Heteroscedastic and Heavy-tailed Regression with Mixtures of Skew Laplace Normal Distributions

Fatma Zehra Doğru[1*], Keming Yu[2] and Olcay Arslan[3]

[1]*Department of Statistics, Faculty of Arts and Science, Giresun University, 28200 Giresun, Turkey*
fatma.dogru@giresun.edu.tr
[2] *Department of Mathematics, College of Engineering Design and Physical Sciences, Brunel University, Uxbridge, Middlesex, UB8 3PH London, UK*
keming.yu@brunel.ac.uk
[3]*Department of Statistics, Faculty of Science, Ankara University, 06100 Ankara, Turkey*
oarslan@ankara.edu.tr

Abstract

Joint modelling skewness and heterogeneity is challenging in data analysis, particularly in regression analysis which allows a random probability distribution to change flexibly with covariates. This paper, based on a skew Laplace normal (SLN) mixture of location, scale, and skewness, introduces a new regression model which provides a flexible modelling of location, scale and skewness parameters simultaneously. The maximum likelihood (ML) estimators of all parameters of the proposed model via the expectation-maximization (EM) algorithm as well as their asymptotic properties are derived. Numerical analyses via a simulation study and a real data example are used to illustrate the performance of the proposed model.

Keywords: EM algorithm, joint location, scale and skewness models, mixture model, ML estimation, SLN, SN.

1. Introduction

Joint mean and dispersion models have been widely used for modelling heteroscedastic data sets in a homogenous population for many years. For example, there have been a number of studies concentrating on joint mean and dispersion models: Park (1966) introduced a log linear model for the variance parameter and described the Gaussian model using a two stage process to estimate the parameters; Harvey (1976) proposed a likelihood ratio test for heteroscedasticity and investigated the maximum likelihood (ML) estimation of the location and scale effects; modelling of variance heterogeneity in normal regression analysis was offered by Aitkin (1987); Verbyla (1993) estimated the parameters of the normal regression model under the log linear dependence of the variances on explanatory variables via the restricted ML; Engel and Huele (1996) examined an extension of the response surface approach to Taguchi type experiments for robust design by accommodating generalized linear modeling; Taylor and Verbyla (2004) proposed the joint modelling of location and scale parameters of the *t* distribution; Lin and Wang (2009) introduced a robust approach for the joint modelling of mean and scale parameters for longitudinal data; Bayesian inference for the joint modelling of location and scale parameters of the *t* distribution for longitudinal data was investigated by Lin and Wang (2011); Wu and Li (2012) studied the variable selection for joint mean and dispersion models of the inverse Gaussian distribution; Wu et al. (2012) examined the variable selection in joint mean and variance models of Box-Cox transformation; Wu et al. (2013) proposed to use the skew normal (SN) (Azzalini (1985, 1986)) distribution for variable selection in the joint location and scale models; Li and Wu (2014) presented the joint modelling of location and scale parameters of the SN distribution; Wu (2014) proposed variable selection in the joint location and scale models using the skew student-*t*-normal (STN) distribution; and Zhao and Zhang (2015) studied variable selection of varying dispersion student-*t* regression models. Recently, joint location, scale and skewness models are started to use modelling heteroscedastic and skew data sets in a homogenous population as well as joint location and scale models. For instance, Li et al. (2017) explored variable selection in the joint location, scale and skewness models of the SN distribution; Wu et al. (2017) offered variable selection in the joint location, scale and skewness models of the STN distribution; and Doğru and Arslan (2018b) proposed the joint modelling of location, scale and skewness parameters of the skew Laplace normal (SLN) distribution.

Since the estimators of classical regression models under normality assumption are very sensitive to the outliers, heavy-tailedness, and the skewness in the data, the robust mixture regression models have been proposed. It is known that mixture regression models are useful tools for the analysis of heterogeneous data sets. Mixture regression models were first introduced by Quandt (1972) and Quandt and Ramsey (1978) as switching regression models. These models are commonly used in areas such as engineering, genetics, biology, econometrics, and marketing. In addition, these models are used to model the relationship between variables that belong to unknown latent groups. Some of recent work on the topic can be summarized as follows: Wei (2012) and Yao et al. (2014) introduced the robust mixture regression model based on the *t* distribution; Zhang (2013) examined the mixture regression model using the Pearson Type VII distribution; Song et al. (2014) proposed the robust mixture regression model using the Laplace distribution; Liu and Lin (2014) proposed the mixture regression model based on the SN distribution (Azzalini (1985, 1986)); Doğru (2015) and Doğru and Arslan (2017a) proposed the robust mixture regression model based on the skew *t* distribution (Azzalini and Capitanio (2003)) to cope with both heavy-tailedness and skewness in the data; and Doğru and Arslan (2016) investigated the robust mixture model based on a mixture of different distributions. Recently, Doğru and Arslan (2017b) proposed finite mixtures of SLN distributions and finite mixtures of SLN distributions methodology is also applied to the mixture regression problem, and Dai et al. (2019) proposed robust variable selection in finite mixture of regression models based on the t distribution. The SLN distribution is a special case of the skew exponential power distribution proposed by Azzalini (1986) and further studied by Gómez et al. (2007). However, all the mixture regression modelling mentioned above is under the assumption that there is no heteroscedasticity and skewness for different covariates in different subgroups of observations. But Li et al. (2016) have recently considered this problem and proposed a skew-normal mixture of joint location, scale and skewness models to examine the heteroscedastic skew normal data set consisting of a heterogeneous population. This model was a generalization of the mixture regression model based on the SN distribution which was proposed by Liu and Li (2014).

Both SN and SLN distributions have the same number of parameters to accommodate location, scale, and skewness, but SLN distribution has heavier tails, which could be used to model heavy-tailedness along with the skewness in the data. In this paper, we propose the joint modelling of location, scale and skewness parameters of mixtures of SLN distributions for modelling heteroscedastic skew-heavy tailed data set coming from a heterogeneous population. Our proposed model will be also an alternative to the joint modelling of location, scale and skewness parameters of mixtures of SN distributions. Additionally, this newly proposed model can be viewed as a generalization of the mixture regression model based on the SLN distribution which was studied by Doğru and Arslan (2017b).

Furthermore, another approach called Bayesian methods for density regression based on a non-parametric mixture of regression models was proposed by Dunson et al. (2007). This Bayesian method was also used before by Fernández and Steel (1998) for linear regression models to model skew error distributions with fat tails. In addition, Dunson et al. (2007) provided a class of weighted mixture of Dirichlet process priors for the uncountable collection of mixture distributions. On the topic of mixture regression in Statistics, our method is a frequentist approach and different from a Bayesian method such as Dunson et al. (2007) and Fernández and Steel (1998). Given that Bayesian method often gives identical answers to frequentist Statistics, and our EM algorithm does not require as much memory to store the results as MCMC sampling if you live in the big data world, different methods should be available for practitioners.

The rest of the paper is designed as follows: Section 2 details the basic information about SLN distribution. Section 3 gives the joint modelling of location, scale and skewness parameters of mixtures of SLN distributions. Section 4 demonstrates the ML estimation of the joint modelling of location, scale and skewness parameters of mixtures of SLN distributions via the EM algorithm. Sections 5 and 6 present the performance of the proposed model providing a simulation study and a real data example. Section 7 is devoted to some conclusions.

2. Skew Laplace normal distribution

Let Y be a random variable which has the SLN distribution ($Y \sim SLN(\mu, \sigma^2, \lambda)$) with the location parameter $\mu \in \mathbb{R}$, scale parameter $\sigma^2 \in (0, \infty)$ and the skewness parameter $\lambda \in \mathbb{R}$. Its probability density function (pdf) is given by

$$f(y) = 2 f_L(y; \mu, \sigma) \Phi\left(\lambda \frac{y-\mu}{\sigma}\right), \quad -\infty < y < \infty, \tag{1}$$

where $f_L(y; \mu, \sigma)$ represents the pdf of Laplace distribution with

$$f_L(y; \mu, \sigma) = \frac{1}{2\sigma} e^{-\frac{|y-\mu|}{\sigma}},$$

and Φ is the cumulative distribution function of the standard normal distribution.

2.1 Stochastic representation of the SLN distribution

Let $Z \sim SN(0,1,\lambda)$ and V with the pdf $f_V(v) = v^{-3} \exp(-(2v^2)^{-1})$, $v > 0$ be two independent random variables. Then, the random variable $Y \sim SLN(\mu, \sigma^2, \lambda)$ can be written as:

$$Y = \mu + \sigma \frac{Z}{V} \ . \tag{2}$$

Moreover, using the stochastic representation of the SN (Azzalini (1986, p. 201) and Henze (1986, Theorem 1)) distributed random variable Z, the following stochastic representation of the random variable Y is obtained as:

$$Y = \mu + \sigma \left(\frac{\lambda |Z_1|}{\sqrt{V^2 (V^2 + \lambda^2)}} + \frac{Z_2}{\sqrt{V^2 + \lambda^2}} \right), \tag{3}$$

where $Z_1 \sim N(0,1)$ and $Z_2 \sim N(0,1)$ are independent random variables. This stochastic representation leads to the following hierarchical representation of the SLN distribution:

$$\begin{aligned} Y|u, v &\sim N\left(\mu + \frac{\sigma\lambda u}{v^2 + \lambda^2}, \frac{\sigma^2}{v^2 + \lambda^2}\right), \\ U|v &\sim TN\left(\left(0, \frac{v^2 + \lambda^2}{v^2}\right); (0, \infty)\right), \\ V &\sim f_V(v) = v^{-3} \exp(-(2v^2)^{-1}), \end{aligned} \tag{4}$$

where $U = \sqrt{V^{-2}(V^2 + \lambda^2)}|Z_1|$ and $TN(\cdot)$ shows the truncated normal distribution.

To derive an EM algorithm of Section 4, we now need some conditional expectations with the following proposition.

Proposition 1. According to the hierarchical representation given in (4), the following conditional expectations are obtained:

$$E(V^2|y) = \frac{\sigma}{|y-\mu|} \ , \tag{5}$$
$$E(U|y) = \lambda s + \frac{\Phi(\lambda s)}{\phi(\lambda s)} \ , \tag{6}$$
$$E(U^2|y) = 1 + \lambda s E(U|y) \ . \tag{7}$$

3. Joint location, scale and skewness models of mixtures of SLN distributions

Let $y_1, y_2, \dots, y_n$ be a random sample from a g-component mixtures of SLN distributions, then the pdf of this mixture model is given by:

$$f(y_j|\boldsymbol{\Theta}) = \sum_{i=1}^{g} \pi_i \, f_i(y_j; \mu_i, \sigma_i^2, \lambda_i) \ , \tag{8}$$

where π_i is the mixing probability with $\sum_{i=1}^{g} \pi_i = 1$, $0 \le \pi_i \le 1$, $f_i(y_j; \mu_i, \sigma_i^2, \lambda_i)$ represents the pdf of the ith component (pdf of the SLN distribution) given in (1) and $\boldsymbol{\Theta} = (\pi_1, \dots, \pi_g, \mu_1, \dots, \mu_g, \sigma_1^2, \dots, \sigma_g^2, \lambda_1, \dots, \lambda_g)'$ is the unknown parameter vector.

Let us consider the following joint location, scale and skewness models of mixtures of SLN distributions:

$$\begin{cases} y_j \sim \sum_{i=1}^{g} \pi_i \, f_i(y_j; \mu_{ij}, \sigma_{ij}^2, \lambda_{ij}), \ \ j = 1,2, \dots, n, \\ \mu_{ij} = \boldsymbol{x}_j^T \boldsymbol{\beta}_i \ , \\ \log \sigma_{ij}^2 = \boldsymbol{h}_j^T \boldsymbol{\gamma}_i \ , \\ \lambda_{ij} = \boldsymbol{w}_j^T \boldsymbol{\alpha}_i \ , i = 1, \dots, g, \end{cases} \tag{9}$$

where y_j is the jth observed response and $\boldsymbol{x}_j = (x_{j1}, \dots, x_{jp})^T, \boldsymbol{h}_j = (h_{j1}, \dots, h_{jq})^T$ and $\boldsymbol{w}_j = (w_{j1}, \dots, w_{jr})^T$ are observed covariates corresponding to y_j. The covariate vectors $\boldsymbol{x}_j, \boldsymbol{z}_j$ and $\boldsymbol{w}_j$ are not needed to be identical. Also, $\boldsymbol{\beta}_i = (\beta_{i1}, \dots, \beta_{ip})^T$ is a $p \times 1$ vector of unknown parameters in the location model of the ith component, $\boldsymbol{\gamma}_i = (\gamma_{i1}, \dots, \gamma_{iq})^T$ is a $q \times 1$ vector of unknown parameters in the scale model of the ith component, and $\boldsymbol{\alpha}_i = (\alpha_{i1}, \dots, \alpha_{ir})^T$ is a $r \times 1$ vector of unknown parameters in the skewness model of the ith component.

Note that if σ_{ij}^2 and λ_{ij} are constant, then the model (9) reduces to the mixture regression model based on the SLN distribution which was introduced by Doğru and Arslan (2017b). Therefore, model (9) can also be considered as an extension of the existing mixture regression model based on the SLN distribution. We assume that the number of component g is fixed and known through of the paper and deal with the estimation of the parameter vector $\boldsymbol{\Theta} = (\pi_1, \dots, \pi_g, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_g)^T$, where $\boldsymbol{\theta}_i = (\boldsymbol{\beta}_i^T, \boldsymbol{\gamma}_i^T, \boldsymbol{\alpha}_i^T)$ for $i = 1, \dots, g$.

As what pointed out by Li et al. (2016), Hennig (2000) and Wang et al. (1996), the issue of identifiability" from a finite mixture models models needs to be defined, and in our case, we have:

Definition 1. The finite SLN mixture of location, scale and skewness model given in (9) is said to be identifiable if the following equation holds for any two parameter vectors $\mathbf{\Theta} = \left(\pi_1, \dots, \pi_g, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_g\right)^T$ and $\mathbf{\Theta}^* = \left(\pi_1^*, \dots, \pi_g^*, \boldsymbol{\theta}_1^*, \dots, \boldsymbol{\theta}_{g*}^*\right)^T$:

$$\sum_{i=1}^{g} \pi_i \, f_i\left(y; \mu_i, \sigma_i^2, \lambda_i\right) = \sum_{i=1}^{g*} \pi_i^* \, f_i\left(y; \mu_i^*, {\sigma_i^2}^*, \lambda_i^*\right)$$

for each $i = 1, \dots g$ and all possible values of y. This then indicates $g = g^*$ and $\mathbf{\Theta} = \mathbf{\Theta}^*$.

