Bimodal extension based on the skew-t-normal distribution

Mehdi Amiri[1,4], Héctor W. Gómez[2], Ahad Jamalizadeh[3] and Mina Towhidi[4]

Abstract: In this paper, a skew and uni-/bi-modal extension of the Student-t distribution is considered. This model is more flexible and has wider ranges of skewness and kurtosis than the other skew distributions in literature. Fisher information matrix for the proposed model and some submodels are derived. With a simulation study and some real data sets, applicability of the proposed models are illustrated.

Keywords: Skewness, Kurtosis, Bimodal density, Fisher information matrix, Maximum likelihood estimation.

BJPS - Accepted Manuscript

1 Introduction

In practice, we sometimes encounter datasets having high values of skewness and/or kurtosis in their frequency curves which may cause the inadequacy of using the ordinary normal or other symmetric distributions such as the Student-t or the Laplace family of distributions as fitting models. In these cases, there is a tendency towards more flexible distributions to represent features of the data. The first proposals of such non-normal or non-symmetric distributions can be traced back to the nineteenth century. Edgeworth (1886) studied the problem of fitting asymmetric distributions to asymmetric frequency data. A few years later, Pearson (1893) defined a "generalized form of the normal curve of an asymmetrical character". In the second half of the twentieth century, the

[1]Department of Statistics, Faculty of Basic Sciences, University of Hormozgan, Bandar Abbas, Iran, e.mail: m.amiri@hormozgan.ac.ir

[2]Departamento de Matematicas, Facultad de Ciencias Basicas, Universidad de Antofagasta, Antofagasta,Chile, e.mail: hector.gomez@uantof.cl

[3]Department of Statistics, Faculty of Mathematics & Computer, Shahid Bahonar University of Kerman, Kerman, Iran, e.mail: A.Jamalizadeh@uk.ac.ir

[4]Department of Statistics, Shiraz University, Shiraz, Iran, e.mail: Mtowhidi@susc.ac.ir

interest for skew distributions grows even stronger.

A decisive point in the development of skew distributions is the paper by Azzalini (1985), where is introduced the so-called skew-normal distribution. Further Azzalini (1986) extended this class to a general class known in the literature as the skew-symmetric distributions. This class is presented in the following,

Theorem 1.1. *Let f_0 be a probability density function (pdf) symmetric about zero, and G is a cumulative distribution function (cdf) such that $\acute{G}$ exists and is a pdf symmetric about zero, then*

$$f(z;\lambda) = 2f_0(z)\,G(\lambda z) \quad , z \in \mathbb{R},$$

is a pdf for any $\lambda \in \mathbb{R}$. For this family the notation $Z \sim Sf_0(\lambda)$ is used to denote the fact that Z is distributed according to the density above.

Skew normal (Azzalini (1985)) pdf is derived from Theorem 1.1 with replacing f_0 and G by standard normal pdf ϕ and cdf Φ respectively and related random variable Z having this pdf is denoted by $Z \sim SN(\lambda)$. Henze (1986), using a probabilistic representatin for the SN distribution, derived the moments. Problems of inference for SN distribution were studied by Peswey (2000).

In recent years, wide range of skew (unimodal or multimodal) distributions have been discussed in literature. A unimodal family is considered by Gupta et al. (2002) who replaced f_0 with the Laplace, logistic and uniform pdfs and G with the respective cdfs. Nadarajah and Kotz (2003) considered another unimodal family by replacing f_0 with fixed standard normal pdf ϕ and G with the cdf Student-t, Cauchy, Laplace, logistic and uniform, named as the distributions skew-normal-t, skew-normal-cauchy, skew-normal-laplace, skew-normal-logistic and skew-normal-uniform, respectively. Gómez et al. (2007) considered f_0 to be the Student-t, logistic, Laplace, uniform pdfs and G be the fixed as Φ.

By generalizing the SN model by adding a further shape parameter, the extended SN (ESN) family of distributions is presented in the seminal paper of Azzalini (1985) and it is studied subsequently by Arnold et al. (1993). Extensions to the multivariate context are also studied by Arnold and Beaver (2002). Azzalini and Capitanio (2003) presented a multivariate skew-t distribution as scale-mixture of the multivariate SN distribution. The random variable $X \overset{d}{=} W^{-1/2}Z$, where W and Z are independent, $W \sim Gamma(\nu/2, \nu/2)$ (Gamma distribution with shape and scale parameters $\nu/2$) and $Z \sim SN(\lambda)$, has the univariate skew-t distribution with parameters λ

BJPS - Accepted Manuscript

and ν (degrees of freedom) and denoted by $X \sim St(\lambda, \nu)$. The skew generalized normal (SGN) distribution has been introduced by Arellano-Valle et al. (2004). In particular, it was established that the SGN distribution can be represented as a shape mixture of the SN distribution by taking a normal mixing distribution for the shape parameter. Arslan and Genc (2009) studied skew generalized-t (SGT) distribution as the scale mixture of a skew exponential power and generalized gamma distributions. Ma and Genton (2004) proposed a flexible class of skew-symmetric distributions and captured skewness, heavy tails and multimodality systematically. Another generalization of the SN distribution is the Balakrishnan skew-normal (BSN) introduced by Balakrishnan (2002), as a discussant of Arnold and Beaver (2002). Shafiei and Doostparast (2014) proposed a generalization of skew-t distribution of Azzalini and Capitanio (2003), as a scale mixture of the BSN distribution, named Balakraisnan skew-t (BST) distribution. Gómez et al. (2011) extended the class of skew-symmetric distributions by proposing the class of skew flexible elliptical distributions. Ali et al. (2010) introduced skew symmetric inverse reflected distributions. Their proposed distributions are skew-symmetric distributions, defined based on the reflected gamma, reflected Weibull and the reflected Pareto distributions.

The classes mentioned above, include the SN distribution as a particular case. Many authors proposed an asymmetric normal family of distributions with a different structure than the SN class considered by Azzalini (1985). For example: Mudholkar and Huston (2000) (Epsilon-skew-normal distribution), Kim (2005) (Two piece skew-normal distribution), Elal-Olivero (2010) (Alpha-skew-normal distribution), Arellano-Valle et al. (2010) (Extended epsilon skew-normal distribution) and Rosco et al. (2011) (Sinh-arcsinhed t distribution).

With this setup, the rest of the paper is organized as follows. In Section 2, the skew-flexible-t-normal (SFTN) distribution is introduced and some of its statistical properties are discussed. In Section 3, the moments of the SFTN distribution are derived and the additional flexibility of the model in covering skewness and kurtosis with respect to other skew models is shown. In Section 4, the Fisher information matrix is obtained. With a simulation study in Section 5, consistency of the maximum likelihood estimators of the parameters are illustrated. Three famous real data sets in the literature are considered in Section 6 to illustrate the applicability of the proposed models.

BJPS - Accepted Manuscript

2 Skew-flexible-t-normal distribution

An extension of skew-symmetric distributions is proposed by Gómez et al. (2011). They extend Theorem 1.1, as follows,

Theorem 2.1. *Let f be a pdf symmetric about zero, F the cdf of f and G an absolutely continuous cdf such that $G(x) + G(-x) = 1$. Then*

$$g(z; \lambda, \delta) = c_\delta f(|z| + \delta)\, G(\lambda z) \quad , z \in \mathbb{R},$$

is the pdf where $\lambda, \delta \in \mathbb{R}$ and $c_\delta = (F(-\delta))^{-1}$. The random variable Z with the above pdf, is said to have skew-flexible-elliptical distribution, denoted by $SFf(\lambda, \delta)$.

Taking $f = \phi$ and $G = \Phi$, we have the skew-flexible-normal (SFN) distribution (Gómez et al. (2011)) with pdf as

$$f(z; \lambda, \delta) = c_\delta \phi(|z| + \delta)\, \Phi(\lambda z) \quad , z \in \mathbb{R}, \tag{2.1}$$

where $c_\delta = (\Phi(-\delta))^{-1}$. A random variable Z having the above pdf is denoted by $Z \sim SFN(\lambda, \delta)$. With numerical calculations, maximum values of skewness and kurtosis coefficients for this family were computed as 1.995 and 5.967, respectively.

The applied models in this paper are a special case of the SFf distributions and is defined as follows.

Corollary 2.2. *The random variable Z has the skew-flexible-t-normal (SFTN) distribution if its pdf is given by*

$$f(z; \lambda, \delta, \nu) = c_{\delta,\nu} t_\nu(|z| + \delta)\, \Phi(\lambda z) \quad , z \in \mathbb{R}, \tag{2.2}$$

where λ, $\delta \in \mathbb{R}$, $\nu \in \mathbb{R}^+$, $c_{\delta,\nu} = (T_\nu(-\delta))^{-1}$ and $t_\nu(.)$ and $T_\nu(.)$ are respectively the pdf and the cdf of t_ν (Student-t distribution with ν degrees of freedom). This distribution is denoted by $Z \sim SFTN(\lambda, \delta, \nu)$.

Using Corollary 2.2, the following properties are derived.

Remark 2.3. *(a)* $f(z; \lambda, 0, \nu) = 2t_\nu(z)\, \Phi(\lambda z)$,

(b) $$f(z; 0, \delta, \nu) = \begin{cases} \frac{c_{\delta,\nu}}{2} t_\nu(z - \delta) & , z < 0 \\ \frac{c_{\delta,\nu}}{2} t_\nu(z + \delta) & , z \geq 0 \end{cases},$$

BJPS - Accepted Manuscript

(c) *If* $\lambda \to +\infty$ *then* $f(z;\lambda,\delta,\nu) \to c_{\delta,\nu} t_\nu (z+\delta) I(z \geq 0)$,

(d) $f(z;0,0,\nu) = t_\nu(z)$,

(e) *If* $\lambda \to -\infty$ *then* $f(z;\lambda,\delta,\nu) \to c_{\delta,\nu} t_\nu (z-\delta) I(z < 0)$,

(f) *If* $v \to +\infty$ *then* $f(z;\lambda,\delta,\nu) \to (\Phi(-\delta))^{-1} \ \phi(|z|+\delta)\, \Phi(\lambda z)$.

