

Article

Skew Generalized Normal Innovations for the AR(p) Process Endorsing Asymmetry

Ané Neethling , Johan Ferreira , Andriëtte Bekker  and Mehrdad Naderi 

Department of Statistics, Faculty of Natural and Agricultural Sciences, University of Pretoria, Pretoria 0028, South Africa; johan.ferreira@up.ac.za (J.F.); andriette.bekker@up.ac.za (A.B.); m.naderi@up.ac.za (M.N.)

* Correspondence: ane.neethling@up.ac.za

Received: 7 July 2020 ; Accepted: 26 July 2020; Published: 29 July 2020



Abstract: The assumption of symmetry is often incorrect in real-life statistical modeling due to asymmetric behavior in the data. This implies a departure from the well-known assumption of normality defined for innovations in time series processes. In this paper, the autoregressive (AR) process of order p (i.e., the AR(p) process) is of particular interest using the skew generalized normal (SGN) distribution for the innovations, referred to hereafter as the ARSGN(p) process, to accommodate asymmetric behavior. This behavior presents itself by investigating some properties of the SGN distribution, which is a fundamental element for AR modeling of real data that exhibits non-normal behavior. Simulation studies illustrate the asymmetry and statistical properties of the conditional maximum likelihood (ML) parameters for the ARSGN(p) model. It is concluded that the ARSGN(p) model accounts well for time series processes exhibiting asymmetry, kurtosis, and heavy tails. Real time series datasets are analyzed, and the results of the ARSGN(p) model are compared to previously proposed models. The findings here state the effectiveness and viability of relaxing the normal assumption and the value added for considering the candidacy of the SGN for AR time series processes.

Keywords: conditional maximum likelihood estimator; skew- t ; generalized normal; heavy tails; skewness

1. Introduction

The autoregressive (AR) model is one of the simplest and most popular models in the time series context. The AR(p) time series process y_t is expressed as a linear combination of p finite lagged observations in the process with a random innovation structure for $t = \{1, 2, \dots\}$ and is given by:

$$y_t = \varphi_0 + \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} + a_t, \quad (1)$$

where $\varphi_1, \varphi_2, \dots, \varphi_p$ are known as the p AR parameters. The process mean (i.e., mean of y_t) for the AR(p) process in (1) is given by $\mu^* = \varphi_0(1 - \varphi_1 - \varphi_2 - \dots - \varphi_p)^{-1}$. Furthermore, if all roots of the characteristic equation:

$$\varphi(x) = 1 - \varphi_1 x - \varphi_2 x^2 - \dots - \varphi_p x^p = 0$$

are greater than one in absolute value, then the process is described as stationary (which is considered for this paper). The innovation process a_t in (1) represents white noise with mean zero (since the process mean is already built into the AR(p) process) and a constant variance, which can be seen as independent “shocks” randomly selected from a particular distribution. In general, it is assumed that a_t follows the normal distribution, in which case the time series process y_t will be a Gaussian process [1].

This assumption of normality is generally made due to the fact that natural phenomena often appear to be normally distributed (examples include age and weights), and it tends to be appealing due to its symmetry, infinite support, and computationally efficient characteristics. However, this assumption is often violated in real-life statistical analyses, which may lead to serious implications such as bias in estimates or inflated variances. Examples of time series data exhibiting asymmetry include (but are not limited to) financial indices and returns, measurement errors, sound frequency measurements, tourist arrivals, production in the mining sector, and sulphate measurements in water.

To address the natural limitations of normal-behavior, many studies have proposed AR models characterized by asymmetric innovation processes that were fitted to real data illustrating their practicality, particularly in the time series environment. The traditional approach for defining non-normal AR models is to keep the linear model (1) and let the innovation process a_t follow a non-normal process instead. Some early studies include the work of Pourahmadi [2], considering various non-normal distributions for the innovation process in an AR(1) process such as the exponential, mixed exponential, gamma, and geometric distributions. Tarami and Pourahmadi [3] investigated multivariate AR processes with the t distribution, allowing for the modeling of volatile time series data. Other models abandoning the normality assumption have been proposed in the literature (see [4] and the references within). Bondon [5] and, more recently, Sharafi and Nematollahi [6] and Ghasami et al. [7] considered AR models defined by the epsilon-skew-normal (\mathcal{ESN}), skew-normal (\mathcal{SN}), and generalized hyperbolic (\mathcal{GH}) innovation processes, respectively. Finally, AR models are not only applied in the time series environment: Tuuç et al. [8] considered AR models for the error terms in the regression context, allowing for asymmetry in the innovation structures.

This paper considers the innovation process a_t to be characterized by the skew generalized normal (\mathcal{SGN}) distribution (introduced in Bekker et al. [9]). The main advantages gained from the \mathcal{SGN} distribution include the flexibility in modeling asymmetric characteristics (skewness and kurtosis, in particular) and the infinite real support, which is of particular importance in modeling error structures. In addition, the \mathcal{SGN} distribution adapts better to skewed and heavy-tailed datasets than the normal and \mathcal{SN} counterparts, which is of particular value in the modeling of innovations for AR processes [7].

The focus is firstly on the \mathcal{SGN} distribution assumption for the innovation process a_t . Following the skewing methodology suggested by Azzalini [10], the \mathcal{SGN} distribution is defined as follows [9]:

Definition 1. Random variable X is characterized by the \mathcal{SGN} distribution with location, scale, shape, and skewing parameters μ, α, β , and λ , respectively, if it has probability density function (PDF):

$$f_X(x; \mu, \alpha, \beta, \lambda) = \frac{2}{\alpha} \phi(z; \beta) \Phi(\sqrt{2}\lambda z),$$

where $z = (x - \mu)/\alpha \in \mathbb{R}$, $\mu, \lambda \in \mathbb{R}$, and $\alpha, \beta > 0$. This is denoted by $X \sim \mathcal{SGN}(\mu, \alpha, \beta, \lambda)$.

Referring to Definition 1, $\Phi(\cdot)$ denotes the cumulative distribution function (CDF) for the standard normal distribution, with $2\Phi(\sqrt{2}\lambda z)$ operating as a skewing mechanism [10]. The symmetric base PDF to be skewed is given by ϕ , denoting the PDF of the generalized normal (\mathcal{GN}) distribution given by:

$$\phi(x; \mu, \alpha, \beta) = \frac{\beta}{2\alpha\Gamma\left(\frac{1}{\beta}\right)} \exp(-|z|^\beta), \quad (2)$$

where $\Gamma(\cdot)$ denotes the gamma function [11]. The standard case for the \mathcal{SGN} distribution with $\mu = 0$ and $\alpha = 1$ in Definition 1 is denoted as $X \sim \mathcal{SGN}(\beta, \lambda)$. Furthermore, the \mathcal{SGN} distribution results in the standard \mathcal{SN} distribution in the case of $\mu = 0$, $\alpha = \sqrt{2}$ and $\beta = 2$, denoted as $X \sim \mathcal{SN}(\lambda)$ [11]. In addition, the distribution of X collapses to that of the standard normal distribution if $\lambda = 0$ [10].

Following the definition and properties of the $SG\mathcal{N}$ distribution, discussed in [11] and summarized in Section 2 below, the $AR(p)$ process in (1) with independent and identically distributed innovations $a_t \sim SG\mathcal{N}(0, \alpha, \beta, \lambda)$ is presented with its maximum likelihood (ML) procedure in Section 3. Section 4 evaluates the performance of the conditional ML estimator for the $ARSGN(p)$ model through simulation studies. Real financial, chemical, and population datasets are considered to illustrate the relevance of the newly proposed model, which can accommodate both skewness and heavy tails simultaneously. Simulation studies and real data applications illustrate the competitive nature of this newly proposed model, specifically in comparison to the $AR(p)$ process under the normality assumption, as well as the $ARSN(p)$ process proposed by Sharafi and Nematollahi [6]; this is an $AR(p)$ process with the innovation process defined by the \mathcal{SN} distribution such that $a_t \sim \mathcal{SN}(0, \alpha, \lambda)$. In addition, this paper also considers the $AR(p)$ process with the innovation process defined by the skew- t (ST) distribution [12] such that $a_t \sim ST(0, \alpha, \lambda, \nu)$, referred to as an $ARST(p)$ process. With a shorter run time, it is shown that the proposed $ARSGN(p)$ model competes well with the $ARST(p)$ model, thus accounting well for processes exhibiting asymmetry and heavy tails. Final remarks are summarized in Section 5.

2. Review on the Skew Generalized Normal Distribution

Consider a random variable $X \sim SG\mathcal{N}(\mu, \alpha, \beta, \lambda)$ with PDF defined in Definition 1. The behavior of the skewing mechanism $2\Phi(\sqrt{2}\lambda z)$ and the PDF of the $SG\mathcal{N}$ distribution is illustrated in Figures 1 and 2, respectively (for specific parameter structures). From Definition 1, it is clear that β does not affect the skewing mechanism, as opposed to λ . When $\lambda = 0$, the skewing mechanism yields a value of one, and the $SG\mathcal{N}$ distribution simplifies to the symmetric \mathcal{GN} distribution. Furthermore, as the absolute value of λ increases, the range of x values over which the skewing mechanism is applied decreases within the interval $(0, 2)$. As a result, higher peaks are evident in the PDF of the $SG\mathcal{N}$ distribution [11]. These properties are illustrated in Figures 1 and 2.

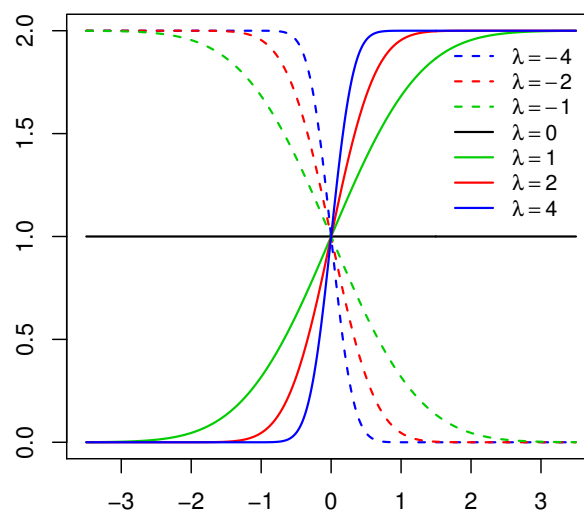


Figure 1. Skewing mechanism $2\Phi(\sqrt{2}\lambda z)$ for the skew generalized normal ($SG\mathcal{N}$) distribution with $\mu = 0$, $\alpha = \sqrt{2}$, $\beta = 20$, and various values of λ .

The main advantage of the $SG\mathcal{N}$ distribution is its flexibility by accommodating both skewness and kurtosis (specifically, heavier tails than that of the \mathcal{SN} distribution); the reader is referred to [11] for more detail. Furthermore, a random variable from the binomial distribution with parameters n and p can be approximated by a normal distribution with mean np and variance $np(1-p)$ if n is large or $p \approx 0.5$ (that is, when the distribution is approximately symmetrical). However, if $p \neq 0.5$, an asymmetric distribution is observed with considerable skewness for both large and small values of

p. Bekker et al. [9] addressed this issue and showed that the $SG\mathcal{N}$ distribution outperforms both the normal and $S\mathcal{N}$ distributions in approximating binomial distributions for both large and small values of p with $n \leq 30$.