4. ML estimation of the joint location, scale and skewness models of mixtures of SLN distributions

Let $\{(\boldsymbol{x}_1, \boldsymbol{h}_1, \boldsymbol{w}_1, y_1), \dots, (\boldsymbol{x}_n, \boldsymbol{h}_n, \boldsymbol{w}_n, y_n)\}$ be a sample to estimate the unknown parameter vector $\mathbf{\Theta}$. The ML estimator of $\mathbf{\Theta}$ for a g-component SLN mixture of joint location, scale and skewness models can be found by maximizing the following log-likelihood function with respect to $\mathbf{\Theta}$:

$$\ell(\mathbf{\Theta}) = \sum_{j=1}^{n} \log\left(\sum_{i=1}^{g} \pi_i \, f_i\left(y_j; \boldsymbol{x}_j^T \boldsymbol{\beta}_i, \boldsymbol{h}_j^T \boldsymbol{\gamma}_i, \boldsymbol{w}_j^T \boldsymbol{\alpha}_i\right)\right). \tag{10}$$

However, a numerical algorithm should be used since this log-likelihood function cannot be directly maximized. Generally, the EM algorithm is used to obtain the ML estimator of $\mathbf{\Theta}$. Here, we will implement the following EM algorithm to estimate the parameters:

Let $Z_j = \left(Z_{1j}, \dots, Z_{gj}\right)^T$ be the latent variables with

$$Z_{ij} = \begin{cases} 1, & if\ j^{th}\ observation\ belongs\ to\ i^{th}\ component \\ 0, & otherwise \end{cases} \tag{11}$$

where $j = 1, \dots, n$ and $i = 1, \dots, g$. To conduct the EM algorithm, we use the stochastic representation of the SLN distribution given in (3). Let V and U be the latent variables. Using the hierarchical representation given in (4), we have the following hierarchical representation for the SLN mixture of joint location, scale and skewness models:

$$\begin{aligned} Y_j \middle| u_j, v_j, Z_{ij} = 1 &\sim N\left(\boldsymbol{x}_j^T \boldsymbol{\beta}_i + \frac{e^{\boldsymbol{h}_j^T \boldsymbol{\gamma}_i/2} \left(\boldsymbol{w}_j^T \boldsymbol{\alpha}_i\right) u_j}{v_j^2 + \left(\boldsymbol{w}_j^T \boldsymbol{\alpha}_i\right)^2}, \frac{e^{\boldsymbol{h}_j^T \boldsymbol{\gamma}_i}}{v_j^2 + \left(\boldsymbol{w}_j^T \boldsymbol{\alpha}_i\right)^2}\right), \\ U_j \middle| v_j, Z_{ij} = 1 &\sim TN\left(\left(0, \frac{v_j^2 + \left(\boldsymbol{w}_j^T \boldsymbol{\alpha}_i\right)^2}{v_j^2}\right); (0, \infty)\right), \\ v_j \middle| Z_{ij} = 1 &\sim f\left(v_j\right) = v_j^{-3} \exp\left(-\left(2v_j^2\right)^{-1}\right). \end{aligned} \tag{12}$$

Let $\boldsymbol{u} = (u_1, \dots, u_n)$, $\boldsymbol{v} = (v_1, \dots, v_n)$ and $\boldsymbol{z} = (z_1, \dots, z_n)$ be the missing data and $(\boldsymbol{y}, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{z})$ be the complete data, where $\boldsymbol{y} = (y_1, \dots, y_n)$. Then, the complete data log-likelihood function of $\mathbf{\Theta}$ can be written using the hierarchical representation given in (12) as follows:

$$\ell_c(\mathbf{\Theta}; \boldsymbol{y}, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{z}) = \sum_{j=1}^{n} \sum_{i=1}^{g} z_{ij} \Big\{ \log \pi_i - \log \pi - \frac{1}{2} \boldsymbol{h}_j^T \boldsymbol{\gamma}_i - 2 \log v_j - \left(2v_j^2\right)^{-1}$$

$$-\frac{1}{2}\left(\frac{\left(y_j-\boldsymbol{x}_j^T\boldsymbol{\beta}_i\right)^2}{e^{\boldsymbol{h}_j^T\boldsymbol{\gamma}_i}}v_j^2+u_j^2-2\frac{\boldsymbol{w}_j^T\boldsymbol{\alpha}_i}{e^{\boldsymbol{h}_j^T\boldsymbol{\gamma}_i/2}}\left(y_j-\boldsymbol{x}_j^T\boldsymbol{\beta}_i\right)u_j+\frac{\left(\boldsymbol{w}_j^T\boldsymbol{\alpha}_i\right)^2}{e^{\boldsymbol{h}_j^T\boldsymbol{\gamma}_i}}\left(y_j-\boldsymbol{x}_j^T\boldsymbol{\beta}_i\right)^2\right)\Bigg\}. \tag{13}$$

The ML estimator of $\boldsymbol{\Theta}$ can be derived by maximizing this function. However, this maximization yields the estimator that will be dependent on the latent variables. Therefore, we have to take the conditional expectation of the complete data log-likelihood function given y_j to cope with this latency problem. Then, we have the conditional expectation (13) as:

$$\begin{aligned}E\left(\ell_c(\boldsymbol{\Theta};\boldsymbol{y},\boldsymbol{u},\boldsymbol{v},\boldsymbol{z})|y_j\right)=\sum_{j=1}^{n}\sum_{i=1}^{g}E\left(Z_{ij}|y_j\right)\Bigg\{&\log\pi_i-\log\pi-\frac{1}{2}\boldsymbol{h}_j^T\boldsymbol{\gamma}_i-2E\left(\log V_j|y_j\right)\\&-E\left(2\left(V_j^2\right)^{-1}\Big|\,y_j\right)-\frac{1}{2}\left(\frac{\left(y_j-\boldsymbol{x}_j^T\boldsymbol{\beta}_i\right)^2}{e^{\boldsymbol{h}_j^T\boldsymbol{\gamma}_i}}E\left(V_j^2|y_j\right)+E\left(U_j^2|y_j\right)\right.\\&\left.-2\frac{\boldsymbol{w}_j^T\boldsymbol{\alpha}_i}{e^{\boldsymbol{h}_j^T\boldsymbol{\gamma}_i/2}}\left(y_j-\boldsymbol{x}_j^T\boldsymbol{\beta}_i\right)E\left(U_j|y_j\right)+\frac{\left(\boldsymbol{w}_j^T\boldsymbol{\alpha}_i\right)^2}{e^{\boldsymbol{h}_j^T\boldsymbol{\gamma}_i}}\left(y_j-\boldsymbol{x}_j^T\boldsymbol{\beta}_i\right)^2\right)\Bigg\}.\end{aligned} \tag{14}$$

The conditional expectation components related to unknown parameters in (14) $only\ have$ $E\left(V_j^2|y_j\right)$, $E\left(U_j|y_j\right)$ and $E\left(U_j^2|y_j\right)$ which can be computed using the conditional expectations given in (5)-(7), and $E\left(Z_{ij}|y_j\right)$ *which* can be calculated using the classical theory of mixture modeling. Let

$$\hat{z}_{ij}=\frac{\hat{\pi}_i f_i\left(y_j;\boldsymbol{x}_j^T\widehat{\boldsymbol{\beta}}_i,\boldsymbol{h}_j^T\widehat{\boldsymbol{\gamma}}_i,\boldsymbol{w}_j^T\widehat{\boldsymbol{\alpha}}_i\right)}{\sum_{i=1}^{n}\hat{\pi}_i f_i\left(y_j;\boldsymbol{x}_j^T\widehat{\boldsymbol{\beta}}_i,\boldsymbol{h}_j^T\widehat{\boldsymbol{\gamma}}_i,\boldsymbol{w}_j^T\widehat{\boldsymbol{\alpha}}_i\right)}, \tag{15}$$

$$\hat{v}_{ij}=E\left(V_j^2|y_j\right)=\frac{e^{\boldsymbol{h}_j^T\widehat{\boldsymbol{\gamma}}_i/2}}{\left|y_j-\boldsymbol{x}_j^T\widehat{\boldsymbol{\beta}}_i\right|}, \tag{16}$$

$$\hat{u}_{1ij}=E\left(U_j|y_j\right)=\hat{\kappa}_{ij}+\frac{\Phi\left(\hat{\kappa}_{ij}\right)}{\phi\left(\hat{\kappa}_{ij}\right)}, \tag{17}$$

$$\hat{u}_{2ij}=E\left(U_j^2|y_j\right)=1+\hat{\kappa}_{ij}\hat{u}_{1ij}, \tag{18}$$

where $\hat{\kappa}_{ij}=\boldsymbol{w}_j^T\widehat{\boldsymbol{\alpha}}_i\frac{\left(y_j-\boldsymbol{x}_j^T\widehat{\boldsymbol{\beta}}_i\right)}{e^{\boldsymbol{h}_j^T\widehat{\boldsymbol{\gamma}}_i/2}}$. Then, we obtain the following objective function after re-writing above conditional expectations in (14):

$$\begin{aligned}Q\left(\boldsymbol{\Theta};\widehat{\boldsymbol{\Theta}}\right)=\sum_{j=1}^{n}\sum_{i=1}^{g}\hat{z}_{ij}\Bigg\{&\log\pi_i-\frac{1}{2}\boldsymbol{h}_j^T\boldsymbol{\gamma}_i-\frac{1}{2}\left(\frac{\left(y_j-\boldsymbol{x}_j^T\boldsymbol{\beta}_i\right)^2}{e^{\boldsymbol{h}_j^T\boldsymbol{\gamma}_i}}\hat{v}_{ij}+\hat{u}_{2ij}\right.\\&\left.-2\frac{\boldsymbol{w}_j^T\boldsymbol{\alpha}_i}{e^{\boldsymbol{h}_j^T\boldsymbol{\gamma}_i/2}}\left(y_j-\boldsymbol{x}_j^T\boldsymbol{\beta}_i\right)\hat{u}_{1ij}+\frac{\left(\boldsymbol{w}_j^T\boldsymbol{\alpha}_i\right)^2}{e^{\boldsymbol{h}_j^T\boldsymbol{\gamma}_i}}\left(y_j-\boldsymbol{x}_j^T\boldsymbol{\beta}_i\right)^2\right)\Bigg\}.\end{aligned} \tag{19}$$

To this end, the steps of the EM algorithm can be organized as follows:

EM algorithm:

1. Take initial value for $\boldsymbol{\Theta}^{(0)}$.

2. E-Step: Compute the following expectations for the $k=0,1,2,\ldots$ iteration

$$\hat{z}_{ij}^{(k)} = \frac{\hat{\pi}_i^{(k)} f_i\left(y_j; \boldsymbol{x}_j^T \widehat{\boldsymbol{\beta}}_i^{(k)}, \boldsymbol{h}_j^T \widehat{\boldsymbol{\gamma}}_{\boldsymbol{i}}^{(k)}, \boldsymbol{w}_j^T \widehat{\boldsymbol{\alpha}}_i^{(k)}\right)}{\sum_{i=1}^{n} \hat{\pi}_i^{(k)} f_i\left(y_j; \boldsymbol{x}_j^T \widehat{\boldsymbol{\beta}}_i^{(k)}, \boldsymbol{h}_j^T \widehat{\boldsymbol{\gamma}}_{\boldsymbol{i}}^{(k)}, \boldsymbol{w}_j^T \widehat{\boldsymbol{\alpha}}_i^{(k)}\right)} \tag{20}$$

$$\hat{v}_{ij}^{(k)} = E\left(V_j^2 \middle| y_j, \widehat{\boldsymbol{\Theta}}^{(k)}\right) = \frac{e^{\boldsymbol{h}_j^T \widehat{\boldsymbol{\gamma}}_{\boldsymbol{i}}^{(k)}/2}}{\left|y_j - \boldsymbol{x}_j^T \widehat{\boldsymbol{\beta}}_i^{(k)}\right|}, \tag{21}$$

$$\hat{u}_{1ij}^{(k)} = E\left(U_j \middle| y_j, \widehat{\boldsymbol{\Theta}}^{(k)}\right) = \hat{\kappa}_{ij}^{(k)} + \frac{\Phi\left(\hat{\kappa}_{ij}^{(k)}\right)}{\phi\left(\hat{\kappa}_{ij}^{(k)}\right)}, \tag{22}$$

$$\hat{u}_{2ij}^{(k)} = E\left(U_j^2 \middle| y_j, \widehat{\boldsymbol{\Theta}}^{(k)}\right) = 1 + \hat{\kappa}_{ij}^{(k)} \hat{u}_{1ij}^{(k)}, \tag{23}$$

where, $\hat{\kappa}_{ij}^{(k)} = \boldsymbol{w}_j^T \widehat{\boldsymbol{\alpha}}_i^{(k)} \frac{\left(y_j - \boldsymbol{x}_j^T \widehat{\boldsymbol{\beta}}_i^{(k)}\right)}{e^{\boldsymbol{h}_j^T \widehat{\boldsymbol{\gamma}}_i^{(k)}/2}}$.

Note that we divide both the numerator and denominator in (20) by the largest term in the sum in the denominator, which was suggested by Wang et al. (1996) to prevent overflow in the computation of $\hat{z}_{ij}^{(k)}$.

3. M-Step: Use the conditional expectations given in (20)-(23) and obtain $Q\left(\boldsymbol{\Theta}; \widehat{\boldsymbol{\Theta}}^{(k)}\right)$. Maximize $Q\left(\boldsymbol{\Theta}; \widehat{\boldsymbol{\Theta}}^{(k)}\right)$ with respect to $\boldsymbol{\Theta}$ to obtain new estimates. The $(k+1)th$ parameter estimates for the ith component can be updated using the following maximization results:

$$\hat{\pi}_i^{(k+1)} = \frac{\sum_{j=1}^{n} \hat{z}_{ij}^{(k)}}{n}, \tag{24}$$

$$\widehat{\boldsymbol{\theta}}_i^{(k+1)} = \widehat{\boldsymbol{\theta}}_i^{(k)} + \left(-H\left(\boldsymbol{\theta}_i^{(k)}\right)\right)^{-1} G\left(\boldsymbol{\theta}_i^{(k)}\right), \tag{25}$$

where $\widehat{\boldsymbol{\theta}}_i^{(k)} = \left(\widehat{\boldsymbol{\beta}}_i^{(k)^T}, \widehat{\boldsymbol{\gamma}}_{\boldsymbol{i}}^{(k)^T}, \widehat{\boldsymbol{\alpha}}_i^{(k)^T}\right)$, $G(\boldsymbol{\theta}_i)$ is the score function of the ith component with

$$G(\boldsymbol{\theta}_i) = \frac{\partial Q\left(\boldsymbol{\theta}_i; \widehat{\boldsymbol{\theta}}_i\right)}{\partial \boldsymbol{\theta}_i} = \left(G_1^T(\boldsymbol{\beta}_i), G_2^T(\boldsymbol{\gamma}_i), G_3^T(\boldsymbol{\alpha}_i)\right)^T,$$

and $H(\boldsymbol{\theta}_i)$ is the observed Fisher information matrix of the ith component with

$$H(\boldsymbol{\theta}_i) = \frac{\partial^2 Q\left(\boldsymbol{\theta}_i; \widehat{\boldsymbol{\theta}}_i\right)}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_i^T} = \begin{bmatrix} \frac{\partial^2 Q\left(\boldsymbol{\theta}_i; \widehat{\boldsymbol{\theta}}_i\right)}{\partial \boldsymbol{\beta}_i \partial \boldsymbol{\beta}_i^T} & \frac{\partial^2 Q\left(\boldsymbol{\theta}_i; \widehat{\boldsymbol{\theta}}_i\right)}{\partial \boldsymbol{\beta}_i \partial \boldsymbol{\gamma}_i^T} & \frac{\partial^2 Q\left(\boldsymbol{\theta}_i; \widehat{\boldsymbol{\theta}}_i\right)}{\partial \boldsymbol{\beta}_i \partial \boldsymbol{\alpha}_i^T} \\ \frac{\partial^2 Q\left(\boldsymbol{\theta}_i; \widehat{\boldsymbol{\theta}}_i\right)}{\partial \boldsymbol{\gamma}_i \partial \boldsymbol{\beta}_i^T} & \frac{\partial^2 Q\left(\boldsymbol{\theta}_i; \widehat{\boldsymbol{\theta}}_i\right)}{\partial \boldsymbol{\gamma}_i \partial \boldsymbol{\gamma}_i^T} & \frac{\partial^2 Q\left(\boldsymbol{\theta}_i; \widehat{\boldsymbol{\theta}}_i\right)}{\partial \boldsymbol{\gamma}_i \partial \boldsymbol{\alpha}_i^T} \\ \frac{\partial^2 Q\left(\boldsymbol{\theta}_i; \widehat{\boldsymbol{\theta}}_i\right)}{\partial \boldsymbol{\alpha}_i \partial \boldsymbol{\beta}_i^T} & \frac{\partial^2 Q\left(\boldsymbol{\theta}_i; \widehat{\boldsymbol{\theta}}_i\right)}{\partial \boldsymbol{\alpha}_i \partial \boldsymbol{\gamma}_i^T} & \frac{\partial^2 Q\left(\boldsymbol{\theta}_i; \widehat{\boldsymbol{\theta}}_i\right)}{\partial \boldsymbol{\alpha}_i \partial \boldsymbol{\alpha}_i^T} \end{bmatrix}.$$

4. Repeat E and M steps until the convergence is obtained.
Remark. See Appendix for the detail expressions of $G(\boldsymbol{\theta}_i)$ and $H(\boldsymbol{\theta}_i)$.