Results (a) and (d) in Remark 2.3 imply that the family of SFTN distributions contains the skew-t-normal distribution, proposed by Gómez et al. (2007), and Student-t distribution. Result (b) indicates that the symmetric form of the SFTN distribution ($\lambda = 0$) coincides with uni-/bi-modal mixture of two truncated Student-t_v distributions $Tt_{(-\infty,0)}(v;\delta,1)$ and $Tt_{(0,+\infty)}(v;-\delta,1)$ where $Tt_I(v;\mu,\sigma)$ denotes the Student-t_ν distribution with location parameter μ, scale parameter σ and ν degrees of freedom truncated to interval $I \subseteq \mathbb{R}$. This mixture form is unimodal if $\delta > 0$ and bimodal if $\delta < 0$. Result (f) establishes that the family of SFTN distributions contains the SFN distribution with pdf (2.1).

A special case of the SFTN model is when $\nu = 1$, which follows from (2.2).

Corollary 2.4. *A random variable* Z *has the skew-flexible-cauchy-normal (SFCN) distribution with parameters* $\lambda, \delta \in \mathbb{R}$*, denoted by* $Z \sim SFCN(\lambda,\delta)$*, if its pdf is given by*

$$f(z;\lambda,\delta) = \frac{\left(\frac{\pi}{2} - \tan^{-1}(\delta)\right)^{-1} \Phi(\lambda z)}{\left(1 + (|z|+\delta)^2\right)} \quad , \ z \in \mathbb{R}. \tag{2.3}$$

2.1 Uni-/bi-modality property

Now, we investigate the properties related to the uni/bi-modality of $SFTN(\lambda,\delta,\nu)$ distribution. It is easy to verify

$$\frac{\partial}{\partial z}\log(f(z;\lambda,\delta,\nu)) = \lambda\frac{\phi(\lambda z)}{\Phi(\lambda z)} - \frac{(\nu+1)(|z|+\delta)}{\nu + (|z|+\delta)^2} sign(z)$$

and hence the pdfs (2.2) and (2.3) is not differentiable at $z = 0$.

Set $\frac{\partial}{\partial z}\log(f(z;\lambda,\delta,\nu))$ to zero, we conclude that zero points $z_1 \in \mathbb{R}^+$ and $z_2 \in \mathbb{R}^-$ (if there exist any) can be found by solving the following equations

$$\begin{aligned}
(\nu+1)\frac{(z_1+\delta)}{\nu + (z_1+\delta)^2} &= \lambda\frac{\phi(\lambda z_1)}{\Phi(\lambda z_1)}, \\
(\nu+1)\frac{(z_2-\delta)}{\nu + (z_2-\delta)^2} &= \lambda\frac{\phi(\lambda z_2)}{\Phi(\lambda z_2)}.
\end{aligned}$$

BJPS - Accepted Manuscript

The second derivative test can be applied to show that z_1 and z_2 are two different modes. For $\delta < 0$ and $\lambda = 0$, the zero points are $z_1 = -\delta$, $z_2 = \delta$ and pdfs (2.2) and (2.3) are bimodal. For $\delta < 0$, if $\lambda \to +\infty$ then zero point $z_1 \to -\delta$ and if $\lambda \to -\infty$ then zero point $z_2 \to \delta$ which proves the unimodality of these pdfs. Numerical calculations show that these pdfs are unimodal for the finite λ and $\delta > 0$. Figure 1 depicts examples of the SFTN and SFCN distributions given in (2.2) and (2.3).

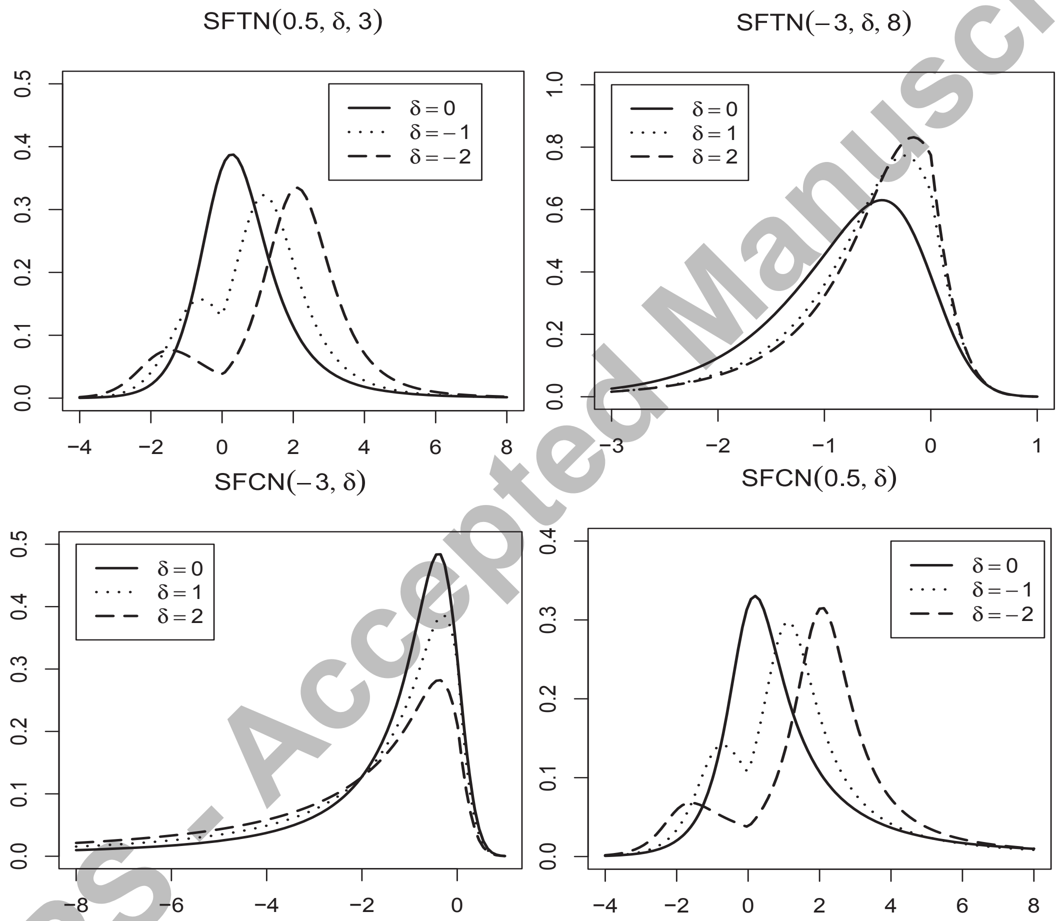

Figure 1: Density plots, First row: $SFTN(0.5, \delta, 3)$(Left figure), $SFTN(-3, \delta, 8)$(Right figure) and Second row: $SFCN(-3, \delta)$ (Left figure) and $SFTN(0.5, \delta)$(right figure).

2.2 Stochastic Representation and data generation

In the following proposition, we give a stochastic representation of the SFTN distributed random variable. This representation will be useful for calculating the moments and random data generation.

BJPS - Accepted Manuscript

Proposition 2.5. *If $T \sim Tt_{(0,+\infty)}(\nu; -\delta, 1)$ and $\frac{S+1}{2}|(T = t) \sim Bernoulli(\Phi(\lambda t))$, then $TS \sim SFTN(\lambda, \delta, \nu)$.*

Proof. Let $Z = TS$. Then the pdf of Z is given by

$$\begin{aligned} f_T(z; \delta, \nu) P(S = 1|T = z) &= c_{\delta,\nu} t_\nu (z + \delta) \Phi(\lambda z), \ for \ z > 0, \\ f_T(-z; \delta, \nu) P(S = -1|T = -z) &= c_{\delta,\nu} t_\nu (-z + \delta)(1 - \Phi(-\lambda z)), \ for \ z < 0, \end{aligned}$$

which is of the form (2.2). □

The p-th quantile of the random variable $T \sim Tt_{(0,+\infty)}(\nu; -\delta, 1)$ is

$$Q_T(p; \delta, \nu) = qt\left(1 + \frac{p-1}{c_{\delta,\nu}}; \nu\right) - \delta \quad , 0 < p < 1, \tag{2.4}$$

where $qt(\alpha; \nu)$ is the α-th quantile of the Student-t_ν distribution. In the special case when $\nu = 1$, it reduces to $Q_T(p; \delta, 1) = \tan\left(\frac{\pi}{2}p + \tan^{-1}(\delta)(1-p)\right) - \delta$.

To simulate the data from $SFTN$ and $SFCN$ distributions, we can use the Proposition 2.5 and the above results and present the following corollary.

Corollary 2.6. *The random variable $Z \sim SFTN(\lambda, \delta, \nu)$ can be simulated as follows: First, generate $U \sim U(0, 1)$ and put $T = Q_T(U; \delta, \nu)$ given by (2.4). Next, generate $B \sim Bernoulli(\Phi(\lambda T))$, then set $Z = (2B - 1)T$.*

3 Moments

To derive the moments of SFTN distribution, the following lemma is useful.