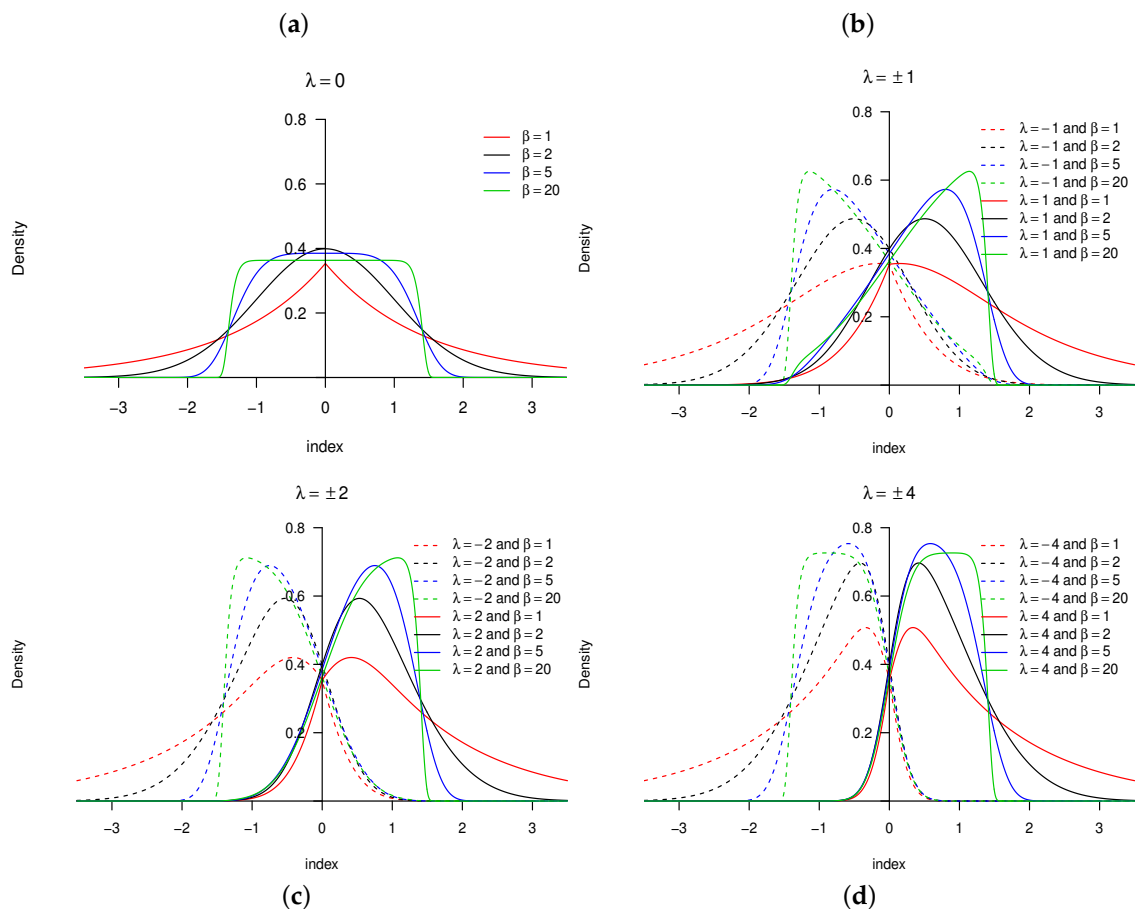


Figure 2. Probability density function (PDF) for the $SG\mathcal{N}$ distribution with $\mu = 0$, $\alpha = \sqrt{2}$ and various values of β for (from left to right, top to bottom): (a) $\lambda = 0$; (b) $\lambda = \pm 1$; (c) $\lambda = \pm 2$; (d) $\lambda = \pm 4$.

In order to demonstrate some characteristics (in particular, the expected value, variance, kurtosis, skewness, and moment generating function (MGF)) of the $SG\mathcal{N}$ distribution, the following theorem from [11] can be used to approximate the k th moment.

Theorem 1. Suppose $X \sim SG\mathcal{N}(\beta, \lambda)$ with the PDF defined in Definition 1 for $\mu = 0$ and $\alpha = 1$, then:

$$\mathbb{E}[X^k] = \begin{cases} \Gamma\left(\frac{k+1}{\beta}\right) / \Gamma\left(\frac{1}{\beta}\right) & \text{for } k \text{ even,} \\ \Gamma\left(\frac{k+1}{\beta}\right) \left\{ 2\mathbb{E}\left[\Phi\left(\sqrt{2}\lambda A^{\frac{1}{\beta}}\right)\right] - 1 \right\} / \Gamma\left(\frac{1}{\beta}\right) & \text{for } k \text{ odd,} \end{cases}$$

where A is a random variable distributed according to the gamma distribution with scale and shape parameters 1 and $(k+1)/\beta$, respectively.

Proof. The reader is referred to [11] for the proof of Theorem 1. \square

Theorem 1 is shown to be the most stable and efficient for approximating the k th moment of the distribution of X , although it is also important to note that the sample size $n > 60,000$ such that significant estimates of these characteristics are obtained. Figure 3 illustrates the skewness and kurtosis characteristics that were calculated using Theorem 1 for various values of β and λ .

When evaluating these characteristics, it is seen that both kurtosis and skewness are affected by β and λ jointly. In particular (referring to Figure 3):

- Skewness is a monotonically increasing function for $\beta \leq 2$ —that is, for $\lambda < 0$, the distribution is negatively skewed, and vice versa.
- In contrast to the latter, skewness is a non-monotonic function for $\beta > 2$.
- Considering kurtosis, all real values of λ and decreasing values of β result in larger kurtosis, yielding heavier tails than that of the normal distribution.

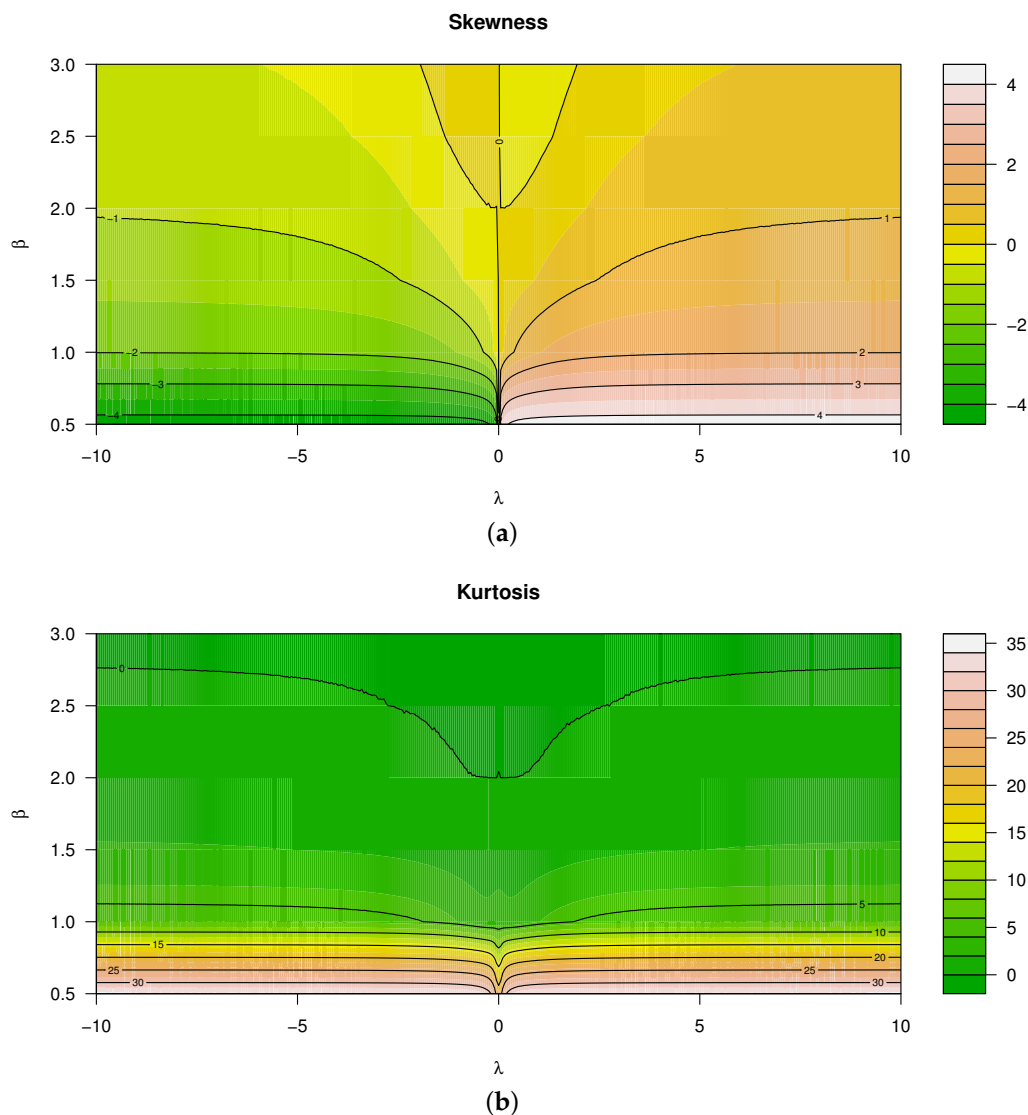


Figure 3. Measures for $X \sim \text{SGN}(\beta, \lambda)$ for various values of β and λ (from top to bottom): (a) Skewness. (b) Kurtosis.

In a more general sense for an arbitrary μ and α , Theorem 1 can be extended as follows:

Theorem 2. Suppose $X \sim \text{SGN}(\beta, \lambda)$ and $Y = \mu + \alpha X$ such that $Y \sim \text{SGN}(\mu, \alpha, \beta, \lambda)$, then:

$$\mathbb{E}[Y^r] = \sum_{k=0}^r \binom{n}{r} \mu^{r-k} \alpha^k \mathbb{E}[X^k]$$

with $\mathbb{E}[X^k]$ defined in Theorem 1.

Proof. The proof follows from Theorem 1 [11]. \square

Theorem 3. Suppose $X \sim \text{SGN}(\beta, \lambda)$ with the PDF defined in Definition 1 for $\mu = 0$ and $\alpha = 1$, then the MGF is given by:

$$M_X(t) = \sum_{j=0}^{\infty} \frac{\Gamma\left(\frac{j+1}{\beta}\right)}{\Gamma\left(\frac{1}{\beta}\right)j!} \left[t^j \mathbb{E} \left[\Phi \left(\sqrt{2\lambda}W \right) \right] + (-t)^j \mathbb{E} \left[\Phi \left(-\sqrt{2\lambda}W \right) \right] \right],$$

where W is a random variable distributed according to the generalized gamma distribution (refer to [11] for more detail) with scale, shape, and generalizing parameters 1, $j + 1$, and β , respectively.

Proof. From Definition 1 and (2), it follows that:

$$\begin{aligned} M_X(t) &= \int_{\mathbb{R}} \frac{\beta}{\Gamma\left(\frac{1}{\beta}\right)} \exp(-|x|^\beta + xt) \Phi \left(\sqrt{2\lambda}x \right) dx \\ &= \int_0^{\infty} \frac{\beta}{\Gamma\left(\frac{1}{\beta}\right)} \exp(-x^\beta + xt) \Phi \left(\sqrt{2\lambda}x \right) dx + \int_0^{\infty} \frac{\beta}{\Gamma\left(\frac{1}{\beta}\right)} \exp(-x^\beta - xt) \Phi \left(-\sqrt{2\lambda}x \right) dx \\ &= I_1 + I_2 \end{aligned}$$

Furthermore, using the infinite series representation of the exponential function:

$$\begin{aligned} I_1 &= \int_0^{\infty} \frac{\beta}{\Gamma\left(\frac{1}{\beta}\right)} \exp(-x^\beta + xt) \Phi \left(\sqrt{2\lambda}x \right) dx \\ &= \frac{\beta}{\Gamma\left(\frac{1}{\beta}\right)} \sum_{j=0}^{\infty} \frac{t^j}{j!} \int_0^{\infty} \Phi \left(\sqrt{2\lambda}x \right) \exp(-x^\beta) x^{(j+1)-1} dx \\ &= \frac{\beta}{\Gamma\left(\frac{1}{\beta}\right)} \sum_{j=0}^{\infty} \frac{t^j}{j!} \frac{\Gamma\left(\frac{j+1}{\beta}\right)}{\beta} \int_0^{\infty} \Phi \left(\sqrt{2\lambda}x \right) \frac{\beta}{\Gamma\left(\frac{j+1}{\beta}\right)} \exp(-x^\beta) x^{(j+1)-1} dx \\ &= \sum_{j=0}^{\infty} \frac{t^j}{j!} \frac{\Gamma\left(\frac{j+1}{\beta}\right)}{\Gamma\left(\frac{1}{\beta}\right)} \mathbb{E} \left[\Phi \left(\sqrt{2\lambda}W \right) \right], \end{aligned}$$

where W is a random variable distributed according to the generalized gamma distribution with scale, shape, and generalizing parameters 1, $j + 1 > 0$ and $\beta > 0$, respectively, and PDF:

$$f(w) = \frac{\beta}{\Gamma\left(\frac{j+1}{\beta}\right)} \exp(-w^\beta) w^{(j+1)-1}.$$

when $w > 0$, and zero otherwise. Similarly, I_2 can be written as:

$$I_2 = \sum_{j=0}^{\infty} \frac{-t^j}{j!} \frac{\Gamma\left(\frac{j+1}{\beta}\right)}{\Gamma\left(\frac{1}{\beta}\right)} \mathbb{E} \left[\Phi \left(-\sqrt{2\lambda}W \right) \right].$$

Thus, the MGF of X can be written as follows:

$$M_X(t) = \sum_{j=0}^{\infty} \frac{\Gamma\left(\frac{j+1}{\beta}\right)}{\Gamma\left(\frac{1}{\beta}\right)j!} \left[t^j \mathbb{E} \left[\Phi \left(\sqrt{2\lambda}W \right) \right] + (-t)^j \mathbb{E} \left[\Phi \left(-\sqrt{2\lambda}W \right) \right] \right]$$

□

The representation of the MGF (and by extension, the characteristic function) of the SGN distribution, defined in Theorem 3 above, can be seen as an infinite series of weighted expected values of generalized gamma random variables.

