


Article

# The Asymmetric Alpha-Power Skew- $t$ Distribution

Roger Tovar-Falón <sup>1,\*</sup> , Heleno Bolfarine <sup>2</sup> and Guillermo Martínez-Flórez <sup>1</sup>

<sup>1</sup> Departamento de Matemáticas y Estadística, Facultad de Ciencias Básicas, Universidad de Córdoba, Montería 230027, Colombia; guillermomartinez@correo.unicordoba.edu.co

<sup>2</sup> Departamento de Estatística, IME, Universidade de São Paulo, São Paulo 1010, Brazil; hbolfar@ime.usp.br

\* Correspondence: rjtoivar@correo.unicordoba.edu.co

Received: 13 November 2019; Accepted: 25 December 2019; Published: 2 January 2020



**Abstract:** In this paper, we propose a new asymmetric and heavy-tail model that generalizes both the skew- $t$  and power- $t$  models. Properties of the model are studied in detail. The score functions and the elements of the observed information matrix are given. The process to estimate the parameters in model is discussed by using the maximum likelihood approach. Also, the observed information matrix is shown to be non-singular at the whole parametric space. Two applications to real data sets are reported to demonstrate the usefulness of this new model.

**Keywords:** alpha-power skew- $t$  distribution; skew- $t$  distribution; power- $t$  distribution; asymmetry; Fisher information matrix; maximum likelihood estimation

## 1. Introduction

In recent years, there has been considerable interest in the statistical literature related to flexible families of distributions able of modeling data that present high degree of asymmetry, with kurtosis index greater or smaller than the captured by normal model. In this context, two proposals that have shown a promising behavior in this type of situations are the skew-normal (SN) distribution of Azzalini [1] and the power-normal (PN) distribution of Durrans [2]. The SN distribution has been widely studied by many authors, and its main drawback is that it presents singular Fisher information matrix, implying the inference is useless from the theory of large samples using the maximum likelihood (ML) approach. Although the PN model has a shorter asymmetry range than SN distribution, it presents non-singular information matrix and can easily be extended to censored scenarios, as it has a simple distribution function, see, for example, in Martínez-Flórez et al. [3].

The PN model is part of a wide family of distributions known as alpha-power, which has been widely studied by many authors. In addition to the normal distribution, the Birnbaum–Saunders (BS) distribution [4] has also been considered, see, for example, in Martínez-Flórez et al. [5], who propose an extension of the BS distribution based on the asymmetric alpha-power family of distributions to illustrate the applicability of the new proposal with a data set is related to the lifetimes in cycles  $\times 10^{-3}$   $n = 101$  aluminum 6061 – T6 pieces cut in parallel angle to the rotation direction of rolling at the rate of 18 cycles per second and maximum stress of 21.000 psi. More details of the PN distribution can be found in Gupta and Gupta [6] and Pewsey et al. [7].

An alternative propose for modeling asymmetric data that unifies the two previous approaches was introduced by Martínez-Flórez et al. [8]. The proposed model, which is called alpha-power skew-normal (APSN), has non-singular Fisher information matrix, and it can fit data with much more asymmetry than PN models it can handle. In addition, symmetry can be tested by using the likelihood ratio statistic, as the properties of large samples are satisfied for the ML estimator.

Another set of distributions with non-singular information matrices, useful for modeling asymmetric and heavy-tailed data, are based on generalizations of the Student- $t$  distribution, see,

for example, in [9–13]. Azzalini and Capitanio [9] for example, introduced a skew- $t$  (ST) distribution as an extension of the SN model for modeling asymmetric and heavy-tailed data as follows; The random variable  $X$  is said to have the ST distribution with parameter  $\lambda$  and degrees of freedom  $\nu$ , if  $X$  has the probability density function (PDF) given by

$$f_{ST}(x; \lambda, \nu) = 2f_T(x; \nu) \mathcal{F}_T\left(\lambda \sqrt{\frac{\nu+1}{x^2+\nu}} x; \nu+1\right), \quad x \in \mathbb{R} \quad (1)$$

where  $\lambda \in \mathbb{R}$  is a parameter that controls the skewness of the distribution, and  $f_T(\cdot; \nu)$  and  $\mathcal{F}_T(\cdot; \nu)$  denote the PDF and the cumulative distribution function (CDF) of a standard Student- $t$  distribution with  $\nu$  degree of freedom, respectively. The ST distribution, like an extension of the SN model, inherits the problem of the singularity of the information matrix and before this inconvenience Zhao and Kim [14] proposed the power Student- $t$  (PT) distribution, whose information matrix is non-singular and for a given degree of freedom, the kurtosis range surpasses the kurtosis range of the skew- $t$  model at all times. The PT distribution is defined as follows. The random variable  $X$  is said to have the PT distribution with parameter  $\alpha$ , and degrees of freedom  $\nu$ , if  $X$  has PDF given by

$$f_{PT}(x; \alpha, \nu) = \alpha f_T(x; \nu) [\mathcal{F}_T(x; \nu)]^{\alpha-1}, \quad x \in \mathbb{R} \quad (2)$$

where  $\alpha > 0$  is a parameter that controls the form of the distribution, and, again,  $f_T(\cdot; \nu)$  and  $\mathcal{F}_T(\cdot; \nu)$  denote the PDF and the CDF of a standard Student- $t$  distribution, respectively.

Based on the properties of the ST model, to fit data with high degree of asymmetry and the characteristic of the PN model to capture kurtosis larger than the normal model, in this paper, we introduce a new distribution for modeling asymmetric and heavy-tailed data. The proposed model possess non-singular information matrix, and it is able to fit data with far more asymmetry than ST and PT models can handle and with large sample properties satisfied for the ML estimator. The model introduced in this paper is named as alpha-power skew- $t$  (APST) model and it extends both, ST and PT models. The APSN model by Martínez-Flórez et al. [8] is also a particular case when  $\nu$  tends to infinite. Note that symmetry can be tested using the likelihood ratio statistics with its large sample chi-square distribution.

The rest of this paper is organized as follows. Section 2 introduces the APST model and some of its properties like moments are studied. In particular, skewness and kurtosis indices are computed showing that their ranges surpass those of the ST and PT models. Section 3 deals with the ML estimation for the location-scale situation and its observed information matrix is derived. The extension to censored data is also presented. Finally, two applications are shown in Section 4, revealing that the model proposed can present much improvement over competitors.

## 2. The Alpha-Power Skew- $t$ Distribution

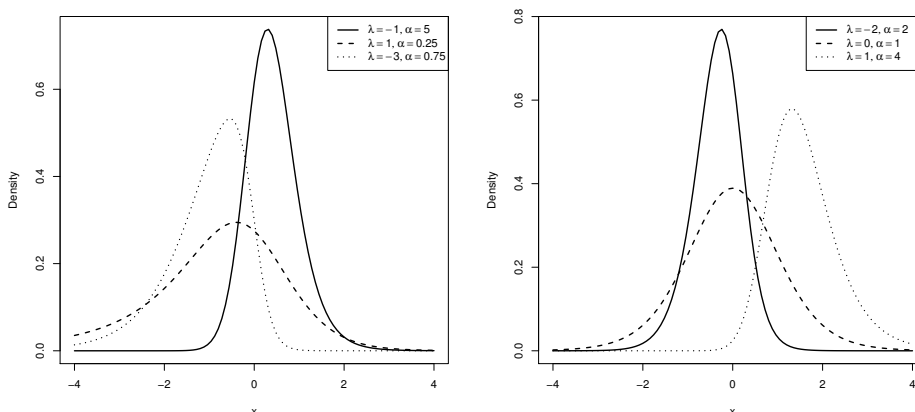
**Definition 1.** The random variable  $X$  is said to have an alpha-power skew- $t$  (APST) distribution, if  $X$  has PDF given by

$$f_{APST}(x; \lambda, \alpha, \nu) = \alpha f_{ST}(x; \lambda, \nu) [\mathcal{F}_{ST}(x; \lambda, \nu)]^{\alpha-1}, \quad (3)$$

for  $x \in \mathbb{R}$ ,  $\lambda \in \mathbb{R}$ , and  $\alpha, \nu \in \mathbb{R}^+$ . Functions  $f_{ST}(\cdot)$  and  $\mathcal{F}_{ST}(\cdot)$  denote the PDF and the CDF of the standard ST distribution. A random variable having  $f_{APST}(x; \lambda, \alpha, \nu)$  distribution is denoted shortly by  $X \sim \text{APST}(\lambda, \alpha, \nu)$ .

Figure 1 displays the form of the APST distribution for some selected values of the parameters  $\lambda$  and  $\alpha$  for  $\nu = 6$ . Note from the figure that the asymmetry and kurtosis of the APST distribution are affected by the parameters  $\alpha$  and  $\lambda$ ; therefore, the APST model is more flexible to model data that can be highly skewed, as well as heavier tails than ST and PT models.

The following result provides some special cases of the model (3), which occur for different values of  $\lambda$ ,  $\alpha$ , and  $\nu$ .



**Figure 1.** Probability density function of APST( $\lambda, \alpha, 10$ ) for some values of  $\lambda$  and  $\alpha$ .

**Proposition 1.** Let  $X \sim \text{APST}(\lambda, \alpha, \nu)$ ,

- (i) if  $\lambda = 0$ , then  $X \sim \text{PT}(\alpha, \nu)$ ,
- (ii) if  $\alpha = 1$ , then  $X \sim \text{ST}(\lambda, \nu)$ ,
- (iii) if  $\lambda = 0$  and  $\alpha = 1$ , then  $X \sim \text{T}(\nu)$ , where  $\text{T}(\nu)$  denotes the Student-t distribution with  $\nu$  degree of freedom.
- (iv) if  $\nu \rightarrow +\infty$ , then  $X \sim \text{APSN}(\lambda, \alpha)$ ,
- (v) if  $\lambda = 0$  and  $\nu \rightarrow +\infty$ , then  $X \sim \text{PN}(\alpha)$ ,
- (vi) if  $\alpha = 1$  and  $\nu \rightarrow +\infty$ , then  $X \sim \text{SN}(\lambda)$ ,
- (vii) if  $\lambda = 0, \alpha = 1$  and  $\nu \rightarrow +\infty$ , then  $X \sim \text{N}(0, 1)$ ,

**Proof.** The proof of (i)–(vii) is immediate from the definition of APST distribution.  $\square$

### 2.1. Moments

The following proposition presents an expression to compute the  $k$ -th moment of a random variable APST( $\lambda, \alpha, \nu$ ).

**Proposition 2.** Let  $X \sim \text{APST}(\lambda, \alpha, \nu)$ , then

$$\mathbb{E}[X^k] = \mathbb{E}\left[\left(\mathcal{F}_{ST}^{-1}(Y; \lambda, \nu)\right)^k\right] \tag{4}$$

where  $Y$  follows a Beta( $\alpha, 1$ ) distribution and  $\mathcal{F}_{ST}^{-1}(\cdot; \lambda, \nu)$  is the inverse of the function  $\mathcal{F}_{ST}(\cdot; \lambda, \nu)$ .