Lemma 3.1. *(Ho et al. (2012)) Let $X \sim Tt_{(\delta,+\infty)}(\nu; 0, 1)$ and $\mu_n = E(X^n)$, then*

$$\mu_{2k+1} = \nu^{k+1} c(\nu) c_{\delta,\nu} \textstyle\sum_{j=0}^{k} \binom{k}{j} \frac{(-1)^{k-j}}{\nu-2j-1} \Delta_\nu(j),$$

$$\mu_{2k} = \nu^k c_{\delta,\nu} \textstyle\sum_{j=0}^{k} \binom{k}{j} (-1)^{k-j} p_j \frac{\Gamma\left(\frac{\nu}{2}-j\right)\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{\nu+1}{2}-j\right)},$$

where $c(\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}}$, $\Delta_\nu(j) = \left(1 + \frac{\delta^2}{\nu}\right)^{-(\nu-2j-1)/2}$ *and* $p_j = T_{\nu-2j}\left(-\delta\sqrt{\frac{\nu-2j}{\nu}}\right)$.

Using Lemma 3.1, the moments of $T \sim Tt_{(0,+\infty)}(\nu; -\delta, 1)$ can be obtained as follows,

$$E(T^n) = \sum_{j=0}^{n} \binom{n}{j} (-1)^{n-j} \delta^{n-j} \mu_j. \tag{3.1}$$

BJPS - Accepted Manuscript

Applying representation given in Proposition 2.5, the moments of SFTN distribution are as follows,

Proposition 3.2. *Let $Z \sim SFTN(\boldsymbol{\theta})$ where $\boldsymbol{\theta} = (\lambda, \delta, \nu)$ and $\nu > n$, then*

$$E_{\boldsymbol{\theta}}(Z^n) = \begin{cases} E_{\boldsymbol{\theta}}\left(T^{2k}\right) & , n = 2k \\ d_{2k+1}(\boldsymbol{\theta}) - E_{\boldsymbol{\theta}}\left(T^{2k+1}\right) & , n = 2k+1 \end{cases} \quad , \ k = 0, 1, ...$$

where $d_j(\boldsymbol{\theta}) = 2c_{\delta,\nu} \int_0^{+\infty} z^j t_\nu(z+\delta) \Phi(\lambda z) dz$, for $j = 0, 1, ...$, should be calculated numerically and $E_{\boldsymbol{\theta}}(T^n)$ is given by (3.1).

Proof. Following Proposition 2.5,

$$\begin{aligned} E_{\boldsymbol{\theta}}(Z^n) &= E_{\boldsymbol{\theta}}(S^n T^n) \\ &= E_{\boldsymbol{\theta}}(T^n E\{S^n|T\}) \\ &= E\{T^n\{\Phi(\lambda T) + (-1)^n \Phi(-\lambda T)\}\} \\ &= \begin{cases} E_{\boldsymbol{\theta}}\left(T^{2k}\right) & , n = 2k \\ 2E_{\boldsymbol{\theta}}\left(T^{2k+1}\Phi(\lambda T)\right) - E_{\boldsymbol{\theta}}\left(T^{2k+1}\right) & , n = 2k+1 \end{cases} , \end{aligned}$$

which completes the proof. □

Using Proposition 3.2, we can obtain the skewness and kurtosis coefficients of $Z \sim SFTN(\boldsymbol{\theta})$, which are defined by

$$\gamma_1(\boldsymbol{\theta}) = \frac{E\left(Z - E(Z)\right)^3}{\sqrt{var(Z)}^3}, \gamma_2(\boldsymbol{\theta}) = \frac{E\left(Z - E(Z)\right)^4}{\sqrt{var(Z)}^4} - 3.$$

Table 1 represents these ranges of $\gamma_i, i = 1, 2$, for different values of ν.

To show the superiority and flexibility of the SFTN model in covering the skewness and kurtosis of the data, we also compute the maximum ranges of skewness and kurtosis for the families, STN (Gómez et al. (2007)), skewed distributions generated by the normal kernel (Nadarajah and Kotz (2003)), alpha-skew-normal (Elal-Olivero (2010)), extended-skew-normal (Arnold et al. (2002)), epsilon-skew-normal (Mudholkar and Huston (2000)), skew-flexible-normal (Gómez et al. (2011)), epsilon-half-normal (Castro et al. (2012)), normal-skew-normal (Gómez et al. (2013)) and our proposed model. Table 2 gives these ranges.

BJPS - Accepted Manuscript

Table 1: Ranges for the measures of skewness and kurtosis coefficients for different values of ν for the SFTN distribution

ν	Range for γ_1		Range for γ_2+3	
	lower	upper	lower	upper
5	-4.648	4.648	6.810	73.799
6	-3.810	3.810	4.888	38.667
7	-3.381	3.381	4.211	27.857
8	-3.118	3.118	3.861	22.725
9	-2.940	2.940	3.648	19.756
10	-2.811	2.811	3.504	17.828
20	-2.344	2.344	3.043	12.127
30	-2.218	2.218	2.934	10.892
40	-2.160	2.160	2.885	10.356
50	-2.126	2.126	2.857	10.056
100	- 2.061	2.061	2.805	9.498
150	-2.040	2.040	2.778	9.331

4 Inference

In the following, inference aspects are discussed for the proposed distribution. The inference procedures are based on the maximum likelihood estimation (MLE) method.

Let $Z \sim SFTN\left(\lambda,\delta,\nu\right)$, then SFTN family of distributions with location-scale parameters is defined as the distribution of $X=\mu+\sigma Z$ for $\mu\in\mathbb{R}$ and $\sigma\in\mathbb{R}^{+}$ and the corresponding pdf is given by

$$f\left(x;\boldsymbol{\eta}\right)=\frac{c_{\delta,\nu}}{\sigma}t_{\nu}\left(\frac{|x-\mu|}{\sigma}+\delta\right)\Phi\left(\lambda\frac{(x-\mu)}{\sigma}\right),\ x\in\mathbb{R}, \tag{4.1}$$

and denoted by $X\sim SFTN\left(\boldsymbol{\eta}\right)$ where $\boldsymbol{\eta}=\left(\mu,\sigma,\lambda,\delta,\nu\right)$. Also location-scale version of the random variable $Z\sim SFCN\left(\lambda,\delta\right)$ with pdf (2.3) is

$$f\left(x;\boldsymbol{\xi}\right)=\frac{\left(\frac{\pi}{2}-\tan^{-1}\left(\delta\right)\right)^{-1}\Phi\left(\lambda\sigma^{-1}\left(x-\mu\right)\right)}{\sigma(1+\left(\sigma^{-1}|x-\mu|+\delta\right)^{2})},\ x\in\mathbb{R}, \tag{4.2}$$

and denoted by $X\sim SFCN\left(\boldsymbol{\xi}\right)$ where $\boldsymbol{\xi}=\left(\mu,\sigma,\lambda,\delta\right)$.

Let $X_1,X_2,...,X_n$ be a random sample drawn from the SFTN distribution. The log-likelihood

BJPS - Accepted Manuscript

Table 2: Maximum of skewness (γ_1) and kurtosis ($\gamma_2 + 3$) coefficients.

Distribution family	Max. γ_1	Max. $\gamma_2 + 3$
Skew-normal	0.995	3.869
Skew-normal-t	0.995	4.124
Skew-normal-cauchy	0.995	4.124
Skew-normal-laplace	0.995	3.869
Skew-normal-logistic	0.995	3.869
Skew-normal-uniform	0.995	3.869
Alpha-skew-normal	0.811	3.749
Epsilon-skew-normal	0.995	3.869
Epsilon-half-normal	1.311	13.077
Extended-skew-normal	1.983	5.607
Normal-skew-normal	1.010	3.956
Skew-flexible-normal	1.995	8.967
Skew-t-normal	2.55	23.108
Skew-flexible-t-normal	4.648	73.799

function for $\boldsymbol{\eta}$ is $\sum_{i=1}^{n} \ell(\boldsymbol{\eta}|X_i)$, where $\ell(\boldsymbol{\eta}|X)$ is the log-likelihood for $\boldsymbol{\eta}$ based on a single observation X from (4.1), that is,

$$\ell(\boldsymbol{\eta}|x) = \log(c(\nu)) + \log(\tfrac{c_{\delta,\nu}}{\sigma}) - \tfrac{\nu+1}{2}\log(w(z)) + \tau(z), \tag{4.3}$$

where $z = \frac{x-\mu}{\sigma}$, $w(z) = 1 + \frac{(|z|+\delta)^2}{\nu}$ and $\tau(z) = \log(\Phi(\lambda z))$. Now, the score function is $\sum_{i=1}^{n} S(\boldsymbol{\eta}|X_i)$, with $S(\boldsymbol{\eta}|X) = \partial\ell(\boldsymbol{\eta}|X)/\partial\boldsymbol{\eta} = (\ell_\mu, \ell_\sigma, \ell_\lambda, \ell_\delta, \ell_\nu)$, where

$$\begin{aligned}
\ell_\mu &= \frac{\nu+1}{\nu\sigma}\frac{z+\delta sign(z)}{w(z)} - \frac{\lambda\phi(\lambda z)}{\sigma\Phi(\lambda z)},\\
\ell_\sigma &= \frac{\nu+1}{\nu\sigma}\frac{z^2+\delta|z|}{w(z)} - \frac{\lambda z\phi(\lambda z)}{\sigma\Phi(\lambda z)} - \frac{1}{\sigma},\\
\ell_\lambda &= z\frac{\phi(\lambda z)}{\Phi(\lambda z)},\\
\ell_\delta &= c_{\delta,\nu} t_\nu(\delta) - \frac{\nu+1}{\nu}\frac{|z|+\delta}{w(z)},\\
\ell_\nu &= H(\nu) - c_{\delta,\nu}\frac{\partial T_\nu(-\delta)}{\partial\nu} + \frac{\nu+1}{2\nu} - \frac{\nu+1}{2\nu w(z)} - \frac{1}{2}\log(w(z)),
\end{aligned} \tag{4.4}$$

where $H(\nu) = \frac{1}{2}\left(\Psi\left(\frac{\nu+1}{2}\right) - \Psi\left(\frac{\nu}{2}\right) - \frac{1}{\nu}\right)$ and Ψ is di-gamma function.