Remark 1. It is clear that β and λ jointly affect the shape of the SGN distribution. In order to distinguish between the two parameters, this paper will refer to λ as the skewing parameter, since the skewing mechanism depends on λ only. β will be referred to as the generalization parameter, as it accounts for flexibility in the tails and generalizing the normal to the GN distribution of [13].

3. The ARSGN(p) Model and Its Estimation Procedure

This section focuses on the model definition and ML estimation procedure of the ARSGN(p) model.

Definition 2. If Y_t is defined by an $AR(p)$ process with independent innovations $a_t \sim SGN(0, \alpha, \beta, \lambda)$ with PDF:

$$f(a_t; \alpha, \beta, \lambda) = \frac{2}{\alpha} \phi\left(\frac{a_t}{\alpha}; \beta\right) \Phi\left(\sqrt{2\lambda} \frac{a_t}{\alpha}\right),$$

then it is said that Y_t is defined by an ARSGN(p) process for time $t = \{1, 2, \dots\}$ and with process mean $\mu^* = \varphi_0 / (1 - \varphi_1 - \varphi_2 - \dots - \varphi_p)^{-1}$.

Remark 2. The process mean for an $AR(p)$ process keeps its basic definition, regardless of the underlying distribution for the innovation process a_t .

With a_t representing the process of independent distributed innovations with the PDF defined in Definition 2, the joint PDF for $(a_{p+1}, a_{p+2}, \dots, a_n)$ is given as:

$$f(a_{p+1}, a_{p+2}, \dots, a_n) = \left(\frac{2}{\alpha}\right)^{n-p} \prod_{t=p+1}^n \phi\left(\frac{a_t}{\alpha}; \beta\right) \Phi\left(\sqrt{2\lambda} \frac{a_t}{\alpha}\right), \tag{3}$$

for $n > p$. Furthermore, from (1), the innovation process can be rewritten as:

$$a_t = y_t - \varphi_0 - \varphi_1 y_{t-1} - \varphi_2 y_{t-2} - \dots - \varphi_p y_{t-p}. \tag{4}$$

Since the distribution for (Y_1, Y_2, \dots, Y_p) is intractable (being a linear combination of SGN variables), the complete joint PDF of (Y_1, Y_2, \dots, Y_n) is approximated by the conditional joint PDF of Y_t , for $t = \{p + 1, p + 2, \dots\}$, which defines the likelihood function $l(\Theta)$ for the ARSGN(p) model. Thus, using (3) and (4), the joint PDF of Y_t given (Y_1, Y_2, \dots, Y_p) is given by:

$$f(y_t | y_1, y_2, \dots, y_p) = \left(\frac{2}{\alpha}\right)^{n-p} \prod_{t=p+1}^n \phi(z_t^*; \beta) \Phi\left(\sqrt{2\lambda} z_t^*\right), \tag{5}$$

where $z_t^* = (y_t - \varphi_0 - \varphi_1 y_{t-1} - \varphi_2 y_{t-2} - \dots - \varphi_p y_{t-p}) / \alpha$. The ML estimator $\Theta = (\alpha, \beta, \lambda, \varphi)$ is based on maximizing the conditional log-likelihood function, where $\varphi = (\varphi_0, \varphi_1, \varphi_2, \dots, \varphi_p)$. Evidently,

the $p + m$ parameters need to be estimated for an AR(p) model, where m represents the number of parameters in the distribution considered for the innovation process.

Theorem 4. *If Y_t is characterized by an ARSGN(p) process, then the conditional log-likelihood function is given as:*

$$\ell_n(\Theta) = \sum_{t=p+1}^n \ln f(y_t - \varphi_0 - \varphi_1 y_{t-1} - \varphi_2 y_{t-2} - \cdots - \varphi_p y_{t-p}; \alpha, \beta, \lambda),$$

for $t = \{p + 1, p + 2, \dots\}$ and $f(\cdot)$ defined in Definition 2.

The conditional log-likelihood in Theorem 4 can be written as:

$$\ell_n(\Theta) = (n - p) \ln \left[\frac{\beta}{\alpha^2 \Gamma\left(\frac{1}{\beta}\right) \sqrt{2\pi}} \right] + \ln \prod_{t=p+1}^n \exp\left(-|z_t^*|^\beta\right) \left[\int_{-\infty}^{\sqrt{2}\lambda z_t^*} \exp\left(-\frac{x^2}{2}\right) dx \right],$$

where $z_t^* = a_t/\alpha$ and a_t is defined in (4). The ML estimation process of the ARSGN(p) process is summarized in Algorithm 1 below.

Algorithm 1:

- 1: Determine the sample mean \bar{y} , variance s^2 , and autocorrelations r_j for $j = 1, 2, \dots, p$.
- 2: Define the p Yule–Walker equations [14] in terms of theoretical autocorrelations ρ_i for an AR(p) process:

$$\rho_1 = \varphi_1 \rho_0 + \varphi_2 \rho_1 + \cdots + \varphi_p \rho_{p-1}$$

$$\rho_2 = \varphi_1 \rho_1 + \varphi_2 \rho_0 + \cdots + \varphi_p \rho_{p-2}$$

$$\vdots$$

$$\rho_p = \varphi_1 \rho_{p-1} + \varphi_2 \rho_{p-2} + \cdots + \varphi_p \rho_0$$

Set the theoretical autocorrelations ρ_i in the Yule–Walker equations equal to the sample autocorrelations r_i , and solve the method of moment estimates (MMEs) for the p AR parameters simultaneously in terms of r_1, r_2, \dots, r_p . Use these MMEs as the starting values for the AR parameters $\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \dots, \varphi_p)^\top$.

- 3: Set starting values for the intercept φ_0 and scale parameter α equal to the MMEs [14] such that:

$$\varphi_0 = \bar{y}(1 - \varphi_1 - \varphi_2 - \cdots - \varphi_p)$$

and

$$\alpha = \sqrt{s^2(1 - \varphi_1 r_1 - \varphi_2 r_2 - \cdots - \varphi_p r_p)}.$$

- 4: Set the starting values for the shaping parameters β and λ equal to two and zero, respectively.
 - 5: Use the `optim()` function in the R software to maximize the conditional log-likelihood function iteratively and yield the most likely underlying distribution with its specified parameters.
-

4. Application

In this section, the performance and robustness of the ML estimator Θ for the ARSGN(p) time series model is illustrated through various simulation studies. The proposed model is also applied to real data in order to illustrate its relevance, in comparison to previously proposed models. All computations were carried out using R 3.5.0 in a Win 64 environment with a 2.30 GHz/Intel(R) Core(TM) i5-6200U CPU Processor and 4.0 GB RAM, and run times are given in seconds. The R code is available from the first author upon request.

4.1. Numerical Studies

The aim of this subsection is to illustrate the performance and robustness of the conditional ML estimator Θ for the proposed model in Definition 2 using various simulation studies. Define the hypothetical AR(5) time series model, which will (partly) be considered in the simulation studies below:

$$y_t = 6 + 0.4y_{t-1} - 0.2y_{t-2} + 0.1y_{t-3} + 0.5y_{t-4} - 0.6y_{t-5} + a_t, \quad (6)$$

where a_t and the sample size n will be defined differently for each simulation study. The simulation studies are algorithmically described in Algorithm 2 below.

Algorithm 2:

- 1: Independent random variables $U \sim \mathcal{GN}(\beta)$ with PDF $\phi(\cdot; 0, 1, \beta)$ and $U_1 \sim \mathcal{N}(0, 1)$ are generated, using the `rgnorm()` and `rnorm()` functions in R, respectively.
- 2: Following [9,15], the innovation process a_t is simulated for $t = \{1, 2, \dots, n\}$ using:

$$a_t = \alpha \begin{cases} U & \text{if } U_1 \leq \sqrt{2}\lambda U \\ -U & \text{if } U_1 > \sqrt{2}\lambda U \end{cases}. \quad (7)$$

- 3: The time series y_t is determined using the `arima.sim()` function in R with the simulated innovations a_t from (7) and theoretical parameters defined in (6), then adding the process mean μ^* to the time series y_t .
 - 4: Algorithm 1 is applied to estimate the parameters of the ARSGN(p) model, considering various values of p .
 - 5: A second simulation study is implemented for $p = 2$, repeating Steps 1 to 4 above 500 times in order to analyze the sampling distributions for the parameters of the ARSGN(p) model.
 - 6: A third simulation study investigates the performance of the ARSGN(p) model when the innovation process a_t is described by an \mathcal{ST} distribution instead, for $p = 2$ in (6).
 - 7: A fourth simulation study extends the latter when the innovation process a_t is described by an \mathcal{ST} distribution for various degrees of freedom, evaluating the performance of the ARSGN(p) model for various levels in the tails of a_t .
-

4.1.1. Simulation Study 1

In order to evaluate the conditional ML estimation performance of the ARSGN(p) model, the time series y_t will be simulated and estimated for various orders of $p = 1, 2, 3, 4, 5$ and sample sizes $n = 100, 500, 1000, 5000$. Assuming $a_t \sim \mathcal{SGN}(0, 4, 3, -10)$ for the hypothetical model defined in (6),

the innovation process a_t is simulated and the time series y_t is estimated with an ARSGN(p) model, as described in Algorithm 2.

Table 1 summarizes the parameter estimates and standard errors obtained from the ARSGN(p) model for all $p = 1, 2, 3, 4, 5$ and $n = 100, 500, 1000, 5000$. In general, it is clear that the model fits the simulated innovation processes (and time series) relatively well, except for λ , which tends to be more volatile with larger standard errors, although these standard errors decrease for larger sample sizes. It is also noted that there are occasional trade-offs in the estimation of β and λ for $p = 4$ and $p = 5$, which decreases the standard error for one parameter, but increases the standard error of the other. These occasional instances of “incorrect” estimations, which consequently have an influence on the estimation of φ_0 as well, may be explained by the fact that β and λ are jointly affecting the asymmetric behavior in the \mathcal{SGN} distribution. In addition to Table 1, Figure 4 illustrates the estimated ARSGN(3) model with the distribution of the residuals, where the residual at time t is defined as:

$$\hat{a}_t = Y_t - \hat{Y}_t,$$

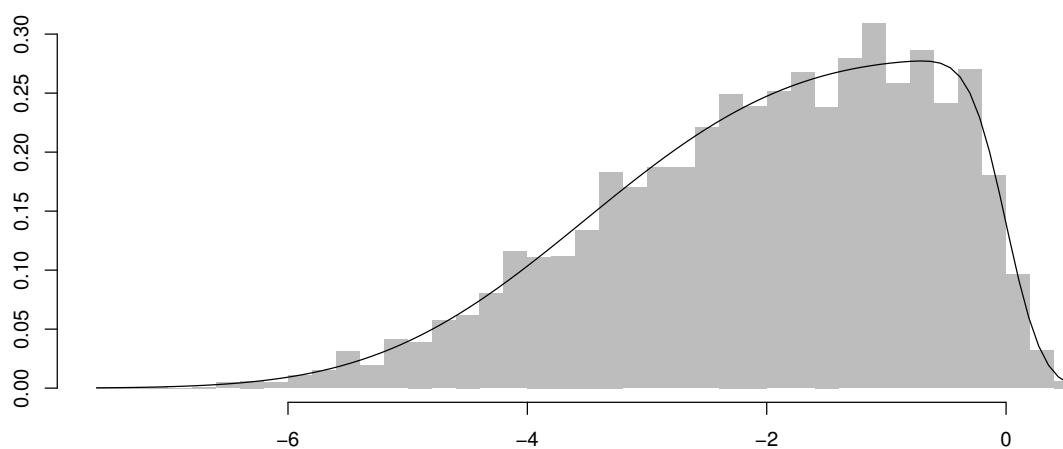
and \hat{Y}_t representing the estimated value at time t . The asymmetric behavior of the \mathcal{SGN} distribution is especially seen from the fitted model in Figure 4.