5. Asymptotic properties

Let $\{(\boldsymbol{x}_1, \boldsymbol{h}_1, \boldsymbol{w}_1, y_1), \dots, (\boldsymbol{x}_n, \boldsymbol{h}_n, \boldsymbol{w}_n, y_n)\}$ be a random sample, Ω be the parameter space, and $\boldsymbol{\Theta} = (\pi_1, \dots, \pi_g, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_g)^T \in \Omega$, where $\boldsymbol{\theta}_i = (\boldsymbol{\beta}_i^T, \boldsymbol{\gamma}_i^T, \boldsymbol{\alpha}_i^T)$, for $i = 1,\dots, g$, be the collection of all parameters in the log-likelihood function given in (10), and $\boldsymbol{\Theta}^0$ is the true value of the parameter $\boldsymbol{\Theta}$, respectively. For the mixture model given in (8),

$$\pi \in A \equiv \left\{(\pi_1, \dots, \pi_g): \pi_i \geq 0, i = 1, \dots, g, \sum_{i=1}^{g} \pi_i = 1\right\},$$

$$\boldsymbol{\theta} \in \boldsymbol{\Theta} \equiv \{(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_g): \boldsymbol{\theta}_i \in \boldsymbol{\Theta}_{\boldsymbol{i}}, i = 1, \dots g\},$$

and the $\boldsymbol{\Theta}_{\boldsymbol{i}}, i = 1, \dots g$, are closed convex sets that belongs to R^p. Let $\Omega = A \times \boldsymbol{\Theta}$. For any given $(\pi^0, \boldsymbol{\Theta}^0) \in \Omega$, it can be defined as

$$\Omega(\pi^0, \boldsymbol{\theta}^0) = \{(\pi, \boldsymbol{\theta}): (\pi, \boldsymbol{\theta}) \in \Omega \; and \; f(.|\pi, \boldsymbol{\theta}) = f(.|\pi^0, \boldsymbol{\theta}^0)\}.$$

Assume that $\widehat{\boldsymbol{\Theta}}_n = (\hat{\pi}_n, \widehat{\boldsymbol{\theta}}_n)$ is the estimate of $\boldsymbol{\Theta}$ obtained by the EM-type algorithm given by the equations (24) and (25), then the asymptotic properties of this estimator and its standard errors of estimation are detailed as follows:

5.1 Consistency and asymptotic distribution

Theorem 1. Let $f(y|\boldsymbol{\Theta})$ be a pdf given in (8). Let $\boldsymbol{\Theta}^0 = (\pi^0, \boldsymbol{\theta}^0)$ be the true value of $\boldsymbol{\Theta} = (\pi, \boldsymbol{\theta})$, which exists at some point in the region Ω, and $\{\widehat{\boldsymbol{\Theta}}_n = (\hat{\pi}_n, \widehat{\boldsymbol{\theta}}_n), \; n = 1,2, \dots\}$ is a sequence. Then, if we assume that Conditions *-* given in Appendix hold, there is a unique strongly consistent solution of the mixture models likelihood equations. Then, $dis\{(\hat{\pi}_n, \widehat{\boldsymbol{\theta}}_n), \Omega(\pi^0, \boldsymbol{\theta}^0)\} \to 0, w.p. 1.$

Proof. See Appendix for the proof of Theorem 1.

Theorem 2. Under Conditions *-*, the asymptotic distribution of $n^{1/2}(\widehat{\boldsymbol{\Theta}}_n - \boldsymbol{\Theta}^0)$ is asymptotically normal with mean zero and covariance matrix $I(\boldsymbol{\Theta}^0)^{-1}$

$$n^{1/2}(\widehat{\boldsymbol{\Theta}}_n - \boldsymbol{\Theta}^0) \underset{d}{\to} N(0, I(\boldsymbol{\Theta}^0)^{-1}),$$

where $I(\boldsymbol{\Theta}^0)^{-1}$ is the inverse of the Fisher information matrix.

Proof. See Appendix for the proof of Theorem 2.

5.2 Estimation of the standard errors

To calculate the standard errors of ML estimators for the parameters of joint location, scale and skewness models of mixtures of SLN distributions, we will use the information based method given by Basford et al. (1997). In this method, the observed information matrix can be approximated by the empirical information matrix. To do so, we use the inverse of the empirical information matrix to get an approximation to the asymptotic covariance matrix of estimators. The empirical information matrix can be defined as:

$$\hat{I}_e(\widehat{\boldsymbol{\Theta}}) = \sum_{j=1}^{n} \hat{\boldsymbol{s}}_j \hat{\boldsymbol{s}}_j^T, \tag{26}$$

where $\hat{\boldsymbol{s}}_j = E_{\widehat{\boldsymbol{\Theta}}}\left(\frac{\partial \ell_{cj}(\boldsymbol{\Theta}; \boldsymbol{y}_j, \boldsymbol{u}_j, \boldsymbol{v}_j, \boldsymbol{z}_j)}{\partial \boldsymbol{\Theta}}\middle| \boldsymbol{y}_j\right)$, $j = 1, \dots, n$ are the individual scores and $\ell_{cj}(\boldsymbol{\Theta}; \boldsymbol{y}_j, \boldsymbol{u}_j, \boldsymbol{v}_j, \boldsymbol{z}_j)$ is the complete data log-likelihood function for the jth observation. The components of the score vector $\hat{\boldsymbol{s}}_j$ are $\left(\hat{\boldsymbol{s}}_{j,\pi_1}, \dots, \hat{\boldsymbol{s}}_{j,\pi_{g-1}}, \hat{\boldsymbol{s}}_{j,\boldsymbol{\beta}_1}, \dots, \hat{\boldsymbol{s}}_{j,\boldsymbol{\beta}_g}, \hat{\boldsymbol{s}}_{j,\boldsymbol{\gamma}_1}, \dots, \hat{\boldsymbol{s}}_{j,\boldsymbol{\gamma}_g}, \hat{\boldsymbol{s}}_{j,\boldsymbol{\alpha}_1}, \dots, \hat{\boldsymbol{s}}_{j,\boldsymbol{\alpha}_g}\right)^T$, where

$$\hat{\boldsymbol{s}}_{j,\pi_r} = \frac{\hat{z}_{rj}}{\hat{\pi}_r} - \frac{\hat{z}_{gj}}{\hat{\pi}_g}, \qquad r = 1, \dots, g-1,$$

$$\hat{\boldsymbol{s}}_{j,\boldsymbol{\beta}_i} = G_1(\widehat{\boldsymbol{\beta}}_i), \hat{\boldsymbol{s}}_{j,\boldsymbol{\gamma}_i} = G_2(\widehat{\boldsymbol{\gamma}}_i), \text{ and } \hat{\boldsymbol{s}}_{j,\boldsymbol{\alpha}_i} = G_3(\widehat{\boldsymbol{\alpha}}_i), i = 1, \dots, g.$$

Here, $G_1(\widehat{\boldsymbol{\beta}}_i)$, $G_2(\widehat{\boldsymbol{\gamma}}_i)$ and $G_3(\widehat{\boldsymbol{\alpha}}_i)$ are given with the equations (28)-(30). Thus, using these equations, we can form the information matrix $\hat{I}_e$ given in (26). After this, the standard errors of $\widehat{\boldsymbol{\Theta}}$ can be found using the square root of the matrix $\hat{I}_e(\widehat{\boldsymbol{\Theta}})^{-1}$.

6. Applications

In this section, we conduct a simulation study and a real data analysis to show the performance of the proposed mixture model over the joint location, scale and skewness models of mixtures of SN distributions. For the computation of the estimators of parameters, we use the EM algorithm given in Section 4. We summarize the computation details as follows:

Details of computation:

i) The simulation study and real data example are conducted using a MATLAB R2017b software.
ii) For all numerical computations, the stopping rule is taken as 10^{-6}.
iii) Initial values for the EM algorithm: the good initial values in the simulation are the true parameter values; the initial values in the real data example are the estimates from the normal mixture regression for the parameters of location models and 6×1 zero vector as initial values for all scale and skewness models.
iv) In the simulation study, we compare the performance of joint location, scale and skewness models of mixtures of SLN distributions with the joint location, scale and skewness models of mixtures of SN distributions under different data sets. The data sets are generated from SLN, SN and STN distributions to compare the behavior of estimators according to the skew and heavy-tailed data sets.

The data set from the SLN distribution can be generated as follows:
- Sample U from the uniform distribution $Uniform(0,1)$ and set $V = \sqrt{-\frac{1}{2\log U}}$.
- Sample Z_1 and Z_2 independently from the standard normal distribution $N(0,1)$.
- After this, setting $Y = \mu + \sigma\left(\frac{\lambda|Z_1|}{\sqrt{V^2(V^2+\lambda^2)}} + \frac{Z_2}{\sqrt{V^2+\lambda^2}}\right)$ with appropriate parameter values gives the SLN distributed sample.

Note that the procedures given in Azzalini and Capitanio (1999) and Cabral et al. (2008) are used for the data generating procedures of SN and STN distributions.

6.1. Simulation study

The simulation study below is based on two scenarios with aim to illustrate the performance of parameter estimates and model fitting of the proposed joint modelling of location, scale and skewness parameters of mixtures of SLN distributions over the joint location, scale and skewness models of mixtures of SN distributions. The performance of the parameter estimators is evaluated via the bias and the mean squared error (MSE). The formulas of the bias and the MSE are given below:

$$\widehat{bias}(\hat{\theta}) = \bar{\theta} - \theta, \;\; \widehat{MSE}(\hat{\theta}) = \frac{1}{N}\sum_{j=1}^{N}\left(\hat{\theta}_j - \theta\right)^2 ,$$

where θ is the true parameter value, $\hat{\theta}_j$ is the estimate of θ for the jth simulated data and $\bar{\theta} = \frac{1}{N}\sum_{j=1}^{N}\hat{\theta}_j$. The number of replications $N = 500$ times. The sample sizes (n) are respectively taken as $200, 400$ and 600 for all simulation configurations.

Scenario 1. We generate the data $\{(x_{1j}, y_j), j = 1, \dots, n\}$ from the following two component mixture of joint location, scale and skewness models

$$\begin{cases} y_j \sim \pi_1 f_1(\mu_{1j}, \sigma_{1j}^2, \lambda_{1j}) + \pi_2 f_2(\mu_{2j}, \sigma_{2j}^2, \lambda_{2j}), \;\; j = 1,2, \dots, n, \\ \mu_{ij} = \boldsymbol{x}_j^T \boldsymbol{\beta}_i , \\ \log \sigma_{ij}^2 = \boldsymbol{h}_j^T \boldsymbol{\gamma}_i , \\ \lambda_{ij} = \boldsymbol{w}_j^T \boldsymbol{\alpha}_i , \qquad i = 1,2, \end{cases} \tag{27}$$

where all covariate vectors $\boldsymbol{x}_j, \boldsymbol{h}_j$ and $\boldsymbol{w}_j$ are independently generated from uniform distribution $Uniform(-1,1)$, $\beta_1 = (0,1,1)^T$, $\gamma_1 = (0,1,1)^T$ and $\alpha_1 = (0,1,1)^T$ for the first component, $\beta_2 = (0,-1,-1)^T$, $\gamma_2 = (0,-1,-1)^T$ and $\alpha_2 = (0,-1,-1)^T$ for the second component, and the mixing proportion $\pi_1 = 0.25$. The considered distributions of $f_1(.)$ and $f_2(.)$ are given with the following cases:

Case I: $f_1 \sim SLN(\mu_{1j}, \sigma_{1j}^2, \lambda_{1j}), f_2 \sim SLN(\mu_{2j}, \sigma_{2j}^2, \lambda_{2j})$.
Case II: $f_1 \sim SN(\mu_{1j}, \sigma_{1j}^2, \lambda_{1j}), f_2 \sim SN(\mu_{2j}, \sigma_{2j}^2, \lambda_{2j})$.
Case III: $f_1 \sim STN(\mu_{1j}, \sigma_{1j}^2, \lambda_{1j}, \nu), f_2 \sim STN(\mu_{2j}, \sigma_{2j}^2, \lambda_{2j}, \nu)$ where ν shows the degrees of freedom parameter, and it is taken as 3.

Scenario 2. We generate the data $\{(x_{1j}, y_j), j = 1, \dots, n\}$ from the two component mixture of joint location, scale and skewness models given in (27) with the true parameters $\beta_1 = (0,1,1)^T, \gamma_1 = (0,1,1)^T$ and $\alpha_1 = (0,1,1)^T$ for the first component, $\beta_2 = (0,-1,-1)^T$, $\gamma_2 = (0,-1,-1)^T$ and $\alpha_2 = (0,-1,-1)^T$ for the second component, and the mixing proportion $\pi_1 = 0.5$.

We consider the following distributions for $f_1(.)$ and $f_2(.)$:
Case I: $f_1 \sim SLN(\mu_{1j}, \sigma_{1j}^2, \lambda_{1j}), f_2 \sim SLN(\mu_{2j}, \sigma_{2j}^2, \lambda_{2j})$.
Case II: $f_1 \sim SN(\mu_{1j}, \sigma_{1j}^2, \lambda_{1j}), f_2 \sim SN(\mu_{2j}, \sigma_{2j}^2, \lambda_{2j})$.
Case III: $f_1 \sim STN(\mu_{1j}, \sigma_{1j}^2, \lambda_{1j}, \nu), f_2 \sim STN(\mu_{2j}, \sigma_{2j}^2, \lambda_{2j}, \nu)$where ν shows the degrees of freedom parameter, and it is taken as 3.

The simulation results for Scenarios 1 and 2 are outlined in Tables 1-3 and Tables 4-6 respectively. The tables contain the bias, MSE values of the parameter estimates, along with the true parameter values. According to the tables, we get the following results: The proposed estimation procedure can accurately estimate all parameters of the SLN mixture of joint location, scale and skewness models. When we are comparing the estimators under the skew and/or heavy-tailed data set, we have similar results for all the cases. For the Case I, II and III for all scenarios, the proposed estimation method fit better than the SN mixture of joint location, scale and skewness models. Further, the MSE values of the SN mixture of joint location, scale and skewness models parameter estimates are larger than the SLN mixture of joint location, scale and skewness models parameter estimates. In summary, the results of our simulation study show that the the SLN mixture of joint location, scale and skewness models should be used when the data set is skew and/or heavy-tailed.

Table 1. The bias and the values of MSE for the different sample sizes for Case I of Scenario 1.

