\left(.|0,1\right)$ is the univariate standard normal density function and $\varphi\left(.|0,1\right)$ is the univariate standard normal distribution function. We denote the normal distribution of the standard deviation with the symbol SN(λ).

Theorem 2.1. *The moment generating function (MGF) of SN* $(0,1,\lambda)$ *distribution is calculated as follows:*

$$M_Z\left(t\right)=2e^{\frac{t^2}{2}}\Phi\left(\frac{\lambda t}{\sqrt{1+\lambda^2}}|0,1\right).$$

Proposition 2.2. *Considering the normal random variable with dimension k+1 as* $\begin{pmatrix} Z_0 \\ Z_1 \end{pmatrix} \sim N_{1+k}\left(\begin{bmatrix} \tau \\ \mu \end{bmatrix}, \begin{bmatrix} 1 & \delta \\ \delta^\top & \Sigma \end{bmatrix}\right)$ *take, which* Z_0 *is a one-dimensional random variable. Then the random variable* $Z = Z_1|Z_0 > \tau$ *has MSN distribution with parameters* (μ, Σ, Λ)*, so that* $\Lambda^\top = \frac{\delta^\top \Sigma^{-1}}{1-\delta\Sigma^{-1}\delta^\top}$

3 Introduction and Investigation of the Limited Skew Normal and Unlimited Skew Normal Distributions

In proposition2.2, we considered the random representation of MSN distribution as $Z = (Z_1 \,|Z_0 > \tau\,)$ so that the random variable Z_0 is univariate, if the dimension of the random variable Z_0 is a) equal to one, b) equal to Z_1 and c) consider any arbitrary positive number, we get new generalizations of the MSN distribution, which are discussed further.

In order to fit multivariate data, Azalini and Dalla-Vale (1996) generalized the univariate SN distribution and introduced the multivariate Skew normal distribution. Since the seminal paper was published by Azalini and Dalla-Valle (1996) on the MSN distribution, several extensions of the Skew normal distribution have been successively presented. [3].

The term limited and unlimited was first considered by Lee and McLachlan [7] to represent multivariate normal distributions. Because by applying this restriction that the hidden variables of skewness are all equal in the form of the family of Skew elliptic distributions presented by Sahu et al. [9], it is obtained. The family of Skew distributions without considering this limitation is called the unbounded family. The restricted multivariate Skew normal distribution (RMSN) is equivalent to the Skew normal distribution proposed by Azalini and Dalla-Valle (1996). Lee and McLachlan [8] provided a systematic overview of existing multivariate Skew distributions and specified their conditional type and convolution type representations. So that if the conditional random variable is limited to a univariate distribution, the presented distribution is called the limited Skew normal distribution, and if the multivariate conditional random variable is considered, it is referred to as the unlimited Skew normal (SUN) [3].

Most of these developments, however, can be considered as special cases of the fundamental fusiform normal distribution (FUSN) introduced by Arellano-Valle and Genton (2005), which is a generalization of the original MSN distribution. MSN distribution generalizations can be systematically classified into three types: a) limited, b) unlimited, and c) extended [1].

FUSN distribution is defined based on conditioning the multidimensional normal variable on another random variable (univariate or multivariate). The comprehensive distribution of FUSN is presented as follows.

Suppose Z_1 and Z_0 have joint distribution as below

$$\begin{bmatrix} Z_0 \\ Z_1 \end{bmatrix} \sim N_{q+p} \left(\begin{bmatrix} \tau \\ \mu \end{bmatrix}, \begin{bmatrix} \Gamma & \Delta \\ \Delta^{\perp} & \Sigma \end{bmatrix} \right), \tag{1}$$

Then the vector $Z = Z_1 | Z_0 > 0_q$ has FUSN distribution, which τ and μ are mean vectors with dimensions q and p respectively, covariance matrices Γ and Σ with dimensions q×q and p×p respectively and symmetric matrix Δ is with dimension q×p.

a) The limited state of the FUSN distribution corresponds to a special form of the equation (1), in which the conditional random variable Z_0 is limited to the univariate state and is therefore limited to the conditions q=1, τ=0, Γ=1 (Z_0 is considered from the univariate standard normal distribution).