Table 1. The autoregressive (AR) process of order p using the skew generalized normal (ARSGN(p)) maximum likelihood (ML) parameter estimates and standard errors (in parentheses) for various sample sizes n and values of p .

Parameter	Sample Size n				
	100	500	1000	5000	
$p = 1$	$\alpha (= 4)$	3.374 (0.668)	3.591 (0.197)	4.054 (0.1367)	4.011 (0.067)
	$\beta (= 3)$	2.021 (0.553)	2.509 (0.272)	3.186 (0.263)	3.042 (0.122)
	$\lambda (= -10)$	-9.188 (7.078)	-14.7189 (3.226)	-11.395 (1.848)	-11.807 (0.916)
	$\varphi_0 (= 6)$	6.199 (0.421)	6.023 (0.156)	5.980 (0.136)	6.018 (0.062)
	$\varphi_1 (= 0.4)$	0.356 (0.064)	0.395 (0.022)	0.412 (0.019)	0.400 (0.009)
$p = 2$	$\alpha (= 4)$	3.277 (0.638)	3.596 (0.197)	4.050 (0.137)	4.009 (0.067)
	$\beta (= 3)$	1.982 (0.544)	2.513 (0.273)	3.180 (0.264)	3.038 (0.122)
	$\lambda (= -10)$	-7.899 (4.560)	-14.8478 (3.286)	-11.377 (1.856)	-11.786 (0.910)
	$\varphi_0 (= 6)$	5.739 (0.430)	6.068 (0.131)	5.958 (0.126)	6.028 (0.056)
	$\varphi_1 (= 0.4)$	0.330 (0.078)	0.392 (0.028)	0.414 (0.022)	0.408 (0.009)
	$\varphi_2 (= -0.2)$	-0.107 (0.061)	-0.208 (0.023)	-0.194 (0.021)	-0.210 (0.009)
$p = 3$	$\alpha (= 4)$	3.104 (0.635)	3.602 (0.196)	4.043 (0.138)	4.010 (0.066)
	$\beta (= 3)$	1.834 (0.497)	2.520 (0.273)	3.166 (0.263)	3.039 (0.122)
	$\lambda (= -10)$	-7.381 (3.565)	-15.297 (3.433)	-11.304 (1.843)	-11.825 (0.913)
	$\varphi_0 (= 6)$	6.234 (0.596)	5.931 (0.177)	5.984 (0.165)	6.090 (0.074)
	$\varphi_1 (= 0.4)$	0.339 (0.076)	0.398 (0.027)	0.409 (0.022)	0.405 (0.009)
	$\varphi_2 (= -0.2)$	-0.116 (0.056)	-0.216 (0.023)	-0.193 (0.022)	-0.208 (0.010)
	$\varphi_3 (= 0.1)$	0.001 (0.076)	0.128 (0.022)	0.096 (0.022)	0.090 (0.009)
$p = 4$	$\alpha (= 4)$	3.484 (0.180)	3.855 (0.072)	4.048 (0.137)	4.007 (0.066)
	$\beta (= 3)$	30.733 (40.892)	31.009 (9.872)	3.180 (0.264)	3.034 (0.121)
	$\lambda (= -10)$	1.541 (0.305)	2.271 (0.191)	-11.496 (1.879)	-11.915 (0.916)
	$\varphi_0 (= 6)$	5.215 (1.265)	1.965 (0.456)	6.160 (0.425)	6.527 (0.194)
	$\varphi_1 (= 0.4)$	0.296 (0.058)	0.402 (0.024)	0.414 (0.019)	0.399 (0.008)
	$\varphi_2 (= -0.2)$	-0.185 (0.058)	-0.209 (0.023)	-0.205 (0.021)	-0.214 (0.009)
	$\varphi_3 (= 0.1)$	0.012 (0.081)	0.086 (0.026)	0.112 (0.022)	0.093 (0.009)
	$\varphi_4 (= 0.5)$	0.550 (0.074)	0.536 (0.027)	0.475 (0.019)	0.497 (0.008)

Table 1. Cont.

Parameter	Sample Size n			
	100	500	1000	5000
α (= 4)	1.273 (0.582)	3.840 (0.070)	3.797 (0.131)	3.686 (0.064)
β (= 3)	0.928 (0.255)	35.158 (15.083)	15.671 (2.697)	14.487 (1.129)
λ (= -10)	-1.190 (0.646)	2.251 (0.189)	2.177 (0.218)	1.962 (0.094)
φ_0 (= 6)	5.986 (0.029)	2.103 (0.207)	2.415 (0.175)	2.572 (0.081)
$p = 5$ φ_1 (= 0.4)	0.352 (0.015)	0.385 (0.024)	0.412 (0.017)	0.406 (0.007)
φ_2 (= -0.2)	-0.138 (0.019)	-0.192 (0.017)	-0.200 (0.016)	-0.204 (0.007)
φ_3 (= 0.1)	-0.012 (0.033)	0.096 (0.018)	0.109 (0.015)	0.092 (0.007)
φ_4 (= 0.5)	0.499 (0.019)	0.517 (0.019)	0.480 (0.016)	0.489 (0.007)
φ_5 (= -0.6)	-0.634 (0.024)	-0.570 (0.019)	-0.607 (0.016)	-0.610 (0.007)

Figure 4. Histogram of the residuals with the fitted ARSGN(3) model overlaid for sample size $n = 5000$.

4.1.2. Simulation Study 2

Sampling distributions for the parameters can be used to evaluate the robustness of the estimator for the proposed model. In order to construct the sampling distributions for the parameters in Table 1, a Monte Carlo simulation study is applied by repeating the simulation for a_t and the estimation procedure 500 times, considering $p = 2$ and $n = 5000$ for the hypothetical model defined in (6) with $a_t \sim \mathcal{SGN}(0, 4, 3, -10)$.

The sampling distributions obtained for the conditional ML parameter estimates are illustrated in Figure 5, all centered around their theoretical values. The occasional trade-offs between β and λ (noted in Simulation Study 1) are evident from these distributions. Furthermore, Table 2 summarizes the 5th, 50th (thus, the median), and 95th percentiles obtained from these sampling distributions, of which the 5th and 95th percentiles can be used for approximating the 95% confidence intervals. It is evident that the theoretical values for all parameters fall within their respective 95% confidence intervals and exclude zero, suggesting that all parameter estimates are significant. It is concluded that the conditional ML estimation of the proposed model is robust since all 50th percentiles are virtually identical to each of the theoretical values, confirming what is depicted in Figure 5.

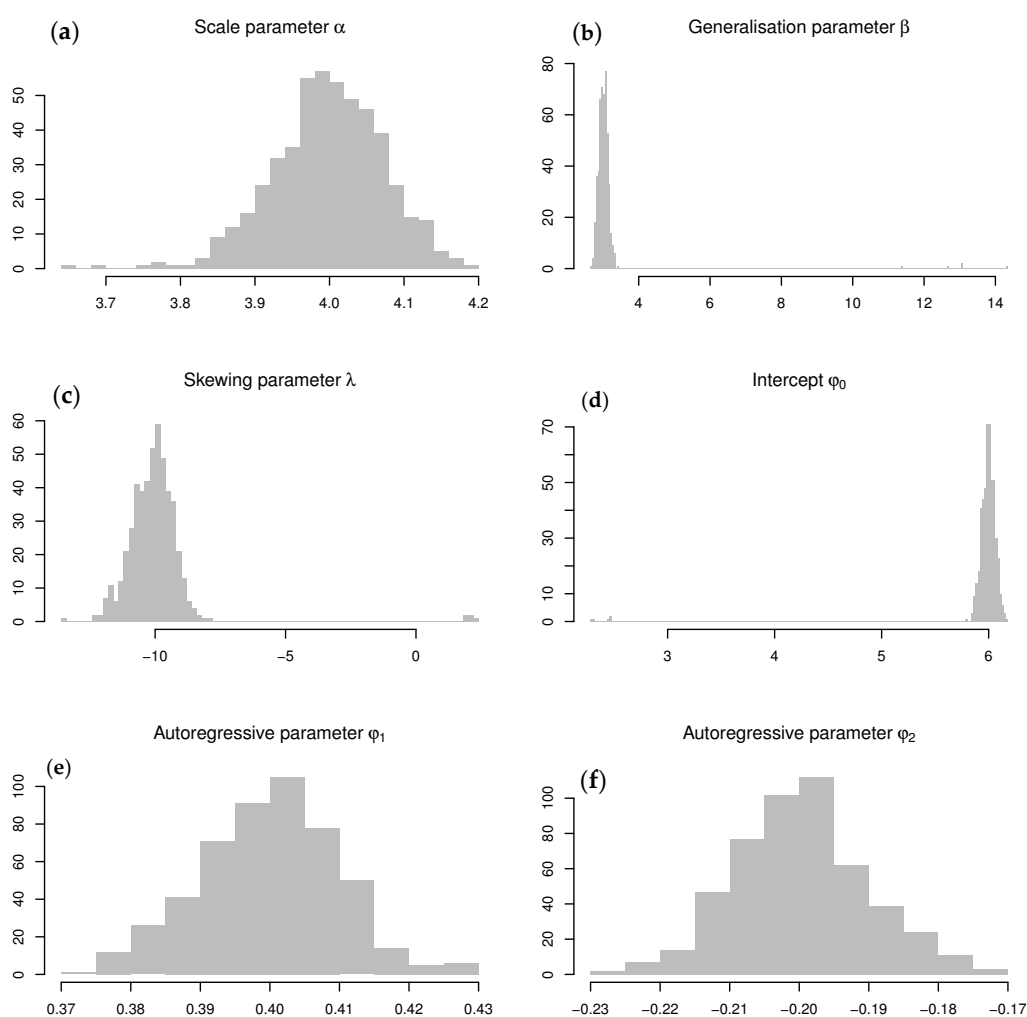


Figure 5. Sampling distributions for the ARSGN(2) ML parameter estimates obtained from a Monte Carlo simulation study (from left to right, top to bottom): (a) Sampling distribution for α . (b) Sampling distribution for β . (c) Sampling distribution for λ . (d) Sampling distribution for φ_0 . (e) Sampling distribution for φ_1 . (f) Sampling distribution for φ_2 .

Table 2. Percentiles for the ARSGN(2) ML parameter estimates obtained from a Monte Carlo simulation study.

	Parameter					
	α	β	λ	φ_0	φ_1	φ_2
P_5	3.869	2.804	-11.510	5.895	0.383	-0.215
P_{50}	4.000	3.009	-10.042	5.998	0.401	-0.200
P_{95}	4.116	3.236	-8.880	6.096	0.415	-0.182

4.1.3. Simulation Study 3

For comparison and completeness' sake, consider the hypothetical time series model defined in (6), for up to $p = 2$ only. This time, the innovation process is simulated from an ST distribution such that $a_t \sim ST(0, 2, 15, 5)$ [12], and thus, the simulated time series y_t is referred to as an ARST(2) process. The aim of this simulation study is to evaluate the fit of the ARSGN(2) model in comparison

to the AR(2), ARSN(2), and ARST(2) models each, even though the true innovation process follows an ST distribution.