**Proof.** We have by definition that

$$\mathbb{E}[X^k] = \int_{\mathbb{R}} x^k \alpha f_{ST}(x) (\mathcal{F}_{ST}(x; \lambda, \nu))^{\alpha-1} dx$$

thus, letting  $y = \mathcal{F}_{ST}(x; \lambda, \nu)$ , then  $x = \mathcal{F}_{ST}^{-1}(y; \lambda, \nu)$ , it follows that

$$\mathbb{E}[X^k] = \int_0^1 \alpha \left(\mathcal{F}_{ST}^{-1}(y; \lambda, \nu)\right)^k y^{\alpha-1} dy$$

which is the expected value of the function  $\left(\mathcal{F}_{ST}^{-1}(Y; \lambda, \nu)\right)^k$ , where  $Y$  follows a beta distribution with parameters  $\alpha$  and 1.  $\square$

The indices of skewness ( $\sqrt{\beta_1}$ ) and kurtosis ( $\beta_2$ ) of APST distribution can be calculated by using the moments (4) as follows,

$$\sqrt{\beta_1} = \frac{\mu_3 - 3\mu_1\mu_2 + 2\mu_1^3}{(\mu_2 - \mu_1^2)^{3/2}} \quad \text{and} \quad \beta_2 = \frac{\mu_4 - 4\mu_1\mu_3 + 6\mu_2\mu_1^2 - 3\mu_1^4}{(\mu_2 - \mu_1^2)^2}$$

where  $\mu_k = \mathbb{E}[X^k]$  for  $k = 1, \dots, 4$ . Table 1 presents the ranges of possible values for the indices of asymmetry and kurtosis for  $ST(\lambda, \nu)$ ,  $PT(\alpha, \nu)$ , and  $APST(\lambda, \alpha, \nu)$  distributions, for values of  $\lambda$  between  $-40$  and  $40$ , values of  $\alpha$  between  $0.5$  and  $50$ , and for values of  $\nu = 2, 3, 4, 5, 6, 7$ . It can be seen from Table 1 that the length of the admissible intervals for the skewness and the kurtosis parameters of the APST distribution are larger than the corresponding intervals of the ST and PT distributions. This is an indicator that the APST model is more flexible in terms of asymmetry and kurtosis than the ST and PT models.

**Table 1.** Skewness and kurtosis for the models  $ST(\lambda, \nu)$ ,  $PT(\alpha, \nu)$ , and  $APST(\lambda, \alpha, \nu)$ , for  $\lambda \in (-40, 40)$ ,  $\alpha \in (0.5, 50)$  and  $\nu = 2, \dots, 7$ .

$\nu$	Skew-t		Power-t		Alpha-Power Skew-t	
	Skewness	Kurtosis	Skewness	Kurtosis	Skewness	Kurtosis
2	(-0.963, 0.963)	(3.170, 3.489)	(-0.119, 3.040)	(1.552, 10.436)	(-2.452, 14.314)	(1.395, 864.385)
3	(-0.950, 0.950)	(3.146, 3.357)	(-0.086, 1.362)	(1.325, 3.223)	(-2.130, 4.902)	(1.628, 114.098)
4	(-1.853, 1.853)	(5.099, 7.824)	(-0.530, 1.178)	(3.461, 5.299)	(-1.898, 3.215)	(3.153, 29.874)
5	(-0.947, 0.947)	(3.051, 3.327)	(-0.475, 0.271)	(1.176, 3.130)	(-1.968, 3.046)	(3.862, 19.925)
6	(-1.681, 1.681)	(4.554, 7.279)	(-0.533, 1.118)	(3.974, 5.173)	(-1.681, 2.145)	(3.892, 11.893)
7	(-0.944, 0.944)	(3.007, 3.367)	(-0.710, 0.243)	(1.264, 3.082)	(-1.535, 2.536)	(3.136, 15.924)

### 2.2. Distribution Function

**Proposition 3.** Let  $X \sim APST(\lambda, \alpha, \nu)$ , then the CDF of  $X$ , namely,  $\mathcal{F}_{APST}(x; \lambda, \alpha, \nu)$  is

$$\mathcal{F}_{APST}(x; \lambda, \alpha, \nu) = [\mathcal{F}_{ST}(x; \lambda, \nu)]^\alpha, \quad x \in \mathbb{R}. \tag{5}$$

**Proof.** The proof is immediate and it follows from results of Durrans [2].  $\square$

The inversion method can be used to generate a random variable with APST distribution. Thus, taking  $\lambda \in \mathbb{R}$ ,  $\alpha, \nu \in \mathbb{R}^+$  and a random variable with uniform distribution, namely,  $U \sim U(0, 1)$ , random variable  $X$  with  $APST(\lambda, \alpha, \nu)$  distribution is generated by taking

$$X = \mathcal{F}_{ST}^{-1}(U^{1/\alpha}; \lambda, \nu).$$

**Remark 1.** We consider a truncated  $APST(\lambda, \alpha)$  distribution to obtain a new and useful lifetime distribution. A random variable  $T$  has a truncated alpha-power skew-t distribution (at zero), denoted by  $TAPST(\lambda, \alpha, \nu)$ , if its PDF is given by

$$f(t) = \frac{\alpha f_{ST}(t, \lambda, \nu) [\mathcal{F}_{ST}(t, \lambda, \nu)]^{\alpha-1}}{1 - [\mathcal{F}_{ST}(0, \lambda, \nu)]^\alpha}; \quad t > 0 \tag{6}$$

The survival and hazard rate functions of a random variable  $T$  following a  $TAPST(\lambda, \alpha, \nu)$  distribution are given by

$$S_T(t) = P(T > t) = \frac{1 - [\mathcal{F}_{ST}(0, \lambda, \nu)]^\alpha - [\mathcal{F}_{ST}(t, \lambda, \nu)]^\alpha}{1 - [\mathcal{F}_{ST}(0, \lambda, \nu)]^\alpha}; \quad t > 0 \tag{7}$$

and

$$h_T(t) = \frac{\alpha f_{ST}(t, \lambda, \nu) [f_{ST}(t, \lambda, \nu)]^{\alpha-1}}{1 - [\mathcal{F}_{ST}(0, \lambda, \nu)]^\alpha - [\mathcal{F}_{ST}(t, \lambda, \nu)]^\alpha}; \quad t > 0 \tag{8}$$

respectively.

### 2.3. Location and Scale Extension

We can also consider a generalization of a APST distribution by adding location and scale parameters. The following definition gives a generalization of the APST model.

**Definition 2.** Let  $X \sim \text{APST}(\lambda, \alpha, \nu)$ . The APST density of location and scale is defined as the distribution of  $Y = \mu + \sigma X$ , for  $\mu \in \mathbb{R}$  and  $\sigma > 0$ . The corresponding PDF is given by

$$f_{\text{APST}}(y; \mu, \sigma, \lambda, \alpha, \nu) = \frac{\alpha}{\sigma} f_{\text{ST}}\left(\frac{y - \mu}{\sigma}; \lambda, \nu\right) \left[ \mathcal{F}_{\text{ST}}\left(\frac{y - \mu}{\sigma}; \lambda, \nu\right) \right]^{\alpha - 1}, \quad x \in \mathbb{R}, \quad (9)$$

for  $\lambda \in \mathbb{R}$  and  $\alpha, \nu \in \mathbb{R}^+$ . A random variable following a APST distribution of location and scale is denoted by  $Y \sim \text{APST}(\mu, \sigma, \lambda, \alpha, \nu)$ .

The  $k$ -th moment of a random variable  $Y \sim \text{APST}(\mu, \sigma, \lambda, \alpha, \nu)$  can be obtained from the formula

$$\mathbb{E}[Y^k] = \sum_{i=0}^k \binom{k}{i} \mu^i \sigma^{k-i} \mathbb{E}[X^{k-i}],$$

where  $X \sim \text{APST}(\lambda, \alpha, \nu)$ .

### 3. Statistical Inference for APST Distribution

This section concerns likelihood inference about the parameter vector  $\boldsymbol{\theta} = (\mu, \sigma, \lambda, \alpha, \nu)^\top$  of the location-scale family defined in Equation (9). Let  $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$  be a random sample of the distribution  $\text{APST}(\mu, \sigma, \lambda, \alpha, \nu)$ . The log-likelihood function for  $\boldsymbol{\theta} = (\mu, \sigma, \lambda, \alpha, \nu)^\top$  can be written as follows,

$$\begin{aligned} \ell(\boldsymbol{\theta}; \mathbf{Y}) &\propto n \log \alpha - n \log \sigma - \frac{n}{2} \log \nu \\ &+ n \log \Gamma\left(\frac{\nu + 1}{2}\right) - n \log \Gamma\left(\frac{\nu}{2}\right) - \frac{\nu + 1}{2} \sum_{i=1}^n \log\left(1 + \frac{z_i^2}{\nu}\right) \\ &+ \sum_{i=1}^n \log \mathcal{F}_T\left(\lambda z_i \sqrt{\frac{\nu + 1}{z_i^2 + \nu}}; \nu + 1\right) + (\alpha - 1) \sum_{i=1}^n \log \mathcal{F}_{\text{ST}}(z_i; \lambda, \nu) \end{aligned} \quad (10)$$

where  $z_i = (y_i - \mu)/\sigma$ . Thus, by differentiating the log-likelihood function, we obtain the following score equations,

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\theta}; \mathbf{Y})}{\partial \mu} &= \frac{\nu + 1}{\sigma \nu} \sum_{i=1}^n z_i \left(1 + \frac{z_i^2}{\nu}\right)^{-1} \\ &- \frac{\lambda}{\sigma} \sum_{i=1}^n w_i \left(1 + \frac{z_i^2}{\nu}\right)^{-1} \frac{f_T(\lambda z_i w_i; \nu + 1)}{\mathcal{F}_T(\lambda z_i w_i; \nu + 1)} - \frac{\alpha - 1}{\sigma} \sum_{i=1}^n \frac{f_{\text{ST}}(z_i; \lambda, \nu)}{\mathcal{F}_{\text{ST}}(z_i; \lambda, \nu)} = 0 \end{aligned} \quad (11)$$

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\theta}; \mathbf{Y})}{\partial \sigma} &= -\frac{n}{\sigma} + \frac{\nu + 1}{\sigma \nu} \sum_{i=1}^n z_i^2 \left(1 + \frac{z_i^2}{\nu}\right)^{-1} - \frac{\lambda}{\sigma} \sum_{i=1}^n z_i w_i \left(1 + \frac{z_i^2}{\nu}\right)^{-1} \frac{f_T(\lambda z_i w_i; \nu + 1)}{\mathcal{F}_T(\lambda z_i w_i; \nu + 1)} \\ &- \frac{\alpha - 1}{\sigma} \sum_{i=1}^n z_i \frac{f_{\text{ST}}(z_i; \lambda, \nu)}{\mathcal{F}_{\text{ST}}(z_i; \lambda, \nu)} = 0 \end{aligned} \quad (12)$$

$$\frac{\partial \ell(\boldsymbol{\theta}; \mathbf{Y})}{\partial \lambda} = \sum_{i=1}^n z_i w_i \frac{f_T(\lambda z_i w_i; \nu + 1)}{\mathcal{F}_T(\lambda z_i w_i; \nu + 1)} - \frac{\alpha - 1}{\pi(1 + \lambda^2)} \sum_{i=1}^n \frac{(1 + (1 + \lambda^2)z_i^2/\nu)^{-\frac{\nu}{2}}}{\mathcal{F}_{ST}(z_i; \lambda, \nu)} = 0, \tag{13}$$

$$\frac{\partial \ell(\boldsymbol{\theta}; \mathbf{Y})}{\partial \alpha} = \frac{n}{\alpha} + \sum_{i=1}^n \log \mathcal{F}_{ST}(z_i; \lambda, \nu) = 0, \tag{14}$$