BJPS - Accepted Manuscript

4.1 Fisher information matrix

In this section, we derive the Fisher information matrix (FIM) of the SFTN distribution. Suppose that $\widehat{\boldsymbol{\eta}}$ represents the vector of MLEs for the model (4.1) based on a random sample $\mathbf{X} = (X_1, X_2, ..., X_n)$ with associated log-likelihood (4.3). Let $\mathbf{I}(\boldsymbol{\eta})$ denotes the FIM for $\boldsymbol{\eta}$ based on a single observation X, i.e.,

$$I(\boldsymbol{\eta}) = E\left(-\frac{\partial^2 \ell(\boldsymbol{\eta}|X)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^T}\right).$$

It is well-known that in most setting, a set of non-restrictive regularity assumtions can be identified which will ensure that $\sqrt{n}\left(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}\right)$ is asymtotically multivariate normal with mean $\mathbf{0}$ and covariance matrix $I^{-1}(\boldsymbol{\eta})$. To estimate $\mathbf{I}(\boldsymbol{\eta})$ for the approximation of $I^{-1}(\boldsymbol{\eta})$, it is common to use either the expected FIM or the observed FIM. The expected FIM is defined as

$$I(\widehat{\boldsymbol{\eta}}) = E\left(-\frac{\partial^2 \ell(\boldsymbol{\eta}|X)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^T}\right)\Big| \boldsymbol{\eta} = \widehat{\boldsymbol{\eta}}, \tag{4.5}$$

whereas the observed FIM is defined as

$$\mathcal{I}(\widehat{\boldsymbol{\eta}}, \mathbf{X}) = \sum_{i=1}^{n}\left(-\frac{\partial^2 \ell(\boldsymbol{\eta}|X_i)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^T}\right)\Big| \boldsymbol{\eta} = \widehat{\boldsymbol{\eta}}.$$

Perhaps the primary advantage of using $I(\widehat{\boldsymbol{\eta}})$ over $n^{-1}\mathcal{I}(\widehat{\boldsymbol{\eta}}, \mathbf{X})$ as an estimator of $I(\boldsymbol{\eta})$ is that $I(\widehat{\boldsymbol{\eta}})$ is the MLE of $I(\boldsymbol{\eta})$. Yet in many instances, evaluating the expectation in (4.5) is either unfeasible or impractical, making $n^{-1}\mathcal{I}(\widehat{\boldsymbol{\eta}}, \mathbf{X})$ the estimator of choise.

For computing the expectations of the second derivatives of (4.3) (See Appendix), it suffices to apply the following expressions,

$$\begin{aligned}
&E\left(Z^{2i+1}R(Z)\right) = 0, \qquad E\left(Z^{2i}R(Z)\right) = \tfrac{2c(\nu)}{\sqrt{2\pi}}K_i,\\
&E\left(Z^{i}R(Z)^2\right) = \tfrac{c(\nu)}{\pi}I_i, \qquad E\left(\tfrac{Z}{w(Z)^j}\right) = 2c(\nu)\,J_j,\\
&a_{i,j} = E\left(\tfrac{|Z|^i}{w(Z)^j}\right) = c_{\delta,\nu}\,p_{-j}\frac{\Gamma\left(\frac{\nu}{2}+j\right)\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu+1}{2}+j\right)\Gamma\left(\frac{\nu}{2}\right)}\left(\tfrac{\nu}{\nu+2j}\right)^{\frac{i}{2}} E_{\boldsymbol{\theta}_j}\left(T^i\right),\\
&b_j = E\left(\tfrac{sign(Z)}{w(Z)^j}\right) = c_{\delta,\nu}\frac{\Gamma\left(\frac{\nu}{2}+j\right)\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu+1}{2}+j\right)\Gamma\left(\frac{\nu}{2}\right)}\left(p_{-j}d_0\left(\boldsymbol{\theta}_j\right) - \tfrac{1}{2}\right),
\end{aligned} \tag{4.6}$$

where $c(\nu)$, $\Delta_\nu(.)$, p_i and $d_j(.)$ are defined in Lemma 3.1, $R(z) = \frac{\phi(\lambda z)}{\Phi(\lambda z)}$, $E_{\boldsymbol{\theta}_j}\left(T^i\right)$ is defined in (3.1), $\boldsymbol{\theta}_j = (\lambda_j, \delta_j, \nu_j)$, for $j = 1, 2$, where $\lambda_j = \lambda\sqrt{\frac{\nu}{\nu+2j}}$, $\delta_j = \delta\sqrt{\frac{\nu+2j}{\nu}}$, $\nu_j = \nu + 2j$.

BJPS - Accepted Manuscript

Also we need the partial derivatives $\frac{\partial^2 \log(c_{\delta,\nu})}{\partial\delta\partial\nu}$, $\frac{\partial^2 \log(c_{\delta,\nu})}{\partial\nu^2}$ and the following integrals

$$
\begin{aligned}
I_i &= \tfrac{c_{\delta,\nu}}{2}\int_{-\infty}^{+\infty} u^i \left(1+\tfrac{(|u|+\delta)^2}{\nu}\right)^{-(\nu+1)/2} \tfrac{\exp\left\{-\lambda^2 u^2\right\}}{\Phi(\lambda u)}du \quad ,i=0,1,2,\\
J_j &= \tfrac{c_{\delta,\nu}}{2}\int_{-\infty}^{+\infty} u \left(1+\tfrac{(|u|+\delta)^2}{\nu}\right)^{-(\nu+2j+1)/2} \Phi\left(\lambda u\right) du \quad ,j=1,2,\\
K_i &= \tfrac{c_{\delta,\nu}}{2}\int_{-\infty}^{+\infty} u^{2i} \left(1+\tfrac{(|u|+\delta)^2}{\nu}\right)^{-(\nu+1)/2} \exp\left\{-\lambda^2 u^2/2\right\} du \ ,i=0,1,
\end{aligned}
\tag{4.7}
$$

which must be computed numerically.

After simple algebraic computations, the FIM, for $\nu > 4$, derived as

$$
\mathbf{I}\left(\boldsymbol{\eta}\right) = \begin{pmatrix}
I_{\mu\mu} & I_{\mu\sigma} & I_{\mu\lambda} & I_{\mu\delta} & I_{\mu\nu}\\
 & I_{\sigma\sigma} & I_{\sigma\lambda} & I_{\sigma\delta} & I_{\sigma\nu}\\
 & & I_{\lambda\lambda} & I_{\lambda\delta} & I_{\lambda\nu}\\
 & & & I_{\delta\delta} & I_{\delta\nu}\\
 & & & & I_{\nu\nu}
\end{pmatrix}
\tag{4.8}
$$

where

$$
\begin{aligned}
I_{\mu\mu} &= \frac{c_{\delta,\nu}}{\sigma^2}\{2p_{-2}\frac{\nu+2}{\nu+3} - p_{-1}\} + \frac{\lambda^2 c\left(\nu\right)}{\pi\sigma^2}I_0,\\
I_{\mu\sigma} &= \frac{c\left(\nu\right)}{\sigma^2}\left\{4\frac{\nu+1}{\nu}J_2 - \frac{2\lambda}{\sqrt{2\pi}}K_0 + \frac{2\lambda^3}{\sqrt{2\pi}}K_1 + \frac{\lambda^2}{\pi}I_1\right\} + \frac{\delta\left(\nu+1\right)}{\nu\sigma^2}b_1,\\
I_{\mu\lambda} &= \frac{2c\left(\nu\right)}{\sigma\sqrt{2\pi}}\left\{K_0 - \lambda^2 K_1 - \frac{\lambda}{\sqrt{2\pi}}I_1\right\}, \quad I_{\mu\delta} = \frac{\nu+1}{\nu\sigma}\left\{b_1 - 2b_2\right\},\\
I_{\mu\nu} &= \frac{\delta}{\nu\sigma}\left\{\frac{\nu+1}{\nu}b_2 - b_1\right\} + \frac{2c\left(\nu\right)}{\nu\sigma}\left\{\frac{\nu+1}{\nu}J_2 - J_1\right\},\\
I_{\sigma\sigma} &= \frac{\nu+1}{\nu\sigma^2}\left\{2a_{2,2} - \left(\nu+\delta^2\right)a_{0,1}\right\} + \frac{\nu\pi + \lambda^2 c\left(\nu\right)I_2}{\pi\sigma^2},\\
I_{\sigma\lambda} &= -\frac{\lambda c\left(\nu\right)}{\pi\sigma}I_2, \qquad I_{\sigma\delta} = \frac{\nu+1}{\nu\sigma}\left(a_{1,1} - 2a_{1,2}\right),\\
I_{\sigma\nu} &= \frac{1}{\nu\sigma}\left(\frac{\nu+1}{\nu}a_{2,2} - a_{2,1}\right) + \frac{\delta}{\nu\sigma}\left(\frac{\nu+1}{\nu}a_{1,2} - a_{1,1}\right),\\
I_{\lambda\lambda} &= \frac{c\left(\nu\right)}{\pi}I_2, \qquad I_{\lambda\delta} = 0, \qquad I_{\lambda\nu} = 0,\\
I_{\delta\delta} &= \delta\frac{c\left(\nu\right)\left(\nu+1\right)c_{\delta,\nu}\Delta_\nu\left(-2\right)}{\nu} - \left(c_{\delta,\nu}t_\nu\left(\delta\right)\right)^2 + c_{\delta,\nu}\{2p_{-2}\frac{\nu+2}{\nu+3} - p_{-1}\},\\
I_{\delta\nu} &= -\frac{\partial^2\log(c_{\delta,\nu})}{\partial\delta\partial\nu} + \frac{\delta c_{\delta,\nu}}{\nu}\left(\frac{\nu}{\nu+1}p_{-1} - \frac{\nu+2}{\nu+3}p_{-2}\right) + \frac{1}{\nu}\left(a_{1,1} - \frac{\nu+1}{\nu}a_{1,2}\right),\\
I_{\nu\nu} &= \frac{1-\nu}{2\nu^2} - \frac{\partial^2\log(c_{\delta,\nu})}{\partial\nu^2} - h\left(\nu\right) + \frac{c_{\delta,\nu}}{2\nu}\left(2p_{-1}\frac{\nu}{\nu+1} - p_{-2}\frac{\nu+2}{\nu+3}\right),
\end{aligned}
$$

where $h\left(\nu\right) = \frac{1}{4}\left(\frac{2}{\nu^2} + \Psi'\left(\frac{\nu+1}{2}\right) - \Psi'\left(\frac{\nu}{2}\right)\right)$.