Considering a sample size of $n = 5000$, the innovation process a_t was simulated using the `rst()` function in R. The AR(2), ARSN(2), ARSGN(2), and ARST(2) models are each fitted to the time series y_t by maximizing the respective conditional log-likelihood functions. Starting values were chosen similar for all models as discussed in Algorithm 1, except for λ , being set equal to the sample skewness for both the ARSN(2) and ARST(2) models and the degrees of freedom ν being initialized at one for the latter. It should be noted that the starting value for λ in the ARSGN(p) model was not set to the sample skewness, since it was noted in Section 2 that λ and β jointly affect the skewness and shape of the distribution.

Table 3 summarizes the conditional ML parameter estimates with the standard errors given in parentheses, as well as the log-likelihood, Akaike information criterion (AIC), and run times for the various AR models, where AIC is defined as:

$$AIC = -2\ell_n(\hat{\Theta}) + 2M,$$

with $\ell_n(\hat{\Theta})$ representing the maximized value of the log-likelihood function defined in Theorem 4 and M denoting the number of parameters in the model [16]. From the log-likelihood and AIC values obtained for the different models, it can be seen that the ARSGN(2) model competes well with the ARST(2) model (see Figure 6). In particular, the run time for the estimation of the ARSGN(2) model is shorter than that of the ARST(2) model, with all parameters being significant at a 95% confidence level. Thus, it can be concluded that the ARSGN(2) model is a valid contender in comparison to other popular models accounting for asymmetry, kurtosis, and heavy tails and performs competitively considering the computational time.

Table 3. AR(2) ML parameter estimates with standard errors (in parentheses), log-likelihood $\ell_n(\hat{\Theta})$, Akaike information criterion (AIC), and run times for sample size $n = 5000$ and $a_t \sim ST(0, 2, 15, 5)$. ST, skew- t .

	Model			
	AR(2)	ARSN(2)	ARSGN(2)	ARST(2)
α	1.732 (0.017)	2.620 (0.027)	2.420 (0.093)	2.068 (0.038)
β	–	–	1.237 (0.038)	–
λ	–	20.140 (1.431)	9.654 (0.973)	16.301 (1.429)
Parameter ν	–	–	–	5.574 (0.424)
φ_0	7.813 (0.162)	5.973 (0.062)	6.041 (0.074)	6.020 (0.059)
φ_1	0.387 (0.014)	0.395 (0.005)	0.394 (0.005)	0.394 (0.005)
φ_2	–0.176 (0.014)	–0.197 (0.006)	–0.194 (0.006)	–0.196 (0.006)
$\ell_n(\hat{\Theta})$	–9839.2	–8612.9	–8456.5	–8426.4
AIC	19,686.5	17,235.8	16,925.1	16,904.0
Run time (s)	3.47583	2.11189	3.88825	5.07767

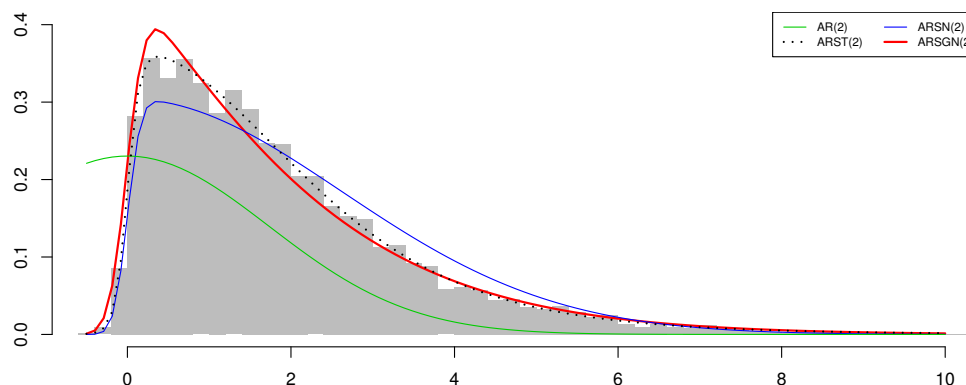


Figure 6. Histogram of simulated innovation process $a_t \sim \mathcal{ST}(0, 2, 15, 5)$ with fitted AR(2) models overlaid for sample size $n = 5000$.

4.1.4. Simulation Study 4

The purpose of this simulation study is to evaluate the estimation performance and adaptability of the proposed ARSGN(p) model (in comparison to some of its competitors) on processes with various tail weights simulated from the \mathcal{ST} distribution. Thus, consider the hypothetical time series model defined in (6), for up to $p = 2$ only and with $a_t \sim \mathcal{ST}(0, 2, 15, \nu)$ where ν represents the different degrees of freedom $\nu = 5, 6, \dots, 35$ under which y_t will be simulated. For a sample size of $n = 10,000$, the AR(2) model is estimated assuming various distributions for the innovation process a_t .

Figures 7 and 8 illustrate the standard errors of the estimates of the parameters of interest and AIC values obtained from the various AR models fitted to the simulated ARST(2) process for degrees of freedom $\nu = 5, 6, \dots, 35$. Figure 7 illustrates that the skewing parameter λ for the ARSGN(2) model is more volatile compared to the other parameters, although it performs with less volatility than λ in the ARSN(2) model for lower degrees of freedom. Observing the AIC values in Figure 8, it is clear that the AR(2) model (under the assumption of normality) performs the worst for all degrees of freedom, whereas the ARSN(2) model performs similar to that of the ARST(2) for degrees of freedom $\nu > 15$. In contrast, the proposed ARSGN(2) model performs almost equivalently to the ARST(2) model for all degrees of freedom indicating that the proposed model adapts well to various levels of skewness and kurtosis.

4.2. Real-World Time Series Analysis

This subsection illustrates the relevance of the ARSGN(p) model in areas such as chemistry, population studies, and economics, in comparison to previously proposed AR models. Descriptive statistics for all time series considered below are summarized in Table 4.

Table 4. Descriptive statistics of real time series data considered below, where * refers to the stationary time series data.

Series	Sample Statistic						
	n	\bar{y}	s	γ	κ	Min	Max
Viscosity	310	9.133	0.603	-0.465	-0.583	7.4	10.4
Australian population *	87	-0.331	11.469	-0.332	2.590	-37.6	41
Insolvencies in South Africa *	238	-0.567	39.609	0.902	9.792	-162	268

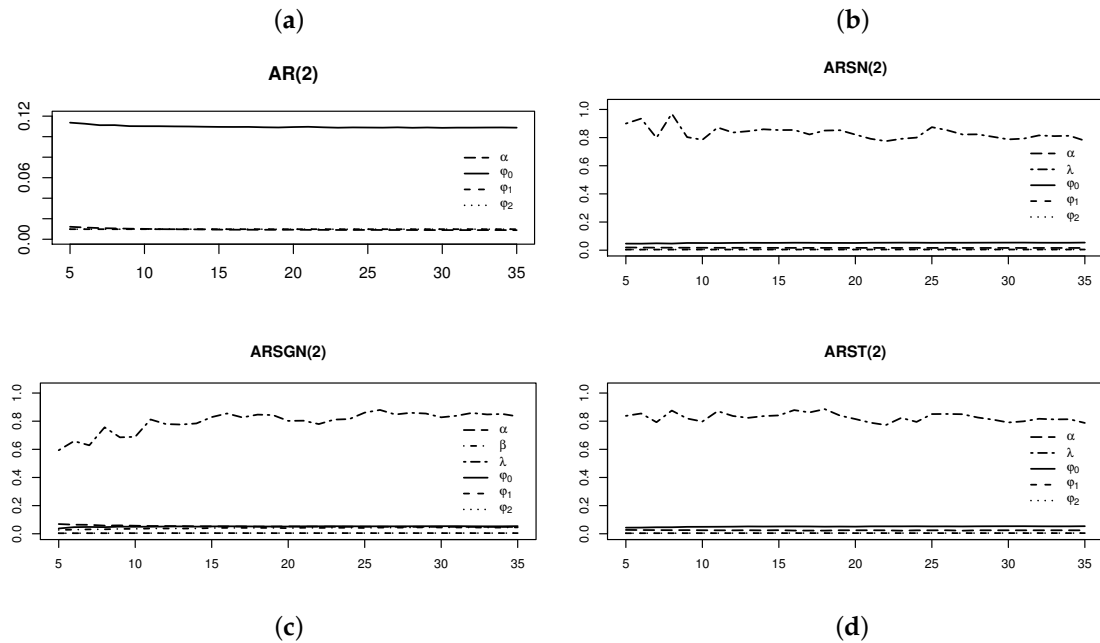


Figure 7. Standard errors of the parameter estimates obtained from AR(2) models fitted to an ARST(2) process simulated with $a_t \sim \mathcal{ST}(0, 2, 15, \nu)$ for different degrees of freedom ν . From left to right, top to bottom: (a) AR(2) model fitted assuming $a_t \sim \mathcal{N}(0, \alpha)$. (b) AR(2) model fitted assuming $a_t \sim \mathcal{SN}(0, \alpha, \lambda)$. (c) AR(2) model fitted assuming $a_t \sim \mathcal{SGN}(0, \alpha, \beta, \lambda)$. (d) AR(2) model fitted assuming $a_t \sim \mathcal{ST}(0, \alpha, \lambda, \nu)$.

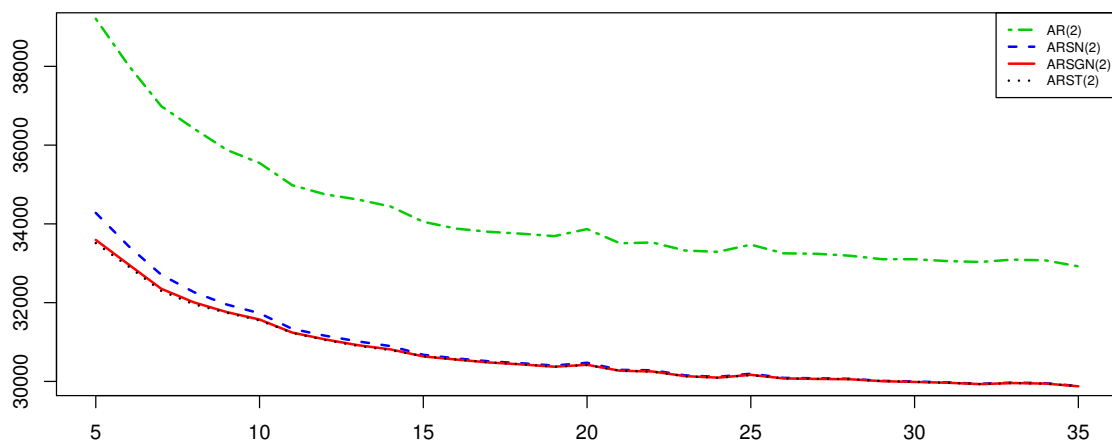


Figure 8. AIC values obtained from AR(2) models fitted (assuming various distributions for a_t) to an ARST(2) process simulated with $a_t \sim \mathcal{ST}(0, 2, 15, \nu)$ for different degrees of freedom ν .

4.2.1. Viscosity during a Chemical Process

In order to compare the ARSGN(p) model, defined in Definition 2, to previously proposed models (i.e., AR models assuming the normal and \mathcal{SN} distributions for the innovation process a_t , respectively), consider the time series data Series D in [17]. This dataset consists of $n = 310$ hourly measurements of viscosity during a chemical process, represented in Figure 9. The Shapiro–Wilk test applied to the time series data suggests that the data are not normally distributed with a p -value < 0.001 ; this is also confirmed by the histogram in Figure 9. From the autocorrelation function (ACF) and partial autocorrelation function (PACF), it is evident that an AR(1) model is a suitable choice for fitting a time series model.