b) In the unlimited state, both random variables Z_0 and Z_1 have p-dimensional normal distribution (in other words, p=q). Therefore, in the unlimited state, the random variable Z_0 is also considered from the normal distribution of the p-variable with the mean vector $\tau = 0_p$ and the covariance matrix Γ. The use of the word restricted here refers to the restrictions of the random vector of the conditional part (Z_0) in the stochastic representation of the Skew normal distribution and is not a restriction on the parameter space. So, the finite form of a Skew distribution is not necessarily nested in its corresponding unbounded counterpart.

c)In the expanded form of the FUSN family, there are no restrictions on the dimensions of Y_0, (q includes any arbitrary number) and τ can be a non-zero vector. Accordingly, in the developed state, random variable Y_0 is considered from normal distribution with arbitrary dimension q with arbitrary mean vector τ and covariance matrix Γ. Extended Skew normal distribution is also called integrated Skew normal distribution.

4 The Normal Distribution is Unbounded

The unrestricted form of the MSN distribution is very similar to the restricted form, except that the scalar latent variable is replaced by a random vector from the p-dimensional normal distribution, such that both components of the conditional representation are assumed to have the same dimension p. Sahu et al. [9]introduced the family of unbounded Skew elliptic distributions by adding a skewness parameter to the elliptic symmetric family. They added the skewness parameter to the elliptic symmetric family by conditioning on a multivariate random variable. One member of the unlimited Skew elliptic distributions family is the unlimited Skew normal distribution (UMSN). Also, Gupta et al. [6] introduced a new UMSN distribution with different parameterization.

Definition 4.1. [9] consider the p-dimensional random vector Z from the multivariate normal distribution with parameters (μ, Σ, Δ) and the symbol $Z{\sim}UMSN_p\left(\mu, \Sigma, \Delta\right)$, when its density function is as follows be shown

$$f\left(z, \mu, \Sigma, \delta\right) = 2^p \varphi_p\left(z|\mu, \Sigma\right) \Phi_p\left(\Delta\Sigma^{-1}\left(z-\mu\right)|0_p, \Lambda\right), z{\in}R^p, \qquad (2)$$

which $\mu{\in}R^p$, $\Sigma{>}0$, Δ is a diagonal matrix in the form Δ=diag(δ), with dimension p×p, $\varphi_p\left(.|\mu, \Sigma\right)$ and $\Phi_p\left(.|\mu, \Sigma\right)$ function The density and normal distribution of p-variable with mean parameter μ and covariance matrix Σ and $\Lambda = I_p - \Delta\Sigma^{-1}\Delta$, so that the same matrix I_p is a diagonal matrix with dimension p×p with diagonal elements equal to one.

Theorem 4.2. *Suppose two random vectors Z_0 and Z_1 have normal distribution as follows*

$$\begin{bmatrix} Z_0 \\ Z_1 \end{bmatrix} {\sim} N_{2p}\left(\begin{bmatrix} 0_p \\ 0_p \end{bmatrix}, \begin{bmatrix} I_p & 0_p \\ 0_p & \Sigma \end{bmatrix}\right),$$

the p-dimensional random vector Z from the UMSN distribution is defined as a random representation of the convolution of the two considered vectors as $Z = \mu + \Delta |Z_0| + Z_1$ *and its PDF is presented as follows*

$$f_Z(z, \mu, \Sigma, \Delta) = 2^p \varphi_p(z|\mu, \Omega)\, \Phi_p\left(\Delta\Omega^{-1}(z-\mu)\,|0_p, \Lambda\right),$$

where μ *is the p-dimensional vector,* Σ *is the diagonal scale matrix with* $p{\times}p$ *dimension and* Δ *is the diagonal matrix of skewness parameters with* $p{\times}p$ *dimension and*

$$\Lambda = \left(\Delta^\top \Sigma^{-1}\Delta + I_p\right)^{-1} = I_p - \Delta\Omega^{-1}\Delta, \qquad \Omega = \Delta\Delta^\top + \varSigma.$$

