Robust Mixture Regression Estimation Based on least trimmed sum of absolute Method by using Several Models

Batool Shaaban[*], Nahed Helmy*,Mervat Elgohary*

Abstract

The present study deals with one of the most important methods of the robust mixture regression estimators,least trimmed sum of absolute deviations LTA method. It is known that mixture regression models are used to investigate the relationship between variables that come from unknown latent groups and to model heterogenous datasets. In general, the error terms are assumed to be normal in the mixture regression model. However, the estimators under normality assumption are sensitive to the outliers. Therefore, we introduce a robust mixture regression procedure based on the LTA-estimation method to combat with the outliers in the data. In this paper, we handle LTA method by using three mixture regression models; Laplace, t and normal distributions. We give a simulation study to illustrate the performance of the proposed estimators over the counterparts in terms of dealing with outliers.

Keywords: *EM algorithm, LTA-estimation method, Mixture regression model, Robust regression.*

1. Introduction

Bassett (1991) and Tableman (1994 a, b) proposed the *least trimmed sum of absolute*(LTA) deviations through minimizing the sum of the smallest absolute residuals:

$$min \sum_{i=1}^{h} |r_i|_{j:n}, \qquad (1)$$

where r shows the residuals which $r_i = y_j - x_j'\beta$ and $|r_i|_{1:n} \leq \cdots \leq |r_i|_{n:n}$ are the ordered absolute residuals, $h = (n(1-\alpha)+1)$, is the number of observations after trimming, and α is the trimming proportion (Dogru and Arslan 2017). It is worth noting that we will use this criterion as a robust criterion. Hawkins and Olive (1999) proved that LTA is an attractive alternative toLeast Median of SquaresLMS and Least Trimmed Squares LTS, particularly for large data sets. It has a statistical efficiency that is not much below that of LTS for outlier-free normal data and better than LTS for more peaked error distributions. They proved that its computational complexity is of a lower order than LMS and LTS. They used very simple calculations for finding exact evaluation of the LTA to outline a "feasible solution algorithm" for sample too large, which provide excellent approximations to the exact LTA solution. Several authors have examined the LTA estimator in the location model (a model including an intercept, but no nontrivial predictors). For the location model, Bassett (1991) gives an algorithm, and Tableman (1994 a, b) derives the influence function and asymptotic. In the regression model, LTA is a special case of the R-estimators of H¨ossjer (1991, 1994). LTA (γ) have breakdown value at $\min(1-\gamma,\gamma)$ [See H¨ossjer (1994)]. Croux, *et al.* (1996) showed that the maxbias curve of LTA is lower than that of LTS.

---

* Al- Azhar Universit Faculty of Commerce - Girls' Branch Department of Statistics.

This paper is organized as follows: Section (2) presents LTA method by using Laplace mixture regression model and EM algorithm for parameters estimation. Section (3) presents LTA method by using t mixture regression model and EM algorithm for parameters estimation. Section (4) shows the LTA method by using normal mixture regression model and EM algorithm for parameters estimation. Section (5) presents a simulation study with the comparisons which are made with some existing procedures in the literature. Conclusions are discussed in Section (6).

2. LTA Method by using Laplace Distribution

2.1 Definition

Let $\boldsymbol{Y}$ is an $n \times 1$ vector of dependent variables, $\boldsymbol{X}$ is an $n \times p$ matrix of predictors, and ϵ_i is an $n \times 1$ vector of errors. The relationship between $\boldsymbol{Y}$ and $\boldsymbol{X}$ is often investigated through a linear regression model. In the mixture linear regression procedure, we assume that with probability π_i,

$i = 1,2,\dots,g$, (X^T, Y) comes from one of the following $g \geq 2$ linear regression models:

$$Y = X^T\beta_i + \sigma_i\epsilon_i, i = 1,2,\dots,g, g \geq 2 \qquad (2)$$

where $\sum_{i=1}^{g}\pi_i = 1$,β_is' are unknown p-dimensional vectors of regression coefficients, σ_is' are unknown positive scalars. The random errors ϵ_is' are assumed to be independent of X_js'. It is commonly assumed that the density functions of ϵ_is' are members in a location-scale family with mean 0 and variance 1. Song *et al.* (2014) proposed a robust estimation procedure for the mixture linear regression models based on Laplacedistribution by using*least absolute deviation*(LAD) method. This research deals with the same equations as in Song *et al.* (2014), but at using trimmed version, the LTA technique, instead of the less robustly LAD, for achieving robustness. Therefore, we assume that ϵ_i follows a Laplace or a double exponential distribution with location 0 and scale parameter $\frac{1}{\sqrt{2}}$, which makes the variance of ϵ_i being 1, $i = 1,2,\dots,g$. Then it is easily seen that for a sample $S = \{(X^T{}_j, Y_j), j = 1,2,\dots,n\}$ from the model (2), the log-likelihood function of
$\theta = (\beta_1, \sigma_1^2, \pi_1, \beta_2, \sigma_2^2, \pi_2, \dots, \beta_g, \sigma_g^2, \pi_g)$ can be written as:

$$L(\theta; S) = \sum_{j=1}^{n} log \sum_{i=1}^{g} \frac{\pi_i}{\sqrt{2}\sigma_i} exp\left(-\frac{\sqrt{2}|Y_j - X_j^T\beta_i|}{\sigma_i}\right) \qquad (3)$$

The maximum likelihood estimators of θ can be obtained by maximizing $L(\theta; S)$ by taking the derivative of $L(\theta; S)$ with respect to θ, and set it equal to 0. Usually, no explicit solution can be obtained, and some numerical methods will be applied. Andrews and Mallows (1974) showed that a Laplace distribution in fact can be expressed as a mixture of normal distribution and another distribution related to exponential distribution. If we assume $Z$and Vbe two random variables, Vhas a distribution with density function:

$$f(v) = \frac{1}{v^3}\exp(-\frac{1}{2v^2}), \quad v > 0 \qquad (4)$$

and given $V = v$, the conditional distribution of $Z$is normal with mean 0 and variance $\frac{\sigma^2}{2v^2}$. Denote $f(z; v)$ the joint density function of $Z$and V, that is;

$f(z,v) = \frac{v}{\sqrt{\pi}\sigma}\exp(-\frac{v^2z^2}{\sigma^2})\frac{1}{v^3}\exp(-\frac{1}{2v^2})$ (5)

Then the marginal distribution of Z will be a Laplace distribution with density function:

$h_z(z) = \exp(-\sqrt{2}|z|/\sigma) / \sqrt{2}\sigma$

Consider Vas a latent variable. If Vcould be observed, then it is easy to see that the log-likelihood function of $\theta = (\beta; \sigma^2)$, based on the sample $P = (X_j, Y_j, V_j)_{j=1}^{n}$ is:

$$L(\theta; P) = -\frac{1}{2}\log\pi\sigma^2 - \frac{1}{\sigma^2}\sum_{j=1}^{n} V_j^2(Y_j - X_j^T\beta)^2 - \sum_{j=1}^{n} logV_j^2 - \frac{1}{2}\sum_{j=1}^{n}\frac{1}{V_j^2} \quad (6)$$

2.2 EM Algorithm for the mixture regression based on the LTA estimation method.

Assume that ϵ_js' follows a Laplace distribution with mean 0 and scale parameter $\sigma_i/\sqrt{2}$.

For$i = 1, 2, \ldots, g,\ j = 1, 2, \ldots, n$, $G_{ij}$are latent Bernoulli variables such that

$$G_{ij} = \begin{cases} 1 & \text{if } j\text{th observation } (X_j, Y_j) \text{ is from } i\text{th component} \\ 0 & \text{otherwise.} \end{cases}$$

Then, if the full data set $T = \{(X_j, Y_j, G_j)\}_{i=1,2,\ldots,g; j=1,2,\ldots,n}$ are observable, then the log likelihood function of $\theta = (\beta_1, \sigma_1^2, \pi_1, \beta_2, \sigma_2^2, \pi_2, \ldots, \beta_g, \sigma_g^2, \pi_g)$ can be written as:

$$L(\theta; T) = \sum_{j=1}^{n}\sum_{i=1}^{g} G_{ij} \log\frac{\pi_i}{\sqrt{2}\sigma_i}\exp\left(-\frac{\sqrt{2}|Y_j - X_j^T\beta_i|}{\sigma_i}\right) \quad (7)$$