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\theta}; \mathbf{Y})}{\partial \nu} &= \frac{n\alpha}{2} \left( \psi\left(\frac{\nu + 1}{2}\right) - \psi\left(\frac{\nu}{2}\right) - \frac{1}{\nu} \right) - \frac{1}{2} \sum_{i=1}^n \log \left( 1 + \frac{z_i^2}{\nu} \right) \\ &\quad + \frac{\nu + 1}{2\nu^2} \sum_{i=1}^n z_i^2 \left( 1 + \frac{z_i^2}{\nu} \right)^{-1} \\ &\quad + \frac{\lambda}{2\nu(\nu + 1)} \sum_{i=1}^n z_i^3 w_i \left( 1 + \frac{z_i^2}{\nu} \right)^{-1} \frac{f_T(\lambda z_i w_i; \nu + 1)}{\mathcal{F}_T(\lambda z_i w_i; \nu + 1)} \\ &\quad - \frac{\lambda}{2\nu(\nu + 1)} \sum_{i=1}^n z_i w_i \left( 1 + \frac{z_i^2}{\nu} \right)^{-1} \frac{f_T(\lambda z_i w_i; \nu + 1)}{\mathcal{F}_T(\lambda z_i w_i; \nu + 1)} \\ &\quad - \frac{(\alpha - 1)}{2\pi(\nu + 1)} \frac{\lambda}{(1 + \lambda^2)} \sum_{i=1}^n \frac{(1 + (1 + \lambda^2)z_i^2/\nu)^{-\frac{\nu}{2}}}{\mathcal{F}_{ST}(z_i; \lambda, \nu)} + \frac{\alpha - 1}{2} \sum_{i=1}^n \frac{g(z_i; \nu)}{\mathcal{F}_{ST}(z_i; \lambda, \nu)} = 0 \end{aligned} \tag{15}$$

where  $\psi(\cdot)$  is the digamma function,  $w_i = \sqrt{\frac{\nu+1}{x_i^2+\nu}}$  for  $i = 1, \dots, n$ , and  $g(x; \nu)$  is the function defined by

$$\begin{aligned} g(x; \nu) &= \int_{-\infty}^x \left\{ \frac{(\nu + 1)}{\nu^2} s^2 \left( 1 + \frac{s^2}{\nu} \right)^{-1} - \log \left( 1 + \frac{s^2}{\nu} \right) \right\} f_{ST}(s; \lambda, \nu) ds \\ &\quad - \frac{\lambda}{\pi\nu} \int_{-\infty}^x s \left( 1 + \frac{s^2}{\nu} \right)^{-1} \left\{ 1 + (1 + \lambda^2) \frac{s^2}{\nu} \right\}^{-\frac{\nu+2}{2}} ds \end{aligned} \tag{16}$$

Equations (11)–(15) include nonlinear functions; therefore, it is not possible to obtain explicit forms of the maximum likelihood estimators (MLEs), and they must be calculated by using numerical methods. In this work, we used the *maxLik* function of R Development Core Team [15] which uses the Newton–Raphson optimization method. The elements of the observed information matrix are easily obtained after calculating the second derivative of the log-likelihood function and multiplying by  $-1$ , that is,

$$j_{\theta_i, \theta_k} = -\frac{\partial^2 \ell(\boldsymbol{\theta}; \mathbf{Y})}{\partial \theta_i \partial \theta_k}, \quad i, k = 1, 2, \dots, 5 \tag{17}$$

where  $\boldsymbol{\theta} = (\mu, \sigma, \lambda, \alpha, \nu)^\top$ . This elements are given in the Appendix A. To find the standard errors (EE) of the MLEs and calculate confidence intervals, the information matrix  $\mathbf{I}$  (or Fisher information) must be calculated, which is defined as the expected value of the second derived from the log-likelihood function or less the expected value of the Hessian matrix; from this matrix, we calculate the EE as the diagonal elements of the inverse of this matrix. The elements of the  $\mathbf{I}$  matrix are obtained as

$$\mathbf{I}(i, k) = -E \left( \frac{\partial^2 \ell(\boldsymbol{\theta}; \mathbf{Y})}{\partial \theta_i \partial \theta_k} \right), \quad i, k = 1, 2, \dots, 5 \tag{18}$$

The role of the Fisher information in the asymptotic theory of maximum-likelihood estimation was emphasized by Ronald Fisher following some initial results by Francis Edgeworth, see Lehman and Casella [16] and Frieden [17] for more details. The Fisher-information matrix is used to calculate the covariance matrices associated with maximum-likelihood estimates, and it can also be used in the formulation of test statistics, such as the Wald test.

As the expected value under the APST distribution and the second-order derivatives are not direct, numerical methods must be used to obtain the explicit form of the information matrix  $\mathbf{I}$ .

Therefore, we use the observed information matrix to calculate the standard errors in the rest of the document.

When  $\nu$  tends to infinite the ST distribution converges to the SN distribution and we recall that the information matrix of a random variable  $X \sim \text{SN}(\mu, \sigma, \lambda)$  which is denoted by  $\mathbf{I}_\lambda(\boldsymbol{\varphi})$ , where  $\boldsymbol{\varphi} = (\mu, \sigma, \lambda)^\top$ , is singular for  $\lambda = 0$ . Therefore, it is convenient to use a centered parameterization of the ST distribution proposed by Arellano-Valle and Azzalini [18].

The centered parameterization of the SN distribution was proposed as an alternative to the problem of singularity of the information matrix of the SN when  $\lambda = 0$ . Arellano-Valle and Azzalini [19] proposed a second representation of the SN by defining a new random variable  $X$  as

$$X = \mu + \sigma \left( \frac{Z - \mathbb{E}[Z]}{\sqrt{\text{Var}[Z]}} \right),$$

where  $\mu \in \mathbb{R}$  and  $\sigma > 0$  are parameters of the random variable  $X$  and  $Z \sim \text{SN}(\lambda)$ . This representation is called centered parameterization, as  $\mathbb{E}[X] = \mu$  and  $\text{Var}[X] = \sigma^2$  and it is denoted by  $\text{CSN}(\mu, \sigma, \gamma_1)$ , where  $-0.9953 < \gamma_1 < 0.9953$ . Under the centered parameterization model,  $\mu$ ,  $\sigma$ , and  $\gamma_1 = \sqrt{\beta_1}$  represent the mean, the standard deviation and the skewness index of  $X$ , respectively. If  $Z \sim \text{SN}(\lambda)$  then  $\mathbb{E}[Z] = b\delta$  and  $\text{Var}[Z] = 1 - (b\delta)^2$ , where  $b = \sqrt{2/\pi}$  and  $\delta = \lambda/\sqrt{1 + \lambda^2}$ ; it has that the random variable  $X$  can be written as  $X = \mu + \sigma Z$  which has  $\text{SN}(\lambda_1, \lambda_2, \lambda)$  distribution, where

$$\lambda_1 = \mu - c\sigma\gamma_1^{1/3}, \quad \lambda_2 = \sigma\sqrt{1 + c^2\gamma_1^{2/3}}, \quad \lambda = \frac{c\gamma_1^{1/3}}{\sqrt{b^2 + c^2(b^2 - 1)\gamma_1^{2/3}}} \tag{19}$$

with  $c = \{2/(4 - \pi)\}^{1/3}$ . Under this denomination, the information matrix can be written as  $\mathbf{I}_{\gamma_1} = \mathbf{D}^\top \mathbf{I}_\lambda \mathbf{D}$ , where  $\mathbf{D}$  is a matrix that represents the derivative of the parameters of the standard representation ( $\lambda_1, \lambda_2$  and  $\lambda$ ) regarding to the new parameters ( $\mu, \sigma$  and  $\gamma_1$ ). It also follows that the information matrix converges to a diagonal matrix  $\boldsymbol{\Sigma}_c^{-1} = \text{diag}(\sigma^2, \sigma^2/2, 6)$  when  $\lambda \rightarrow 0$ . This guarantees the existence and uniqueness of the MLEs of  $\lambda_1$  and  $\lambda_2$  for each fixed value of  $\lambda$ .

Following this same line of thought, we suppose that  $Y$  follows the model (1) with location parameter  $\mu \in \mathbb{R}$  and scale parameter  $\sigma > 0$ , that is,

$$f_{ST}(y; \mu, \sigma, \lambda, \nu) = \frac{2}{\sigma} f_T \left( \frac{y - \mu}{\sigma}; \nu \right) \mathcal{F}_T \left( \lambda \sqrt{\frac{\nu + 1}{Q_y + \nu}} \left( \frac{y - \mu}{\sigma} \right); \nu + 1 \right), \quad y \in \mathbb{R} \tag{20}$$

where  $\lambda \in \mathbb{R}$  and  $Q_y = ((y - \mu)/\sigma)^2$ . This representation relates to the direct parameterization of the ST distribution with parameter vector  $\boldsymbol{\rho} = (\mu, \sigma, \lambda, \nu)^\top$ . It follows that  $Z_T = (Y - \mu)/\sigma \sim \text{ST}(\lambda, \nu)$ , and by the stochastic representation of the ST distribution is given by  $Z_T = Z/\sqrt{V}$ , where  $Z \sim \text{SN}(\lambda)$  and  $V \sim \chi_\nu^2/\nu$ . This entails to compute the first four cumulants of  $Z_T$  denoted by  $\mu_1(\delta, \nu)$ ,  $\mu_2(\delta, \nu)$ ,  $\mu_3(\delta, \nu)$  and  $\mu_4(\delta, \nu)$ , see [18]. The centered parameterization of the ST distribution of a random variable  $Y$  comes by defining

$$\begin{aligned} \mu_t &= \mathbb{E}[Y] = \mu + \sigma\mu_1(\delta, \nu) = \mu + \sigma b_\nu \delta \\ \sigma_t^2 &= \text{Var}[Y] = \sigma^2\mu_2(\delta, \nu) = \eta^2 \left\{ \frac{\nu}{\nu - 2} - b_\nu^2 \delta^2 \right\}, \end{aligned}$$

$$\gamma_{1t} = \frac{\mu_3(\delta, \nu)}{\mu_2(\delta, \nu)^{3/2}} = \frac{b_\nu \delta}{\mu_2(\delta, \nu)^{3/2}} \left\{ \frac{\nu(3 - \delta^2)}{\nu - 3} - \frac{3\nu}{\nu - 2} + 2b_\nu^2 \delta^2 \right\}$$

$$\gamma_{2t} = \frac{\mu_4(\delta, \nu)}{\mu_2(\delta, \nu)^2} = \frac{1}{\mu_2(\delta, \nu)^2} \left\{ \frac{3\nu^2}{(\nu - 2)(\nu - 4)} - \frac{4b_\nu^2 \delta^2 \nu(3 - \delta^2) + \frac{6b_\nu^2 \delta^2 \nu}{\nu - 2} - 4b_\nu^4 \delta^4}{\nu - 3} \right\} - 3.$$