BJPS - Accepted Manuscript

4.1.1 Some special cases

Now for some special submodels from SFTN family, FIMs are derived. In the particular case, when $\delta = 0$ and $\lambda \neq 0$ in (4.1), skew-t-normal model (Gómez et al. (2007)) is obtained. In this case $\boldsymbol{\theta}_i = (\lambda_i, 0, \nu_i)$, the integrals K_i J_i, k_i, defined in (4.7), should be calculated numerically with $\delta = 0$ and expectations $a_{i,j}$ and b_j in (4.6) can be written as:

$$a_{i,j} = E\left(\frac{|Z|^i}{w(Z)^j}\right) = c(\nu)\frac{\Gamma\left(\frac{i+1}{2}\right)\Gamma\left(\frac{\nu-i}{2}+j\right)}{\Gamma\left(\frac{\nu+1}{2}+j\right)}\nu^{\frac{i+1}{2}},$$
$$b_j = E\left(\frac{sign(Z)}{w(Z)^j}\right) = \frac{\Gamma\left(\frac{\nu}{2}+j\right)\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu+1}{2}+j\right)\Gamma\left(\frac{\nu}{2}\right)}\left(d_0\left(\boldsymbol{\theta}_j\right) - 1\right).$$

Thus the elements of the FIM (4.8) reduce to

$$I_{\mu\mu} = \frac{\nu+1}{(\nu+3)\sigma^2} + \frac{\lambda^2 c(\nu)}{\pi\sigma^2} I_0,$$
$$I_{\mu\sigma} = \frac{c(\nu)}{\sigma^2}\left\{4\frac{\nu+1}{\nu}J_2 - \frac{2\lambda}{\sqrt{2\pi}}K_0 + \frac{2\lambda^3}{\sqrt{2\pi}}K_1 + \frac{\lambda^2}{\pi}I_1\right\},$$
$$I_{\mu\lambda} = \frac{2c(\nu)}{\sigma\sqrt{2\pi}}\left\{K_0 - \lambda^2 K_1 - \frac{\lambda}{\sqrt{2\pi}}I_1\right\},$$
$$I_{\mu\delta} = \frac{1}{\sigma}\left\{\left(d_0\left(\boldsymbol{\theta}_1\right) - 1\right) - 2\frac{\nu+2}{\nu+3}\left(d_0\left(\boldsymbol{\theta}_2\right) - 1\right)\right\},$$
$$I_{\mu\nu} = \frac{2c(\nu)}{\nu\sigma}\left\{\frac{\nu+1}{\nu}J_2 - J_1\right\}, \quad I_{\sigma\sigma} = \frac{2\nu}{(\nu+3)\sigma^2} + \frac{\lambda^2 c(\nu)}{\pi\sigma^2}I_2,$$
$$I_{\sigma\lambda} = -\frac{\lambda c(\nu)}{\pi\sigma}I_2, \quad I_{\sigma\delta} = -\frac{2(\nu-1)c(\nu)}{\sigma(\nu+3)}, \quad I_{\sigma\nu} = -\frac{2}{\sigma(\nu+1)(\nu+3)},$$
$$I_{\lambda\lambda} = \frac{c(\nu)}{\pi}I_2, \quad I_{\lambda\delta} = 0, \quad I_{\lambda\nu} = 0, \quad I_{\delta\delta} = \frac{\nu+1}{\nu+3} - (2c(\nu))^2,$$
$$I_{\delta\nu} = \frac{2c(\nu)(\nu-1)}{\nu(\nu+1)(\nu+3)}, \quad I_{\nu\nu} = -\frac{\nu-3}{2\nu^2(\nu+1)(\nu+3)} - h(\nu).$$

In this case the FIM is nonsingular for finite values of ν ($\nu > 4$). Also, when $\lambda = \delta = 0$, the model SFTN reduces to Student-t_ν. In this case $d_0\left(\boldsymbol{\theta}_j\right) = 1$ and the elements of the FIM (4.8) are as follows,

$$I_{\mu\mu} = \frac{\nu+1}{(\nu+3)\sigma^2}, \quad I_{\mu\sigma} = 0, \quad I_{\mu\lambda} = \frac{2}{\sigma\sqrt{2\pi}},$$
$$I_{\mu\delta} = 0, \quad I_{\mu\nu} = 0, \quad I_{\sigma\sigma} = \frac{2\nu}{(\nu+3)\sigma^2}, \quad I_{\sigma\lambda} = 0,$$
$$I_{\sigma\delta} = -\frac{2(\nu-1)c(\nu)}{\sigma(\nu+3)}, \quad I_{\sigma\nu} = -\frac{2}{\sigma(\nu+1)(\nu+3)}, \quad I_{\lambda\lambda} = \frac{2\nu}{\pi(\nu-2)},$$
$$I_{\lambda\delta} = 0, \quad I_{\lambda\nu} = 0, \quad I_{\delta\delta} = \frac{\nu+1}{\nu+3} - (2c(\nu))^2,$$
$$I_{\delta\nu} = \frac{2c(\nu)(\nu-1)}{\nu(\nu+1)(\nu+3)}, \quad I_{\nu\nu} = -\frac{\nu-3}{2\nu^2(\nu+1)(\nu+3)} - h(\nu).$$

In this case, the FIM is also nonsingular for finite ν ($\nu > 4$). Notice that the FIM of the SFN, SN and normal models are derived from the results for models SFTN, submodels STN and Student-t_ν as $\nu \to \infty$, respectively (See Gómez et al. (2011)).

BJPS - Accepted Manuscript

Profile FIM for the model (4.1) with parameters μ, σ, λ and δ with fixed ν reduces to

$$\mathbf{I}\left(\boldsymbol{\eta}|\nu\right) = \begin{pmatrix} I_{\mu\mu} & I_{\mu\sigma} & I_{\mu\lambda} & I_{\mu\delta} \\ & I_{\sigma\sigma} & I_{\sigma\lambda} & I_{\sigma\delta} \\ & & I_{\lambda\lambda} & I_{\lambda\delta} \\ & & & I_{\delta\delta} \end{pmatrix}.$$

Further for the standard model (2.2), profile FIM reduces to

$$\mathbf{I}\left(\boldsymbol{\theta}|\nu\right) = \begin{pmatrix} I_{\lambda\lambda} & 0 \\ 0 & I_{\delta\delta} \end{pmatrix}. \tag{4.9}$$

5 Numerical illustration

Now to illustrate the consistency of the MLEs of the parameters in the models $SFTN$ and $SFCN$, we apply the methodology discussed in Corollary 2.6 to simulate the data from these models. We consider standard cases $SFTN(\lambda,\delta,\nu)$ and $SFCN(\lambda,\delta)$. The MLEs of the parameters for $n = 50, 100, 200, 300, 500, 1000$ simulated data are evaluated by the function "optim" available in software R. For the function "optim", we use the method "L-BFGS-B", which use a limited-memory modification of the quasi-Newton method.

The simulations from the models $SFTN(0.5, -0.5, 4)$ and $SFTN(2, 0, 8)$ are performed 15000 times and the average and the estimated mean square errors (EMSE) are reported in Tables 3 and 4 respectively. Also simulations from the models $SFCN(-0.3, 0.2)$ and $SFCN(1, -1)$ are

Table 3: Average and EMSE (in parentheses) of the estimated parameters in 15000 simulated path from $SFTN(0.5, -0.5, 4)$ model.

n	$\widehat{\lambda}$	$\widehat{\delta}$	$\widehat{\nu}$
50	0.56188(0.06525)	-0.46199(0.06898)	4.04953(0.00340)
100	0.53068(0.02346)	-0.48260(0.02876)	4.04962(0.00292)
200	0.51523(0.00891)	-0.49320(0.01022)	4.04980(0.00268)
300	0.51178(0.00581)	-0.49838(0.00584)	4.04969(0.00256)
500	0.50879(0.00349)	-0.50000(0.00344)	4.04988(0.00256)
1000	0.50319(0.00153)	-0.50693(0.00184)	4.03036(0.00094)

BJPS - Accepted Manuscript

Table 4: Average and EMSE (in parentheses) of the estimated parameters in 15000 simulated path from $SFTN(2,0,8)$ model.

n	$\widehat{\lambda}$	$\widehat{\delta}$	$\widehat{\nu}$
50	2.04410(0.15109)	0.11041(0.21323)	8.0500(0.00375)
100	2.02846(0.09003)	0.07521(0.11683)	8.05005(0.00315)
200	2.01899(0.04529)	0.03474(0.04519)	8.05040(0.00275)
300	2.01028(0.02653)	0.02257(0.02454)	8.05013(0.00265)
500	2.00270(0.01396)	0.01351(0.01205)	8.05021(0.00259)
1000	1.996(0.00575)	0.00811(0.00526)	7.96954(0.00096)

performed and the results are reported in Table 5. Note that in the simulation examples, the EMSE of the estimators becomes smaller when the sample size increases. For large sample size, the EMSE tends to zero and this illustrates the consistency of the estimators.

