Supplementary Material for

"Shape mixtures of skew-t-normal distributions: characterizations and estimation"

Mostafa Tamandi[a], Ahad Jamalizadeh[a], Tsung-I Lin[b,c]

[a]*Department of Statistics, Faculty of Mathematics & Computer, Shahid Bahonar University of Kerman, Kerman, Iran*

[b]*Institute of Statistics, National Chung Hsing University, Taichung 402, Taiwan*

[c]*Department of Public Health, China Medical University, Taichung 404, Taiwan*

This supporting information is a longer version of the printed paper. It contains the proof of Eq. (18), some details associated with the ECME estimating equations together with a small simulation study to demonstrate the adaptability and flexibility when using the SMSTN distribution.

A. Proof of Eq. (18)

From (10), we have

$$
\begin{aligned}
f(y_j, x_j, \gamma_j, \tau_j) &= f(y_j \mid x_j, \gamma_j, \tau_j) f(\gamma_j \mid x_j, \tau_j) f(x_j) f(\tau_j) \\
&= \frac{\sqrt{\tau_j + x_j^2}}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{\tau_j + x_j^2}{2\sigma^2} \left(y_j - \xi - \frac{\sigma x_j}{\tau_j + x_j^2} \gamma_j \right)^2 \right\} \\
&\quad \times \sqrt{\frac{2}{\pi}} \frac{\tau_j^{1/2}}{\sqrt{\tau_j + x_j^2}} \exp\left\{ -\frac{\tau_j \gamma_j^2}{2(\tau_j + x_j^2)} \right\} \frac{1}{\sqrt{2\pi\alpha}} \exp\left\{ -\frac{(x_j - \lambda)^2}{2\alpha} \right\} \\
&\quad \times \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)} \tau_j^{\frac{\nu}{2}-1} e^{-\frac{\nu}{2}\tau_j}.
\end{aligned}
$$

where $u_j = (y_j - \xi)/\sigma$. After some algebraic operations, the complete data log-likelihood function, excluding additive constants not related to parameters, is given by

$$\begin{aligned}
\ell_c(\boldsymbol{\theta} \mid \boldsymbol{y}_c) &= \sum_{j=1}^{n} \log f(y_j, x_j, \gamma_j, \tau_j) \\
&= -\frac{n}{2}(\log \sigma^2 + \log \alpha) - \frac{1}{2}\sum_{j=1}^{n} \tau_j u_j^2 - \frac{1}{2\alpha}\sum_{j=1}^{n} (x_j - \lambda)^2 - \frac{1}{2}\sum_{j=1}^{n} (\gamma_j - u_j x_j)^2 \\
&\quad + \frac{n\nu}{2}\log\left(\frac{\nu}{2}\right) - n\log\Gamma\left(\frac{\nu}{2}\right) + \frac{\nu}{2}\sum_{j=1}^{n}(\log \tau_j - \tau_j).
\end{aligned}$$

This completes the proof. □

B. Some details associated with the ECME estimating equations

(a) Differentiating the Q-function with respect to ξ leads to

$$\frac{\partial}{\partial \xi} Q(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}^{(k)}) = \frac{1}{\sigma^2}\sum_{j=1}^{n}(y_j - \xi)(\hat{\tau}_j^{(k)} + \hat{s}_{2j}^{(k)}) - \frac{1}{\sigma}\sum_{j=1}^{n} \hat{s}_{3j}^{(k)}. \tag{S.1}$$

On equating (S.1) to zero and substituting σ with the current estimate $\hat{\sigma}^{(k+1)}$, this gives

$$\hat{\xi}^{(k+1)} = \frac{\sum_{j=1}^{n} y_j(\hat{\tau}_j^{(k)} + \hat{s}_{2j}^{(k)}) - \hat{\sigma}^{(k)} \sum_{j=1}^{n} \hat{s}_{3j}^{(k)}}{\sum_{j=1}^{n}(\hat{\tau}_j^{(k)} + \hat{s}_{2j}^{(k)})}.$$