5 Introducing the SUN Distribution

The m-dimensional random vector Z has the infinite normal distribution with the m-dimensional location vector μ and the positive definite scale matrix Σ of order m×m and the order m×n matrix of skewness Λ and is denoted by $Z \sim SUN_{m,n}(\mu, \Sigma, \Lambda)$ is displayed. The density function of SUN random variable is as follows

$$f_{sun}(z\,|\mu, \Sigma, \Lambda) = 2^n \varphi_m(z\,|\mu, \psi)\, \Phi_n\left(\Lambda^\top \psi^{-1}(z-\mu)\,|0_n, \Gamma\right), z{\in}R^m, \tag{3}$$

in this relation $\mathbf{\Psi} = \mathbf{\Sigma}{+}\mathbf{\Lambda}\mathbf{\Lambda}^\top$, $\mathbf{\Gamma} = \left(\mathbf{I}_n + \mathbf{\Lambda}^\top\mathbf{\Sigma}^{-1}\mathbf{\Lambda}\right)^{-1} = \mathbf{I}_n{-}\mathbf{\Lambda}^\top\mathbf{\Psi}^{-1}\mathbf{\Lambda}$, $\varphi_m(\cdot\,|\boldsymbol{\mu}, \mathbf{\Psi})$ m-variable density function with mean vector μ and covariance matrix ψ and $\Phi_n(\cdot\,|\mathbf{0}_n, \mathbf{\Gamma})$ are n-variate normal distribution function with mean vector 0n and covariance matrix Γ. If the skewness matrix is zero $\mathbf{\Lambda} = \mathbf{0}_n$, the unlimited Skew normal distribution becomes the well-known normal distribution, and in the case which m=n=1, the well-known univariate Skew normal distribution of Azalini (1985) is obtained [2].

6 Autoregressive Normal Moving Average Process

Considering the strictly stationary process of SUN as a random variable, we introduce the autoregressive process of unbounded normal moving average SUN-ARMA(p,q) as follows

$$X_t - \sum_{i=1}^{p} \varphi_i X_{t-i} = Z_t + \sum_{j=1}^{q} \theta_j Z_{t-j}, t = 0, \pm 1, \pm 2, \ldots, \tag{4}$$

where the sequence $\{Z_t\}$ is independent of the sequence $\{X_s\}$ for $s < t$ and $\{Z_t\}$ is an independent sequence and identically distributed (iid) of one-variable SUN process with zero mean as follows

$$\{Z_t\} \sim SUN_{1,1}\left(b\lambda, \sigma^2, \lambda\right),$$

which $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_p)^\top$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)^\top$ are respectively called auto reversion coefficients and moving average coefficients of the SUN-ARMA process.

In the continuation of the article, the self-reversion process of the unbounded normal moving average $SUN - ARMA(p,q)$ with the order of auto reversion p and the order of the moving average q, as $\{X_t\} \sim SUN - ARMA(p,q)$ with parameters $\boldsymbol{\Theta} = \left(\boldsymbol{\varphi}^\top, \boldsymbol{\theta}^\top, \sigma^2, \lambda\right)^\top$ is symbolized.

Definition 6.1. Considering the process $\{X_t\} \sim SUN - ARMA(p,q)$, its MA($\infty$) moving average representation is $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$ and if $\sum_{j=0}^{\infty} |\psi_j| < \infty$, then the process converges in the mean.

Proposition 6.2. *In the strictly stationary process $\{X_t\} \sim SUN - ARMA(p,q)$, its mean and covariance functions are presented as follows*

$$\mu_X(t) = E(X_t) = 0, \qquad \gamma_X(h) = Cov(X_t, X_{t+h}) = \sigma_z^2 \zeta(h), \tag{5}$$

in which

$$\sigma_z^2 = Var(Z_t) = \sigma^2 + \left(1 - \frac{2}{\pi}\right)\lambda^2 \qquad , \zeta(h) = \sum_{j=0}^{\infty} \psi_{j+|h|}\psi_j,$$

and for $h \to \infty$, $\gamma_X(h)$ converges to zero.

7 Estimating the Parameters of the ARMA Model with Unlimited Normal Coefficients

In this part, we estimate the parameters of SUN-ARMA model with the maximum likelihood method. In so doing, the EM algorithm is used.

Symbols $\mathfrak{x}_{t-1} = (X_{t-1}, \ldots, X_{t-p})^\top$ and $\mathbf{z}_{t-1} = (Z_{t-1}, \ldots, Z_{t-q})^\top$ For random data and t=1,...,n, based on initial observations $\mathbf{X}_0 = (X_0, \ldots, X_{-p+1})^\top$ and $\mathbf{Z}_0 = (Z_0, \ldots, Z_{-q+1})^\top$ We consider conditioned. Also, considering the vector of observations in the form $\mathbf{X} = (X_1, \ldots, X_n)^\top$ and the Markovian property of ARMA(p,q) processes, the conditional likelihood function is shown as follows