					SLN		SN	
n		Model	Parameter	True	Bias	MSE	Bias	MSE
200	Component 1	Location	β_{10}	0	0.001609	0.000523	0.011240	0.280015
			β_{11}	1	-0.000866	0.001937	-0.129769	0.942425
			β_{12}	1	0.000342	0.001835	-0.208158	0.768263
		Scale	γ_{10}	0	-0.060628	0.018006	0.781018	0.890588
			γ_{11}	1	-0.065036	0.025605	-0.562109	1.637458
			γ_{12}	1	-0.074615	0.027414	-0.500881	1.501806
		Skewness	α_{10}	0	0.001422	0.001843	-0.010485	0.007808
			α_{11}	1	-0.042347	0.012478	-0.947892	0.939000
			α_{12}	1	-0.038905	0.010831	-0.960368	0.977903
			π_1	0.25	0.002221	0.002511	0.024569	0.010822
	Component 2	Location	β_{20}	0	-0.001010	0.000247	-0.009864	0.011726
			β_{21}	-1	-0.000618	0.000746	0.026629	0.042916
			β_{22}	-1	0.000028	0.000769	0.025152	0.047674
		Scale	γ_{20}	0	-0.048273	0.008344	0.326883	0.177632
			γ_{21}	-1	0.030339	0.009403	0.038803	0.224439
			γ_{22}	-1	0.004941	0.009548	-0.056450	0.215279
		Skewness	α_{20}	0	-0.002709	0.001166	0.001246	0.002856
			α_{21}	-1	0.023648	0.004822	0.655526	0.446036
			α_{22}	-1	0.027575	0.005621	0.662656	0.455720
400	Component 1	Location	β_{10}	0	-0.001250	0.000127	0.036724	0.697977
			β_{11}	1	0.000215	0.000448	-0.189467	0.651509
			β_{12}	1	-0.003555	0.000623	-0.283836	0.706189
		Scale	γ_{10}	0	-0.042464	0.008039	0.974165	1.160705
			γ_{11}	1	-0.063019	0.011630	-0.401689	0.524403
			γ_{12}	1	-0.061203	0.011875	-0.581877	0.802200
		Skewness	α_{10}	0	-0.001549	0.000768	-0.001310	0.017263
			α_{11}	1	-0.051244	0.007253	-0.967317	0.952731
			α_{12}	1	-0.032932	0.004571	-0.963397	0.952814
			π_1	0.25	0.007604	0.001144	0.030443	0.007137
	Component 2	Location	β_{20}	0	0.001108	0.000063	0.008083	0.004737
			β_{21}	-1	-0.000025	0.000268	0.035137	0.022689
			β_{22}	-1	-0.001237	0.000313	0.026257	0.025561
		Scale	γ_{20}	0	-0.024183	0.003017	0.402017	0.211105
			γ_{21}	-1	0.024396	0.004814	-0.018353	0.094900
			γ_{22}	-1	0.027671	0.004721	0.022027	0.106480
		Skewness	α_{20}	0	0.002321	0.000425	0.001715	0.001219
			α_{21}	-1	0.020529	0.002304	0.677500	0.467822
			α_{22}	-1	0.015959	0.002323	0.673768	0.463623
600	Component 1	Location	β_{10}	0	0.001199	0.000063	0.042770	0.048973
			β_{11}	1	0.003195	0.000387	-0.130805	0.202193
			β_{12}	1	-0.000065	0.000424	-0.226833	0.194833
		Scale	γ_{10}	0	-0.013283	0.003021	1.045563	1.180564
			γ_{11}	1	-0.041744	0.006662	-0.374963	0.582185
			γ_{12}	1	-0.065954	0.009973	-0.460080	0.504019
		Skewness	α_{10}	0	0.005057	0.000598	-0.000277	0.003087
			α_{11}	1	-0.032114	0.003190	-0.970636	0.951767
			α_{12}	1	-0.026726	0.002891	-0.959954	0.934348
			π_1	0.25	0.008317	0.000757	0.026867	0.004964
	Component 2	Location	β_{20}	0	-0.000839	0.000034	-0.002218	0.002368
			β_{21}	-1	-0.004053	0.000125	0.006510	0.009653
			β_{22}	-1	-0.001872	0.000107	0.031415	0.010706
		Scale	γ_{20}	0	-0.024593	0.002368	0.415573	0.193907
			γ_{21}	-1	0.030459	0.003143	-0.020431	0.086157
			γ_{22}	-1	0.030171	0.003184	0.039385	0.069028
		Skewness	α_{20}	0	0.001464	0.000311	0.000417	0.000940
			α_{21}	-1	0.023689	0.002314	0.689811	0.480334
			α_{22}	-1	0.018305	0.001650	0.685139	0.472959

Table 2. The bias and the values of MSE for the different sample sizes for Case II of Scenario 1

					SLN		SN	
n		Model	Parameter	True	Bias	MSE	Bias	MSE
200	Component 1	Location	β_{10}	0	0.001109	0.000769	-0.012841	0.066953
			β_{11}	1	-0.016551	0.002856	-0.000516	0.178959
			β_{12}	1	-0.013614	0.003198	0.001712	0.203035
		Scale	γ_{10}	0	-0.235791	0.067199	-0.170880	0.151085
			γ_{11}	1	-0.087698	0.028067	-0.188926	0.526102
			γ_{12}	1	-0.086925	0.029040	-0.145863	0.508284
		Skewness	α_{10}	0	-0.001093	0.002182	0.002699	0.002998
			α_{11}	1	-0.089190	0.018228	-0.824070	0.707861
			α_{12}	1	-0.080104	0.015585	-0.819849	0.702039
			π_1	0.25	-0.005405	0.002303	0.003829	0.003348
	Component 2	Location	β_{20}	0	0.002140	0.000281	0.006141	0.006055
			β_{21}	-1	0.006230	0.001136	0.001387	0.021180
			β_{22}	-1	0.004524	0.001148	-0.006446	0.021494
		Scale	γ_{20}	0	-0.239452	0.062775	-0.099366	0.031959
			γ_{21}	-1	0.040523	0.011194	0.016009	0.063325
			γ_{22}	-1	0.039572	0.011416	0.021868	0.064961
		Skewness	α_{20}	0	0.001430	0.001928	0.000334	0.002344
			α_{21}	-1	0.106507	0.019447	0.637837	0.422275
			α_{22}	-1	0.108943	0.019563	0.637480	0.421178
400	Component 1	Location	β_{10}	0	-0.000349	0.000221	-0.001654	0.021374
			β_{11}	1	-0.008243	0.000978	0.006227	0.064144
			β_{12}	1	-0.009139	0.001029	0.006050	0.068607
		Scale	γ_{10}	0	-0.211798	0.050122	-0.015247	0.049273
			γ_{11}	1	-0.094044	0.017843	-0.184842	0.185828
			γ_{12}	1	-0.087006	0.016954	-0.146736	0.174489
		Skewness	α_{10}	0	0.000254	0.001128	-0.000867	0.000975
			α_{11}	1	-0.084456	0.011285	-0.816848	0.677439
			α_{12}	1	-0.087060	0.012280	-0.824266	0.689911
			π_1	0.25	-0.003995	0.001204	-0.001563	0.001633
	Component 2	Location	β_{20}	0	0.000080	0.000105	0.001491	0.003003
			β_{21}	-1	0.004448	0.000445	0.006867	0.009911
			β_{22}	-1	0.003268	0.000457	-0.002297	0.010119
		Scale	γ_{20}	0	-0.227366	0.054167	-0.054554	0.013307
			γ_{21}	-1	0.043732	0.006399	0.042027	0.035319
			γ_{22}	-1	0.045252	0.006347	0.039654	0.033329
		Skewness	α_{20}	0	0.000499	0.000866	0.000736	0.001093
			α_{21}	-1	0.108800	0.015154	0.657821	0.438906
			α_{22}	-1	0.107182	0.015169	0.658440	0.441105
600	Component 1	Location	β_{10}	0	-0.000560	0.000134	-0.000940	0.013028
			β_{11}	1	-0.007480	0.000516	0.009413	0.042451
			β_{12}	1	-0.007855	0.000514	0.001101	0.036672
		Scale	γ_{10}	0	-0.201989	0.044573	0.043493	0.040170
			γ_{11}	1	-0.091518	0.014095	-0.200900	0.142848
			γ_{12}	1	-0.088876	0.013752	-0.191354	0.139637
		Skewness	α_{10}	0	-0.000809	0.000688	-0.000859	0.000664
			α_{11}	1	-0.084658	0.010356	-0.820236	0.679964
			α_{12}	1	-0.080771	0.009500	-0.818160	0.676026
			π_1	0.25	-0.006843	0.000847	-0.005198	0.001117
	Component 2	Location	β_{20}	0	0.000088	0.000055	0.002358	0.001909
			β_{21}	-1	0.004066	0.000248	0.004324	0.006555
			β_{22}	-1	0.002990	0.000240	-0.001375	0.006185
		Scale	γ_{20}	0	-0.218468	0.049523	-0.034601	0.008564
			γ_{21}	-1	0.047229	0.004806	0.050445	0.023689
			γ_{22}	-1	0.046040	0.004750	0.050332	0.022725
		Skewness	α_{20}	0	0.000434	0.000577	-0.000833	0.000660
			α_{21}	-1	0.105586	0.013574	0.661310	0.441874
			α_{22}	-1	0.104923	0.013422	0.662377	0.443240

Table 3. The bias and the values of MSE for the different sample sizes for Case III of Scenario 1.

					SLN		SN	
n		Model	Parameter	True	Bias	MSE	Bias	MSE
200	Component 1	Location	β_{10}	0	-0.000397	0.000616	0.018731	0.135352
			β_{11}	1	-0.017140	0.003016	-0.180409	0.452858
			β_{12}	1	-0.016814	0.002614	-0.128414	0.425086
		Scale	γ_{10}	0	-0.310774	0.120583	0.357399	0.556072
			γ_{11}	1	-0.104443	0.047028	-0.591648	1.712985
			γ_{12}	1	-0.103863	0.047722	-0.539700	1.769955
		Skewness	α_{10}	0	-0.001789	0.003271	-0.004973	0.010617
			α_{11}	1	-0.129420	0.033552	-0.932381	0.914104
			α_{12}	1	-0.127279	0.031659	-0.930578	0.913499
			π_1	0.25	-0.011568	0.002692	0.021119	0.008344
	Component 2	Location	β_{20}	0	0.000607	0.000227	0.003627	0.008094
			β_{21}	-1	0.007019	0.000883	0.014838	0.038019
			β_{22}	-1	0.007826	0.000918	0.014923	0.021441
		Scale	γ_{20}	0	-0.373253	0.150206	-0.337700	0.214894
			γ_{21}	-1	0.054963	0.019771	0.015238	0.203828
			γ_{22}	-1	0.060928	0.020446	0.011487	0.234463
		Skewness	α_{20}	0	0.001061	0.002267	-0.001227	0.003831
			α_{21}	-1	0.183551	0.043735	0.709829	0.520136
			α_{22}	-1	0.185624	0.045361	0.710655	0.524452
400	Component 1	Location	β_{10}	0	0.001467	0.000226	0.001841	0.047356
			β_{11}	1	-0.015171	0.001109	-0.182278	0.302034
			β_{12}	1	-0.013767	0.001026	-0.161205	0.275155
		Scale	γ_{10}	0	-0.291731	0.103225	0.675286	0.794572
			γ_{11}	1	-0.080879	0.032578	-0.669896	1.174824
			γ_{12}	1	-0.106559	0.034302	-0.747111	1.401166
		Skewness	α_{10}	0	0.001539	0.003459	-0.003024	0.003464
			α_{11}	1	-0.136283	0.026396	-0.971675	0.974001
			α_{12}	1	-0.141216	0.035989	-0.958415	0.946797
			π_1	0.25	-0.012628	0.001533	0.008994	0.008108
	Component 2	Location	β_{20}	0	-0.000443	0.000092	-0.001880	0.003777
			β_{21}	-1	0.006435	0.000330	0.059456	0.055267
			β_{22}	-1	0.007003	0.000383	0.058185	0.055851
		Scale	γ_{20}	0	-0.360828	0.135552	-0.232327	0.133715
			γ_{21}	-1	0.062898	0.011724	0.092793	0.159561
			γ_{22}	-1	0.069038	0.013493	0.115241	0.159555
		Skewness	α_{20}	0	-0.000137	0.001246	0.001095	0.001364
			α_{21}	-1	0.183425	0.038827	0.732075	0.544502
			α_{22}	-1	0.182880	0.039148	0.726522	0.537019
600	Component 1	Location	β_{10}	0	-0.000301	0.000088	-0.049243	0.524876
			β_{11}	1	-0.013445	0.000588	-0.228720	0.976806
			β_{12}	1	-0.011839	0.000468	-0.275677	0.261613
		Scale	γ_{10}	0	-0.281397	0.086396	0.770466	0.822284
			γ_{11}	1	-0.088600	0.019999	-0.745200	0.918232
			γ_{12}	1	-0.104672	0.020942	-0.811049	1.082750
		Skewness	α_{10}	0	-0.001545	0.001552	0.008113	0.011606
			α_{11}	1	-0.138424	0.024753	-0.972339	0.959915
			α_{12}	1	-0.137585	0.026023	-0.980113	0.973411
			π_1	0.25	-0.010421	0.000913	0.018112	0.009126
	Component 2	Location	β_{20}	0	-0.000354	0.000055	0.024777	0.170531
			β_{21}	-1	0.005619	0.000197	0.075376	0.142800
			β_{22}	-1	0.005965	0.000208	0.071837	0.081430
		Scale	γ_{20}	0	-0.349705	0.125276	-0.178856	0.176302
			γ_{21}	-1	0.056785	0.009037	0.073615	0.128455
			γ_{22}	-1	0.059668	0.008717	0.114503	0.134638
		Skewness	α_{20}	0	-0.002565	0.000750	-0.005638	0.011737
			α_{21}	-1	0.186865	0.038309	0.750850	0.574010
			α_{22}	-1	0.174977	0.033985	0.733508	0.551020

Table 4. The bias and the values of MSE for the different sample sizes for Case I of Scenario 2.

					SLN		SN	
n		Model	Parameter	True	Bias	MSE	Bias	MSE
200	Component 1	Location	β_{10}	0	0.000576	0.000279	-0.007611	0.043553
			β_{11}	1	0.000085	0.000996	-0.077230	0.177593
			β_{12}	1	-0.001168	0.001003	-0.086459	0.217542
		Scale	γ_{10}	0	-0.047044	0.011284	0.502081	0.387389
			γ_{11}	1	-0.030875	0.012611	-0.072009	0.398917
			γ_{12}	1	-0.035084	0.011440	-0.102232	0.358123
		Skewness	α_{10}	0	0.001772	0.001240	-0.000974	0.006147
			α_{11}	1	-0.028360	0.006700	-0.791869	0.662071
			α_{12}	1	-0.023361	0.006035	-0.776437	0.621219
			π_1	0.5	0.000425	0.002541	0.003687	0.011088
	Component 2	Location	β_{20}	0	-0.001577	0.000293	0.001691	0.044979
			β_{21}	-1	-0.001644	0.001039	0.029027	0.186494
			β_{22}	-1	-0.000176	0.000886	0.052016	0.124408
		Scale	γ_{20}	0	-0.047673	0.010704	0.507695	0.413876
			γ_{21}	-1	0.043480	0.013851	0.101352	0.531926
			γ_{22}	-1	0.036514	0.013561	0.105172	0.508195
		Skewness	α_{20}	0	-0.000825	0.001265	0.001548	0.002572
			α_{21}	-1	0.025550	0.006343	0.782613	0.633634
			α_{22}	-1	0.029111	0.006238	0.786309	0.634687
400	Component 1	Location	β_{10}	0	-0.000291	0.000104	0.008866	0.015111
			β_{11}	1	-0.001074	0.000416	-0.117399	0.141666
			β_{12}	1	-0.001204	0.000344	-0.099611	0.150786
		Scale	γ_{10}	0	-0.033043	0.004771	0.629843	0.469345
			γ_{11}	1	-0.037432	0.007730	-0.182876	0.354569
			γ_{12}	1	-0.048396	0.006422	-0.207732	0.266259
		Skewness	α_{10}	0	0.001814	0.000654	0.001749	0.001040
			α_{11}	1	-0.028906	0.003516	-0.810502	0.668093
			α_{12}	1	-0.027801	0.003830	-0.802324	0.654907
			π_1	0.5	-0.001597	0.001356	-0.006254	0.007064
	Component 2	Location	β_{20}	0	0.000510	0.000106	0.007376	0.025332
			β_{21}	-1	-0.001176	0.000373	0.117996	0.147458
			β_{22}	-1	-0.000416	0.000391	0.107499	0.117666
		Scale	γ_{20}	0	-0.028700	0.004814	0.619026	0.461890
			γ_{21}	-1	0.039896	0.006090	0.190832	0.256190
			γ_{22}	-1	0.028728	0.005998	0.135455	0.294544
		Skewness	α_{20}	0	-0.000018	0.000468	-0.000837	0.000985
			α_{21}	-1	0.024436	0.003584	0.803397	0.654470
			α_{22}	-1	0.027482	0.003377	0.807567	0.662511
600	Component 1	Location	β_{10}	0	-0.000412	0.000045	-0.003796	0.006020
			β_{11}	1	0.000365	0.000195	-0.112051	0.051485
			β_{12}	1	0.001325	0.000208	-0.099321	0.042431
		Scale	γ_{10}	0	-0.033643	0.035067	0.699558	0.546034
			γ_{11}	1	-0.029745	0.005856	-0.178957	0.190350
			γ_{12}	1	-0.031489	0.005236	-0.168946	0.176055
		Skewness	α_{10}	0	0.003812	0.000761	0.004628	0.000809
			α_{11}	1	-0.036921	0.023650	-0.809107	0.659929
			α_{12}	1	-0.039523	0.040666	-0.800467	0.646277
			π_1	0.5	0.000497	0.001032	-0.000007	0.004664
	Component 2	Location	β_{20}	0	0.000744	0.000065	0.003492	0.009150
			β_{21}	-1	-0.000409	0.000209	0.096864	0.042253
			β_{22}	-1	-0.000083	0.000192	0.079985	0.045150
		Scale	γ_{20}	0	-0.013648	0.002221	0.696584	0.542612
			γ_{21}	-1	0.043055	0.004942	0.201331	0.159944
			γ_{22}	-1	0.039937	0.005640	0.179940	0.168125
		Skewness	α_{20}	0	-0.000130	0.000360	-0.003756	0.000564
			α_{21}	-1	0.020215	0.001864	0.798827	0.642789
			α_{22}	-1	0.024863	0.002338	0.806310	0.654582

Table 4. The bias and the values of MSE for the different sample sizes for Case II of Scenario 2.