(b) Differentiating the Q-function with respect to σ leads to

$$\frac{\partial}{\partial \sigma} Q(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}^{(k)}) = -\frac{n}{\sigma} - \frac{1}{\sigma^2}\sum_{j=1}^{n}(y_j - \xi)\hat{s}_{3j}^{(k)} + \frac{1}{\sigma^3}\sum_{j=1}^{n}(y_j - \xi)^2(\hat{\tau}_j^{(k)} + \hat{s}_{2j}^{(k)}). \tag{S.2}$$

Fixing $\xi = \hat{\xi}^{(k+1)}$ and equating (S.2) to zero gives

$$\sigma^2 + b\sigma - c = 0,$$

where $b = n^{-1}\sum_{j=1}^{n}(y_j - \hat{\xi}^{(k+1)})\hat{s}_{3j}^{(k)}$ and $c = n^{-1}\sum_{j=1}^{n}(y_j - \hat{\xi}^{(k+1)})^2(\hat{\tau}_j^{(k)} + \hat{s}_{2j}^{(k)})$. Therefore, we can get

$$\hat{\sigma}^{(k+1)} = \frac{\sqrt{b^2 + 4c} - b}{2}.$$

Note that the above solution always remains positive.

(c) Differentiating the Q-function with respect to λ leads to

$$\frac{\partial}{\partial \lambda} Q(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}^{(k)}) = \frac{1}{\alpha} \sum_{j=1}^{n} \hat{s}_{1j}^{(k)} - \frac{n}{\alpha} \lambda. \tag{S.3}$$

On equating (S.3) to zero gives

$$\hat{\lambda}^{(k+1)} = \frac{1}{n} \sum_{j=1}^{n} \hat{s}_{1j}^{(k)}.$$

(d) Differentiating the Q-function with respect to α leads to

$$\frac{\partial}{\partial \alpha} Q(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}^{(k)}) = -\frac{n}{2\alpha} - \frac{\lambda}{\alpha^2} \sum_{j=1}^{n} \hat{s}_{1j}^{(k)} + \frac{1}{2\alpha^2} \sum_{j=1}^{n} \hat{s}_{2j}^{(k)} + \frac{n\lambda^2}{2\alpha^2}. \tag{S.4}$$

Fixing $\lambda = \hat{\lambda}^{(k+1)}$ and equating (S.4) to zero gives

$$n\alpha + 2\hat{\lambda}^{(k+1)} \sum_{j=1}^{n} \hat{s}_{1j}^{(k)} - \sum_{j=1}^{n} \hat{s}_{2j}^{(k)} - n\hat{\lambda}^{2(k+1)} = 0. \tag{S.5}$$

Substituting $\sum_{j=1}^{n} \hat{s}_{1j}^{(k)} = n\hat{\lambda}^{(k+1)}$ into (S.5) yields the desired result

$$\hat{\alpha}^{(k+1)} = \frac{1}{n} \sum_{j=1}^{n} \hat{s}_{2j}^{(k)} - \hat{\lambda}^{2(k+1)}.$$

(e) Differentiating the Q-function with respect to ν leads to

$$\frac{\partial}{\partial \nu} Q(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}^{(k)}) = \frac{n}{2} \log\left(\frac{\nu}{2}\right) + \frac{n\nu}{2}\frac{1}{\nu} - \frac{n}{2}\frac{\Gamma'(\nu/2)}{\Gamma(\nu/2)} + \frac{1}{2} \sum_{j=1}^{n} (\hat{\kappa}_j^{(k)} - \hat{\tau}_j^{(k)}).$$

Then, $\hat{\nu}^{(k+1)}$ can be obtained as the solution of

$$\log\left(\frac{\nu}{2}\right) + 1 - \mathrm{DG}\left(\frac{\nu}{2}\right) + \frac{1}{n} \sum_{j=1}^{n} (\hat{\kappa}_j^{(k)} - \hat{\tau}_j^{(k)}) = 0,$$

where $\mathrm{DG}(x) = d \log \Gamma(x)/dx$ is the digamma function.