$$L(\Theta \mid X) = f_X\left(X \mid X_0, Z_0, \Theta\right) = \prod_{t=1}^{n} f_Z\left(Z_t \mid X_0, Z_0, \Theta\right)$$

where the sequence $\{Z_1, \ldots, Z_n\}$ based on X, X_0 and Z_0 with repetition $Z_t = X_t - \boldsymbol{\varphi}^\top \mathfrak{X}_{t-1} - \boldsymbol{\theta}^\top \mathbf{z}_{t-1}$ for t=1,2, ..., n is obtained. In order to calculate the initial values of Z_0 and X_0, Bax et al. (1976) suggested that Z_t be equal to zero and X_t be equal to the observations themselves. Therefore, the desired repetitions start from time t=p+1. We consider the initial values of X_0 equal to the observations themselves and

$$\ell(\Theta \mid X) = \sum_{t=1}^{n} \ell_t(\Theta \mid X) = \sum_{t=1}^{n} \log f_{\text{sun}}\ \left(X_t - \varphi^\top \mathfrak{X}_{t-1} - \theta^\top z_{t-1} \mid b\lambda, \sigma^2, \lambda\right) \tag{6}$$

In this article, we will use the hierarchical representation of the SUN distribution and its desirable features to estimate the parameters of the SUN-ARMA(p,q) model by using the EM algorithm.

To apply the EM algorithm, we use the hierarchical representation of the SUN distribution in the autoregressive normal moving average model of the unrestricted Skew. for t=1,. . . ,n

$$Z_t\,|W_t \sim N\left(\lambda\left(W_t + b\right), \sigma^2\right)$$

$$W_t \sim TN\left(0,1\right) I\left(W_t > 0\right)$$

where, in the SUN-ARMA process is equivalent

$$\left(X_t - \varphi^\top \mathfrak{X}_{t-1} - \theta^\top z_{t-1}\right) \mid W_t \sim \mathrm{N}\left(\lambda W_t, \sigma^2\right)$$
$$W_t \sim \mathrm{TN}(b,1)\mathrm{I}\left(w_t > b\right)$$

and in these relationships, $b = -\sqrt{\frac{2}{\pi}}$.

We consider the set of complete observations as $\mathbf{D} = \left(\mathbf{X}^\top, \mathbf{W}^\top\right)^\top$ which $X = (X_1, \ldots, X_n)^\top$ is the visible part of the observations and W is the hidden part of the observations (from the distribution normal cut). The conditional likelihood function based on complete observations is presented as follows

$$\begin{aligned} L(\Theta \mid D) = \textstyle\prod_{t=1}^n \phi_1\left(X_t - \varphi^\top \mathfrak{X}_{t-1} - \theta^\top z_{t-1}; \lambda W_t, \sigma^2\right) \\ \mathrm{T}\phi_1\left(W_t; b, 1\right)\mathrm{I}\left(W_t > b\right) \end{aligned} \tag{7}$$

The second part of the likelihood function (7) does not include model parameters. Therefore, the logarithm of the conditional likelihood function that depends on the model parameters is:

$$\ell(\Theta \mid D) = -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^n \left(X_t - \varphi^\top \mathfrak{X}_{t-1} - \theta^\top z_{t-1} - \lambda W_t\right)^2 \tag{8}$$

Therefore, in the expectation step (E-Step) in the (k+1)-th repetition of the EM algorithm, it is necessary to calculate the following mathematical expectation

$$Q\left(\Theta \middle| \Theta^{(k)}\right) = E_{\Theta^{(k)}}\left[l\left(\Theta \,|D\right)|X\right].$$

Therefore, taking into account the relation (8), the considered mathematical hope is presented as follows

$$\begin{aligned} Q\left(\Theta \middle| \Theta^{(k)}\right) &= \frac{-n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\{\sum_{t=1}^n \left(X_t - \varphi^\top X_{t-1} - \theta^\top z_{t-1}\right)^2 + \lambda^2\beta_t\} \\ & \frac{2\lambda}{2\sigma^2}\sum_{t=1}^n \left(X_t - \varphi^\top \mathfrak{X}_{t-1} - \theta^\top z_{t-1}\right)\alpha_t, \end{aligned} \tag{9}$$

In this function, $\alpha_t = E_\Theta\left[W_t \,|\mathbf{X}\right]$ and $\beta_t = E_\Theta\left[W_t^2 \,|\mathbf{X}\right]$ are the first and second order moments of the hidden variable W_t. As it was shown before, the random variable $W_t\,|\mathbf{X}$ has a truncated normal distribution in the interval (b,+∞) with parameters $\left(\mu_w, \sigma_w^2\right)$, so that

$$\mu_w = \frac{\lambda\left(X_t - \varphi^\top \mathfrak{X}_{t-1} - \theta^\top z_{t-1} - b\lambda\right) + b\sigma^2}{\sigma^2 + \lambda^2},$$

$$\sigma_w^2 = \frac{\sigma^2}{\sigma^2 + \lambda^2}. \tag{10}$$

Therefore, α_t and β_t are the first and second order moments of normal distribution cut in the interval (b, $+\infty$).