The new representation is defined as the centered skew- $t$  distribution with parameter vector  $\tilde{\rho} = (\mu, \sigma^2, \gamma_1, \gamma_2)^\top$ . According to Arellano-Valle and Azzalini [18], the information matrix of this representation can be written as

$$\mathbf{I}(\tilde{\rho}) = \mathbf{B}^\top \mathbf{I}(\rho) \mathbf{B},$$

where  $\mathbf{B}$  is a matrix representing the derivative of the parameter vector  $\rho$  with respect to the new vector  $\tilde{\rho}$ . It can be shown that  $b_\nu \rightarrow b$  when  $\nu \rightarrow \infty$ , see [18]. Therefore, the parameters of the centered ST model converge to  $\mu_t \rightarrow \mu$ ,  $\sigma_t^2 \rightarrow \sigma^2$ , and  $\gamma_{1t} \rightarrow \gamma_1$  when  $\nu \rightarrow \infty$ , that is, the parameters of the CSN. As  $Z_T \rightarrow \text{SN}(\lambda)$  when  $\nu \rightarrow \infty$ , it follows that the random variable  $Y$  converges to a distribution with information matrix

$$\mathbf{I}(\mu, \sigma^2, \gamma_1, \alpha) = \begin{pmatrix} \mathbf{I}_{\theta_1, \theta_1} & \mathbf{I}_{\theta_1, \alpha} \\ \mathbf{I}_{\theta_1, \alpha}^\top & I_{\alpha, \alpha} \end{pmatrix}, \tag{21}$$

where the elements of the diagonal correspond to the information of the parameter vector  $\theta_1 = (\mu, \sigma^2, \gamma_1)$  and  $\alpha$ , and  $\mathbf{I}_{\theta_1, \alpha}$  is the joint information of  $\theta_1 = (\mu, \sigma^2, \gamma_1)^\top$  and  $\alpha$ . Now, when  $\lambda \rightarrow 0$  and  $\alpha = 1$ , it can be shown that  $\mathbf{I}_{\theta_1, \theta_1} \rightarrow \text{diag}(\sigma^2, \sigma^2/2, 6)$ , with determinant equal to  $0.3333/\sigma^4$ , and  $\mathbf{I}_{\theta_1, \alpha} = (0.9031/\sigma, -0.5956/\sigma, 0.7206)^\top$ ; therefore, the determinant  $|\mathbf{I}(\mu, \sigma^2, \gamma_1, \alpha)| \neq 0$ , and it concludes that the random variable  $Y$  converges to a distribution with information matrix non-singular when  $\nu$  tends to infinite.

### 3.1. Extension to Censored Data

Based on the goodness of the APST distribution to fit asymmetric and heavy-tailed data, in this section we introduce the censored APST model which we will be denote by CAPST.

**Definition 3.** Suppose that the random variable  $Y$  follows APST distribution, and consider a random sample  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)$  where only the  $Y_i$  values greater than a constant  $k$  are recorded. In addition, for values  $Y_i \leq k$  only the value of  $k$  is recorded. Therefore, for  $i = 1, 2, \dots, n$ , the observed values  $Y_i^o$  can be written as

$$Y_i^o = \begin{cases} k, & \text{if } Y_i \leq k, \\ Y_i, & \text{if } Y_i > k. \end{cases}$$

The resulting sample is said to be a censored APST, and we say that  $Y$  is a censored random variable APST. We will use the notation  $Y \sim \text{CAPST}(\theta)$ , where  $\theta = (\mu, \sigma, \lambda, \alpha, \nu)^\top$ .

From Definition 3 it follows that  $P(Y_i^o = k) = P(Y_i \leq k) = \{\mathcal{F}_{ST}((k - \mu)/\sigma)\}^\alpha$  and for the observations  $Y_i^o = Y_i$ , the distribution of  $Y_i^o$  is the same of  $Y_i$ , i.e.,  $Y_i^o \sim \text{APST}(\theta)$ . For convenience, we choose to work with the case of left-censored data; however, the followings results can be extended to other types of censorship.

### 3.2. Properties of the CAPST Model

Let  $Y \sim \text{CAPST}(\mu, \sigma, \lambda, \alpha, \nu)$ ,

1. If  $\alpha = 1$ , then  $Y \sim \text{CST}(\mu, \sigma, \lambda, \nu)$ , where CST indicates the censored skew- $t$  model.
2. If  $\lambda = 0$ , then  $Y \sim \text{CPT}(\mu, \sigma, \alpha, \nu)$ , where CPT indicates the censored power- $t$  model.
3. If  $\alpha = 1$  and  $\lambda = 0$ , then  $Y \sim \text{CT}(\mu, \sigma, \nu)$ , that is, the censored Student- $t$  model follows.
4. If  $\nu \rightarrow +\infty$ , then  $Y \sim \text{CAPSN}(\mu, \sigma, \lambda, \alpha)$ , where CAPSN indicates the censored alpha-power skew-normal model.



5. If  $\alpha = 1$  and  $\nu \rightarrow +\infty$ , then  $Y \sim \text{CSN}(\mu, \sigma, \lambda)$ , that is, the censored skew-normal model follows.
6. If  $\lambda = 0$  and  $\nu \rightarrow +\infty$ , then  $Y \sim \text{CPN}(\mu, \sigma, \alpha)$ , that is, the censored power-normal model follows.
7. If  $\alpha = 1, \lambda = 0$  and  $\nu \rightarrow +\infty$ , then  $Y \sim \text{CN}(\mu, \sigma^2)$ , that is, the censored normal model follows.

The estimates of the parameters of the model can be obtained via maximum likelihood method, where the log-likelihood function is given by

$$\begin{aligned} \ell(\theta; \mathbf{Y}) \propto & \alpha \sum_0 \log \mathcal{F}_{ST} \left( \frac{k - \mu}{\sigma}; \lambda, \nu \right) + n_1 \log \alpha - n_1 \log \sigma - \frac{n_1}{2} \log \nu \\ & + n_1 \log \Gamma \left( \frac{\nu + 1}{2} \right) - n_1 \log \Gamma \left( \frac{\nu}{2} \right) - \frac{\nu + 1}{2} \sum_1 \log \left( 1 + \frac{x_i^2}{\nu} \right) \\ & + \sum_1 \log \mathcal{F}_T \left( \lambda x_i \sqrt{\frac{\nu + 1}{x_i^2 + \nu}}; \nu + 1 \right) + (\alpha - 1) \sum_1 \log \mathcal{F}_{ST}(x_i; \lambda, \nu) \end{aligned}$$

where  $x_i = (y_i - \mu) / \sigma$ ;  $\sum_1$  and  $\sum_0$  are the sum over censored individuals and uncensored individuals, respectively; and  $n_1$  is the number of uncensored individuals.

#### 4. Real Data Applications

In this section, we illustrate the applicability of the proposed model in Section 2 by analyzing two data sets. We use the statistical software R [15], version 3.5.3 with the package *maxLike* for maximizing the corresponding likelihood functions. For comparing purposes of various models, the AIC Akaike [20], BIC Schwarz [21], and corrected AIC (CAIC) Bozdogan [22] information criteria were used.

##### 4.1. Application 1: Volcano Heights Data

Consider the data set related to heights of 1520 volcanoes in the world which is available in website [dx.doi.org/10.5479/si.GVP.VOTW4-2013](https://dx.doi.org/10.5479/si.GVP.VOTW4-2013) [23]. Table 2 presents the summary statistics for the data set. It can be noted that the asymmetry and kurtosis indices seem to indicate that the use of an asymmetric and heavy-tailed model is appropriate to analyze this data set. We analyzed these data by fitting the Student-*t*, ST, PT, and APST distributions.

**Table 2.** Volcano heights data: Statistical summary.

<i>n</i>	Mean	Variance	$\sqrt{b_1}$	<i>b</i> <sub>2</sub>
1520	16.7760	15.6682	0.6461	4.3809

Table 3 shows the parameter estimates, together with their corresponding standard errors (SE). Note that the values of the standard errors of the  $\mu$  and  $\sigma$  estimates for the APST model are smaller than the corresponding standard errors of the respective parameters for the Student-*t*, ST, and PT models. Table 3 also presents some model selection criteria, together with the values of the log-likelihood. The AIC, BIC, and CAIC criteria indicate that the APST model seems to provide better fit to the volcanoes heights data than the T, ST, and PT models, supporting the asymmetry assertion of the volcano’s heights variable. Figure 2 shows the graphs *QQplot* of the fitted models. It can be clearly seen from the figure that the APST model fits the data better than the Student-*t*, ST, and PT models. In addition, we can use the likelihood ratio (LR) test statistic to conform our claim. To do this, we consider the following hypotheses,

$$H_0 : (\lambda, \alpha) = (0, 1) \text{ (T}(\mu, \sigma, \nu)) \quad \text{v.s.} \quad H_1 : (\lambda, \alpha) \neq (0, 1) \text{ (APST}(\mu, \sigma, \lambda, \alpha, \nu)),$$

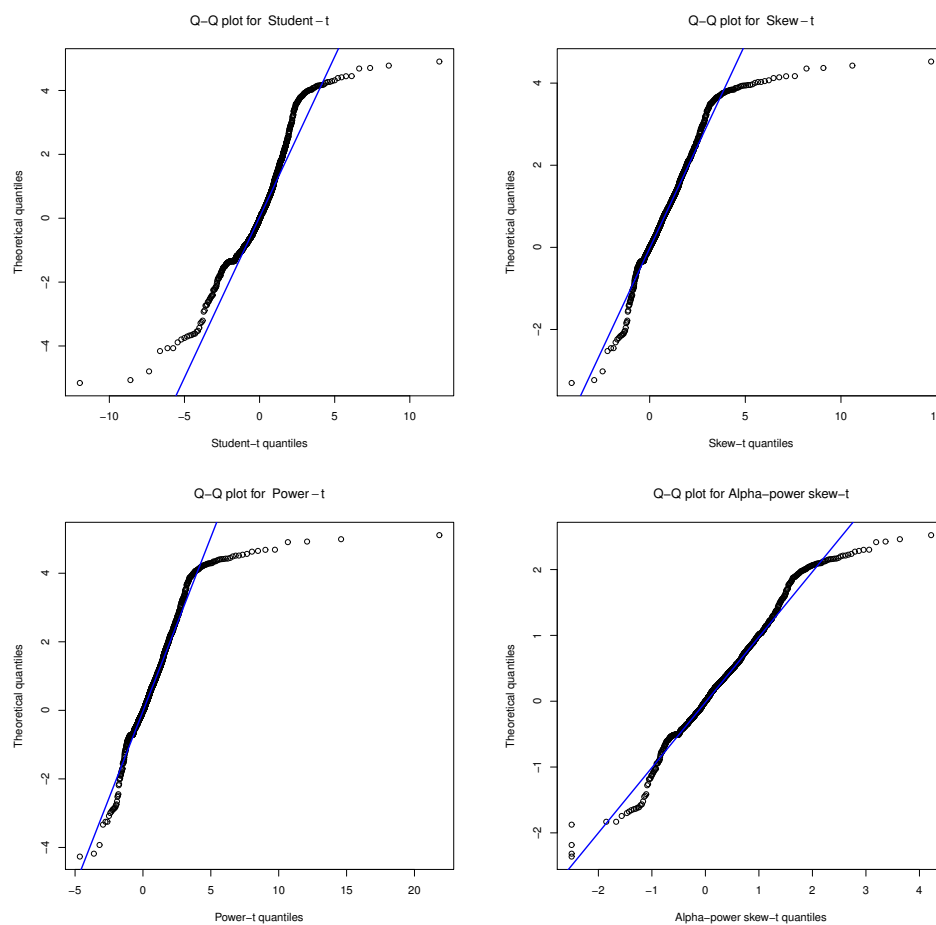
The value of the LR test statistic is  $-2 \log(\Lambda) = -2(\ell_T(\hat{\theta}) - \ell_{APST}(\hat{\theta})) = 134.823$  and comparing this quantity with  $\chi_2^2 = 5.9914$ , the null hypothesis is rejected. The APST model is also compared with the ST and PT models by considering the hypotheses

$$H_{01} : \alpha = 1 \text{ (ST}(\mu, \sigma, \lambda, \nu)) \quad \text{v.s.} \quad H_{11} : \alpha \neq 1 \text{ (APST}(\mu, \sigma, \lambda, \alpha, \nu)),$$