Table 5: Average and EMSE (in parentheses) of the estimated parameters in 15000 simulated path from $SFCN(-0.3,0.2)$ (with $\widehat{\lambda}_1, \widehat{\delta}_1$) and $SFCN(1,-1)$ (with $\widehat{\lambda}_2, \widehat{\delta}_2$) models.

n	$\widehat{\lambda}_1$	$\widehat{\delta}_1$	$\widehat{\lambda}_2$	$\widehat{\delta}_2$
50	-0.3823(0.0460)	0.2072(0.0124)	1.0666(0.0981)	-0.9729(0.09610)
100	-0.3408(0.0136)	0.2085(0.0065)	1.0295(0.0468)	-0.9928(0.0341)
200	-0.3242(0.0061)	0.2091(0.0030)	1.0106(0.0185)	-0.9982(0.0130)
300	-0.3179(0.0040)	0.2093(0.0019)	1.0059(0.0104)	-0.9973(0.0084)
500	-0.3124(0.0025)	0.2084(0.0011)	1.0015(0.0060)	-0.9986(0.0047)
1000	-0.3040(0.0009)	0.2173(0.0008)	0.9997(0.0028)	-0.9995(0.0023)

Using FIM (4.9), with different sample sizes, we can derive the 95% confidence interval for the shape/skewness parameters λ and δ with fixed $\nu > 4$. We generate the samples of sizes $n = 50, 100, 200, 300, 500, 1000$ from the model $SFTN(\lambda, \delta, \nu)$. Our simulation are done 15000 times and coverage probability (CP) of the 95% confidence intervals and average length (AL) of simulated 95% confidence intervals for the parameters are computed and the results are given in Table 6. Histograms of the standardized MLEs of parameters λ and δ for the simulated samples of

BJPS - Accepted Manuscript

size $n = 1000$ are shown in Figure 2. This figure shows the asymptotic normality of the distribution of MLEs.

Table 6: CP and AL (in parentheses) of parameters λ and δ in 15000 simulated path from $SFTN(\lambda, \delta, \nu)$.

Model	SFTN(-1,0.2,8)		SFTN(1,-1,6)	
n	$\lambda = -1$	$\delta = 0.2$	$\lambda = 1$	$\delta = -1$
50	0.9604(1.19)	0.9029(1.32)	0.9673(1.06)	0.9531(0.77)
100	0.9548(0.81)	0.9283(0.95)	0.9564(0.69)	0.9523(0.54)
200	0.9522(0.56)	0.9378(0.67)	0.9529(0.48)	0.9514(0.38)
300	0.9527(0.45)	0.9431(0.55)	0.9521(0.39)	0.9486(0.31)
500	0.9497(0.35)	0.9469(0.43)	0.9471(0.29)	0.9502(0.24)
1000	0.9514(0.24)	0.9472(0.30)	0.9538(0.21)	0.9517(0.17)

6 Illustrations with real data sets

To illustrate the applicability of the proposed models, we analyze three real data sets available from different sources. To compare the fitting of various models, we use the Akaike (AIC) information criteria and the Bayesian information criteria (BIC) which defined as $AIC = 2m - 2\ell\left(\widehat{\boldsymbol{\theta}}|\mathbf{X}\right)$ and $BIC = m\, Ln(n) - 2\ell\left(\widehat{\boldsymbol{\theta}}|\mathbf{X}\right)$ where $\ell\left(\widehat{\boldsymbol{\theta}}|\mathbf{X}\right)$ is the maximized log-likelihood, n is the sample size and m is the number of the model parameters. In this section, for fitting SFTN distribution with pdf introduced in (4.1) to a set of data, the parameters $\boldsymbol{\theta} = (\mu, \sigma, \lambda, \delta, \nu)$ are estimated by maximizing $\ell(\boldsymbol{\theta}|\mathbf{X})$, using the method of "Profile maximum likelihood ". That is for some fixed values of ν $(\nu = 1, 2, 3, ...)$, the likelihood function is maximized with respect to the other parameters. In Examples 1-2, using the profile method, the parameter ν is estimated as 6 and 5, respectively, and in the last example, the model SFCN with pdf (4.2) is fitted. The observed standard errors (SE) of the estimates $\widehat{\boldsymbol{\theta}}$ are extracted from the square root of the diagonal elements of the inverse of the observed FIM. For the 4 distributions skew-normal (SN), skew-flexible-normal (SFN), skew-t-normal (STN) and SFTN, in each example, the related table lists the MLEs for parameters with their SEs (in parenthesis) and the information criteria AIC and BIC. Also, in the examples, we use

BJPS - Accepted Manuscript

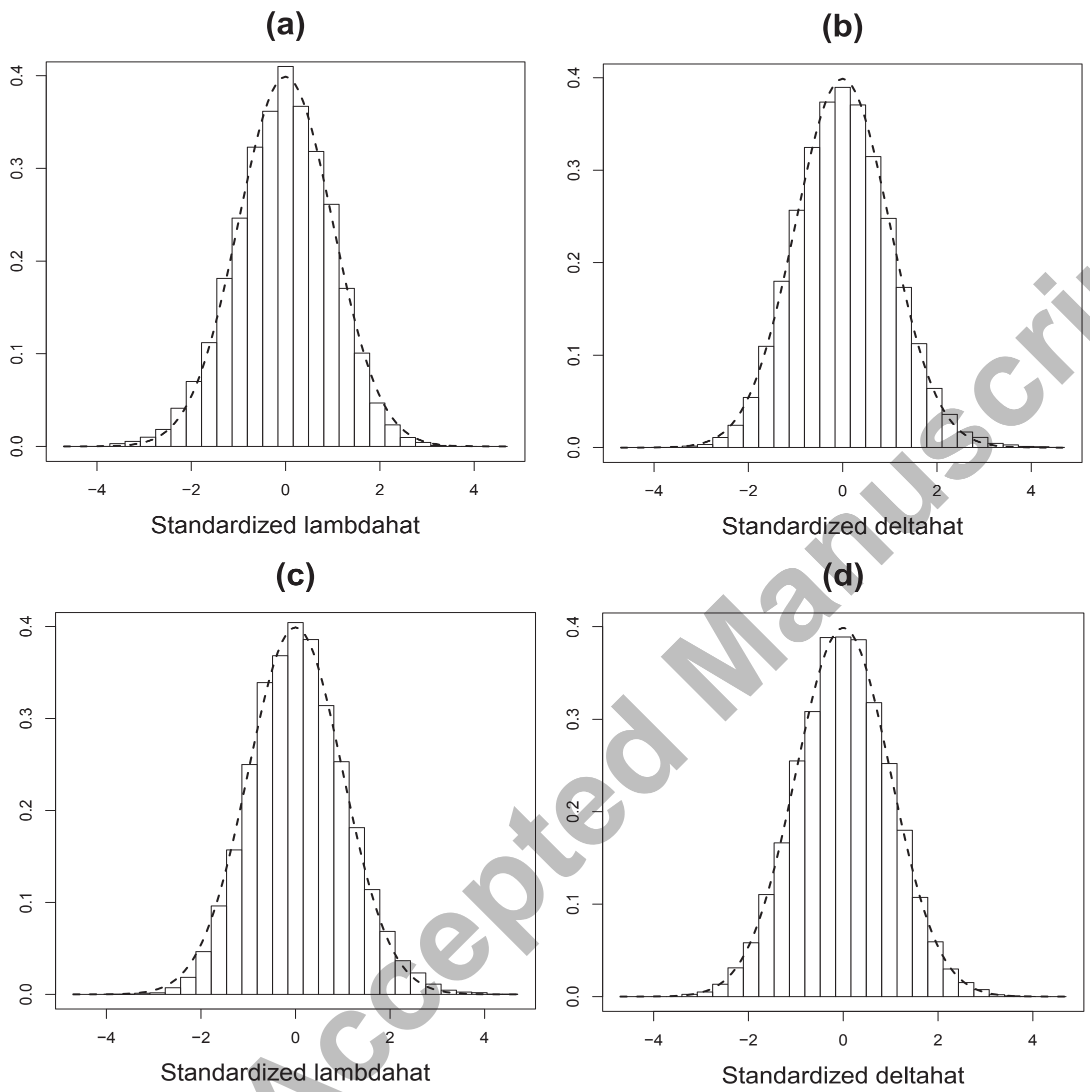

Figure 2: Histograms of standardized $\widehat{\lambda}$ (left column) and standardized $\widehat{\delta}$ (right column)for the models: $SFTN(-1, 0.2, 8)$(first row) and $SFTN(1, -1, 6)$(second row) in 15000 simulated samples of size $n = 1000$ with the standard normal pdf plot (dashed line).

BJPS - Accepted Manuscript

the Kolmogorov-Smirnov (K-S) and Anderson-Darling (A-D) (Anderson and Darling (1954)) tests of goodness-of-fit of the proposed model. Furthermore, the likelihood ratio test (LRT) statistic, which is a comparison of likelihood scores between two competitive models, is used to judge which of the two models is more appropriate for this data set. For testing the null hypothesis H_{0i} versus the alternative hypothesis H_{1i} , for $i = 1, 2, 3$, in the following cases

$$H_{01} : \delta = 0, \upsilon = +\infty \text{ (SN) versus } H_{11} : \delta \neq 0, \upsilon < +\infty \text{ (SFTN)},$$
$$H_{02} : \upsilon = +\infty \text{ (SFN) versus } H_{12} : \upsilon < +\infty \text{ (SFTN)},$$
$$H_{03} : \delta = 0 \text{ (STN) versus } H_{13} : \delta \neq 0 \text{ (SFTN)},$$

the LRT statistics are as $\Lambda_i = -2(\ell_{0i} - \ell_1)$, where ℓ_{0i} are the maximaized log-likelihood value for the model under the the null hypothesis H_{0i} and ℓ_1 is the maximaized log-likelihood value for the SFTN model. For the enough large sample size n, at the significant level of 0.05, H_{0i} is rejected if $\Lambda_i > \chi^2_{df_i,0.05}$, for $df_1 = 2$ and $df_2 = df_3 = 1$.