It should be noted that $\alpha_t^{(k)}$ and $\beta_t^{(k)}$ are the updated moments in the k-th stage of the EM algorithm with the estimated value of $\mathbf{\Theta}^{(k)}$ of the Θ parameter in the current stage. The next steps of this algorithm are the conditional maximization steps (CM-Steps), which are actually estimated by maximizing the Q function of the parameters.

8 Numerical Studies of SUN-ARMA Model

In this part, numerical studies and classical inference of the SUN-ARMA process are discussed. First, it is shown that not only the classical maximum likelihood estimators in the SUN-ARMA model provide adequate performance but also the obtained estimates are consistent and converge to the true value of the parameter. Then, the efficiency of the SUN-ARMA process in fitting two real data series is measured. In the implementation of numerical calculations, R statistical software is used and the tolerance threshold of EM-type algorithm equal to $\tau = 10^{-3}$.

8.1 Classical Simulation Studies

Using simulated data, we investigate the performance of SUN-ARMA process parameter estimators.
In order to check the performance of ML estimators, we use two different models from the family of SUN-ARMA processes with autoregressive orders and moving average equal to p=1, q=1 and p=2, q=2 respectively, So that in the first process that Denoted by the symbol SUN-ARMA (1,1), the parameters are as follows

$$\varphi_1 = 0.5, \theta_1 = 0.5, \qquad \sigma^2 = 1, \lambda = -0.01, 2.$$

And in the second process, with the SUN-ARMA (2,2) symbol, the parameters are

$$\varphi_1 = 0.3, \theta_1 = 0.3, \qquad \varphi_2 = 0.4, \theta_2 = 0.2, \qquad \sigma^2 = 2, \lambda = -0.01, 2,$$

we will consider.
In the definition of the parameters φ_1 and φ_2, attention has also been paid to checking the static conditions. So that in the SUN-ARMA(1,1) process, for stationarity it is necessary that the condition $|\varphi_1| \neq 1$ be satisfied, which is considered stationary due to the value of $\varphi_1 = 0.5$.
In the SUN-ARMA(2,2) process, to check the stationary condition, it is necessary that the roots of the equation $\boldsymbol{\varphi}(B) = 1 - \varphi_1 B - \varphi_2 B^2 = 0$ be outside the unit circle. So,

$$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 = 1 - 0.3B - 0.4B^2 = 0,$$

$$\varphi_1 + \varphi_2 < 1$$

$$\varphi_2 - \varphi_1 < 1$$

$$|\varphi_2| < 1$$

and

$$B = \frac{-\varphi_1 \pm \sqrt{\varphi_1^2 + 4\varphi_2}}{2\varphi_2} = \frac{-0.3 \pm \sqrt{0.09 + 1.6}}{0.8} = 1.25, -2,$$

Both roots of the equation $\boldsymbol{\varphi}(B) = 0$, which are equal to 1.25 and -2, are not located on the unit circle, so the considered SUN-ARMA(2,2) process is also stationary.

In general, the two processes SUN-ARMA (1,1) and SUN-ARMA (2,2) are shown as follows.

$$X_t - 0.5X_{t-1} = Z_t + 0.5Z_{t-1}, \qquad \{Z_t\} \sim SUN\left(-\lambda\sqrt{\frac{2}{\pi}}, \sigma^2 = 1, \lambda\right),$$

and

$$X_t - 0.3X_{t-1} + 0.4X_{t-2} = Z_t + 0.3Z_{t-1} + 0.2Z_{t-2},$$

$$\{Z_t\} \sim SUN\left(-\lambda\sqrt{\frac{2}{\pi}}, \sigma^2 = 2, \lambda\right),$$

It should be noted that for the skewness parameter λ, two different values -0.01 and 2 have been considered, The first selection indicates a weak negative skewness and is close to normal, while the second case shows a strong positive skewness