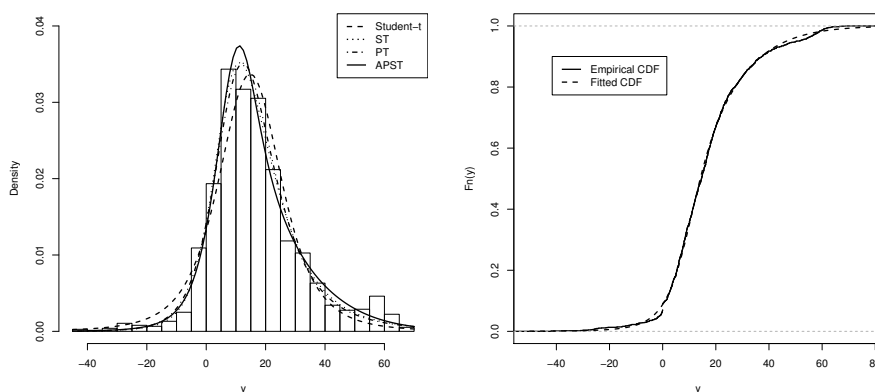
and

$$H_{02} : \lambda = 0 \text{ (PT}(\mu, \sigma, \alpha, \nu)) \quad \text{v.s.} \quad H_{12} : \lambda \neq 0 \text{ (APST}(\mu, \sigma, \lambda, \alpha, \nu)),$$

respectively. The respective values of the LR test statistic are given by  $-2 \log(\Lambda_1) = -2(\ell_{ST}(\hat{\theta}) - \ell_{APST}(\hat{\theta})) = 26.620$  and  $-2 \log(\Lambda_2) = -2(\ell_{PT}(\hat{\theta}) - \ell_{APST}(\hat{\theta})) = 45.660$  and comparing these quantities with  $\chi_1^2 = 3.8414$ , both null hypotheses are rejected. Finally, Figure 3left shows the histogram of the volcano heights variable, whereas Figure 3right presents the empirical CDF (solid line) together with the CDF of the fitted APST model (dotted line).



**Figure 2.** Volcano height data: QQplot for Student-*t*, ST, PT, and APST models.



**Figure 3.** (Left) Graph of fitted densities to volcano height data. (Right) Empirical CDF and CDF of fitted APST model.

**Table 3.** Parameter estimates (SE) for the fitted models to the volcano height data.

Estimates	Distribution			
	Student- <i>t</i>	ST	PT	APST
$\hat{\mu}$	14.7835(0.3615)	4.7469(0.6892)	8.4027(0.7923)	11.5509(0.1337)
$\hat{\sigma}$	11.0045(0.3975)	14.1532(0.7237)	11.8146(0.4707)	22.6885(0.0792)
$\hat{\lambda}$	–	1.5673(0.1838)	–	5.2347(0.2870)
$\hat{\alpha}$	–	–	1.7912(0.1147)	0.3205(0.0347)
$\hat{\nu}$	3.4156(0.3601)	3.4075(0.3454)	2.7473(0.2566)	12.8734(2.9729)
$\hat{\ell}$	–6273.35	–6219.25	–6228.77	–6205.94
AIC	12,552.70	12,446.49	12,465.53	12,421.87
BIC	12,568.68	12,467.79	12,486.53	12,448.50
CAIC	12,571.68	12,471.79	12,490.83	12,453.50

#### 4.2. Application 2: Stellar Abundances Data

The second data set is related to measurements for 68 solar-type stars, which are available in the package *astrodatR* of the software *R* [24] under the name *Stellar abundances*. These data were previously analyzed Mattos et al. [25] by using the Scale Mixture of Skew Normal Censored Regression (SMSNCR) models. We take only the response variable:  $\log N(Be)$ , which represents the log of the abundance of beryllium scaled to Sun’s abundance (i.e., the Sun has  $\log N(Be) = 0.0$ )

In astronomical research, a previously identified sample of objects (stars, galaxies, quasars, X-ray sources, etc.) is observed at some new wavebands. According to Feigelson [24], due to limited sensitivities, some objects may be undetected, leading to upper limits in their derived luminosities. For this dataset we have 12 left-censored data points, i.e., 12 undetected beryllium measurement, that represents 19.35% of observations. Table 4 presents the ML estimates for the parameters of the censored Student-*t* (CT), censored skew-*t* (CST), censored power-*t* (CPT), and censored alpha-power skew-*t* (CAPST) models, together with their corresponding standard errors. Table 4 also compares the fit of the four models using the model selection criteria (AIC, CAIC and BIC). Note that, again, the CAPST model with heavy tails have better fit than the CT, CST, and CPT models.

To identify atypical observations and/or model misspecification, we analyzed the transformation of the martingale residual,  $r_{MT_i}$ , proposed in Barros et al. [26]. These residuals are defined by

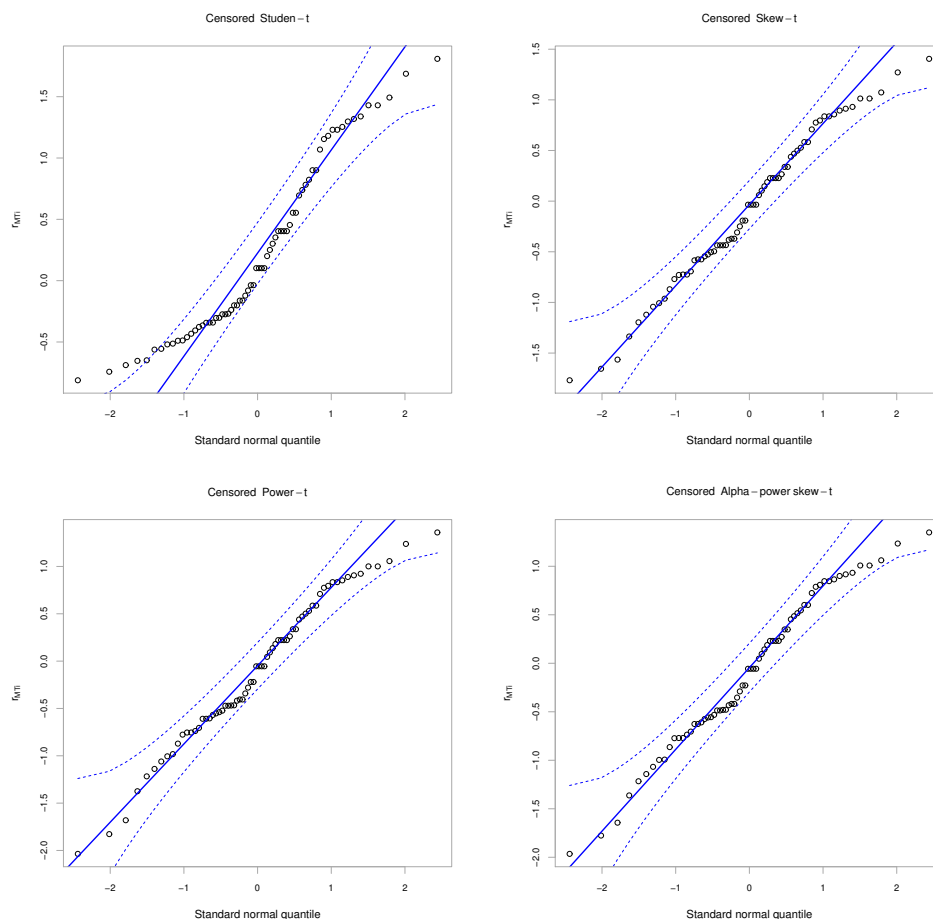
$$r_{MT_i} = \text{sign}(r_{Mi}) \sqrt{-2[r_{Mi} + \delta_i \log(\delta_i - r_{Mi})]}, \quad i = 1, \dots, n$$

where  $r_{Mi} = \delta_i + \log S(y_i; \hat{\theta})$  is the martingale residual proposed by Ortega et al. [27], where  $\delta_i = 0, 1$  indicates whether the *i*-th observation is censored or not, respectively;  $\text{sign}(r_{Mi})$  denotes the sign of

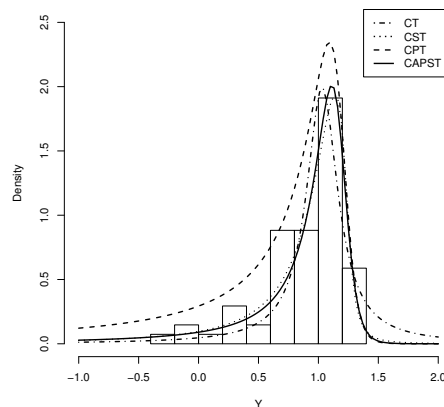
$r_{Mi}$ ; and  $S(y_i; \hat{\theta}) = P_{\hat{\theta}}(Y_i > y_i)$  represents the survival function evaluated at  $y_i$ , where  $\hat{\theta}$  are the MLE for  $\theta$ . The plots of  $r_{MT_i}$  with generated confidence envelopes are presented in Figure 4. From this figure, we can see clearly that the CST, CPT, and CAPST models fit better the data than the CT model, since, in that cases, there are not observations which lie outside the envelopes. The Figure 5 shows the graph of the densities of the different models fitted to the stellar abundances data. From the figure, the CAPST model seems to fit better the stellar abundances data than CT, CST and CPT models.

**Table 4.** Parameter estimates (SE) for the fitted models to the stellar abundances data.

Estimates	Distribution			
	CT	CST	CPT	CAPST
$\hat{\mu}$	1.0314(0.0010)	1.2306(0.0018)	1.2098(0.0052)	1.1761(0.0054)
$\hat{\sigma}$	0.1596(0.0012)	0.2712(0.0058)	0.0818(0.0008)	0.0905(0.0020)
$\hat{\lambda}$	–	–3.5655(3.7748)	–	0.6580(0.5031)
$\hat{\alpha}$	–	–	0.1705(0.0208)	0.1518(0.0251)
$\hat{\nu}$	0.9974(0.0884)	1.2501(0.1774)	6.0927(0.7501)	6.0999(0.7326)
$\hat{\ell}$	–29.50743	–18.87016	–17.67113	–14.80241
AIC	65.01487	45.74033	43.34227	39.60482
BIC	71.67339	54.61836	52.22030	50.70236
CAIC	59.38987	38.37525	35.97719	30.57256



**Figure 4.** Stellar abundances data. Envelopes of transformed martingale residuals for CT, CST, CPT, and CAPST models.



**Figure 5.** Graph of fitted densities to stellar abundances data.

## 5. Conclusions

In this work, a new asymmetric model has been introduced. It is based on the combination of skew- $t$  [1] and power- $t$  [2] models. The new model presents greater ranges of asymmetry and kurtosis, which is very useful for modeling skewed and heavy-tailed data. The problem of estimating the parameters in the model is dealt by using the maximum likelihood approach which is also used for developing large sample properties for the estimators. The elements of the observed information matrix are analytically obtained. The likelihood ratio statistics can be used for testing the APST null hypothesis since the Student- $t$ , ST, and PT models are special cases of the model entertained. Two applications to volcano heights data and stellar abundances data indicate that the proposed model can be a useful alternative to the ST and PT models.