Example 1. Strength of glass fibres data set

Smith and Naylor (1987) presented an experimental data set on the strength of 63 glass fibres of length 1.5 cm. This data set has been considered by several authors in the literature. The results for the mentioned models are presented in Table 7. Graphical results are shown in Figure 3. By assuming the SFTN distribution for the strength of glass fibers, the K-S statistic and the A-D statistic obtained. The corresponding p-values are 0.93 and 0.98, respectively. For testing the null hypothesis H_{0i} versus the alternative hypothesis H_{1i}, for $i = 1, 2, 3$, the LRT statistic gives the values $\Lambda_1 = 10.27$, $\Lambda_2 = 3.86$ and $\Lambda_3 = 5.93$ which are significant, indicating that the null hypothesises are not acceptable for the Strength of glass fibres data.

Table 7: MLEs with SEs and Information Criteria for the Strength of glass fibres in Example 1.

Distributions	$\widehat{\mu}(SE)$	$\widehat{\sigma}(SE)$	$\widehat{\lambda}(SE)$	$\widehat{\delta}(SE)$	AIC	BIC
SN	1.85(0.05)	0.47(0.06)	-2.68(0.80)	–	33.91	40.34
SFN	1.09(0.05)	0.23(0.02)	0.72(0.17)	-1.99(0.31)	29.50	38.07
STN($\widehat{\nu}$ =1.96)	1.65(0.06)	0.18(0.03)	-0.36(0.27)	–	31.57	40.14
SFTN (ν =6)	1.16(0.04)	0.19(0.02)	0.53(0.13)	-2.20(0.32)	27.64	36.21

Example 2. Fibers data set

To illustrate more, we used a set of data that were part of an extensive study on the association of plasma retinol and beta-carotene levels with the risk of developing certain types of cancer (see "http ://lib.stat.cmu.edu/datasets/Plasma_Retinol"). The data consist of 315 observations

BJPS - Accepted Manuscript

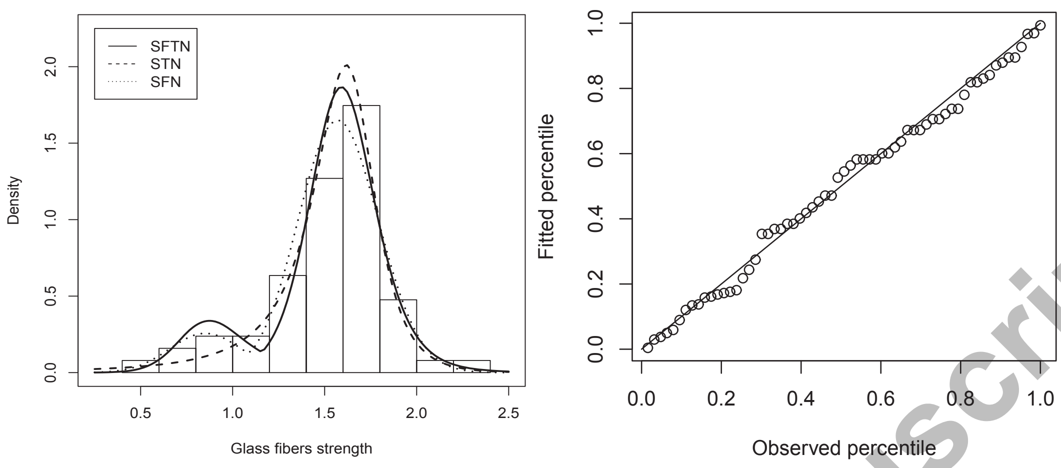

Figure 3: Histogram of the strength of glass fibres and fitted STN, SFN and SFTN models on them (left figure) and PP-plot based on the fitted SFTN model (right figure).

of grams of fibre consumed per day taken from patients who had an elective surgical procedure during a 3-year period to biopsy or removal of a lesion of the lung, colon, breast, skin, ovary or uterus that was found to be non-cancerous. By assuming the above mentioned distributions for the random variable of grams of fibre consumed per day, the MLEs of the parameters are obtained and the results are made out in Table 8 and Figure 4. By assuming the SFTN distribution for the Grams of fibre observations, the K-S and the A-D tests of goodness-of-fit have the p-values 0.91 and 0.99, respectively. For testing H_{0i} versus the H_{1i}, for $i = 1, 2, 3$, we obtain the LRT statistic values $\Lambda_1 = 12.69$, $\Lambda_2 = 12.29$ and $\Lambda_3 = 9.26$ which are significant, indicating that the null hypothesises are not acceptable for the Grams of fiber data.

Table 8: MLEs with SEs and Information Criteria for the Fibers data in Example 2.

Distributions	$\widehat{\mu}(SE)$	$\widehat{\sigma}(SE)$	$\widehat{\lambda}(SE)$	$\widehat{\delta}(SE)$	AIC	BIC
SN	6.25(0.43)	8.43(0.47)	4.54(1.14)	–	1902.85	1914.11
SFN	6.70(0.85)	9.29(1.58)	4.25(1.21)	0.35(0.64)	1904.45	1919.46
STN($\widehat{\nu} = 10.6$)	6.89(0.42)	7.15(0.47)	3.07(0.81)	–	1901.42	1916.43
SFTN($\nu = 5$)	4.41(0.52)	5.19(0.38)	6.83(2.51)	-1.32(0.24)	1894.16	1909.17

Example 3. Nickel concentration data set

The data set is related to nickel concentration in 86 soil samples analyzed at the Mining

BJPS - Accepted Manuscript

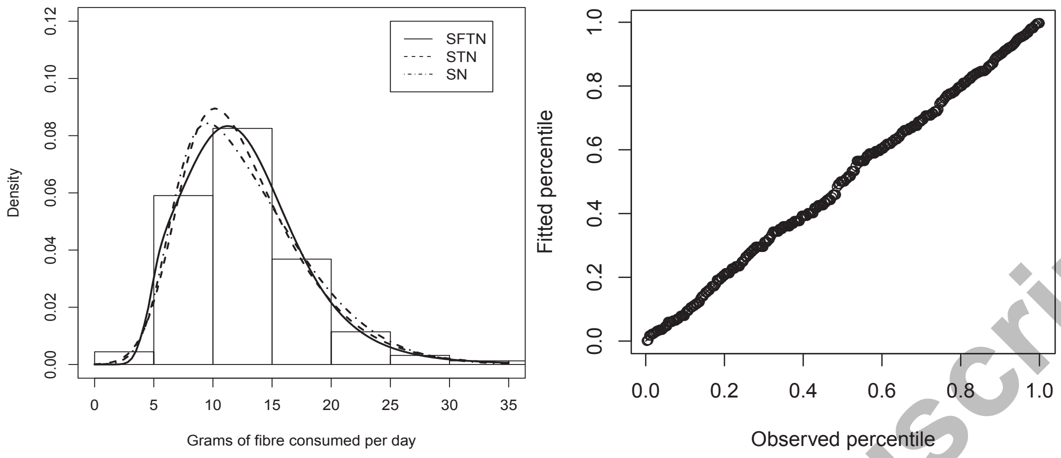

Figure 4: Histogram of the Grams of fibre and fitted SN, STN and SFTN models on them (left figure) and PP-plot based on the fitted SFTN model (right figure).

Department in University of Atacama-Chile. Table 9 shows the MLEs of the parameters of SN, SFN and STN distributions and our proposed SFCN distribution. Graphical fitness of the models are shown in Figure 5. By assuming the SFCN distribution for the nickel concentration, the K-S and the A-D tests of goodness-of-fit have the p-values 0.90 and 0.91, respectively. For testing H_{0i} versus the H_{1i}, for $i = 1, 2, 3$, we obtain the LRT statistic values $\Lambda_1 = 19.97$, $\Lambda_2 = 8.59$ and $\Lambda_3 = 4.86$ which are significant, indicating that the H_{0i}, $i = 1, 2, 3$, are not acceptable for the Nickel concentration data.

Table 9: MLEs with SEs and Information Criteria for the Nickel data in Example 3.

Distributions	$\widehat{\mu}(SE)$	$\widehat{\sigma}(SE)$	$\widehat{\lambda}(SE)$	$\widehat{\delta}(SE)$	AIC	BIC
SN	2.62(2.07)	24.97(2.46)	10.21(9.53)	–	695.52	702.89
SFN	7.18(2.35)	150(246)	27.23(44.16)	9.84(17.07)	686.14	695.96
STN($\widehat{\nu}$ =2.42)	8.47(2.59)	11.59(2.27)	1.81(1.24)	–	682.41	692.23
SFCN	9.32(1.43)	5.74(1.04)	0.80(0.31)	-1.01(0.30)	679.55	689.37

BJPS - Accepted Manuscript

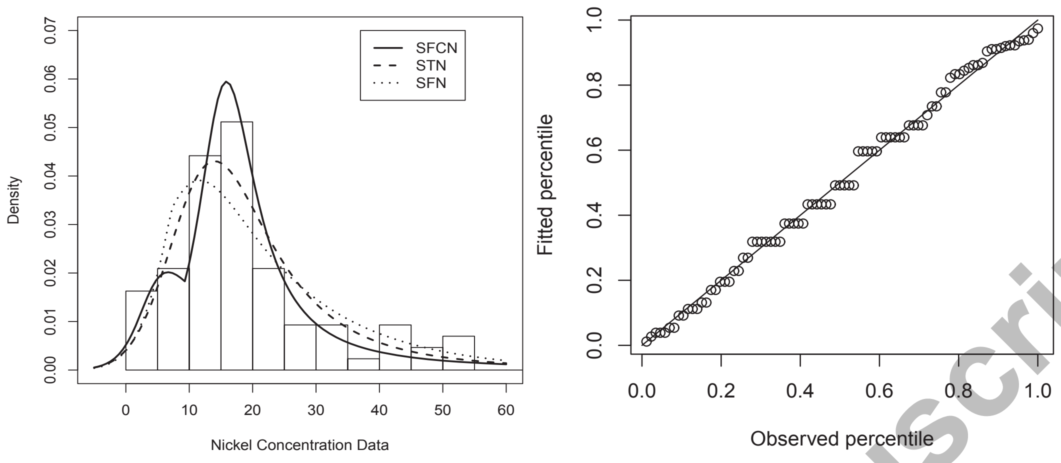

Figure 5: Histogram of the Nickel data set and fitted SFCN, STN and SFN models on them (left figure) and PP-plot based on the fitted SFCN model(right figure).