Simulated data with different sample sizes are generated as $n = 50, 150, 350$, and each simulation is repeated 1000 times. In order to compare and check the performance of ML estimates, the criteria of average estimates and mean squared error (MSE) have been used, Each of the criteria is defined as follows

$$Mean = \sum_{i=1}^{n} \frac{\widehat{\varrho}_i}{n}, \qquad MSE = \sum_{i=1}^{n} \frac{(\varrho - \widehat{\varrho}_i)^2}{n},$$

in which $\widehat{\varrho}_i$ is equivalent to the i-th ML estimate of parameter ϱ and n is the sample volume.

The simulation results of two SUN-ARMA (1,1) and SUN-ARMA (2,2) processes are summarized in Table 1 and Table 2, respectively. The results of Table 1 and Table 2 confirm that with the increase of the sample size n, the MSE values decrease and the average estimates approach the real values of the parameters used in the simulation. Therefore, the obtained estimates are consistent and converge to the real values of the parameters with the increase of the sample size n. Therefore, the ML estimators provide good performance in estimating the parameters of the SUN-ARMA process.

Also, by examining the values of the skewness parameter, the value of weak or strong skewness does not have a significant effect on the estimation accuracy of other parameters, and the MSE values of the corresponding parameters for λ=-0.01 and λ=2 show almost similar performance.

8.2 Real Data Studies

In this part, researchers fit the SUN-ARMA model on a series of real data and evaluate the performance of this model in fitting real data in

Table 1: Average ML and MSE estimators in the SUN-ARMA (1,1) process

Distribution shape	Parameter	$n = 50$		$n = 150$		$n = 350$	
		Mean	MSE	Mean	MSE	Mean	MSE
weak skewness	$\varphi_1 = 0.5$	0.5403	0.0038	0.5202	0.0017	0.5187	0.0015
	$\theta_1 = 0.5$	0.5371	0.0033	0.5198	0.0018	0.5144	0.0013
	$\sigma^2 = 1$	1.2401	0.0044	1.0452	0.0011	0.9815	0.0008
	$\lambda = -0.01$	-0.0241	0.0013	-0.0132	0.0010	-0.0097	0.0009
Strong skewness	$\varphi_1 = 0.5$	0.5374	0.0032	0.5197	0.0017	0.5190	0.0015
	$\theta_1 = 0.5$	0.5311	0.0030	0.5140	0.0011	0.5135	0.0012
	$\sigma^2 = 1$	1.1746	0.0052	1.0298	0.0009	1.0253	0.0009
	$\lambda = 2$	2.2153	0.0125	2.1036	0.0093	1.9901	0.0091

Table 2: Average ML and MSE estimators in the SUN-ARMA $(2, 2)$ process

Distribution shape	Parameter	$n = 50$		$n = 150$		$n = 350$	
		Mean	MSE	Mean	MSE	Mean	MSE
weak skewness	$\varphi_1 = -0.3$	-0.2753	0.0024	-0.2916	0.0010	-0.2931	0.0010
	$\varphi_2 = 0.4$	0.4274	0.0021	0.4105	0.0009	0.4090	0.0008
	$\theta_1 = 0.3$	0.3201	0.0029	0.3064	0.0010	0.3057	0.0009
	$\theta_2 = 0.2$	0.2303	0.0034	0.2087	0.0013	0.2066	0.0010
	$\sigma^2 = 2$	2.2731	0.0041	1.9103	0.0007	2.0734	0.0007
	$\lambda = -0.01$	-0.0092	0.0009	-0.0122	0.0008	-0.0119	0.0007
Strong skewness	$\varphi_1 = -0.3$	-0.2645	0.0032	-0.2902	0.0012	-0.2920	0.0010
	$\varphi_2 = 0.4$	0.4187	0.0043	0.4065	0.0011	0.4061	0.0011
	$\theta_1 = 0.3$	0.3191	0.0037	0.3046	0.0009	0.3043	0.0009
	$\theta_2 = 0.2$	0.2289	0.0041	0.2056	0.0011	0.2057	0.0009
	$\sigma^2 = 2$	2.1645	0.0053	2.0856	0.0013	2.0649	0.0009
	$\lambda = 2$	2.1083	0.0032	1.9532	0.0008	1.9736	0.0006