**Author Contributions:** Individual contributions to this article: conceptualization, R.T.-F., H.B., and G.M.-F.; methodology, R.T.-F., H.B., and G.M.-F.; software, R.T.-F., H.B., and G.M.-F.; validation, R.T.-F., H.B., and G.M.-F.; formal analysis, R.T.-F., H.B., and G.M.-F.; investigation, R.T.-F., H.B., and G.M.-F.; resources, R.T.-F., H.B., and G.M.-F.; writing-original draft preparation, R.T.-F., H.B., and G.M.-F.; writing-review and editing, R.T.-F., H.B., and G.M.-F. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding. The research of R. Tovar-Falón was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq/Brazil) under grant 140831/2014-2.

**Acknowledgments:** We thank the anonymous referees for helpful suggestions which improved the article.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

In this section, expressions for the elements of the observed information matrix of the alpha-power skew- $t$  model are provided. Initially we suppose that  $Y \sim \text{APST}(\mu, \sigma, \lambda, \alpha, \nu)$ , and for  $i = 1, \dots, n$  we define  $z_i = (y_i - \mu)/\sigma$ ,  $w_i = \sqrt{(v+1)/(z_i^2 + v)}$ ,  $r_1(z; \nu) = f_T(z; \nu)/\mathcal{F}_T(z; \nu)$ ,  $r_2(z; \lambda, \nu) = f_{ST}(z; \lambda, \nu)/\mathcal{F}_{ST}(z; \lambda, \nu)$ , and  $r_3(z; \lambda, \nu) = (1 + (1 + \lambda^2)z^2/v)^{-\nu/2}/\mathcal{F}_{ST}(z; \lambda, \nu)$ . Denoting the elements of the observed information matrix of the APST model by  $j_{\mu\mu}, j_{\mu\sigma}, \dots, j_{\alpha\alpha}$ , and after some algebraic manipulations, we obtain

$$\begin{aligned}
 j_{\mu\mu} = & -\frac{1}{\sigma^2} \frac{\nu+1}{\nu^2} \sum_{i=1}^n z_i^2 \left(1 + \frac{z_i^2}{\nu}\right)^{-2} + \frac{1}{\sigma^2} \frac{\nu+1}{\nu} \sum_{i=1}^n \left(1 + \frac{z_i^2}{\nu}\right)^{-2} \\
 & + \frac{\lambda}{\sigma^2} \sum_{i=1}^n z_i w_i \left(1 + \frac{z_i^2}{\nu}\right)^{-1} r_1(\lambda z_i w_i; \nu + 1) \\
 & + \frac{2}{\sigma^2} \frac{\lambda}{\nu} \sum_{i=1}^n z_i w_i \left(1 + \frac{z_i^2}{\nu}\right)^{-2} r_1(\lambda z_i w_i; \nu + 1) \\
 & + \frac{\lambda^3}{\sigma^2} \frac{\nu+2}{\nu} \sum_{i=1}^n z_i w_i \left(1 + \frac{z_i^2}{\nu}\right)^{-2} \left(1 + (1 + \lambda^2) \frac{z_i^2}{\nu}\right)^{-1} r_1(\lambda z_i w_i; \nu + 1) \\
 & + \frac{\lambda^2}{\sigma^2} \frac{\nu+1}{\nu} \sum_{i=1}^n \left(1 + \frac{z_i^2}{\nu}\right)^{-3} [r_1(\lambda z_i w_i; \nu + 1)]^2
 \end{aligned}$$

$$\begin{aligned}
 j_{\mu\sigma} = & -\frac{\lambda}{\sigma^2} \frac{\alpha-1}{\pi} \sum_{i=1}^n \left(1 + \frac{z_i^2}{\nu}\right)^{-1} \left(1 + (1 + \lambda^2) \frac{z_i^2}{\nu}\right)^{-1} r_3(z_i; \lambda, \nu) \\
 & + \frac{\alpha-1}{\sigma^2} \frac{\nu+1}{\nu} \sum_{i=1}^n z_i \left(1 + \frac{z_i^2}{\nu}\right)^{-1} r_2(z_i; \lambda, \nu) + \frac{\alpha-1}{\sigma^2} \sum_{i=1}^n [r_2(z_i; \lambda, \nu)]^2 \\
 & \frac{2}{\sigma^2} \frac{\nu+1}{\nu} \sum_{i=1}^n z_i \left(1 + \frac{z_i^2}{\nu}\right)^{-2} \\
 & + \frac{\lambda^3}{\sigma^2} \frac{\nu+2}{\nu} \sum_{i=1}^n z_i^2 w_i \left(1 + \frac{z_i^2}{\nu}\right)^{-2} \left(1 + (1 + \lambda^2) \frac{z_i^2}{\nu}\right)^{-1} r_1(\lambda z_i w_i; \nu + 1) \\
 & + \frac{2}{\sigma^2} \frac{\lambda}{\nu} \sum_{i=1}^n z_i^2 w_i \left(1 + \frac{z_i^2}{\nu}\right)^{-2} r_1(\lambda z_i w_i; \nu + 1) \\
 & - \frac{\lambda}{\sigma^2} \sum_{i=1}^n w_i \left(1 + \frac{z_i^2}{\nu}\right)^{-2} r_1(\lambda z_i w_i; \nu + 1) \\
 & + \frac{\lambda^2}{\sigma^2} \frac{\nu+1}{\nu} \sum_{i=1}^n z_i \left(1 + \frac{z_i^2}{\nu}\right)^{-3} [r_1(\lambda z_i w_i; \nu + 1)]^2 \\
 & - \frac{\lambda}{\sigma^2} \frac{\alpha-1}{\pi} \sum_{i=1}^n z_i \left(1 + \frac{z_i^2}{\nu}\right)^{-1} \left(1 + (1 + \lambda^2) \frac{z_i^2}{\nu}\right)^{-1} r_3(z_i; \lambda, \nu) \\
 & - \frac{\alpha-1}{\sigma^2} \sum_{i=1}^n r_2(z_i; \lambda, \nu) + \frac{\alpha-1}{\sigma^2} \frac{\nu+1}{\nu} \sum_{i=1}^n z_i^2 \left(1 + \frac{z_i^2}{\nu}\right)^{-1} r_2(z_i; \lambda, \nu) \\
 & + \frac{\alpha-1}{\sigma^2} \sum_{i=1}^n z_i [r_2(z_i; \lambda, \nu)]^2
 \end{aligned}$$

$$\begin{aligned}
j_{\mu\lambda} = & -\frac{\lambda^2 \nu + 2}{\sigma} \frac{1}{\nu} \sum_{i=1}^n z_i^2 w_i \left(1 + \frac{z_i^2}{\nu}\right)^{-1} \left(1 + (1 + \lambda^2) \frac{z_i^2}{\nu}\right)^{-1} r_1(\lambda z_i w_i; \nu + 1) \\
& + \frac{1}{\sigma} \sum_{i=1}^n w_i \left(1 + \frac{z_i^2}{\nu}\right)^{-1} r_1(\lambda z_i w_i; \nu + 1) \\
& - \frac{\lambda \nu + 1}{\sigma} \frac{1}{\nu} \sum_{i=1}^n z_i \left(1 + \frac{z_i^2}{\nu}\right)^{-2} [r_1(\lambda z_i w_i; \nu + 1)]^2 \\
& + \frac{\alpha - 1}{\pi\sigma} \sum_{i=1}^n z_i^2 \left(1 + \frac{z_i^2}{\nu}\right)^{-1} r_3(z_i; \lambda, \nu) \\
& + \frac{\alpha - 1}{\pi\sigma} \frac{1}{1 + \lambda^2} \sum_{i=1}^n r_2(z_i; \lambda, \nu) r_3(z_i; \lambda, \nu)
\end{aligned}$$

$$j_{\mu\alpha} = \frac{1}{\sigma} \sum_{i=1}^n r_2(z_i; \lambda, \nu)$$

$$\begin{aligned}
j_{\mu\nu} = & -\frac{1}{\sigma\nu} \sum_{i=1}^n z_i \left(1 + \frac{z_i^2}{\nu}\right)^{-1} + \frac{\nu + 1}{\sigma\nu^2} \sum_{i=1}^n z_i \left(1 + \frac{z_i^2}{\nu}\right)^{-2} \\
& + \frac{\lambda}{\sigma\nu^2} \sum_{i=1}^n z_i^2 w_i \left(1 + \frac{z_i^2}{\nu}\right)^{-2} r_1(\lambda z_i w_i; \nu + 1) \\
& + \frac{\lambda}{2\sigma} \frac{1}{\nu(\nu + 1)} \sum_{i=1}^n (z_i^2 - 1) w_i \left(1 + \frac{z_i^2}{\nu}\right)^{-2} r_1(\lambda z_i w_i; \nu + 1) \\
& + \frac{\lambda^3}{2\sigma} \frac{\nu + 2}{\nu^2(\nu + 1)} \sum_{i=1}^n z_i^2 (z_i^2 - 1) w_i \left(1 + \frac{z_i^2}{\nu}\right)^{-2} \left(1 + (1 + \lambda^2) \frac{z_i^2}{\nu}\right)^{-1} r_1(\lambda z_i w_i; \nu + 1) \\
& - \frac{\lambda^2}{2\sigma} \frac{1}{\nu^2} \sum_{i=1}^n z_i (z_i^2 - 1) \left(1 + \frac{z_i^2}{\nu}\right)^{-3} [r_1(\lambda z_i w_i; \nu + 1)]^2 \\
& + \frac{\lambda}{2\pi\sigma} \frac{\alpha - 1}{\nu + 1} \sum_{i=1}^n z_i \left(1 + (1 + \lambda^2) \frac{z_i^2}{\nu}\right)^{-1} r_3(z_i; \lambda, \nu) \\
& + \frac{\alpha - 1}{2\pi\sigma(\nu + 1)} \frac{\lambda}{1 + \lambda^2} \sum_{i=1}^n r_2(z_i; \lambda, \nu) r_3(z_i; \lambda, \nu) + \frac{\alpha - 1}{2\sigma} \frac{\nu + 1}{\nu^2} \sum_{i=1}^n z_i^2 \left(1 + \frac{z_i^2}{\nu}\right)^{-1} r_2(z_i; \lambda, \nu) \\
& - \frac{\alpha - 1}{2\sigma} \sum_{i=1}^n \log\left(1 + \frac{z_i^2}{\nu}\right) r_2(z_i; \lambda, \nu) - \frac{\alpha - 1}{2\sigma} \sum_{i=1}^n \frac{g(z_i, \nu)}{\mathcal{F}_{ST}(z_i; \lambda, \nu)} r_2(z_i; \lambda, \nu) \\
& + \frac{\lambda}{2\pi\sigma} \frac{\alpha - 1}{\nu} \sum_{i=1}^n z_i \left(1 + \frac{z_i^2}{\nu}\right)^{-1} \left(1 + (1 + \lambda^2) \frac{z_i^2}{\nu}\right)^{-1} r_3(z_i; \lambda, \nu)
\end{aligned}$$