C. An illustration with simulated data

In this experiment, 300 samples of size $n = 500$ are generated from the SMSTN distribution with $\xi = 1$, $\sigma = 2$, $\lambda = 3$, $\alpha = 10$ and $\nu = 4$. Then, we fit each simulated dataset with the

Table S.1: Simulation results, based on 300 replications, for comparing the performance of four skew models.

Criterion		SN	ST	STN	SMSTN
$\ell_{\max}$	Mean	−1155.52	−1102.00	−1105.18	−1097.41
	Std	37.73	23.97	24.61	24.09
	Freq	0	5	0	295
AIC	Mean	2317.04	2212.00	2218.36	2204.82
	Std	75.46	47.94	49.22	48.18
	Freq	0	25	1	274
BIC	Mean	2329.69	2228.86	2235.22	2225.90
	Std	75.46	47.94	49.22	48.18
	Freq	0	109	1	190

SN, ST, STN and SMSTN distributions. Table S.1 summarizes the average log-likelihood maxima, AIC and BIC values and their standard deviations together with the frequencies (Freq) supported by criteria as a guide to select the most plausible model. Observing the table, all criteria tend to select the SMSTN model, indicating that it cannot be replaced by other competitive models. Figure S.1 displays 300 empirical fitted density curves. It is clearly seen that only the SMSTN model adapt the true underlying distribution well. This gives further evidence that the SMSTN distribution can be taken a prominent alternative to several other skew distributions as it is more capable of capturing distinct non-normal features.

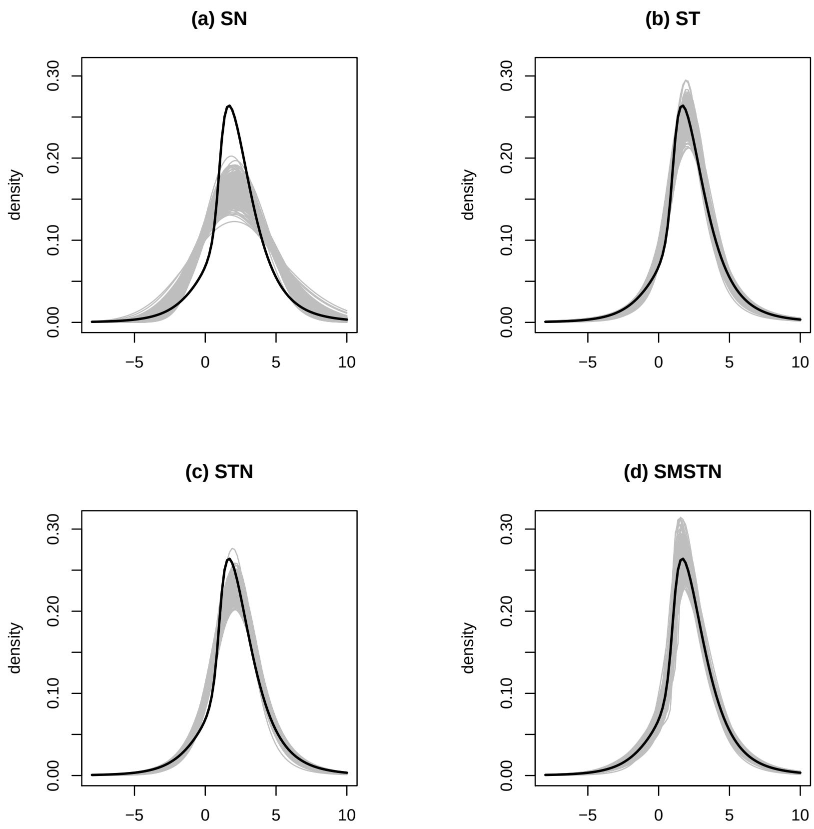

Figure S.1: The true density of the SMSTN distribution (solid line) and 300 estimated densities (grey lines).