comparison to the symmetric ARMA model that shows the characteristics of the normal distribution and is associated with the abbreviation Gaussian-ARMA. The primary data related to the import of Australian service goods in terms of million dollars is a quarterly average from September 1959 to December 1990 with a number of n=126 observations (the data is also used in Brockwell and Davis, 1996) [5]. Also, the file of the data is available in the ITSM software with the name "imports.tsm". According to the time series graph of the observations in Figure 1 (first row panel on the left), an upward trend is seen in the observations, and as a result, the data are not static. . The graph of observations shows both the trend and the increase of dispersion during a specific period of time. Therefore, in order to obtain static data, three different transformations are used to remove all the followings; namely, changes in dispersion, the upward trend and the constant value of the average. In the first step, Box-Cox power transformations are used (see Box and Cox (1964) in which, the parameter λ is defined and a value between -5$\leq\lambda\leq$5 is considered. All values of λ are tried and the optimal value for the data is chosen (the value that best approximates the normal distribution curve). Box-Cox power transformations on X data are defined as follows [4]

$$X\left(\lambda\right)=\begin{cases}\frac{X^{\lambda}-1}{\lambda}, & \lambda\neq0\\ \ln\left(X\right), & \lambda=0\end{cases},$$

which X(λ) is the transformed data.

By examining the different values of λ for the real data related to the import of Australian service goods, the optimal value of λ=0.8 is obtained, and therefore, to remove the changes in the dispersion of the observations, the transformation $x_t^*=\frac{x_t^{0.8}-1}{0.8}$ is used and the changes We confirm the observation that the results can be seen in Figure 1 (the first right row panel). In the second step, using differential transformations, we remove the upward trend of observations is removed. Therefore, by applying the first-order differentiation operator in the form

$$\nabla x_t^*=x_t^*-x_{t-1}^*,$$

The upward trend is removed from the observations. The results of removing the upward trend from the observations can be seen in Figure pic.png

Figure 1: Chart of primary data of Australian imports (top-left) and stationary step transformations (top-right and middle panel) along with ACF (bottom-right) and PACF (bottom-left) charts

1 (the second left panel). In the third step, we plot the data around the central mean and the resulting data become completely stationary, as shown in Figure 1 (second row panel on the right). Now the obtained observations are static. ACF autocorrelation and partial PACF autocorrelation diagrams of the data are also drawn in the third row of Figure 1, respectively. AIC model selection criterion for different values of p and q orders for SUN-ARMA and Gaussian-ARMA processes has been calculated. As we know, we consider the model that provides the lowest AIC value as the appropriate model. Therefore, according to the AIC criterion, the order of auto regression p=7 and the order of mov-

Table 3: ML estimators of SUN-ARMA $(7,1)$ and Gaussian-ARMA $(7,1)$ models for import data

Model		SUN-ARMA(7,1)	Gaussian-ARMA(7,1)
	φ_1	0.49261	-0.55570
	φ_2	-0.08377	-0.11780
	φ_3	-0.09124	-0.13759
	φ_4	0.27605	0.09773
	φ_5	-0.37126	-0.15687
	φ_6	-0.07007	-0.32009
	φ_7	-0.07495	-0.27816
	θ_1	-0.60259	0.49496
	λ	1.09403	.—
	σ^2	5093.32291	4865.78830
Log-Like		-671.04522	-708.56526
AIC		1362.09013	1435.13493
BIC		1389.79701	1460.06747

ing average q=1 are suggested for SUN-ARMA and Gaussian-ARMA processes. Accordingly, in the following, we consider the SUN-ARMA (7,1) and Gaussian-ARMA (7,1) processes. In Table 3, the parameter estimation values of two SUN-ARMA (7,1) and Gaussian-ARMA (7,1) processes are presented. Likewise, the model selection criteria including the logarithm of likelihood, AIC and BIC have also been calculated, where the lowest absolute value of the model selection criteria belongs to the SUN-ARMA (7,1) process, which has a better performance compared to the SUN-ARMA (7,1)process is observed .This can result in a symmetric Gaussian-ARMA (7,1) process is observed.