					SLN		SN	
n		Model	Parameter	True	Bias	MSE	Bias	MSE
200	Component 1	Location	β_{10}	0	0.000400	0.000394	0.002326	0.013972
			β_{11}	1	-0.007895	0.001607	0.007523	0.044839
			β_{12}	1	-0.005967	0.001589	0.018829	0.045107
		Scale	γ_{10}	0	-0.228935	0.059330	-0.118107	0.058020
			γ_{11}	1	-0.067914	0.016118	-0.040547	0.136666
			γ_{12}	1	-0.061875	0.017052	-0.028742	0.143562
		Skewness	α_{10}	0	0.000616	0.001829	-0.002269	0.002161
			α_{11}	1	-0.089356	0.015649	-0.726383	0.543608
			α_{12}	1	-0.094598	0.016173	-0.729759	0.549161
			π_1	0.25	0.000624	0.002814	-0.001949	0.004484
	Component 2	Location	β_{20}	0	0.001630	0.000375	0.010066	0.013562
			β_{21}	-1	0.006655	0.001506	-0.012123	0.044529
			β_{22}	-1	0.005553	0.001416	-0.012485	0.049339
		Scale	γ_{20}	0	-0.227995	0.058990	-0.100487	0.054764
			γ_{21}	-1	0.065551	0.015969	0.061956	0.137695
			γ_{22}	-1	0.061699	0.015080	0.053632	0.130893
		Skewness	α_{20}	0	0.001083	0.002060	-0.001189	0.001897
			α_{21}	-1	0.094128	0.016714	0.730903	0.550750
			α_{22}	-1	0.092514	0.016357	0.728376	0.545504
400	Component 1	Location	β_{10}	0	0.000311	0.000153	0.002448	0.006716
			β_{11}	1	-0.004760	0.000515	0.004439	0.019050
			β_{12}	1	-0.004910	0.000549	0.004550	0.020130
		Scale	γ_{10}	0	-0.211273	0.048133	-0.029562	0.022327
			γ_{11}	1	-0.065168	0.009642	-0.069854	0.066099
			γ_{12}	1	-0.066527	0.010520	-0.084333	0.077434
		Skewness	α_{10}	0	0.001008	0.000968	0.001322	0.000896
			α_{11}	1	-0.094340	0.012150	-0.739738	0.554068
			α_{12}	1	-0.090583	0.011694	-0.731974	0.542744
			π_1	0.25	0.000250	0.001496	0.000198	0.002163
	Component 2	Location	β_{20}	0	0.000117	0.000134	0.002080	0.006601
			β_{21}	-1	0.006370	0.000536	-0.005312	0.018605
			β_{22}	-1	0.004913	0.000511	-0.001585	0.020120
		Scale	γ_{20}	0	-0.211647	0.048038	-0.037857	0.021406
			γ_{21}	-1	0.068266	0.010291	0.077592	0.065552
			γ_{22}	-1	0.073354	0.010531	0.096149	0.071954
		Skewness	α_{20}	0	0.000172	0.000918	-0.001406	0.000911
			α_{21}	-1	0.089055	0.011358	0.731854	0.542872
			α_{22}	-1	0.093250	0.012096	0.737231	0.550062
600	Component 1	Location	β_{10}	0	-0.000346	0.000081	-0.017651	0.305096
			β_{11}	1	-0.007140	0.000337	-0.006654	0.030378
			β_{12}	1	-0.006254	0.000343	-0.011107	0.038824
		Scale	γ_{10}	0	-0.208837	0.045863	-0.009180	0.123349
			γ_{11}	1	-0.064483	0.007912	-0.074416	0.055170
			γ_{12}	1	-0.067862	0.007913	-0.092083	0.064276
		Skewness	α_{10}	0	-0.000028	0.000608	0.013843	0.199475
			α_{11}	1	-0.093164	0.011070	-0.728692	0.559871
			α_{12}	1	-0.090784	0.010699	-0.723645	0.587739
			π_1	0.25	-0.000429	0.001045	0.000656	0.002116
	Component 2	Location	β_{20}	0	-0.000450	0.000083	-0.001506	0.005639
			β_{21}	-1	0.005443	0.000336	-0.001007	0.013321
			β_{22}	-1	0.006094	0.000353	0.004707	0.014236
		Scale	γ_{20}	0	-0.207342	0.045030	-0.021564	0.017919
			γ_{21}	-1	0.066245	0.008025	0.077442	0.043951
			γ_{22}	-1	0.064901	0.007717	0.071057	0.047627
		Skewness	α_{20}	0	-0.001343	0.000572	-0.002810	0.001889
			α_{21}	-1	0.093732	0.011206	0.734284	0.552744
			α_{22}	-1	0.092198	0.010544	0.733229	0.549518

Table 6. The bias and the values of MSE for the different sample sizes for Case III of Scenario 2.

					SLN		SN	
n		Model	Parameter	True	Bias	MSE	Bias	MSE
200	Component 1	Location	β_{10}	0	0.000369	0.000320	0.009921	0.103064
			β_{11}	1	-0.011891	0.001455	-0.069605	0.175460
			β_{12}	1	-0.010132	0.001192	-0.074488	0.181750
		Scale	γ_{10}	0	-0.327350	0.120565	-0.009022	0.282907
			γ_{11}	1	-0.066364	0.024149	-0.125100	0.568277
			γ_{12}	1	-0.061852	0.025730	-0.057684	0.621618
		Skewness	α_{10}	0	0.000000	0.002518	-0.000699	0.024888
			α_{11}	1	-0.151068	0.034839	-0.795599	0.663874
			α_{12}	1	-0.156801	0.035730	-0.809014	0.720977
			π_1	0.25	-0.000261	0.003117	-0.005383	0.012187
	Component 2	Location	β_{20}	0	-0.000458	0.000346	-0.002735	0.033440
			β_{21}	-1	0.008763	0.001203	0.057113	0.138349
			β_{22}	-1	0.009424	0.001263	0.052256	0.132388
		Scale	γ_{20}	0	-0.336142	0.144971	-0.011376	0.209101
			γ_{21}	-1	0.093657	0.042260	0.179458	0.581779
			γ_{22}	-1	0.090997	0.044242	0.180798	0.685933
		Skewness	α_{20}	0	-0.001632	0.003017	0.001293	0.002923
			α_{21}	-1	0.154233	0.053127	0.800751	0.662200
			α_{22}	-1	0.151986	0.043501	0.802162	0.663834
400	Component 1	Location	β_{10}	0	-0.000528	0.000115	-0.009012	0.083574
			β_{11}	1	-0.008848	0.000516	-0.122024	0.206062
			β_{12}	1	-0.008340	0.000541	-0.103521	0.202619
		Scale	γ_{10}	0	-0.304696	0.134916	0.119134	0.192750
			γ_{11}	1	-0.077674	0.015681	-0.229209	0.337922
			γ_{12}	1	-0.083622	0.023190	-0.264490	0.408142
		Skewness	α_{10}	0	0.001182	0.001560	0.002757	0.006455
			α_{11}	1	-0.150202	0.044357	-0.822553	0.698361
			α_{12}	1	-0.147265	0.039669	-0.806582	0.661021
			π_1	0.25	0.001665	0.001536	0.003344	0.009825
	Component 2	Location	β_{20}	0	-0.000055	0.000104	0.046246	0.167193
			β_{21}	-1	0.008269	0.000507	0.074121	0.121277
			β_{22}	-1	0.010226	0.000593	0.097022	0.206187
		Scale	γ_{20}	0	-0.318470	0.108036	0.127403	0.193529
			γ_{21}	-1	0.073210	0.015581	0.204215	0.364476
			γ_{22}	-1	0.073182	0.015678	0.238675	0.477829
		Skewness	α_{20}	0	0.001753	0.001097	-0.002325	0.003442
			α_{21}	-1	0.154353	0.029965	0.818965	0.688081
			α_{22}	-1	0.157860	0.030440	0.815663	0.678820
600	Component 1	Location	β_{10}	0	0.001204	0.000066	-0.029192	0.183424
			β_{11}	1	-0.007370	0.000300	-0.058255	0.123120
			β_{12}	1	-0.007523	0.000319	-0.086713	0.169070
		Scale	γ_{10}	0	-0.304066	0.097874	0.228212	0.283121
			γ_{11}	1	-0.063873	0.012985	-0.262691	0.376105
			γ_{12}	1	-0.063463	0.011449	-0.244429	0.386810
		Skewness	α_{10}	0	-0.001203	0.000975	0.014326	0.019875
			α_{11}	1	-0.155690	0.029917	-0.820609	0.698134
			α_{12}	1	-0.151315	0.028539	-0.830182	0.723467
			π_1	0.25	0.001196	0.000958	-0.004667	0.010895
	Component 2	Location	β_{20}	0	-0.000436	0.000068	0.011472	0.110872
			β_{21}	-1	0.008005	0.000317	0.083733	0.100941
			β_{22}	-1	0.009288	0.000312	0.108439	0.067094
		Scale	γ_{20}	0	-0.306306	0.100791	0.231837	0.273336
			γ_{21}	-1	0.076241	0.012435	0.327785	0.378631
			γ_{22}	-1	0.070747	0.011809	0.315733	0.461707
		Skewness	α_{20}	0	-0.001390	0.001475	-0.004572	0.007106
			α_{21}	-1	0.156591	0.032026	0.823369	0.697389
			α_{22}	-1	0.153986	0.030491	0.820885	0.688152

6.2. Real data example

We apply the proposed method for the analysis of the “Pinus Nigra” tree data set. This data set was given by García-Escudero et al. (2010) for the robust clusterwise linear regression using trimming. Also, this data set was investigated by Doğru and Arslan (2018a) for the robust mixture regression modelling based on the least trimmed squares estimation method. The data set includes heights (in meters) and diameters (in millimeters) of 362 trees, which form in a cultivated forest of Pinus Nigra located in the north of Palencia (Spain). Figure 1 gives the scatter plot of the “Pinus Nigra” tree data set and the histogram of the heights. It was pointed out by García-Escudero et al. (2010) that there are three groups in the data set and also some outliers on the top right corner and one isolated point on the bottom right corner. We can also observe this from Figure 1(a). Overall, Figure 1 shows that this data set may contain heteroscedasticity and skewness in different subgroups because of its heterogeneous structure. Therefore, there is desirable to analyze this data by the joint location, scale and skewness models of mixtures of SLN distributions or SN distributions. We then compare the performance of joint location, scale and skewness models of mixtures of SLN distributions with the joint location, scale and skewness models of mixtures of SN distributions, based on the following information criteria:

$$-2\ell(\widehat{\boldsymbol{\theta}}) + mc_n ,$$

where $\ell(\cdot)$ represents the maximized log-likelihood, m is the number of free parameters to be estimated in the model and c_n is the penalty term. Here, we take $c_n = 2$ for the Akaike information criteria (AIC) (Akaike (1973)), $c_n = \log(n)$ for the Bayesian information criteria (BIC) (Schwarz (1978)) and $c_n = 0.2\sqrt{n}$ for the efficient determination criteria (EDC) (Bai et al. (1989)).

Table 3 shows the estimates and the corresponding standard errors (SEs) for the parameters of the three components obtained from the joint location, scale and skewness models of mixtures of SN and SLN distributions, respectively. The SEs of estimators are computed using the Fisher information-based method given by Basford et al. (1997), see the details of computation of the SEs for the ML estimators of joint location, scale and skewness models of mixtures of SLN distributions in section 5.2. In the table, we also provide the information criteria to assess the performance of fitted models. We observe that the results obtained from the joint location, scale and skewness models of mixtures of SLN distributions are significantly superior to the results obtained from the joint location, scale and skewness models of mixtures of SN distributions. In addition, Figure 2 displays the fitted regression lines on the scatter plot of the data. These fitted lines also confirm the superiority of the SLN fits over the SN fits. It can be seen that unlike the SLN fits, the SN fits are ruined by the outliers.

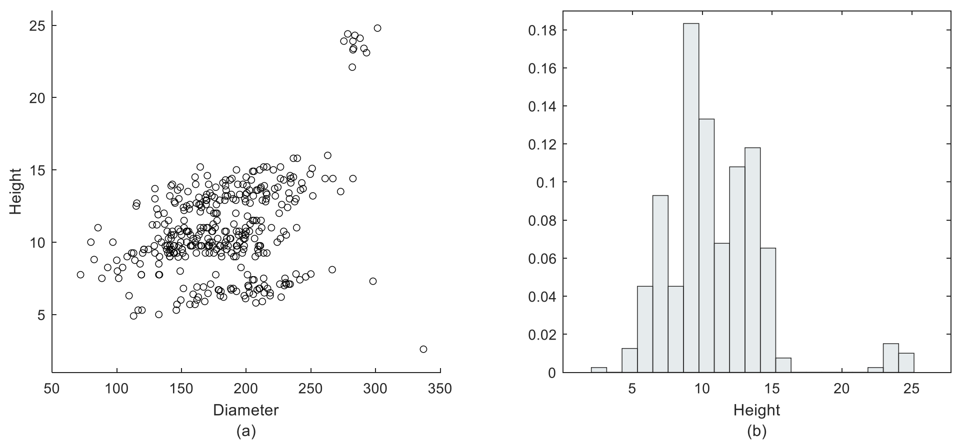

Figure 1. (a) The scatterplot of the “Pinus Nigra” tree data set. (b) Histogram of the height.