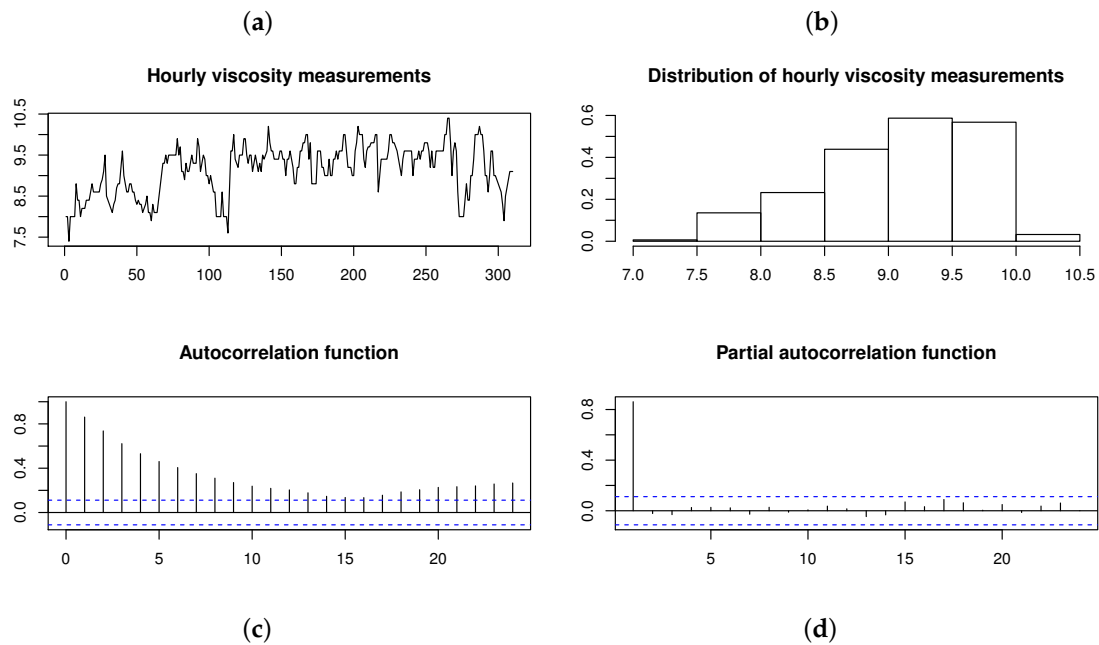


Figure 9. Viscosity measured hourly during a chemical process [17]. From left to right, top to bottom: (a) Time plot. (b) Histogram. (c) Autocorrelation function (ACF). (d) Partial autocorrelation function (PACF).

Previously, Box and Jenkins [17] fit an AR(1) model to this time series, assuming that the innovations are normally distributed. Sharafi and Nematollahi [6] relaxed this normality assumption and allowed for asymmetry by fitting an ARSN(1) model. Table 5 summarizes the conditional ML parameter estimates (with the standard errors given in parentheses) for both of these models, together with those obtained for the newly proposed ARSGN(1) model and the ARST(1) model. In addition, the maximized log-likelihood, AIC values, Kolmogorov–Smirnov (KS) test statistics, and run times are also represented for the various models, where the KS test statistic is defined as:

$$KS = \max_{a_t} |F(\hat{a}_t) - F(a_t)| ,$$

where $F(\hat{a}_t)$ and $F(a_t)$ represent the empirical and estimated distribution functions for the residuals and innovation process, respectively [18]. From the log-likelihood, AIC values, and KS test statistics calculated for the four models, it can be concluded that the ARSGN(1) model fits this time series the best, with a competitive estimation run time. Take note that from the estimated intercept $\hat{\varphi}_0$, the process mean is estimated as $\hat{\mu}^* = 9.492$, which is evident from the time plot in Figure 9.

Evaluating the standard errors for the parameter estimates obtained for the ARSGN(1) model, it is observed that all parameters differ significantly from zero at a 95% confidence level, except for the skewing parameter λ , suggesting that the innovation process does not contain significant skewness. This is confirmed by Figure 10, from which it is clear that only slight skewness is present when observing the distribution for the residuals obtained for all four models. In this case, it is also clear to see that the $S\mathcal{GN}$ distribution captures the kurtosis well. Finally, Figure 11 illustrates the CDFs for the various estimated models together with the empirical CDFs for the residuals obtained from the ARSGN(1) model, suggesting that the ARSGN(1) model fits the innovation process best.

Table 5. ML parameter estimates with standard errors (in parentheses), log-likelihood $\ell_n(\hat{\Theta})$, AIC, Kolmogorov–Smirnov (KS) test statistics, and run times for the hourly viscosity measurements during a chemical process [17].

Parameter	Model			
	AR(1)	ARSN(1)	ARSGN(1)	ARST(1)
α	0.300 (0.012)	0.386 (0.030)	0.134 (0.038)	0.228 (0.025)
β	–	–	0.771 (0.103)	–
λ	–	–1.324 (0.313)	–0.033 (0.027)	–0.186 (0.340)
ν	–	–	–	4.177 (1.409)
φ_0	1.197 (0.258)	1.515 (0.263)	0.598 (0.005)	0.871 (0.283)
φ_1	0.869 (0.028)	0.861 (0.028)	0.937 (0.001)	0.910 (0.028)
$\ell_n(\hat{\Theta})$	–67.8	–62.7	–43.3	–55.7
AIC	141.5	133.4	96.6	121.5
KS test statistic	0.233	0.223	0.181	0.230
Run time (s)	0.15462	0.37018	0.52313	0.69936

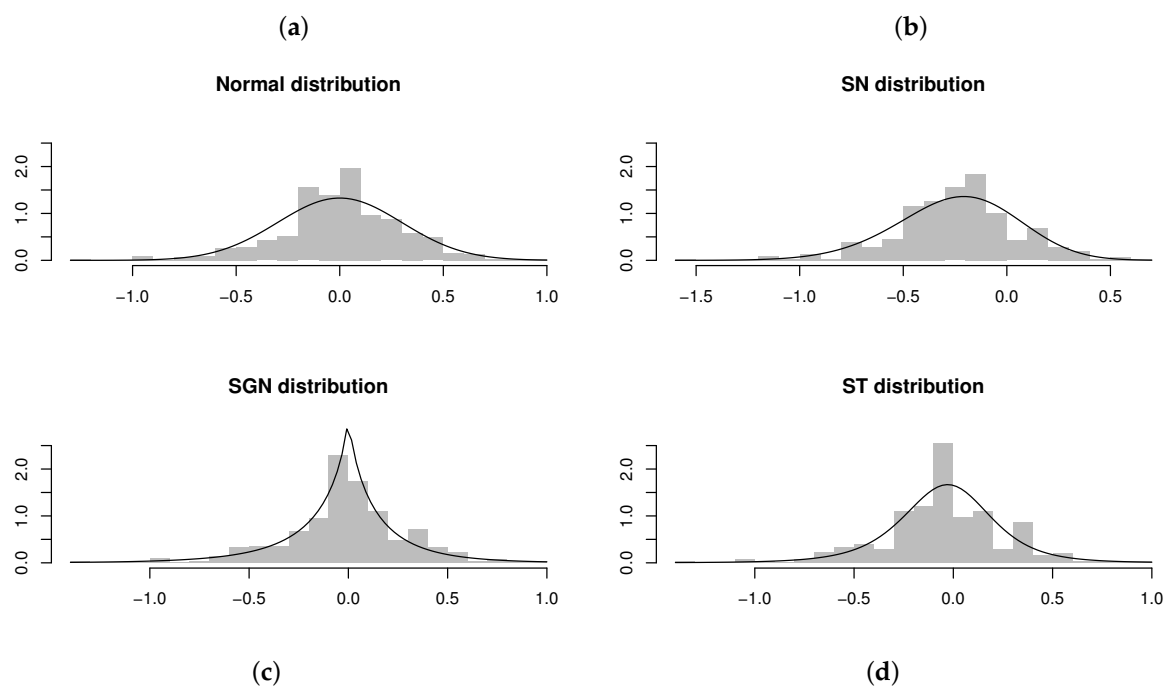


Figure 10. Residuals and estimated models obtained from AR(1) models fitted to the viscosity time series [17]. From left to right, top to bottom: (a) AR(1) model fitted assuming $a_t \sim \mathcal{N}(0, \alpha)$. (b) AR(1) model fitted assuming $a_t \sim \mathcal{SN}(0, \alpha, \lambda)$. (c) AR(1) model fitted assuming $a_t \sim \mathcal{SGN}(0, \alpha, \beta, \lambda)$. (d) AR(1) model fitted assuming $a_t \sim \mathcal{ST}(0, \alpha, \lambda, \nu)$.

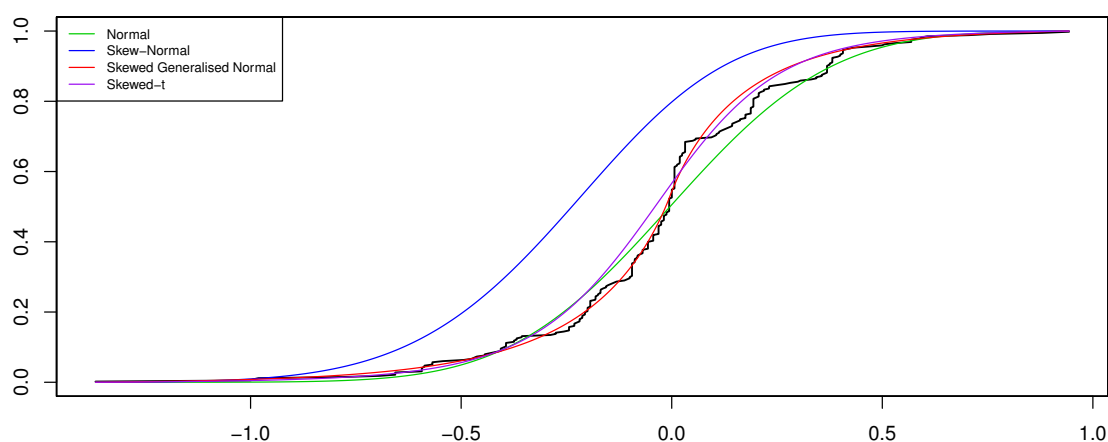


Figure 11. Cumulative distribution function (CDF) for the estimated AR(1) models under various distributions assumed for the innovation process a_t , with the empirical CDF for the residuals obtained from the estimated ARSGN(1) model for the viscosity time series [17].

4.2.2. Estimated Resident Population for Australia

In order to illustrate the relevance of the proposed model for higher orders of p , consider the quarterly estimated Australian resident population data (in thousands), which consists of $n = 89$ observations from June 1971 to June 1993. Figure 12 shows the time plot and ACF for the original time series from which it is clear that nonstationarity is present since the process mean and autocorrelations depend on time. This is also confirmed by the augmented Dickey–Fuller (ADF) test, which yields a p -value = 0.3493, suggesting that the time series exhibits a unit root [19].

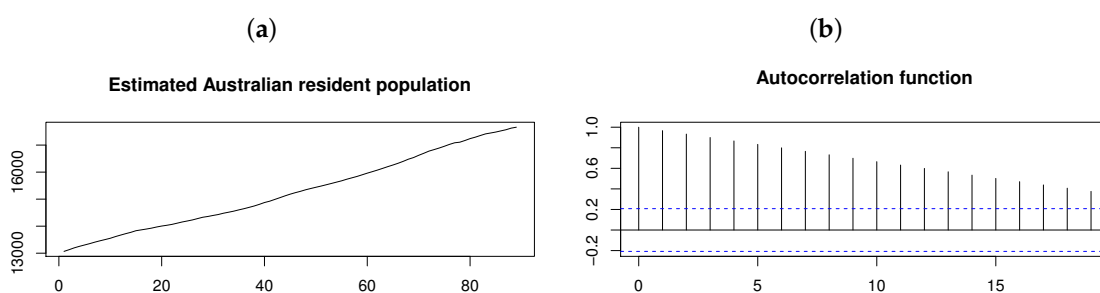


Figure 12. Australian resident population on a quarterly basis from June 1971 to June 1993 (estimated in thousands). From left to right: (a) Time plot. (b) ACF.