$$\begin{aligned}
j_{\sigma\sigma} = & -\frac{n}{\sigma^2} + \frac{1}{\sigma^2} \frac{\nu+1}{\nu} \sum_{i=1}^n z_i^2 \left(1 + \frac{z_i^2}{\nu}\right)^{-1} + \frac{2}{\sigma^2} \frac{\nu+1}{\nu} \sum_{i=1}^n z_i^2 \left(1 + \frac{z_i^2}{\nu}\right)^{-2} \\
& - \frac{2\lambda}{\sigma^2} \sum_{i=1}^n z_i w_i \left(1 + \frac{z_i^2}{\nu}\right)^{-2} r_1(\lambda z_i w_i; \nu+1) \\
& - \frac{\lambda}{\sigma^2 \nu} \sum_{i=1}^n z_i^3 w_i \left(1 + \frac{z_i^2}{\nu}\right)^{-2} r_1(\lambda z_i w_i; \nu+1) \\
& + \frac{\lambda^3}{\sigma^2} \frac{\nu+2}{\nu} \sum_{i=1}^n z_i^3 w_i \left(1 + \frac{z_i^2}{\nu}\right)^{-2} \left(1 + (1+\lambda^2) \frac{z_i^2}{\nu}\right)^{-1} r_1(\lambda z_i w_i; \nu+1) \\
& + \frac{\lambda^2}{\sigma^2} \frac{\nu+1}{\nu} \sum_{i=1}^n z_i^2 \left(1 + \frac{z_i^2}{\nu}\right)^{-3} r_1(\lambda z_i w_i; \nu+1) \\
& - \frac{\lambda}{\sigma^2} \frac{\alpha-1}{\pi} \sum_{i=1}^n z_i^2 \left(1 + \frac{z_i^2}{\nu}\right)^{-1} \left(1 + (1+\lambda^2) \frac{z_i^2}{\nu}\right)^{-1} r_3(z_i; \lambda, \nu) \\
& - \frac{2(\alpha-1)}{\sigma^2} \sum_{i=1}^n z_i r_2(z_i; \lambda, \nu) - \frac{\alpha-1}{\sigma^2} \frac{\nu+1}{\nu} \sum_{i=1}^n z_i^3 \left(1 + \frac{z_i^2}{\nu}\right)^{-1} r_2(z_i; \lambda, \nu) \\
& + \frac{\alpha-1}{\sigma^2} \sum_{i=1}^n z_i^2 r_2(z_i; \lambda, \nu)
\end{aligned}$$

$$\begin{aligned}
j_{\sigma\lambda} = & \frac{1}{\sigma} \sum_{i=1}^n z_i w_i \left(1 + \frac{z_i^2}{\nu}\right)^{-1} r_1(\lambda z_i w_i; \nu+1) \\
& + \frac{\lambda^2}{\sigma} \frac{\nu+2}{\nu} \sum_{i=1}^n z_i^3 w_i \left(1 + \frac{z_i^2}{\nu}\right)^{-1} \left(1 + (1+\lambda^2) \frac{z_i^2}{\nu}\right)^{-1} r_1(\lambda z_i w_i; \nu+1) \\
& + \frac{\lambda}{\sigma} \frac{\nu+1}{\nu} \sum_{i=1}^n z_i \left(1 + \frac{z_i^2}{\nu}\right)^{-2} [r_1(\lambda z_i w_i; \nu+1)]^2 \\
& - \frac{\alpha-1}{\pi\sigma} \sum_{i=1}^n z_i^2 \left(1 + \frac{z_i^2}{\nu}\right)^{-1} r_3(z_i; \lambda, \nu) \\
& + \frac{\alpha-1}{\pi\sigma} \frac{1}{1+\lambda^2} \sum_{i=1}^n r_2(z_i; \lambda, \nu) r_3(z_i; \lambda, \nu)
\end{aligned}$$

$$j_{\sigma\alpha} = \frac{1}{\sigma} \sum_{i=1}^n z_i r_2(z_i; \lambda, \nu)$$



$$\begin{aligned}
 j_{\sigma\nu} = & -\frac{1}{\sigma\nu} \sum_{i=1}^n z_i^2 \left(1 + \frac{z_i^2}{\nu}\right)^{-1} + \frac{1}{\sigma} \frac{\nu+1}{\nu^2} \sum_{i=1}^n z_i^2 \left(1 + \frac{z_i^2}{\nu}\right)^{-2} \\
 & + \frac{\lambda}{\sigma\nu^2} \sum_{i=1}^n z_i^3 w_i \left(1 + \frac{z_i^2}{\nu}\right)^{-2} r_1(\lambda z_i w_i; \nu+1) \\
 & + \frac{\lambda}{2\sigma} \frac{1}{\nu(\nu+1)} \sum_{i=1}^n z_i (z_i^2 - 1) w_i \left(1 + \frac{z_i^2}{\nu}\right)^{-2} r_1(\lambda z_i w_i; \nu+1) \\
 & - \frac{\lambda^3}{2\sigma} \frac{\nu+2}{\nu^2(\nu+1)} \sum_{i=1}^n z_i^3 (z_i^2 - 1) w_i \left(1 + \frac{z_i^2}{\nu}\right)^{-2} \left(1 + (1 + \lambda^2) \frac{z_i^2}{\nu}\right)^{-1} r_1(\lambda z_i w_i; \nu+1) \\
 & - \frac{\lambda^2}{2\sigma\nu^2} \sum_{i=1}^n z_i^2 (z_i^2 - 1) \left(1 + \frac{z_i^2}{\nu}\right)^{-3} [r_1(\lambda z_i w_i; \nu+1)]^2 \\
 & + \frac{\lambda}{2\pi\sigma} \frac{\alpha-1}{\nu(\nu+1)} \sum_{i=1}^n z_i^2 (z_i^2 - 1) \left(1 + \frac{z_i^2}{\nu}\right)^{-1} \left(1 + (1 + \lambda^2) \frac{z_i^2}{\nu}\right)^{-1} r_3(z_i; \lambda, \nu) \\
 & + \frac{\alpha-1}{2\pi\sigma(\nu+1)} \frac{\lambda}{1 + \lambda^2} \sum_{i=1}^n z_i r_2(z_i; \lambda, \nu) r_3(z_i; \lambda, \nu) \\
 & + \frac{\alpha-1}{2\sigma} \frac{\nu+1}{\nu^2} \sum_{i=1}^n z_i^3 \left(1 + \frac{z_i^2}{\nu}\right)^{-1} r_1(\lambda z_i w_i; \nu+1) \\
 & - \frac{\alpha-1}{2\sigma} \sum_{i=1}^n z_i \log\left(1 + \frac{z_i^2}{\nu}\right) r_1(\lambda z_i w_i; \nu+1) \\
 & - \frac{\alpha-1}{2\sigma} \sum_{i=1}^n z_i \frac{g(z_i, \nu)}{\mathcal{F}_{ST}(z_i, \lambda, \nu)} r_2(z_i; \lambda, \nu)
 \end{aligned}$$

$$\begin{aligned}
 j_{\lambda\lambda} = & \frac{\lambda(\nu+2)}{\nu} \sum_{i=1}^n z_i^3 w_i \left(1 + (1 + \lambda^2) \frac{z_i^2}{\nu}\right)^{-1} r_1(\lambda z_i w_i; \nu+1) \\
 & + \frac{\nu+1}{\nu} \sum_{i=1}^n z_i \left(1 + \frac{z_i^2}{\nu}\right)^{-1} [r_1(\lambda z_i w_i; \nu+1)]^2 \\
 & - \frac{2(\alpha-1)}{\pi} \frac{\lambda}{(1 + \lambda^2)^2} \sum_{i=1}^n \left(1 + (1 + \lambda^2) \frac{z_i^2}{\nu}\right)^{-1} r_3(z_i; \lambda, \nu) \\
 & - \frac{\alpha-1}{\pi} \frac{\lambda}{1 + \lambda^2} \frac{\nu+2}{\nu} \sum_{i=1}^n z_i^2 \left(1 + (1 + \lambda^2) \frac{z_i^2}{\nu}\right)^{-1} r_3(z_i; \lambda, \nu) \\
 & + \frac{\alpha-1}{\pi^2(1 + \lambda^2)^2} \sum_{i=1}^n [r_3(z_i; \lambda, \nu)]^2
 \end{aligned}$$

$$j_{\lambda\alpha} = \frac{1}{\pi(1+\lambda^2)} \sum_{i=1}^n r_3(z_i; \lambda, \nu)$$

$$\begin{aligned} j_{\lambda\nu} = & -\frac{1}{2\nu(\nu+1)} \sum_{i=1}^n z_i(z_i^2-1)w_i \left(1 + \frac{z_i^2}{\nu}\right)^{-1} r_1(\lambda z_i w_i; \nu+1) \\ & + \frac{\lambda^2}{2\nu^2} \frac{\nu+2}{(\nu+1)^2} \sum_{i=1}^n z_i^3(z_i^2-1)w_i \left(1 + \frac{z_i^2}{\nu}\right)^{-1} \left(1 + (1+\lambda^2)\frac{z_i^2}{\nu}\right)^{-1} r_1(\lambda z_i w_i; \nu+1) \\ & + \frac{\lambda}{2\nu^2} \sum_{i=1}^n z_i^2(z_i^2-1) \left(1 + \frac{z_i^2}{\nu}\right)^{-2} [r_1(\lambda z_i w_i; \nu+1)]^2 \\ & + \frac{\alpha-1}{2\pi(\nu+1)} \frac{1-\lambda^2}{(1+\lambda^2)^2} \sum_{i=1}^n r_3(z_i; \lambda, \nu) + \frac{\alpha-1}{2\pi(\nu+1)} \frac{\lambda}{(1+\lambda^2)^2} \sum_{i=1}^n [r_3(z_i; \lambda, \nu)]^2 \\ & - \frac{\alpha-1}{2\pi(\nu+1)} \frac{\lambda^2}{1+\lambda^2} \sum_{i=1}^n z_i^2 \left(1 + (1+\lambda^2)\frac{z_i^2}{\nu}\right)^{-1} r_3(z_i; \lambda, \nu) \\ & - \frac{\alpha-1}{2\pi(1+\lambda^2)} \sum_{i=1}^n \frac{g(z_i, \nu)}{\mathcal{F}_{ST}(z_i, \lambda, \nu)} r_3(z_i; \lambda, \nu) \\ & - \frac{\alpha-1}{2\pi} \sum_{i=1}^n \frac{g_1(z_i, \nu)}{\mathcal{F}_{ST}(z_i, \lambda, \nu)} \end{aligned}$$

$$j_{\alpha\alpha} = \frac{n}{\alpha^2}$$

$$\begin{aligned} j_{\alpha\nu} = & -\frac{n}{2}\psi\left(\frac{\nu+1}{2}\right) + \frac{n}{2}\psi\left(\frac{\nu}{2}\right) + \frac{n}{2\nu} \\ & + \frac{1}{2\pi(\nu+1)} \frac{\lambda}{1+\lambda^2} \sum_{i=1}^n r_3(z_i; \lambda, \nu) - \frac{1}{2} \sum_{i=1}^n \frac{g(z_i, \nu)}{\mathcal{F}_{ST}(z_i, \lambda, \nu)} \end{aligned}$$