Acknowledgments

We are grateful an associate editor and anonymous referees for their valuable comments and suggestions on earlier version of this article.

References

Ali, M., M., Woo, J. and Nadarajah, S. (2010). Some skew symmetric inverse reflected distributions. Brazilian Journal of Probability and Statistics, **24**(1), 1–23.

Anderson, T.W. and Darling, D.A. (1954). A test of goodness-of-fit. Journal of the American Statistical Association, **49**(628), 765–769.

Arellano-Valle R. B., Cortés M.A. and Gómez H.W. (2010). An extension of the epsilon-skew-normal distribution. Communications in Statistics-Theory and Methods, **39**, 912-922.

Arellano-Valle, R.B., Gómez, H.W. and Quintana, F.A. (2004). A new class of skew-normal distributions. Communications in statistics-Theory and Methods, **33**, 1465–1480.

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

BJPS - Accepted Manuscript

Arnold, B.C. and Beaver, R.J. (2002). Skew multivariate models related to hidden truncation and/or selective reporting. Test, **11** ,7-54.

Arnold, B.C., Castillo, E. and Sarabia, J.M. (2002). Conditionally specified multivariate skewed distributions. Sankhya, **64**(Serie A), 206-226.

Arslan, O. and Genc, A.I. (2009). The skew generalized t distribution as the scale mixture of a skew exponential power distribution and its applications in robust estimation. Statistics, **43**, 481-498.

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

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

Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution. Journal of the Royal Statistical Society, Series B, **65**, 367-389.

Balakrishnan, N. (2002). Discussion of "Skewed multivariate models related to hidden truncation and/or selective reporting". Test, **11**, 37–39.

Castro, L. M., Gómez, H. W. and Valenzuela, M. (2012). Epsilon half-normal model: Properties and inference. Computational Statistics and Data Analysis, **56(12)**, 4338-4347.

Cavanaugh, J. E. and Shumway, R. H. (1996). On computing the expected Fisher information matrix for state-space model parameters. Statistics and probability letters, **26**(4), 347–355.

Edgeworth, F.Y. (1886). The law of error and the elimination of chance. Philosophical Magazine, **21**, 308–324.

Elal-Olivero, D. (2010). Alpha-skew-normal distribution. Proyecciones (Antofagasta), 29(3), 224–240.

Gómez, H. W., Venegas, O. and Bolfarine, H. (2007). Skew-symmetric distributions generated by the distribution function of the normal distribution. Environmetrics, **18** , 395-407.

BJPS - Accepted Manuscript

Gómez, H. W., Elal-Olivero, D., Salinas, H.S. and Bolfarine, H. (2011). Bimodal extension based on the skew-normal distribution with application to pollen data. Environmetrics, **22**, 50-62.

Gómez, H. W., Varela, V. and Vidal, I. (2013). A new class of skew-symmetric distributions and related families. *Statistics*, **47(2)**, 411-421.

Gupta, A.K., Chang, F.C., and Huang, W.J. (2002). Some skew-symmetric model. Randon Operators and Stochastic Equations, **10**, 113-140.

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

Ho, H. J. , Tsung, I. L., YuChen, H. and Wang, W. L. (2012). Some results on the truncated multivariate t distribution. Journal of Statistical Planning and Inference, **142**, 25-40.

Mahmoudi, E. and Sepahdar, A. (2013). Exponentiated Weibull-Poisson distribution: Model, properties and applications. Mathematics and Computers in Simulation, **92**, 76-97.

Mudholkar, G. S. and Hutson A. B. (2000). The epsilon-skew-normal distribution for analyzing near-normal data. Journal of Statistical Planning and Inference, **83**, 291-309.

Kim, H.J. (2005). On a class of two-piece skew-normal distributions. Statistics, **39(6)**, 537-553.

Ma, Y. and Genton, M. G. (2004). A flexible class of skew-symmetric distributions. Scandinavian Journal of Statistics, **31**, 459-468.

Nadarajah S. and Kotz, S. (2003). Skewed distributions generated by the normal kernel. Statistics & Probability Letters, **65**, 269-277.

Pearson, K. (1893). Asymmetrical frequency curves. Nature, **427**, 299–300.

Pewsey, A. (2000). Problems of inference for Azzalini's skew-normal distribution. Journal of Applied Statistics, **27(7)**, 859-870.

Rosco, J. F., Jones, M. C. and Pewsey, A. (2011). Skew t distributions via the sinh-arcsinh transformation. Test, **20**, 630-652.

Shafiei, S. and Doostparast, M. (2014). Balakrishnan skew-t distribution and associated statistical characteristics. Communication in Statistics-theory and Methods, **43(19)**, 4109-4122.

BJPS - Accepted Manuscript

Smith, R. L. and Naylor, J. C. (1987). A comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution. Journal of the Royal Statistical Society, Series C, **36**, 358-369.

Appendix

The second partial derivatives of (4.3) are as following expressions:

$$
\begin{aligned}
\ell_{\mu\mu} &= \frac{\nu+1}{\nu\sigma^2}\left(\frac{1}{w(z)}-\frac{2}{w(z)^2}\right)+\tau_{\mu\mu},\\
\ell_{\mu\sigma} &= -\frac{\nu+1}{\nu\sigma^2}\left(\frac{2z}{w(z)^2}+\frac{\delta sign(z)}{w(z)}\right)+\tau_{\mu\sigma},\\
\ell_{\mu\lambda} &= \tau_{\mu\lambda},\quad \ell_{\mu\delta}=\frac{\nu+1}{\nu\sigma}sign(z)\left(\frac{2}{w(z)^2}-\frac{1}{w(z)}\right),\\
\ell_{\mu\nu} &= \frac{1}{\nu\sigma}sign(z)(|z|+\delta)\left(\frac{1}{w(z)}-\frac{\nu+1}{\nu w(z)^2}\right),\\
\ell_{\sigma\sigma} &= -\frac{\nu}{\sigma^2}-\frac{\nu+1}{\nu\sigma^2}\left(2\frac{z^2}{w(z)^2}-\frac{\nu+\delta^2}{w(z)}\right)+\tau_{\sigma\sigma},\\
\ell_{\sigma\lambda} &= \tau_{\sigma\lambda},\qquad \ell_{\sigma\delta}=\frac{\nu+1}{\nu\sigma}|z|\left(\frac{2}{w(z)^2}-\frac{1}{w(z)}\right),\\
\ell_{\sigma\nu} &= \frac{1}{\nu\sigma}|z|\left(|z|+\delta\right)\left(\frac{1}{w(z)}-\frac{\nu+1}{\nu w(z)^2}\right),\\
\ell_{\lambda\lambda} &= \tau_{\lambda\lambda},\qquad \ell_{\lambda\delta}=0,\qquad \ell_{\lambda v}=0,\\
\ell_{\delta\delta} &= \left(c_{\delta,\nu}t_\nu\left(\delta\right)\right)^2-\frac{\nu+1}{\nu}c\left(\nu\right)\delta c_{\delta,\nu}\Delta_\nu\left(-2\right)-\frac{\nu+1}{\nu}\left(\frac{2}{w(z)^2}-\frac{1}{w(z)}\right),\\
\ell_{\delta\nu} &= \frac{\partial^2\log\left(c_{\delta,\nu}\right)}{\partial\delta\partial\nu}-\frac{|z|+\delta}{\nu}\left(\frac{1}{w(z)}-\frac{\nu+1}{\nu w(z)^2}\right),\\
\ell_{\nu\nu} &= \frac{\partial^2\log\left(c_{\delta,\nu}\right)}{\partial\nu^2}+h\left(\nu\right)+\frac{\nu-1}{\nu^2}-\frac{1}{\nu w(z)}+\frac{\nu+1}{2\nu^2 w(z)^2},
\end{aligned}
$$

where $R(z)=\frac{\phi(\lambda z)}{\Phi(\lambda z)}$, $h\left(\nu\right)=\frac{1}{4}\left(\frac{2}{\nu^2}+\Psi'\left(\frac{\nu+1}{2}\right)-\Psi'\left(\frac{\nu}{2}\right)\right)$ and

$$
\begin{aligned}
\tau_{\mu\mu} &= -\frac{\lambda^3 zR(z)+\lambda^2R(z)^2}{\sigma^2},\\
\tau_{\mu\sigma} &= \frac{\lambda R(z)-\lambda^3z^2R(z)-\lambda^2zR(z)^2}{\sigma^2},\\
\tau_{\mu\lambda} &= \frac{-R(z)+\lambda zR(z)^2+\lambda^2z^2R(z)}{\sigma},\\
\tau_{\sigma\sigma} &= \frac{2\lambda zR(z)-\lambda^3z^3R(z)-\lambda^2z^2R(z)^2}{\sigma^2},\\
\tau_{\sigma\lambda} &= \frac{\lambda^2z^3R(z)+\lambda z^2R(z)^2-zR(z)}{\sigma},\\
\tau_{\lambda\lambda} &= -\lambda z^3R(z)-z^2R(z)^2.
\end{aligned}
$$

BJPS - Accepted Manuscript