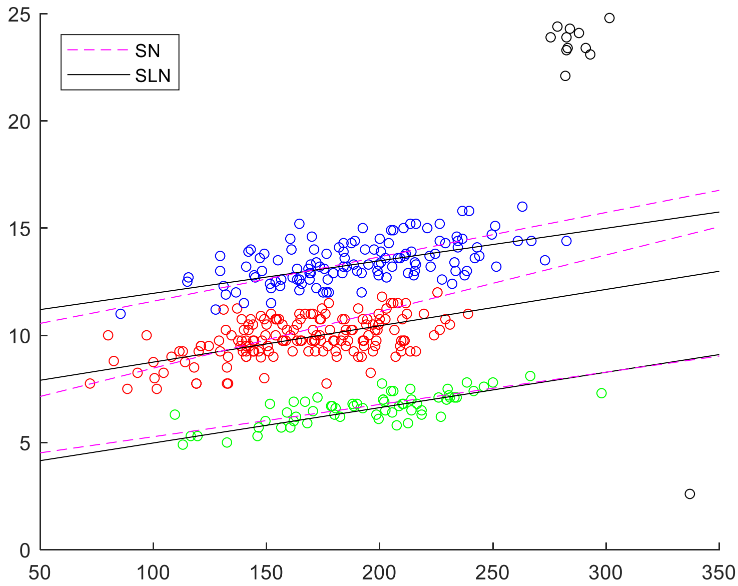

Figure 2. The scatter plot of the data set along with the fitted regression lines gained from joint location, scale and skewness models of mixtures of SLN and SN distributions

Table 3. Estimation results for the "Pinus Nigra" tree data set

		SN		SLN	
Model	Parameter	Estimate	SE	Estimate	SE
Location	π_1	0.156582	0.031161	0.164801	0.024763
	π_2	0.584023	0.050864	0.563643	0.056175
	β_{10}	3.764876	37.131904	3.327843	0.507334
	β_{11}	0.015053	0. 251693	0.016494	0.002568
	β_{20}	5.829494	0. 415441	7.055084	0.332883
	β_{21}	0.026380	0.008871	0.016949	0.002414
	β_{30}	9.521251	16.989675	10.440033	0.723771
	β_{31}	0.020691	0.055243	0.015172	0.003698
Scale	γ_{10}	-2.790566	1.866088	-2.263616	1.754836
	γ_{11}	0.012334	0.010113	0.002901	0.009652
	γ_{20}	-4.588224	1.140637	-5.128381	0.886256
	γ_{21}	0.029300	0.006261	0.030586	0.004570
	γ_{30}	-0.573953	2.300905	0.619260	1.860537
	γ_{31}	0.003117	0.012462	-0.006081	0.009573
Skewness	α_{10}	-0.000159	266.001222	0.712341	1.802666
	α_{11}	0.0000006	0.523874	-0.002423	0.008874
	α_{20}	-0.237599	3.544586	-0.765644	0.654246
	α_{21}	0.001462	0.011849	0.006205	0.003989
	α_{30}	0.039622	33.521359	-0.327859	1.185471
	α_{31}	-0.000223	0.085429	0.001229	0.006333
Information criteria	$\ell(\hat{\boldsymbol{\Theta}})$	-809.1278		**-796.1866**	
	AIC	1634.2556		**1608.3731**	
	BIC	1653.8883		**1639.5063**	
	EDC	1648.6977		**1622.8152**	

7. Conclusions

In this paper, we propose the joint modelling of location, scale and skewness parameters of mixtures of SLN distributions for modelling heteroscedastic skew-heavy tailed data set coming from a heterogeneous population., which could be regarded as an alternative mixture model to the joint modelling of location, scale and skewness parameters of mixtures of SN distributions. We obtain the ML estimates of parameters using the EM algorithm and investigated the asymptotic properties of the estimates. Simulation study and a real data analysis show that the proposed model and method is applicable in practice and the derived estimators of parameters are superior to the estimators obtained from the joint modelling of location, scale and skewness parameters of mixtures of SN distributions, as well as better model fitting. In general, we may conclude this newly proposed model is useful for modelling heterogeneous data sets that may face with heteroscedasticity, asymmetry and heavy-tailedness problems.

Acknowledgments

This study is supported by "The Scientific and Technological Research Council of Turkey (TUBITAK)" (Grant Number: 1059B191700233) as part of "2219-International Postdoctoral Research Scholarship Programme". This research was conducted when the first author visited Brunel University London. The authors would like to thank the university which hosted the visit. Finally, the authros thank two anonymous referees and the associate editor for their thoughtful suggestions that greatly improved the paper.

Appendix

A1. Score function and Fisher information matrix:
Using the objective function given in (19), we obtain the score function of the ith component

$$G(\boldsymbol{\theta}_i) = \frac{\partial Q(\boldsymbol{\theta}_i; \widehat{\boldsymbol{\theta}}_i)}{\partial \boldsymbol{\theta}_i} = \left(G_1^T(\boldsymbol{\beta}_i), G_2^T(\boldsymbol{\gamma}_i), G_3^T(\boldsymbol{\alpha}_i)\right)^T,$$

where

$$G_1(\boldsymbol{\beta}_i) = \sum_{j=1}^{n} \hat{z}_{ij} \left(\frac{(y_j - \boldsymbol{x}_j^T \boldsymbol{\beta}_i)\boldsymbol{x}_j}{e^{\boldsymbol{h}_j^T \boldsymbol{\gamma}_i}} \left(\hat{v}_{ij} + (\boldsymbol{w}_j^T \boldsymbol{\alpha}_i)^2\right) - \frac{(\boldsymbol{w}_j^T \boldsymbol{\alpha}_i)\boldsymbol{x}_j \hat{u}_{1ij}}{e^{\boldsymbol{h}_j^T \boldsymbol{\gamma}_i/2}} \right), \tag{28}$$

$$G_2(\boldsymbol{\gamma}_i) = \sum_{j=1}^{n} \hat{z}_{ij} \left(-\frac{1}{2}\boldsymbol{h}_j + \frac{1}{2} \frac{(y_j - \boldsymbol{x}_j^T \boldsymbol{\beta}_i)^2 \boldsymbol{h}_j}{e^{\boldsymbol{h}_j^T \boldsymbol{\gamma}_i}} \left(\hat{v}_{ij} + (\boldsymbol{w}_j^T \boldsymbol{\alpha}_i)^2\right), - \frac{1}{2} \frac{(\boldsymbol{w}_i^T \boldsymbol{\alpha}_i)(y_j - \boldsymbol{x}_j^T \boldsymbol{\beta}_i)\boldsymbol{h}_j \hat{u}_{1ij}}{e^{\boldsymbol{h}_j^T \boldsymbol{\gamma}_i/2}} \right) \tag{29}$$

$$G_3(\boldsymbol{\alpha}_i) = \sum_{j=1}^{n} \hat{z}_{ij} \left(\frac{(y_j - \boldsymbol{x}_j^T \boldsymbol{\beta}_i)\boldsymbol{w}_j \hat{u}_{1ij}}{e^{\boldsymbol{h}_j^T \boldsymbol{\gamma}_i/2}} - \frac{(\boldsymbol{w}_j^T \boldsymbol{\alpha}_i)(y_j - \boldsymbol{x}_j^T \boldsymbol{\beta}_i)^2 \boldsymbol{w}_j}{e^{\boldsymbol{h}_j^T \boldsymbol{\gamma}_i}} \right), \tag{30}$$

and the observed Fisher information matrix of the ith component

$$H(\boldsymbol{\theta}_i) = \frac{\partial^2 Q(\boldsymbol{\theta}_i; \widehat{\boldsymbol{\theta}}_i)}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_i^T} = \begin{bmatrix} \frac{\partial^2 Q(\boldsymbol{\theta}_i; \widehat{\boldsymbol{\theta}}_i)}{\partial \boldsymbol{\beta}_i \partial \boldsymbol{\beta}_i^T} & \frac{\partial^2 Q(\boldsymbol{\theta}_i; \widehat{\boldsymbol{\theta}}_i)}{\partial \boldsymbol{\beta}_i \partial \boldsymbol{\gamma}_i^T} & \frac{\partial^2 Q(\boldsymbol{\theta}_i; \widehat{\boldsymbol{\theta}}_i)}{\partial \boldsymbol{\beta}_i \partial \boldsymbol{\alpha}_i^T} \\ \frac{\partial^2 Q(\boldsymbol{\theta}_i; \widehat{\boldsymbol{\theta}}_i)}{\partial \boldsymbol{\gamma}_i \partial \boldsymbol{\beta}_i^T} & \frac{\partial^2 Q(\boldsymbol{\theta}_i; \widehat{\boldsymbol{\theta}}_i)}{\partial \boldsymbol{\gamma}_i \partial \boldsymbol{\gamma}_i^T} & \frac{\partial^2 Q(\boldsymbol{\theta}_i; \widehat{\boldsymbol{\theta}}_i)}{\partial \boldsymbol{\gamma}_i \partial \boldsymbol{\alpha}_i^T} \\ \frac{\partial^2 Q(\boldsymbol{\theta}_i; \widehat{\boldsymbol{\theta}}_i)}{\partial \boldsymbol{\alpha}_i \partial \boldsymbol{\beta}_i^T} & \frac{\partial^2 Q(\boldsymbol{\theta}_i; \widehat{\boldsymbol{\theta}}_i)}{\partial \boldsymbol{\alpha}_i \partial \boldsymbol{\gamma}_i^T} & \frac{\partial^2 Q(\boldsymbol{\theta}_i; \widehat{\boldsymbol{\theta}}_i)}{\partial \boldsymbol{\alpha}_i \partial \boldsymbol{\alpha}_i^T} \end{bmatrix},$$

where

$$\frac{\partial^2 Q(\boldsymbol{\theta}_i;\widehat{\boldsymbol{\theta}}_i)}{\partial \boldsymbol{\beta}_i \partial \boldsymbol{\beta}_i{}^T} = -\sum_{j=1}^{n} \hat{z}_{ij}\left(\frac{\boldsymbol{x}_j\boldsymbol{x}_j^T}{e^{\boldsymbol{h}_j^T\boldsymbol{\gamma}_i}}\hat{v}_{ij} + \frac{(\boldsymbol{w}_i^T\boldsymbol{\alpha}_i)^2}{e^{\boldsymbol{h}_j^T\boldsymbol{\gamma}_i}}\boldsymbol{x}_j\boldsymbol{x}_j^T\right),$$

$$\frac{\partial^2 Q(\boldsymbol{\theta}_i;\widehat{\boldsymbol{\theta}}_i)}{\partial \boldsymbol{\beta}_i \partial \boldsymbol{\gamma}_i{}^T} = -\sum_{i=1}^{n} \hat{z}_{ij}\left(\frac{(y_j - \boldsymbol{x}_j^T\boldsymbol{\beta}_i)\boldsymbol{x}_j\boldsymbol{h}_j^T}{e^{\boldsymbol{h}_j^T\boldsymbol{\gamma}_i}}\left(\hat{v}_{ij} + (\boldsymbol{w}_j^T\boldsymbol{\alpha}_i)^2\right) - \frac{1}{2}\frac{(\boldsymbol{w}_j^T\boldsymbol{\alpha}_i)\boldsymbol{x}_j\boldsymbol{h}_j^T\hat{u}_{1ij}}{e^{\boldsymbol{h}_j^T\boldsymbol{\gamma}_i/2}}\right),$$

$$\frac{\partial^2 Q(\boldsymbol{\theta}_i;\widehat{\boldsymbol{\theta}}_i)}{\partial \boldsymbol{\beta}_i \partial \boldsymbol{\alpha}_i{}^T} = -\sum_{i=1}^{n} \hat{z}_{ij}\left(\frac{\boldsymbol{x}_j\boldsymbol{w}_j^T\hat{u}_{1ij}}{e^{\boldsymbol{h}_j^T\boldsymbol{\gamma}_i/2}} - 2\frac{(\boldsymbol{w}_i^T\boldsymbol{\alpha}_i)(y_j - \boldsymbol{x}_j^T\boldsymbol{\beta}_i)\boldsymbol{x}_j\boldsymbol{w}_j^T}{e^{\boldsymbol{h}_j^T\boldsymbol{\gamma}_i}}\right),$$

$$\frac{\partial^2 Q(\boldsymbol{\theta}_i;\widehat{\boldsymbol{\theta}}_i)}{\partial \boldsymbol{\gamma}_i \partial \boldsymbol{\beta}_i{}^T} = -\sum_{i=1}^{n} \hat{z}_{ij}\left(\frac{(y_j - \boldsymbol{x}_j^T\boldsymbol{\beta}_i)\boldsymbol{h}_j\boldsymbol{x}_j^T}{e^{\boldsymbol{h}_j^T\boldsymbol{\gamma}_i}}\left(\hat{v}_{ij} + (\boldsymbol{w}_j^T\boldsymbol{\alpha}_i)^2\right) - \frac{1}{2}\frac{(\boldsymbol{w}_j^T\boldsymbol{\alpha}_i)\boldsymbol{h}_j\boldsymbol{x}_j^T\hat{u}_{1ij}}{e^{\boldsymbol{h}_j^T\boldsymbol{\gamma}_i/2}}\right),$$

$$\frac{\partial^2 Q(\boldsymbol{\theta}_i;\widehat{\boldsymbol{\theta}}_i)}{\partial \boldsymbol{\gamma}_i \partial \boldsymbol{\gamma}_i{}^T} = -\sum_{i=1}^{n} \hat{z}_{ij}\left(\frac{1}{2}\frac{(y_j - \boldsymbol{x}_j^T\boldsymbol{\beta}_i)^2\boldsymbol{h}_j\boldsymbol{h}_j^T}{e^{\boldsymbol{h}_j^T\boldsymbol{\gamma}_i}}\left(\hat{v}_{ij} + (\boldsymbol{w}_j^T\boldsymbol{\alpha}_i)^2\right) - \frac{1}{4}\frac{(\boldsymbol{w}_j^T\boldsymbol{\alpha}_i)(y_j - \boldsymbol{x}_j^T\boldsymbol{\beta}_i)\boldsymbol{h}_j\boldsymbol{h}_j^T\hat{u}_{1ij}}{e^{\boldsymbol{h}_j^T\boldsymbol{\gamma}_i/2}}\right),$$

$$\frac{\partial^2 Q(\boldsymbol{\theta}_i;\widehat{\boldsymbol{\theta}}_i)}{\partial \boldsymbol{\gamma}_i \partial \boldsymbol{\alpha}_i{}^T} = -\sum_{i=1}^{n} \hat{z}_{ij}\left(\frac{1}{2}\frac{(y_j - \boldsymbol{x}_j^T\boldsymbol{\beta}_i)\boldsymbol{h}_j\boldsymbol{w}_j^T\hat{u}_{1ij}}{e^{\boldsymbol{h}_j^T\boldsymbol{\gamma}_i/2}} - \frac{(\boldsymbol{w}_j^T\boldsymbol{\alpha}_i)(y_j - \boldsymbol{x}_j^T\boldsymbol{\beta}_i)^2\boldsymbol{h}_j\boldsymbol{w}_j^T}{e^{\boldsymbol{h}_j^T\boldsymbol{\gamma}_i}}\right),$$

$$\frac{\partial^2 Q(\boldsymbol{\theta}_i;\widehat{\boldsymbol{\theta}}_i)}{\partial \boldsymbol{\alpha}_i \partial \boldsymbol{\beta}_i{}^T} = -\sum_{i=1}^{n} \hat{z}_{ij}\left(\frac{\boldsymbol{w}_j\boldsymbol{x}_j^T\hat{u}_{1ij}}{e^{\boldsymbol{h}_j^T\boldsymbol{\gamma}_i/2}} - 2\frac{(\boldsymbol{w}_j^T\boldsymbol{\alpha}_i)(y_j - \boldsymbol{x}_j^T\boldsymbol{\beta}_i)\boldsymbol{w}_j\boldsymbol{x}_j^T}{e^{\boldsymbol{h}_j^T\boldsymbol{\gamma}_i}}\right),$$

$$\frac{\partial^2 Q(\boldsymbol{\theta};\widehat{\boldsymbol{\theta}}_i)}{\partial \boldsymbol{\alpha}_i \partial \boldsymbol{\gamma}_i{}^T} = -\sum_{i=1}^{n} \hat{z}_{ij}\left(\frac{1}{2}\frac{(y_j - \boldsymbol{x}_j^T\boldsymbol{\beta}_i)\boldsymbol{w}_j\boldsymbol{h}_j^T\hat{u}_{1ij}}{e^{\boldsymbol{h}_j^T\boldsymbol{\gamma}_i/2}} - \frac{(\boldsymbol{w}_j^T\boldsymbol{\alpha}_i)(y_j - \boldsymbol{x}_j^T\boldsymbol{\beta}_i)^2\boldsymbol{w}_j\boldsymbol{h}_j^T}{e^{\boldsymbol{h}_j^T\boldsymbol{\gamma}_i}}\right),$$

$$\frac{\partial^2 Q(\boldsymbol{\theta};\widehat{\boldsymbol{\theta}}_i)}{\partial \boldsymbol{\alpha}_i \partial \boldsymbol{\alpha}_i{}^T} = -\sum_{i=1}^{n} \hat{z}_{ij}\left(\frac{(y_j - \boldsymbol{x}_j^T\boldsymbol{\beta}_i)^2\boldsymbol{w}_j\boldsymbol{w}_j^T}{e^{\boldsymbol{h}_j^T\boldsymbol{\gamma}_i}}\right).$$

A2. Proof of theorems:

In this part, we summarize the necessary conditions for the consistency and asymptotic distribution of $\widehat{\boldsymbol{\Theta}}$. See Kiefer (1978), Peters and Walker (1978), Redner and Walker (1984), McLachlan and Peel (2000), Cheng and Liu (2001) and Tan et al. (2007) for details about the consistency and asymptotic properties of mixture models. We also give the proofs of Theorems 1 and 2. We follow the consistency procedure for the mixture models given in Cheng and Liu (2001) which they extended the classic consistency inferences given in Wald (1949). Also, we follow Tan et al. (2007) for the proof of Theorem 2.