9 Summary and Conclusion

Time series models have been studied and investigated by applying them to data per unit of time in different sciences. Due to suitable features such as ability of autoregressive moving average models, it has been interesting subject for a long time. Moving average self-reversal processes are defined based on previous values of the same process and results in current and previous times. Classical time series modeling is often based on symmetric distributions such as normal, which are not practical for asymmetric observations. Therefore, in this article, we present a new moving average self-regressive time series model based on the asymmetric multivariate Skew normal distribution. Using the maximum likelihood estimation method with the EM algorithm approach, the process parameters are estimated. Based on the simulation studies, the estimates obtained from the EM method are consistent and converge with the actual value of the parameters. Two real time series of goods imports and population estimation in Australia confirm a more accurate performance of the SUN-ARMA process compared to its symmetric Gaussian-ARMA counterpart. It should be noted that the two real series in question have a skew structure. The goodness of fit of the SUN-ARMA model has also been confirmed using residual analysis.

According to the theoretical and simulation content in this article, the following topics can be mentioned as the future of research:

- Studying and investigating nonlinear regression models using errors from the asymmetric distribution of the unbounded normal distribution and calculating different methods of parameter estimation and checking the applicability and efficiency of the model.
- Examining the application of the unbounded normal distribution in fitting censored data and estimating model parameters under different censors such as first type, second type, incremental and hybrid censoring.
- Examining different parametric and non-parametric methods of estimating the parameters of regression and time series models with unlimited Skew normal errors and introducing the best estimation method.
- Examining non-stationary time series models such as ARIMA process based on samples from the unlimited Skew normal distribution and fitting real data with the help of the new process.
- Presenting new time series models for asymmetric and censored data based on new asymmetric distributions and calculating the best estimator for unknown model parameters.

References

[1] Arellano-Valle R.B. , H. Gómez and F.A. Quintana, Statistical inference for a general class of asymmetric distributions. *Journal of Statistical Planning and Inference*, 128 (2005), 427-443.

[2] Azalini.A , A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12 (1985), 171-178.

[3] Azzalini A. and Dalla-Val A. le, The multivariate skew-normal distribution. *Biometrika*, 83 (1996), 715-726.

[4] Box G.E.P. and D.R. Cox, An analysis of transformations. *Journal of the Royal Statistical Society: Series B*, 26 (1964), 211-252.

[5] Brockwell P.J. and R.A. Davis, *Time Series and Forecasting*, Springer-Verlag New York, Inc. (1996).

[6] Gupta A.K. , G. Gonzalez-Farias and J.A. Dominguez-Molina, A multivariate skew normal distribution. *Journal of Multivariate Analysis*, 89 (2004), 181–190.

[7] Lee S.X. and G.J. McLachlan, On the fitting of mixtures of multivariate skew t-distributions via the EM algorithm. *arXiv: Methodology*, (2011), arXiv:11094706.

[8] Lee S.X. and G.J. McLachlan, On mixtures of skew normal and skew t-distributions. *Advances in Data Analysis and Classification*, 7 (2013), 241-266.

[9] Sahu S.K. , D.K. Dey and M.D. Branco, A new class of multivariate skew distributions with applications to Bayesian regression models. *Canadian Journal of Statistics*, 31(2) (2003), 129–150.

[10] Camassa, R., Kilic, Z., McLaughlin, R.M.. On the symmetry properties of a random passive scalar with and without boundaries, and their connection between hot and cold states. *Physica D: Nonlinear Phenomena* ,(2019), 400, 132124.

[11] Resnick, S.G., *Dynamical problems in non-linear advective partial differential equations*. 1996.

[12] Vanden Eijnden, E. Non-gaussian invariant measures for the majda model of decaying turbulent transport. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 54 (2001), 1146–1167.

[13] Bronski, J.C., Camassa, R., Lin, Z., McLaughlin, R.M., Scotti, A. An explicit family of probability measures for passive scalar diffusion in a random flow. *Journal of Statistical Physics*, 128 (2007), 927–968.

Parisa Zarin
Ph.D. Student
Department of Statistics
Marvdasht Branch
Islamic Azad University
Marvdasht, Iran.
E-mail: p.zarin256@gmail.com

Zahra Khodadadi
Assistant Professor of Statistics
Department of Statistics
Marvdasht Branch
Islamic Azad University
Marvdasht, Iran.
zkhodadadi@iau.ac.ir

Mohsen Maleki
Assistant Professor of Statistics
Department of Statistics
Faculty of Mathematics and Statistics
University of Isfahan
Isfahan, Iran.
E-mail: m.maleki.stat@gmail.com

Hadieh Jafarpour
Assistant Professor of Statistics
Department of Mathematics and Statistics
Shiraz Branch
Islamic Azad University
Shiraz, Iran.

E-mail: Hedieh.jafarpour@gmail.com