Transforming the original time series by differencing the time series twice yields a stationary time series with a p -value < 0.01 for the ADF test (suggesting no unit roots). This stationary time series and its distribution are illustrated in Figure 13, with the ACF and PACF suggesting that an AR(3) model is the appropriate choice for fitting a time series model. Previously, Brockwell and Davis [20] fitted an AR(3) model to the differenced (i.e., stationary) time series, assuming that the innovations are normally distributed. However, both the histogram and Shapiro–Wilk test (with p -value < 0.001) applied to this stationary time series suggest that the innovation process a_t is not normally distributed. Instead, SGN is considered as a distribution for the innovation process—that is, $a_t \sim SGN(0, \alpha, \beta, \lambda)$.

Table 6 summarizes the estimation results for AR models under the SGN distribution in comparison to the normal, SN , and ST distributions. From the maximized log-likelihood, AIC values, and KS test statistics calculated for the four models, it can be concluded that the ARSGN(3) model fits the best, with an estimated process mean $\hat{\mu}^* = 0.339$ (also evident from Figure 13). Furthermore, evaluating the standard errors of the parameters obtained for the ARSGN(3) model, it is evident that all parameters differ significantly from zero at a 95% confidence level, except for the skewing parameter

λ , suggesting that the innovation process does not exhibit significant skewness. Figure 14 illustrates these estimated models, confirming that the proposed ARSGN(3) model adapts well to various levels of asymmetry.

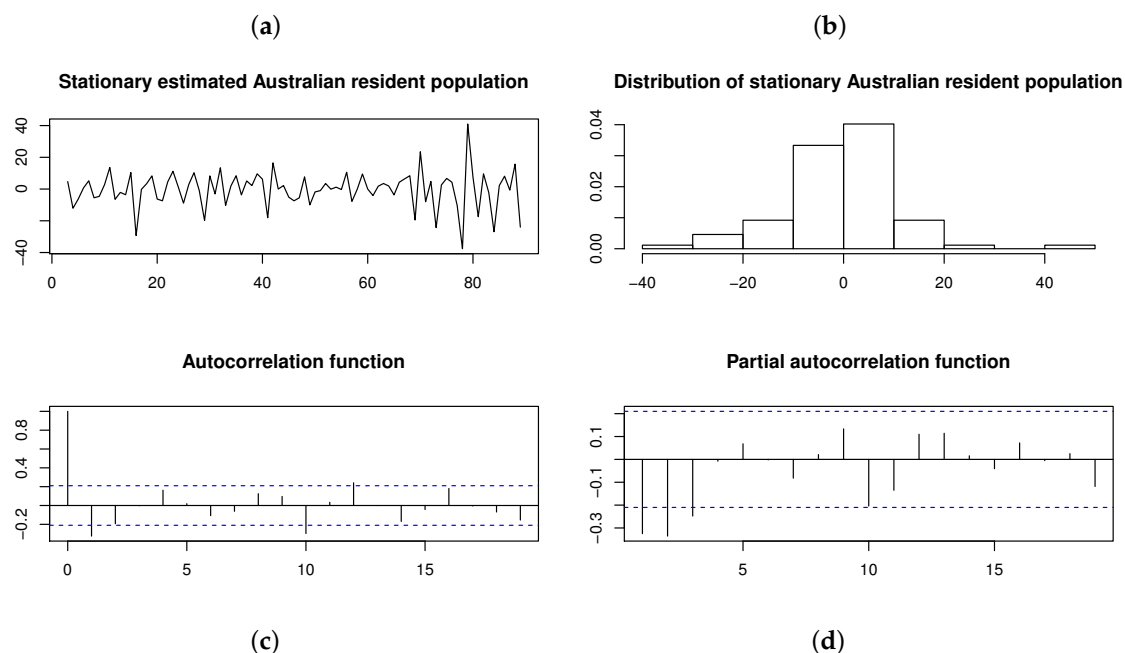


Figure 13. Differenced (i.e., stationary) Australian resident population on a quarterly basis from June 1971 to June 1993 (estimated in thousands). From left to right, top to bottom: (a) Time plot. (b) Histogram. (c) ACF. (d) PACF.

Table 6. ML parameter estimates with standard errors (in parentheses), log-likelihood $\ell_n(\hat{\Theta})$, AIC, KS test statistics, and run times for the differenced time series of estimated Australian resident population on a quarterly basis from June 1971 to June 1993.

	Model			
	AR(3)	ARSN(3)	ARSGN(3)	ARST(3)
α	9.648 (0.731)	9.821 (0.796)	5.150 (2.436)	8.835 (3.804)
β	–	–	0.834 (0.205)	–
λ	–	0.027 (1.451)	–0.040 (0.055)	–0.899 (1.154)
ν	–	–	–	5.158 (4.154)
Parameter				
φ_0	–0.362 (1.054)	–0.573 (11.414)	0.497 (0.029)	5.347 (5.410)
φ_1	–0.544 (0.105)	–0.544 (0.107)	–0.545 (0.011)	–0.545 (0.153)
φ_2	–0.458 (0.108)	–0.458 (0.110)	–0.378 (0.010)	–0.420 (0.111)
φ_3	–0.262 (0.105)	–0.262 (0.107)	–0.327 (0.008)	–0.239 (0.101)
$\ell_n(\hat{\Theta})$	–320.7	–311.1	–304.0	–305.8
AIC	651.3	634.2	622.0	625.7
KS test statistic	0.286	0.274	0.226	0.262
Run time (s)	0.88435	0.74169	2.40199	2.48347

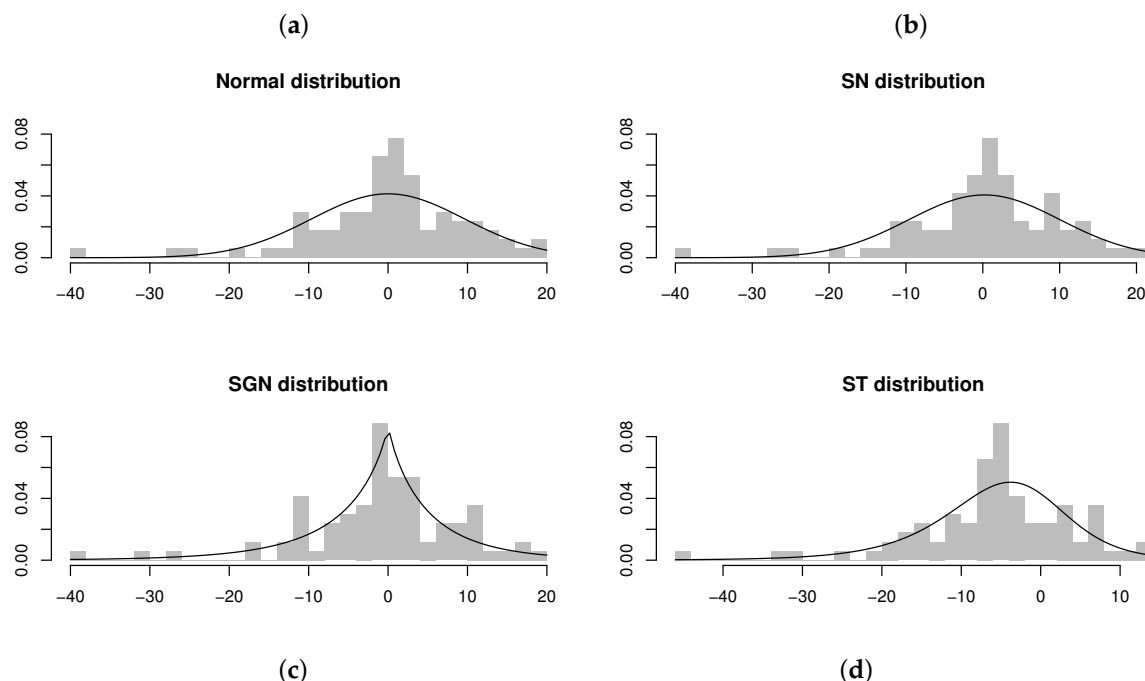


Figure 14. Residuals and estimated models obtained from AR(3) models fitted to the differenced time series of the estimated Australian resident population on a quarterly basis (from June 1971 to June 1993). From left to right, top to bottom: (a) AR(3) model fitted assuming $a_t \sim \mathcal{N}(0, \alpha)$. (b) AR(3) model fitted assuming $a_t \sim \mathcal{SN}(0, \alpha, \lambda)$. (c) AR(3) model fitted assuming $a_t \sim \mathcal{SGN}(0, \alpha, \beta, \lambda)$. (d) AR(3) model fitted assuming $a_t \sim \mathcal{ST}(0, \alpha, \lambda, \nu)$.

4.2.3. Insolvencies in South Africa

As a final application, consider the monthly (seasonally adjusted) number of insolvencies in South Africa for January 2000 to November 2019 (retrieved from Stats SA [21]). The time plot and ACF for the time series in Figure 15 suggests that the time series is nonstationary (which is also confirmed by the ADF test with p -value = 0.5807). A stationary time series is obtained by differencing the original time series once. This stationary time series and its distribution is illustrated in Figure 16, with the ACF and PACF suggesting that an AR(2) model is the appropriate choice for fitting a time series model. Although the Shapiro–Wilk test yields a p -value < 0.001 (suggesting non-normality), the AR(2) model is fitted assuming each of the normal, \mathcal{SN} , \mathcal{SGN} , and \mathcal{ST} distributions for the innovation process a_t , respectively; Table 7 and Figure 17 show the results obtained.

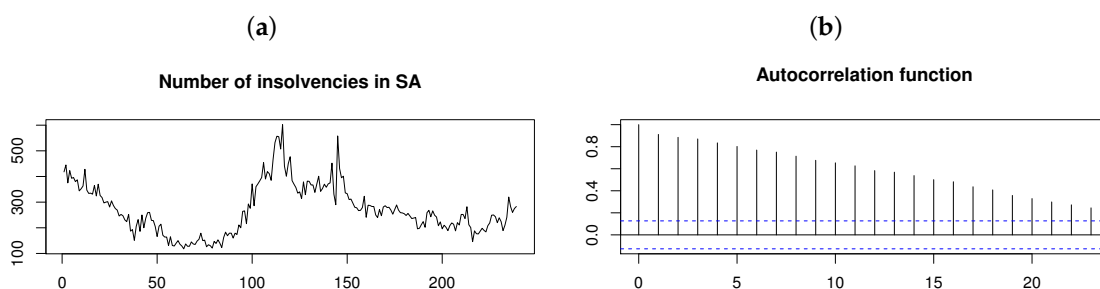


Figure 15. Number of insolvencies per month in South Africa from January 2000 to November 2019 [21]. From left to right: (a) Time plot. (b) ACF.

Referring to Table 7, it appears as if the ARST(2) model fits the best, with the proposed ARSGN(2) model as the runner-up (by comparing the maximized log-likelihood and AIC values). However, when comparing the standard errors for the parameter estimates, it is suggested that the estimates for the ARSGN(2) model generally exhibit less volatility compared to its competitors, with all parameters

being significant at a 95% confidence level. Figure 17 confirms that the SGN distribution adapts well to various levels of skewness and kurtosis.