Let L^1 and B^+ be the spaces of integrable functions on the interval $(-\infty,\infty)$ as given below:

$$L^1 = \left\{f: f\ measureable, \|f\| = \int_{-\infty}^{\infty}|f| < \infty\right\},$$
$$B^+ = \{f: f \in L^1, \|f\| = 1, f \geq 0\}.$$

Let $f_1, f_2 \in L^1$. Then, $f_1 = f_2$ in L^1 if and only if $f_1(x) = f_2(x)$ almost everywhere in R^1. Let A_1 and A_2 be two closed sets in R^m. A metric between the two sets can be defined as:

$$dis(\mathrm{A}_1, \mathrm{A}_2) = dis(\mathrm{A}_2, \mathrm{A}_1) = \inf_{y\in \mathrm{A}_2}\inf_{x\in \mathrm{A}_1}|x - y|.$$

We note that if A_1 and A_2 are singleton sets (i.e. single points), this metric turns the Euclidian distance.

Property 1. i) $dis(\mathrm{A}_1, \mathrm{A}_2) = 0$ if and only if there are sequences of points, $\{x_n\}$ in A_1 and $\{y_n\}$ in A_2, such that $|x_n - y_n| \to 0$ as $n \to \infty$.
ii) $dis(x_n, \mathrm{A}) \to 0$ if and only if there is a sequence $\{y_n\}$ of points in A, such that $|x_n - y_n| \to 0$ as $n \to \infty$.
Note that Property 1 will be used in the proof of Theorem 1.

Conditions:
1. The sample is independent and identically distributed from $(\boldsymbol{x}, \boldsymbol{h}, \boldsymbol{w}, y)$. The density $f(y|\boldsymbol{\Theta})$ given in (8) is identifiable. See Definition 1 for identifiability.
2. There is a neighborhood Ω of $\boldsymbol{\Theta}^0$ that for all $\boldsymbol{\Theta} \in \Omega$ and for almost all $y \in \mathbb{R}^n$. Then, the partial derivatives $\partial \mathrm{f}(y|\boldsymbol{\Theta})/\partial \boldsymbol{\Theta}_i$, $\partial^2 \mathrm{f}(y|\boldsymbol{\Theta})/\partial \boldsymbol{\Theta}_i\, \partial \boldsymbol{\Theta}_j$ and $\partial^3 \mathrm{f}(y|\boldsymbol{\Theta})/\partial \boldsymbol{\Theta}_i\, \partial \boldsymbol{\Theta}_j\, \partial \boldsymbol{\Theta}_k$ exist and satisfy

$$\left|\frac{\partial f(y|\boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}_i}\right| \leq f_i(y), \left|\frac{\partial^2 f(y|\boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}_i \partial \boldsymbol{\Theta}_j}\right| \leq f_{ij}(y), \left|\frac{\partial^3 f(y|\boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}_i \partial \boldsymbol{\Theta}_j \partial \boldsymbol{\Theta}_k}\right| \leq f_{ijk}(y),$$

where f_i and f_{ij} are integrable and f_{ijk} satisfies

$$\int_{\mathbb{R}^n} f_{ijk}(y)\, f(y|\boldsymbol{\Theta}^0) dy < \infty.$$

3. The Fisher information matrix

$$I(\boldsymbol{\Theta}) = \int_{R^n} \frac{\partial f(y|\boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}_i} \frac{\partial f(y|\boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}_j} f(y|\boldsymbol{\Theta}) dy$$

is well defined and positive definite at $\boldsymbol{\Theta}^0$.

4. $f_i(.,\boldsymbol{\theta}_i) \in B^+$, for any $\boldsymbol{\theta}_i \in \boldsymbol{\Theta}$, $1 \leq i \leq k$, and the support of f_i is independent of θ_i. Furthermore, $f_i(.,\boldsymbol{\theta}_i^1) = f_i(.,\boldsymbol{\theta}_i^2)$ in B^+ only if $\boldsymbol{\theta}_i^1 = \boldsymbol{\theta}_i^2$.
5. Let $1 \leq i \leq k$, and $\eta_i(y, \boldsymbol{\theta}_i) = \max\{f_i(y, \boldsymbol{\theta}_i), 1\}$. For any $\boldsymbol{\theta}_i \in \boldsymbol{\Theta}_i$,

$$E_{\boldsymbol{\theta}_i^0}[\log\{f_i(y, \boldsymbol{\theta}_i)\}] > -\infty,$$

on the support of f_i, and

$$E_{\boldsymbol{\theta}_i^0}[\log\{\eta_i(y, \boldsymbol{\theta}_i)\}] < \infty.$$

Also,

$$E_{\boldsymbol{\theta}_i^0}\left[\log\left\{\sup_{\boldsymbol{\theta}_i \in \Theta_i, |\boldsymbol{\theta}_i - \boldsymbol{\theta}_i^0| \leq \rho} \eta_i(y, \boldsymbol{\theta}_i)\right\}\right] < \infty,$$

for $\rho > 0$ sufficiently small, and

$$E_{\boldsymbol{\theta}_i^0}\left[\log\left\{\sup_{\boldsymbol{\theta}_i \in \Theta_i, |\boldsymbol{\theta}_i| > r > 0} \eta_i(y, \boldsymbol{\theta}_i)\right\}\right] < \infty,$$

for r sufficiently large.
6. Let $1 \leq i \leq k$. For almost every fixed $x \in R$, $\lim_{|\boldsymbol{\theta}_i| \to \infty} \eta_i(y, \boldsymbol{\theta}_i) = 0$. If $\boldsymbol{\theta}_i, \boldsymbol{\theta}_i^0 \in \boldsymbol{\Theta}_i$,

$$\lim_{\boldsymbol{\theta}_i \to \boldsymbol{\theta}_i^0} \eta_i(y, \boldsymbol{\theta}_i) = \eta(y, \boldsymbol{\theta}_i^0),$$

For any $\eta \in L^1$, let $E_{(\pi^0, \boldsymbol{\theta}^0)}\{\eta(y)\} = \int_{-\infty}^{\infty} \eta(y)\, f(y|\pi^0, \boldsymbol{\theta}^0) dy$. The following lemmas will be used in the proof of Theorem 1.

Lemma 1. If Condition 5 holds with $k = 1$ for any $(\pi, \boldsymbol{\theta}) \in \Omega$, $\boldsymbol{\theta}_1$ changed by $(\pi, \boldsymbol{\theta})$, $\boldsymbol{\theta}_1^0$ by $(\pi^0, \boldsymbol{\theta}^0)$ and $f_1(., \boldsymbol{\theta}_1)$ by $f(.|\pi, \boldsymbol{\theta})$.

Lemma 2 . Let $C = \{f \in L^1 : \|f\| < 1, f > 0\}$. For any $f \in C$ and $\eta \in B^+$

$$\int_{-\infty}^{\infty} \log(f/\eta)\eta dy < 0.$$

Note that for the proofs of these lemmas see Cheng and Liu (2001).

Proof of Theorem 1:
It is assumed that Ω is compact whole of the paper. Then, it should be shown that

$$P\left\{\lim_{n\to\infty} \sup_{(\pi,\boldsymbol{\theta})\in S} \left(\frac{f(y_1|\pi,\boldsymbol{\theta}) f(y_2|\pi,\boldsymbol{\theta}) \dots f(y_n|\pi,\boldsymbol{\theta})}{f(y_1|\pi^0,\boldsymbol{\theta}^0) f(y_2|\pi^0,\boldsymbol{\theta}^0) \dots f(y_n|\pi^0,\boldsymbol{\theta}^0)}\right) = 0\right\} = 1, \tag{31}$$

where S is any closed subset of Ω such that $dis\{S, \Omega(\pi^0, \boldsymbol{\theta}^0)\} > 0$. We have to approve for each point $(\pi^*, \boldsymbol{\theta}^*) \in S$, there is always a neighborhood called $N(\pi^*, \boldsymbol{\theta}^*)$ of the point that

$$E_{(\pi^0,\boldsymbol{\theta}^0)} \log\left(\sup_{(\pi,\boldsymbol{\theta})\in N(\pi^*,\boldsymbol{\theta}^*)} f(y|\pi,\boldsymbol{\theta})\right) < E_{(\pi^0,\boldsymbol{\theta}^0)} \log\big(f(y|\pi^0,\boldsymbol{\theta}^0)\big). \tag{32}$$

We suppose that $(\pi^*, \boldsymbol{\theta}^*)$ is a finite point. Then, let $\{N_i(\pi^*, \boldsymbol{\theta}^*), i = 1,2, \dots\}$ be a sequence of decreasing neighborhoods of the point $(\pi^*, \boldsymbol{\theta}^*)$ that $\cap_{i\geq 1} N_i(\pi^*, \boldsymbol{\theta}^*) = (\pi^*, \boldsymbol{\theta}^*)$. It can be assumed that $E_{(\pi^0,\boldsymbol{\Theta}^0)} \log\left(\sup_{\pi,\boldsymbol{\theta}\in N_i(\pi^*,\boldsymbol{\theta}^*)} f(y|\pi,\boldsymbol{\theta})\right)$ exists for $i = 1,2, \dots$ according to the Condition (5). Then, using the conditions, we get

$$\lim_{i\to\infty} \log\left(\sup_{(\pi,\boldsymbol{\theta})\in N_i(\pi^*,\boldsymbol{\theta}^*)} f(y|\pi,\boldsymbol{\theta})\right) = \log\big(f(y|\pi^*,\boldsymbol{\theta}^*)\big).$$

We also have

$$\lim_{n\to\infty} E_{(\pi^0,\boldsymbol{\theta}^0)} \log\left(\sup_{(\pi,\boldsymbol{\theta})\in N_i(\pi^*,\boldsymbol{\theta}^*)} f(y|\pi,\boldsymbol{\theta})\right) \geq E_{(\pi^0,\boldsymbol{\theta}^0)} \log\big(f(y|\pi^*,\boldsymbol{\theta}^*)\big). \tag{33}$$

It is clear that the sequence $\left\{E_{(\pi^0,\boldsymbol{\theta}^0)} \log\left(\sup_{(\pi,\boldsymbol{\theta})\in N_i(\pi^*,\boldsymbol{\theta}^*)} f(y|\pi,\boldsymbol{\theta})\right)\right\}$ is decreasing that

$$\log\left(\sup_{(\pi,\boldsymbol{\theta})\in N_1(\pi^*,\boldsymbol{\theta}^*)} f(y|\pi,\boldsymbol{\theta})\right) - \log\left(\sup_{(\pi,\boldsymbol{\theta})\in N_i(\pi^*,\boldsymbol{\theta}^*)} f(y|\pi,\boldsymbol{\theta})\right) \geq 0.$$

Then, via the Fatou's lemma and (33), we obtain that

$$\lim_{i\to\infty} E_{(\pi^0,\boldsymbol{\theta}^0)} \log\left(\sup_{(\pi,\boldsymbol{\theta})\in N_i(\pi^*,\boldsymbol{\theta}^*)} f(y|\pi,\boldsymbol{\theta})\right) = E_{(\pi^0,\boldsymbol{\theta}^0)} \log\big(f(y|\pi^*,\boldsymbol{\theta}^*)\big) < E_{(\pi^0,\boldsymbol{\theta}^0)} \log\big(f(y|\pi^0,\boldsymbol{\theta}^0)\big).$$

The inequality (32) results if $(\pi^*,\boldsymbol{\theta}^*)$ is a finite point.

If $(\pi^*,\boldsymbol{\theta}^*)$ is an infinite point, we have to prove that (32) is true when $N(\pi^*,\boldsymbol{\theta}^*)$ degenerates into the single point $(\pi^*,\boldsymbol{\theta}^*)$. It is known that the form of $f(y|\pi^*,\boldsymbol{\theta}^*)$ is:

$$f(y|\pi^*,\boldsymbol{\theta}^*) = \sum_{i=1}^{g} \pi^*_{m_i} f_{m_i}\big(y;\boldsymbol{\theta}^*_{m_i}\big),$$

where $0 \le g \le k-1$ and $\pi^*_{m_i} f_{m_i}\big(y;\boldsymbol{\theta}^*_{m_i}\big) > 0$. If $\sum_{i=1}^{g} \pi^*_{m_i} < 1$, and according to Lemma 2, we get

$$E_{(\pi^0,\boldsymbol{\theta}^0)} \log\big(f(y|\pi^*,\boldsymbol{\theta}^*)\big) < E_{(\pi^0,\boldsymbol{\theta}^0)} \log\big(f(y|\pi^0,\boldsymbol{\theta}^0)\big).$$

On the other hand, we have to verify that $f(y|\pi^*,\boldsymbol{\theta}^*) \ne f(y|\pi^0,\boldsymbol{\theta}^0)$. First we suppose that this is not true. Thus, $(\pi^*,\boldsymbol{\theta}^*) \in \Omega(\pi^0,\boldsymbol{\theta}^0)$, and the limiting point of the sequence $\{(\pi_1^s,\dots,\pi_k^s)(\boldsymbol{\theta}_1^s,\dots,\boldsymbol{\theta}_k^s)\} \in \Omega(\pi^0,\boldsymbol{\theta}^0)$, where

$$\pi_j^s = \pi_j^* \text{ if } j = m_i, \text{ otherwise } \pi_j^s = 0,$$
$$\boldsymbol{\theta}_j^s = \boldsymbol{\theta}_j^* \text{ if } j = m_i, \text{ otherwise } \boldsymbol{\theta}_j^s \to \infty.$$

It is not possible to have $dis\{S,\Omega(\pi^0,\boldsymbol{\theta}^0)\} > 0$. Then, let $N_i(\pi^*,\boldsymbol{\theta}^*)$ be a sequence of decreasing neighborhoods of the point $(\pi^*,\boldsymbol{\theta}^*)$ that $\cap_i N_i(\pi^*,\boldsymbol{\theta}^*) = (\pi^*,\boldsymbol{\theta}^*)$. As per Lemma 1 and Fatou's Lemma,

$$\lim_{i\to\infty} E_{(\pi^0,\boldsymbol{\theta}^0)} \log\left(\sup_{(\pi,\boldsymbol{\theta})\in N_i(\pi^*,\boldsymbol{\theta}^*)} f(y|\pi,\boldsymbol{\theta})\right) \le E_{(\pi^0,\boldsymbol{\theta}^0)} \log\big(f(y|\pi^*,\boldsymbol{\theta}^*)\big) < E_{(\pi^0,\boldsymbol{\theta}^0)} \log\big(f(y|\pi^0,\boldsymbol{\theta}^0)\big).$$