$$\begin{aligned}
 j_{\nu\nu} = & -\frac{n\alpha}{2\nu^2} - \frac{n\alpha}{4}\psi_1\left(\frac{\nu+1}{2}\right) + \frac{n\alpha}{4}\psi_1\left(\frac{\nu}{2}\right) - \frac{\nu-1}{2\nu^3} \sum_{i=1}^n z_i^2 \left(1 + \frac{z_i^2}{\nu}\right)^{-1} \\
 & + \frac{\nu+1}{2\nu^3} \sum_{i=1}^n z_i^2 \left(1 + \frac{z_i^2}{\nu}\right)^{-2} \\
 & + \frac{\lambda}{4(\nu+1)^2} \frac{1}{\nu^2} \sum_{i=1}^n z_i(z_i^2-1)(z_i^2+4\nu+3)w_i \left(1 + \frac{z_i^2}{\nu}\right)^{-2} r_1(\lambda z_i w_i; \nu+1) \\
 & - \frac{\lambda}{4\nu(\nu+1)} \left(\psi\left(\frac{\nu+2}{2}\right) - \psi\left(\frac{\nu+1}{2}\right) - \frac{1}{\nu+1}\right) \\
 & \quad \sum_{i=1}^n z_i(z_i^2-1)w_i \left(1 + \frac{z_i^2}{\nu}\right)^{-1} r_1(\lambda z_i w_i; \nu+1) \\
 & - \frac{\lambda^3}{4(\nu+1)} \frac{\nu+2}{\nu^3} \sum_{i=1}^n z_i^3(z_i^2-1) \left(1 + \frac{z_i^2}{\nu}\right)^{-2} \left(1 + (1+\lambda^2)\frac{z_i^2}{\nu}\right)^{-1} r_1(\lambda z_i w_i; \nu+1) \\
 & + \frac{\lambda}{4\nu(\nu+1)} \sum_{i=1}^n z_i(z_i^2-1) \log\left(1 + \frac{\lambda^2 z_i^2}{\nu+z_i^2}\right) r_1(\lambda z_i w_i; \nu+1) \\
 & + \frac{\lambda^2}{4\nu^3(\nu+1)} \sum_{i=1}^n z_i^2(z_i^2-1)^2 \left(1 + \frac{z_i^2}{\nu}\right)^{-3} [r_1(\lambda z_i w_i; \nu+1)]^2 \\
 & - \frac{\alpha-1}{2\pi(\nu+1)^2} \frac{\lambda}{1+\lambda^2} \sum_{i=1}^n r_3(z_i; \lambda, \nu) \\
 & + \frac{\alpha-1}{4\pi(\nu+1)} \frac{\lambda}{\nu} \sum_{i=1}^n z_i^2 \left(1 + (1+\lambda^2)\frac{z_i^2}{\nu}\right)^{-1} r_3(z_i; \lambda, \nu) \\
 & - \frac{\alpha-1}{4\pi(\nu+1)} \frac{\lambda}{1+\lambda^2} \sum_{i=1}^n \log\left(1 + (1+\lambda^2)\frac{z_i^2}{\nu}\right) r_3(z_i; \lambda, \nu) \\
 & + \frac{\alpha-1}{4\pi(\nu+1)} \frac{\lambda}{1+\lambda^2} \left(\psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right) - \frac{1}{\nu}\right) \sum_{i=1}^n r_3(z_i; \lambda, \nu) \\
 & - \frac{\alpha-1}{2\pi(\nu+1)} \frac{\lambda}{1+\lambda^2} \sum_{i=1}^n \frac{g(z_i, \nu)}{\mathcal{F}_{ST}(z_i, \lambda, \nu)} r_3(z_i; \lambda, \nu) \\
 & + \frac{\alpha-1}{4\pi^2(\nu+1)^2} \frac{\lambda^2}{(1+\lambda^2)^2} \sum_{i=1}^n [r_3(z_i; \lambda, \nu)]^2 \\
 & + \frac{\alpha-1}{4} \sum_{i=1}^n \left(\frac{g(z_i, \nu)}{\mathcal{F}_{ST}(z_i, \lambda, \nu)}\right)^2 - \frac{\alpha-1}{2} \sum_{i=1}^n \frac{g_2(z_i, \nu)}{\mathcal{F}_{ST}(z_i, \lambda, \nu)}
 \end{aligned}$$

where  $g(z; \nu)$  is given in Equation (16), and  $g_1(z; \nu)$  and  $g_2(z; \nu)$  are given in Equations (A1) and (A3), respectively.

$$\begin{aligned}
g_1(x; \nu) = & \int_{-\infty}^x \left\{ \frac{(\nu+1)s^2}{\nu(s^2+\nu)} - \log \left( 1 + \frac{s^2}{\nu} \right) \right\} \left\{ 1 + \frac{(1+\lambda^2)s^2}{\nu} \right\}^{-\frac{\nu+2}{2}} ds \\
& - \int_{-\infty}^x \frac{s}{s^2+\nu} \left\{ 1 + \frac{(1+\lambda^2)s^2}{\nu} \right\}^{-\frac{\nu+2}{2}} ds \\
& + \int_{-\infty}^x \frac{\lambda^2(\nu+2)s^3}{(s^2+\nu)(\nu+(1+\lambda^2)s^2)} \left\{ 1 + \frac{(1+\lambda^2)s^2}{\nu} \right\}^{-\frac{\nu+2}{2}} ds
\end{aligned} \tag{A1}$$

$$\begin{aligned}
g_2(x; \nu) = & \int_{-\infty}^x \left\{ \frac{s^2(s^2\nu - 2\nu - s^2)}{\nu^2(s^2+\nu)^2} + \frac{1}{2} \left[ \frac{(\nu+1)s^2}{\nu(s^2+\nu)} - \log \left( 1 + \frac{s^2}{\nu} \right) \right]^2 \right\} f_{ST}(s; \lambda, \nu) ds \\
& + \frac{\lambda}{2\pi(\nu+1)} \int_{-\infty}^x \frac{s(s^2-1)}{(s^2+\nu)} \left\{ \frac{(\nu+1)s^2}{\nu(s^2+\nu)} - \log \left( 1 + \frac{s^2}{\nu} \right) \right\} \\
& \quad \times \left\{ 1 + \frac{(1+\lambda^2)s^2}{\nu} \right\}^{-\frac{\nu+2}{2}} ds \\
& + \frac{\lambda}{\pi} \int_{-\infty}^x \frac{s}{(s+\nu)^2} \left\{ 1 + \frac{(1+\lambda^2)s^2}{\nu} \right\}^{-\frac{\nu+2}{2}} ds
\end{aligned} \tag{A2}$$

$$\begin{aligned}
& + \frac{\lambda}{2\pi} \left( \psi \left( \frac{\nu+1}{2} \right) - \psi \left( \frac{\nu}{2} \right) - \frac{1}{\nu} \right) \int_{-\infty}^x \frac{s}{s^2+\nu} \left\{ 1 + \frac{(1+\lambda^2)s^2}{\nu} \right\}^{-\frac{\nu+2}{2}} ds \\
& - \frac{\lambda}{2\pi} \int_{-\infty}^x \frac{s}{s^2+\nu} \left\{ \frac{(\nu+2)(1+\lambda^2)s^2}{\nu(\nu+(1+\lambda^2)s^2)} - \log \left( 1 + \frac{(1+\lambda^2)s^2}{\nu} \right) \right\} \\
& \quad \times \left\{ 1 + \frac{(1+\lambda^2)s^2}{\nu} \right\}^{-\frac{\nu+2}{2}} ds
\end{aligned} \tag{A3}$$

## References

1. Azzalini, A. A class of distributions which includes the normal ones. *Scand. J. Stat.* **1985**, *12*, 171–178.
2. Durrans, S.R. Distributions of fractional order statistics in hydrology. *Water Resour. Res.* **1992**, *28*, 1649–1655. [\[CrossRef\]](#)
3. Martínez-Flórez, G.; Bolfarine, H.; Gómez, H.W. The alpha–power tobit model. *Commun. Stat. Theory Methods* **2013**, *42*, 633–643. [\[CrossRef\]](#)
4. Birnbaum, Z.W.; Saunders, S.C. A new family of life distributions. *J. Appl. Probab.* **1969**, *6*, 319–327. [\[CrossRef\]](#)
5. Martínez-Flórez, G.; Bolfarine, H.; Gómez, H.W. An alpha-power extension for the Birnbaum-Saunders distribution. *Statistics* **2014**, *48*, 896–912. [\[CrossRef\]](#)
6. Gupta, R.D.; Gupta, R.C. Analyzing skewed data by power-normal model. *Test* **2008**, *17*, 197–210. [\[CrossRef\]](#)
7. Pewsey, A.; Gómez, H. W.; Bolfarine, H. Likelihood–based inference for power distributions. *Test* **2012**, *21*, 775–789. [\[CrossRef\]](#)
8. Martínez-Flórez, G.; Bolfarine, H.; Gómez, H.W. Skew-normal alpha-power model. *Stat. J. Theor. Appl. Stat.* **2014**, *48*, 1414–1428. [\[CrossRef\]](#)
9. Azzalini, A.; Capitanio, A. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew-*t* distribution. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2003**, *65*, 367–389. [\[CrossRef\]](#)
10. Branco, M.D.; Dey, D.K. General class of multivariate skew-elliptical distributions. *J. Multivar. Anal.* **2001**, *79*, 99–113. [\[CrossRef\]](#)
11. Durrans, S.R. Multivariate skew *t*-distribution. *Stat. J. Theor. Appl. Stat.* **2003**, *37*, 359–363.
12. Sahu, S.K.; Dey, D.K.; Branco, M.D. A new class of multivariate skew distributions with applications to Bayesian regression models. *Can. J. Stat.* **2003**, *31*, 129–150. [\[CrossRef\]](#)
13. Jones, M.C.; Faddy, M.J. A skew extension of the *t*-distribution, with Applications. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2003**, *65*, 159–174. [\[CrossRef\]](#)

14. Zhao, J.; Kim, H.M. Power  $t$  distributions. *Commun. Stat. Appl. Methods* **2016**, *23*, 321–334.
15. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2018. Available online: <http://www.R-project.org> (accessed on 10 October 2019).
16. Lehman, E.L.; Casella, G. *Theory of Point Estimation*, 2nd ed.; Springer: New York, NY, USA, 1998.
17. Frieden, B.R. *Science from Fisher Information: A Unification*; Cambridge University Press: Cambridge, UK, 2004.
18. Arellano-Valle, R.B.; Azzalini, A. The centered parameterization and related quantities of the skew- $t$  distribution. *J. Multivar. Anal.* **2013**, *113*, 73–90. [[CrossRef](#)]
19. Arellano-Valle, R.B.; Azzalini, A. The centered parametrization for the multivariate skew-normal distribution. *J. Multivar. Anal.* **2008**, *99*, 1362–1382. [[CrossRef](#)]
20. Akaike, H. A new look at statistical model identification. *IEEE Trans. Autom. Contr.* **1974**, *19*, 716–722. [[CrossRef](#)]
21. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **1978**, *6*, 461–464. [[CrossRef](#)]
22. Bozdogan, H. Model selection and akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika* **2010**, *52*, 345–370. [[CrossRef](#)]
23. Siebert, L.; Simkin, T.; Kimberly, P. Global Volcanism Program. In *Volcanoes of the World*; v. 4.6.0.; Venzke, E., Ed.; Smithsonian Institution: Washington, DC, USA, 2013. Available online: <https://doi.org/10.5479/si.GVP.VOTW4-2013> (accessed on 10 October 2019).
24. Feigelson, E.D. astrodatR: Astronomical Data. R Package v. 0.1. Available online: <http://CRAN.R-project.org/package=astrodatR> (accessed on 10 October 2019).
25. Mattos, T.; Garay, A.M.; Lachos, V.H. Likelihood-based inference for censored linear regression models with scale mixtures of skew-normal distributions. *J. Appl. Stat.* **2018**, *45*, 2019–2066. [[CrossRef](#)]
26. Barros, M.; Galea, M.; González, M.; Leiva, V. Influence diagnostics in the tobit censored response model. *Stat. Methods Appl.* **2010**, *19*, 379–397. [[CrossRef](#)]
27. Ortega, E.M.; Bolfarine, H.; Paula, G.A. Influence diagnostics in generalized log-gamma regression models. *Comput. Stat. Data Anal.* **2003**, *42*, 165–186. [[CrossRef](#)]



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