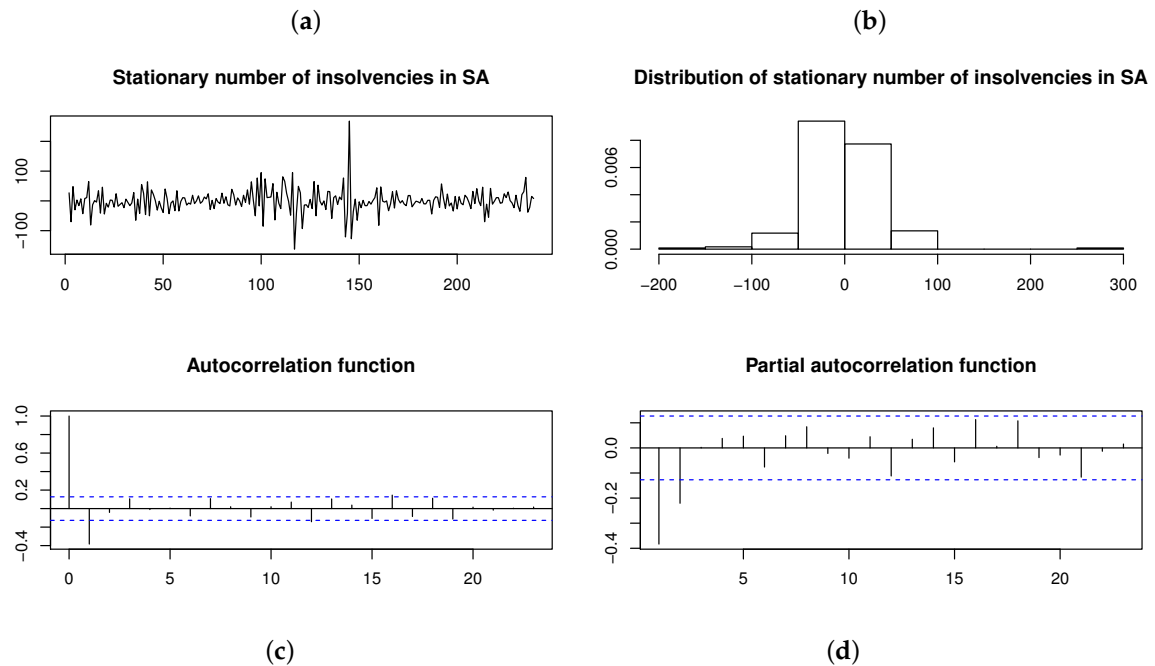


Figure 16. Differenced (i.e., stationary) number of insolvencies per month in South Africa from January 2000 to November 2019 [21]. From left to right, top to bottom: (a) Time plot. (b) Histogram. (c) ACF. (d) PACF.

Table 7. ML parameter estimates with standard errors (in parentheses), log-likelihood $\ell_n(\hat{\Theta})$, AIC, KS test statistics, and run times for the differenced time series of number of insolvencies per month in South Africa from January 2000 to November 2019 [21].

	Model			
	AR(2)	ARSN(2)	ARSGN(2)	ARST(2)
α	35.370 (1.621)	35.531 (1.661)	24.448 (4.386)	25.258 (3.067)
β	–	–	0.987 (0.113)	–
λ	–	–0.009 (1.395)	0.119 (0.048)	0.472 (0.360)
ν	–	–	–	3.590 (0.909)
Parameter	–0.870 (2.304)	–0.606 (39.502)	–7.399 (0.205)	–12.058 (6.827)
φ_0	–0.463 (0.063)	–0.463 (0.063)	–0.421 (0.031)	–0.418 (0.059)
φ_1	–0.219 (0.063)	–0.458 (0.063)	–0.210 (0.031)	–0.169 (0.061)
φ_2				
$\ell_n(\hat{\Theta})$	–1186.4	–1177.4	–1154.3	–1153.3
AIC	2380.8	2364.8	2320.6	2318.6
KS test statistic	0.356	0.356	0.305	0.331
Run time (s)	0.42257	0.29679	1.32080	0.87772

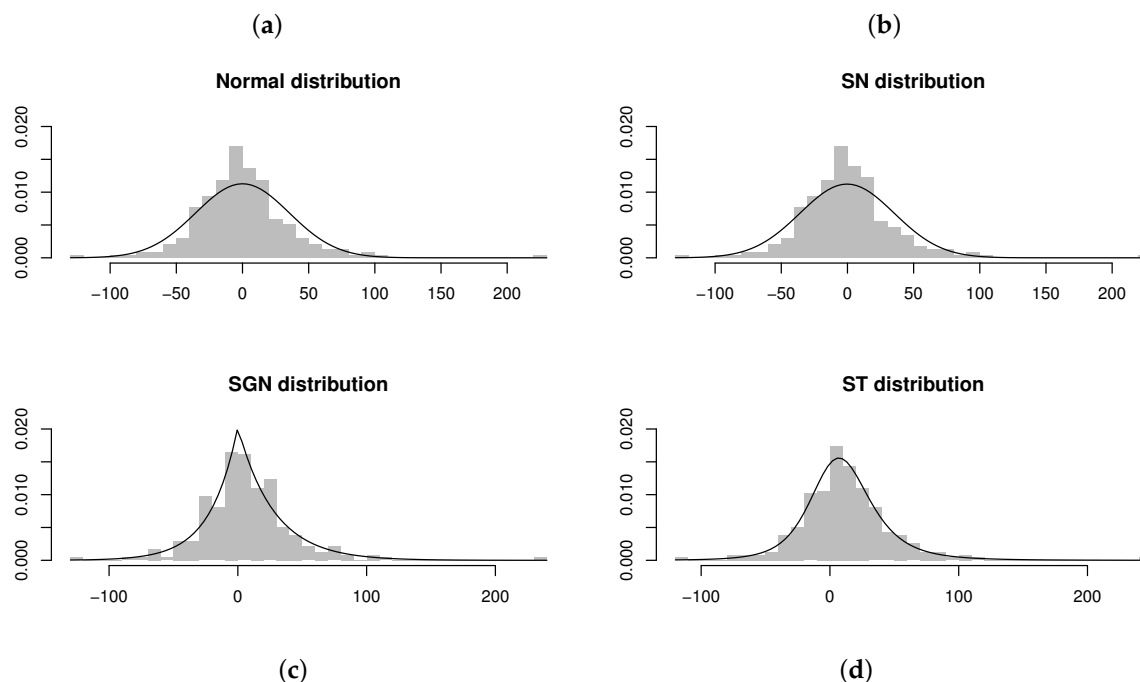


Figure 17. Residuals and estimated models obtained from AR(2) models fitted to the differenced time series of number of insolvencies per month in South Africa (from January 2000 to November 2019) [21]. From left to right, top to bottom: (a) AR(2) model fitted assuming $a_t \sim \mathcal{N}(0, \alpha)$. (b) AR(2) model fitted assuming $a_t \sim \mathcal{SN}(0, \alpha, \lambda)$. (c) AR(2) model fitted assuming $a_t \sim \mathcal{SGN}(0, \alpha, \beta, \lambda)$. (d) AR(2) model fitted assuming $a_t \sim \mathcal{ST}(0, \alpha, \lambda, \nu)$.

5. Conclusions

In this paper, the AR(p) time series model with skewed generalized normal (\mathcal{SGN}) innovations was proposed, i.e., ARSGN(p). The main advantage of the \mathcal{SGN} distribution is its flexibility by accommodating asymmetry, kurtosis, as well as heavier tails than the normal distribution. The conditional ML estimator for the parameters of the ARSGN(p) model was derived and the behavior was investigated through various simulation studies, in comparison to previously proposed models. Finally, real-time series datasets were fitted using the ARSGN(p) model, of which the usefulness was illustrated by comparing the estimation results for the proposed model to some of its competitors. In conclusion, the ARSGN(p) is a meaningful contender in the AR(p) environment compared to other often-considered models. A stepping-stone for future research includes the extension of the ARSGN(p) model for the multivariate case. Furthermore, an alternative method for defining the non-normal AR model by discarding the linearity assumption may be explored; see, for example, the work done on non-linear time series models in [22,23].

Author Contributions: Conceptualization, J.F. and M.N.; methodology, A.N., J.F., A.B., and M.N.; software, A.N.; formal analysis, A.N.; investigation, A.N.; writing, original draft preparation, A.N.; writing, review and editing, A.N., J.F., A.B., and M.N.; visualization, A.N.; supervision, J.F. and M.N.; funding acquisition, J.F. and A.B. All authors read and agreed to the published version of the manuscript.

Funding: This work was based on research supported by the National Research Foundation, South Africa (SRUG190308422768 nr. 120839), the South African NRF SARChI Research Chair in Computational and Methodological Statistics (UID: 71199), the South African DST-NRF-MRC SARChI Research Chair in Biostatistics (Grant No. 114613), and the Research Development Programme at UP 296/2019.

Acknowledgments: Hereby, the authors would like to acknowledge the support of the StatDisT group based at the University of Pretoria, Pretoria, South Africa. The authors also thank the anonymous reviewers for their comments, which improved the quality of the paper.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; nor in the decision to publish the results.

References

1. Pourahmadi, M. *Foundation of Time Series Analysis and Prediction Theory*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2001.
2. Pourahmadi, M. Stationarity of the solution of $(X_t = A_t X_{t-1} + \varepsilon_t)$ and analysis of non-Gaussian dependent random variables. *J. Time Ser. Anal.* **1988**, *9*, 225–239. [[CrossRef](#)]
3. Tarami, B.; Pourahmadi, M. Multi-variate t autoregressions: Innovations, prediction variances and exact likelihood equations. *J. Time Ser. Anal.* **2003**, *24*, 739–754. [[CrossRef](#)]
4. Grunwald, G.K.; Hyndman, R.J.; Tedesco, L.; Tweedie, R.L. Theory & methods: Non-Gaussian conditional linear AR(1) models. *Aust. N. Z. J. Stat.* **2000**, *42*, 479–495.
5. Bondon, P. Estimation of autoregressive models with epsilon-skew-normal innovations. *J. Multivar. Anal.* **2009**, *100*, 1761–1776. [[CrossRef](#)]
6. Sharafi, M.; Nematollahi, A. AR(1) model with skew-normal innovations. *Metrika* **2016**, *79*, 1011–1029. [[CrossRef](#)]
7. Ghasami, S.; Khodadadi, Z.; Maleki, M. Autoregressive processes with generalized hyperbolic innovations. *Commun. Stat. Simul. Comput.* **2019**, 1–13. [[CrossRef](#)]
8. Tuuç, Y.; Güney, Y.; Arslan, O. Parameter estimation of regression model with AR(p) error terms based on skew distributions with EM algorithm. *Soft Comput.* **2020**, *24*, 3309–3330. [[CrossRef](#)]
9. Bekker, A.; Ferreira, J.; Arashi, M.; Rowland, B. Computational methods applied to a skewed generalized normal family. *Commun. Stat. Simul. Comput.* **2018**, 1–14. [[CrossRef](#)]
10. Azzalini, A. A class of distributions which includes the normal ones. *Scand. J. Stat.* **1985**, *12*, 171–178.
11. Rowland, B.W. Skew-Normal Distributions: Advances in Theory with Applications. Master's Thesis, University of Pretoria, Pretoria, South Africa, 2017.
12. Jones, M.; Faddy, M. A skew extension of the t-distribution, with applications. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2003**, *65*, 159–174. [[CrossRef](#)]
13. Subbotin, M.T. On the law of frequency of error. *Math. Collect.* **1923**, *31*, 296–301.
14. Hamilton, J.D. *Time Series Analysis*, 2nd ed.; Princeton University Press: Princeton, NJ, USA, 1994.
15. Elal-Olivero, D.; Gómez, H.W.; Quintana, F.A. Bayesian modeling using a class of bimodal skew-elliptical distributions. *J. Stat. Plan. Inference* **2009**, *139*, 1484–1492. [[CrossRef](#)]
16. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974**, *19*, 716–723. [[CrossRef](#)]
17. Box, G.E.; Jenkins, G.M. *Time Series Analysis: Forecasting and Control*; Holden Day: San Francisco, CA, USA, 1976.
18. Massey, F.J., Jr. The Kolmogorov-Smirnov test for goodness of fit. *J. Am. Stat. Assoc.* **1951**, *46*, 68–78. [[CrossRef](#)]
19. Cheung, Y.W.; Lai, K.S. Lag order and critical values of the Augmented Dickey–Fuller test. *J. Bus. Econ. Stat.* **1995**, *13*, 277–280.
20. Brockwell, P.J.; Davis, R.A. *Introduction to Time Series and Forecasting*; New York Springer: New York, NY, USA, 1996.
21. Statistics South Africa. Available online: http://www.statssa.gov.za/?page_id=1854&PPN=P0043&SCH=72647 (accessed on 11 December 2019).
22. Tong, H. *Nonlinear Time Series*, 6th ed.; Oxford Statistical Science Series; The Clarendon Press Oxford University Press: New York, NY, USA, 1990.
23. Fan, J.; Yao, Q. *Nonlinear Time Series*; Springer Series in Statistics; Springer: New York, NY, USA, 2003.