Thus, the inequality (32) was proved. According to the Heine-Borel finite open cover theorem and the same way given in the proof of Theorem 1 in Wald (1949), the equation (31) results.
Let $(\bar{\pi}_n,\overline{\boldsymbol{\theta}}_n)$ be a function of the observations $y_1,\dots,y_n$ that

$$\frac{f(y_1|\bar{\pi}_n,\overline{\boldsymbol{\theta}}_n)f(y_2|\bar{\pi}_n,\overline{\boldsymbol{\theta}}_n)\dots f(y_n|\bar{\pi}_n,\overline{\boldsymbol{\theta}}_n)}{f(y_1|\pi^0,\boldsymbol{\theta}^0)f(y_2|\pi^0,\boldsymbol{\theta}^0)\dots f(y_n|\pi^0,\boldsymbol{\theta}^0)} \ge c > 0$$

for all n and for all $y_1,\dots,y_n$. Now, we show that $dis\{(\pi_n,\theta_n),\Omega(\pi^0,\boldsymbol{\theta}^0)\} \to 0$ w.p. 1 by the help of proof of Theorem 2 given in Wald (1949). To prove this, we have to demonstrate that all limit points $(\bar{\pi},\overline{\boldsymbol{\theta}})$ of the sequence $\{\bar{\pi}_n,\overline{\boldsymbol{\theta}}_n\}$ hold $dis\{(\bar{\pi},\overline{\boldsymbol{\theta}}),\Omega(\pi^0,\boldsymbol{\theta}^0)\} \le \epsilon$ for any $\epsilon > 0$, and this probability equals to 1. Otherwise, there is a limit point $(\bar{\pi},\overline{\boldsymbol{\theta}})$ of the sequence $\{\bar{\pi}_n,\overline{\boldsymbol{\theta}}_n\}$ that $dis\{(\bar{\pi},\overline{\boldsymbol{\theta}}),\Omega(\pi^0,\boldsymbol{\theta}^0)\} > \epsilon$ states

$$\sup_{dis\{(\bar{\pi},\overline{\boldsymbol{\theta}}),\Omega(\pi^0,\boldsymbol{\theta}^0)\}>\epsilon} f(y_1|\pi,\boldsymbol{\theta})f(y_2|\pi,\boldsymbol{\theta})\dots f(y_n|\pi,\boldsymbol{\theta}) \ge f(y_1|\bar{\pi}_n,\overline{\boldsymbol{\theta}}_n)f(y_2|\bar{\pi}_n,\overline{\boldsymbol{\theta}}_n)\dots f(y_n|\bar{\pi}_n,\overline{\boldsymbol{\theta}}_n)$$

for infinitely many n. However,

$$\frac{\sup_{dis\{(\bar{\pi},\overline{\boldsymbol{\theta}}),\Omega(\pi^0,\boldsymbol{\theta}^0)\}>\epsilon} f(y_1|\pi,\boldsymbol{\theta})f(y_2|\pi,\boldsymbol{\theta})\dots f(y_n|\pi,\boldsymbol{\theta})}{f(y_1|\pi^0,\boldsymbol{\theta}^0)f(y_2|\pi^0,\boldsymbol{\theta}^0)\dots f(y_n|\pi^0,\boldsymbol{\theta}^0)} \ge c > 0$$

for infinitely many n. Since the probability of this event is 0 according to the equation (31), now we can say that all limit points $(\bar{\pi}, \bar{\boldsymbol{\theta}})$ of the sequence $\{\bar{\pi}_n, \bar{\boldsymbol{\theta}}_n\}$ hold $dis\{(\bar{\pi}, \bar{\boldsymbol{\theta}}), \Omega(\pi^0, \boldsymbol{\theta}^0)\} \leq \epsilon$. Therefore, if the maximum likelihood estimator $\widehat{\boldsymbol{\Theta}}_n = (\hat{\pi}_n, \widehat{\boldsymbol{\theta}}_n)$ exists, it is an consistent estimator of $\boldsymbol{\Theta} = (\pi, \boldsymbol{\theta})$.

Proof of Theorem 2:
It was shown that $\widehat{\boldsymbol{\Theta}}_n$ is consistent; therefore, this estimator will be an interior point of Ω if n is large. Then, we have to prove that

$$\frac{\partial \ell(\widehat{\boldsymbol{\Theta}}_n)}{\partial \boldsymbol{\Theta}} = 0.$$

It can be written by the help of Taylor's expansion such that

$$0 = \frac{\partial \ell(\widehat{\boldsymbol{\Theta}}_n)}{\partial \boldsymbol{\Theta}} = \frac{\partial \ell(\boldsymbol{\Theta}^0)}{\partial \boldsymbol{\Theta}} + \frac{\partial^2 \ell(\boldsymbol{\Theta}^0)}{\partial \boldsymbol{\Theta} \partial \boldsymbol{\Theta}^T}(\widehat{\boldsymbol{\Theta}}_n - \boldsymbol{\Theta}^0) + \frac{1}{2}(\widehat{\boldsymbol{\Theta}}_n - \boldsymbol{\Theta}^0)^T \frac{\partial^3 \ell(\boldsymbol{\Theta}^{*i})}{\partial \boldsymbol{\Theta}^3}(\widehat{\boldsymbol{\Theta}}_n - \boldsymbol{\Theta}^0),$$

where $\frac{\partial^3 \ell(\boldsymbol{\Theta}^{*i})}{\partial \boldsymbol{\Theta}^3}$ is a three dimensional array with its ith $(i = 1, \ldots, 3g-1)$ component whose $(j,k)th$ element will be

$$\frac{\partial^3 \ell(\boldsymbol{\Theta}^{*i})}{\partial \boldsymbol{\Theta}_i \partial \boldsymbol{\Theta}_j \partial \boldsymbol{\Theta}_k}, \qquad j,k = 1, \ldots, 3g-1,$$

where $\boldsymbol{\Theta}^{*i}$ is a mixing distribution between $\widehat{\boldsymbol{\Theta}}_n$ and $\boldsymbol{\Theta}^0$. Then, using the expansion given above, we have

$$\frac{1}{2}\left[(\widehat{\boldsymbol{\Theta}}_n - \boldsymbol{\Theta}^0)^T \frac{\partial^3 \ell(\boldsymbol{\Theta}^{*i})}{\partial \boldsymbol{\Theta}^3} + \frac{\partial^2 \ell(\boldsymbol{\Theta}^0)}{\partial \boldsymbol{\Theta} \partial \boldsymbol{\Theta}^T}\right](\widehat{\boldsymbol{\Theta}}_n - \boldsymbol{\Theta}^0) = -\frac{\partial \ell(\boldsymbol{\Theta}^0)}{\partial \boldsymbol{\Theta}}.$$

where $\frac{1}{n}\frac{\partial^3 \ell(\boldsymbol{\Theta}^{*i})}{\partial \boldsymbol{\Theta}^3} = O(1)$, and $\frac{1}{n}\frac{\partial^2 \ell(\boldsymbol{\Theta}^0)}{\partial \boldsymbol{\Theta} \partial \boldsymbol{\Theta}^T} = I(\boldsymbol{\Theta}^0) + o(1)$. Then, we get

$$\left[(\widehat{\boldsymbol{\Theta}}_n - \boldsymbol{\Theta}^0)^T \frac{\partial^3 \ell(\boldsymbol{\Theta}^{*i})}{\partial \boldsymbol{\Theta}^3} + \frac{\partial^2 \ell(\boldsymbol{\Theta}^0)}{\partial \boldsymbol{\Theta} \partial \boldsymbol{\Theta}^T}\right](\widehat{\boldsymbol{\Theta}}_n - \boldsymbol{\Theta}^0) = n\{I(\boldsymbol{\Theta}^0) + o_p(1)\}(\widehat{\boldsymbol{\Theta}}_n - \boldsymbol{\Theta}^0).$$

After rearranging the equation we obtain

$$\sqrt{n}(\widehat{\boldsymbol{\Theta}}_n - \boldsymbol{\Theta}^0) = -\{I(\boldsymbol{\Theta}^0)^{-1} + o(1)\}\left(\frac{1}{\sqrt{n}}\frac{\partial \ell(\boldsymbol{\Theta}^0)}{\partial \boldsymbol{\Theta}}\right).$$

Via the central limit theorem, it can be written as:

$$\frac{1}{\sqrt{n}}\frac{\partial \ell(\boldsymbol{\Theta}^0)}{\partial \boldsymbol{\Theta}} \to N(0, I(\boldsymbol{\Theta}^0))$$

and, we have the desired result as follow:

$$\sqrt{n}(\widehat{\boldsymbol{\Theta}}_n - \boldsymbol{\Theta}^0) \underset{d}{\to} N(0, I(\boldsymbol{\Theta}^0)^{-1}).$$

References

Aitkin, M. (1987), "Modelling variance heterogeneity in normal regression using GLIM", Applied statistics, 332-339.

Akaike, H. (1973), “Information theory and an extension of the maximum likelihood principle”, Proceeding of the Second International Symposium on Information Theory, B.N. Petrov and F. Caski, eds., 267-281, Akademiai Kiado, Budapest.

Azzalini, A. (1985), “A class of distributions which includes the normal ones”, Scandinavian Journal of Statistics, 12(2), 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. (1999), “Statistical applications of the multivariate skew normal distribution. Journal of the Royal Statistical Society: Series B (Statistical Methodology)”, 61(3), 579-602.

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 (Statistical Methodology), 65(2), 367-389.

Bai, Z.D., Krishnaiah, P.R., and Zhao, L.C. (1989), “On rates of convergence of efficient detection criteria in signal processing with white noise”, IEEE Transactions on Information Theory, 35, 380-388.

Basford, K.E., Greenway, D.R., McLachlan, G.J., and Peel, D. (1997), “Standard errors of fitted means under normal mixture”, Computational Statistics, 12, 1-17.

Cabral, C.R.B., Bolfarine, H., Pereira, J.R.G. (2008), “Bayesian density estimation using skew student-t-normal mixtures”, Computational Statistics and Data Analysis, 52, 5075-5090.

Cheng, R. C. H., and Liu, W. B. (2001), “The consistency of estimators in finite mixture models”, Scandinavian Journal of Statistics, 28(4), 603-616.

Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977), “Maximum likelihood from incomplete data via the EM algorithm”, Journal of the Royal Statistical Society, Series B, 39, 1-38.

Dai, L.,Yin, J.H., Xie, Z.F. and Wu,L.C. (2019). Robust variable selection in finite mixture of regression models using the t distribution, Communications in Statistics-Theory and Methods, DOI: 10.1080/03610926.2018.1513143.

Doğru, F.Z. (2015), “Robust parameter estimation in mixture regression models”, Ph.D. thesis, Ankara University.

Doğru, F.Z., and Arslan, O. (2016), “Robust mixture regression using the mixture of different distributions”, Recent Advances in Robust Statistics: Theory and Applications. C. Agostinelli et al. (eds.), Springer India.

Doğru, F.Z., and Arslan, O. (2017a), “Robust mixture regression based on the skew *t* distribution”, Revista Colombiana de Estadística, 40(1), 45-64.

Doğru, F.Z., and Arslan, O. (2017b), “Parameter estimation for mixtures of skew Laplace normal distributions and application in mixture regression modeling”, Communications in Statistics-Theory and Methods, 46(21), 10879-10896.

Doğru, F.Z., and Arslan, O. (2018a), “Robust mixture regression modeling using the least trimmed squares (LTS)-estimation method”, Communications in Statistics-Simulation and Computation, 47(7), 2184-2196.

Doğru, F.Z., and Arslan, O. (2018b), “Joint modelling of the location, scale and skewness parameters of the skew Laplace normal distribution”, Iranian Journal of Science and Technology, Transactions A: Science (just accepted).

Dunson, D.B., Pillai, N., and Park, J.H. (2007), “Bayesian density regression”, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 69(2), 163-183.

Engel, J., and Huele, A.F. (1996), “A generalized linear modeling approach to robust design”, Technometrics, 38(4), 365-373.

Fernández, C., and Steel, M.F. (1998), “On Bayesian modeling of fat tails and skewness”, Journal of the American Statistical Association, 93(441), 359-371.

García-Escudero, L. A., Gordaliza, A.,Mayo-Iscar, A., and SanMartín, R. (2010), “Robust clusterwise linear regression through trimming”, Computational Statistics and Data Analysis, 54, 3057-3069.

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.

Harvey, A.C. (1976), “Estimating regression models with multiplicative heteroscedasticity”, Econometrica: Journal of the Econometric Society, 44(3), 461-465.
Hennig, C. (2000), “Identifiability of models for clusterwise linear regression”, Journal of Classification, 17(2), 273-296.
Henze, N. (1986), “A probabilistic representation of the skew-normal distribution”, Scandinavian Journal of Statistics, 13(4), 271-275.
Kiefer, N.M. (1978), “Discrete parameter variation: Efficient estimation of a switching regression model”, Econometrica: Journal of the Econometric Society, 427-434.
Li, H.Q., and Wu, L.C. (2014), “Joint modelling of location and scale parameters of the skew-normal distribution”, Applied Mathematics-A Journal of Chinese Universities, 29(3), 265-272.
Li, H. Q., Wu, L.C., and Yi, J.Y. (2016), “A skew–normal mixture of joint location, scale and skewness models”, Applied Mathematics-A Journal of Chinese Universities, 31(3), 283-295.
Li, H., Wu, L., and Ma, T. (2017), “Variable selection in joint location, scale and skewness models of the skew-normal distribution”, Journal of Systems Science and Complexity, 30(3), 694-709.
Lin, T.I., and Wang, Y.J. (2009), “A robust approach to joint modeling of mean and scale covariance for longitudinal data”, Journal of Statistical Planning and Inference, 139(9), 3013-3026.
Lin, T.I., and Wang, W.L. (2011), “Bayesian inference in joint modelling of location and scale parameters of the t distribution for longitudinal data”, Journal of Statistical Planning and Inference, 141(4), 1543-1553.
Liu, M., and Lin, T.I. (2014), “A skew-normal mixture regression model”, Educational and Psychological Measurement, 74(1), 139-162.
McLachlan, G.J. and Peel, D. (2000), “Finite Mixture Models”, Wiley, New York.
Park, R.E. (1966), “Estimation with heteroscedastic error terms”, Econometrica, 1966, 34, 888.
Peters, Jr, B.C., and Walker, H.F. (1978), “An iterative procedure for obtaining maximum-likelihood estimates of the parameters for a mixture of normal distributions”, SIAM Journal on Applied Mathematics, 35(2), 362-378.
Quandt, R.E. (1972), “A new approach to estimating switching regressions”, Journal of the American Statistical Association, 67(338), 306-310.
Quandt, R.E., and Ramsey, J.B. (1978), “Estimating mixtures of normal distributions and switching regressions”, Journal of the American Statistical Association, 73(364), 730-738.
Redner, R.A., and Walker, H.F. (1984), “Mixture densities, maximum likelihood and the EM algorithm”, SIAM review, 26(2), 195-239.
Schwarz, G. (1978), “Estimating the dimension of a model”, The Annals of Statistics, 6(2), 461-464.
Song, W., Yao, W., and Xing, Y. (2014), “Robust mixture regression model fitting by Laplace distribution”, Computational Statistics and Data analysis, 71, 128-137.
Tan, X., Chen, J., and Zhang, R. (2007). “Consistency of the constrained maximum likelihood estimator in finite normal mixture models”, Proceedings of the American Statistical Association, American Statistical Association, Alexandria, VA, 2113-2119.
Taylor, J., and Verbyla, A. (2004), “Joint modelling of location and scale parameters of the t distribution”, Statistical Modelling, 4(2), 91-112.
Verbyla, A.P. (1993), “Modelling variance heterogeneity: residual maximum likelihood and diagnostics”, Journal of the Royal Statistical Society. Series B (Methodological), 55(2), 493-508.
Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. The Annals of Mathematical Statistics, 20(4), 595-601.
Wang, P., Puterman, M. L., Cockburn, I., and Le, N. (1996), “Mixed Poisson regression models with covariate dependent rates”, Biometrics, 381-400.
Wei, Y. (2012), “Robust mixture regression models using t-distribution”, Master Report, Kansas State University.
Wu, L., and Li, H. (2012), “Variable selection for joint mean and dispersion models of the inverse Gaussian distribution”, Metrika, 75(6), 795-808.
Wu, L.C, Zhang, Z.Z, Xu D.K. (2012). Variable selection in joint mean and variance models of Box-Cox transformation. Journal of Applied Statistics, 39(12):2543-2555.

Wu, L.C, Zhang, Z.Z., and Xu, D.K. (2013), "Variable selection in joint location and scale models of the skew-normal distribution", Journal of Statistical Computation and Simulation, 83(7), 1266-1278.

Wu, L.C. (2014), "Variable selection in joint location and scale models of the skew-t-normal distribution", Communications in Statistics - Simulation and Computation, 43(3), 615 - 630.

Wu, L., Tian, G. L., Zhang, Y. Q., and Ma, T. (2017), "Variable selection in joint location, scale and skewness models with a skew-t-normal distribution", Statistics and Its Interface, 10(2), 217-227.

Yao, W., Wei, Y., and Yu, C. (2014), "Robust mixture regression using the t-distribution", Computational Statistics and Data Analysis, 71, 116–127.

Zhang, J. (2013), "Robust mixture regression modeling with Pearson Type VII distribution", Master Report, Kansas State University.

Zhao, W., and Zhang, R. (2015), "Variable selection of varying dispersion student-t regression models", Journal of Systems Science and Complexity, 28(4), 961-977.