From Andrews and Mallows (1974), we know that a Laplace distributed random variable is a scale mixture of a normal random variable and another variable related to exponential distribution. Denote V_j , coupled with (X_j ; Y_j), as the latent scale variable, $j = 1, 2, \ldots, n$ and it will be regarded as missing observations because they cannot be observable. Then, the complete data log-likelihood function of θ, based on $D = \{X_j, Y_j, V_j, G_{ij}\}_{i=1,2,\ldots,g;j=1,2,\ldots,n}$, has the form

$$L_c(\theta; D) = \sum_{j=1}^{n}\sum_{i=1}^{g} G_{ij} \log \frac{V_j}{\sqrt{\pi\sigma_i}} \exp\left(-\frac{V_j^2(Y_j - X_j^T\beta_i)^2}{\sigma_i^2}\right)\frac{1}{V_j^3}\exp\left(-\frac{1}{2V_j^2}\right)$$

$$= \sum_{j=1}^{n}\sum_{i=1}^{g} G_{ij}\log\pi_i - \frac{1}{2}\sum_{j=1}^{n}\sum_{i=1}^{g} G_{ij}\log\pi_i\sigma_i^2 - \sum_{j=1}^{n}\sum_{i=1}^{g}\frac{G_{ij}V_j^2(Y_j - X_j^T\beta_i)^2}{\sigma_i^2} - \sum_{j=1}^{n}\sum_{i=1}^{g} G_{ij}\log V_j^2 - \frac{1}{2}\sum_{j=1}^{n}\sum_{i=1}^{g}\frac{G_{ij}}{V_j^2}$$

(8)

From Equation (8), we find that the third part of the right-hand side like a least square criterion which can be replaced by the robust criterion LTA Equation (1) and then we can apply EM steps (Dogru and Arslan 2017). As the last two terms in Equation (8) do not involve the unknown regression parameters, we can simply drop them from the analysis. Based on EM algorithm principle, in E-step, we have to calculate the conditional expectation $E\left[L(\theta; D)|S, \theta^{(0)}\right]$, where

$S = \{(X_j, Y_j)\}_{j=1}^{h}$, $h = (n(1-\alpha)+1)$, is the number of observations after trimming, and α is the trimming proportion and $\theta^{(0)} = (\beta_1^{(0)}, \sigma_1^{2(0)}, \pi_1^{(0)}, \beta_2^{(0)}, \sigma_2^{2(0)}, \pi_2^{(0)}, \ldots, \beta_g^{(0)}, \sigma_g^{2(0)}, \pi_g^{(0)})$ are initial values for θ. Thus, to find $E\left[L(\theta; D)|S, \theta^{(0)}\right]$ we only have to calculate the following two terms:

$\tau_{ij} = E\left[G_{ij}|S, \theta^{(0)}\right]$, $\delta_{ij} = E[V_j^2|S, \theta^{(0)}, G_{ij} = 1]$

One can show that

$$\tau_{ij} = \frac{\pi_i^{(0)}\sigma_i^{-1(0)}\exp\left(-\frac{|Y_j - X_j^T\beta_i^{(0)}|}{\sigma_i^{(0)}}\right)}{\sum_{m=1}^{g}\pi_m^{(0)}\sigma_m^{-1(0)}\exp\left(-\frac{|Y_j - X_j^T\beta_m^{(0)}|}{\sigma_m^{(0)}}\right)}, \quad \delta_{ij} = \frac{\sigma_i^{(0)}}{\sqrt{2}|Y_j - X_j^T\beta_i^{(0)}|} \quad (9)$$

The calculation for δ_{ij} follows the same thread as in Phillips (2002). In M-step, the following expression will be maximized with respect to $\pi_i's, \beta_i's\ and\ \sigma_i^2 s$,

$$\sum_{j=1}^{n}\sum_{i=1}^{g}\tau_{ij}\log\pi_i - \frac{1}{2}\sum_{j=1}^{n}\sum_{i=1}^{g}\tau_{ij}\log\sigma_i^2 - \sum_{j=1}^{h}\sum_{i=1}^{g}\frac{\tau_{ij}\delta_{ij}(Y_j - X_j^T\beta_i)^2}{\sigma_i^2} \quad (10)$$

and the maximizer will be used for the next iteration
We propose the following EM algorithm to maximize (3).

- **EM Algorithm:**

1. Choose initial values for $\boldsymbol{\theta} = (\beta_1, \sigma_1^2, \pi_1, \ldots, \beta_g, \sigma_g^2, \pi_g)$,
2. E-Step: at the $(k+1)^{th}$ iteration, calculate $\tau_{ij}{}^{(k+1)}$ and $\delta_{ij}{}^{(k+1)}$ from Equation (9) with (0) replaced by (k).
3. M-Step: at the $(k+1)^{th}$ iteration, use the following formulas to calculate the maximizer of (10):

$$\pi_i^{(k+1)} = \frac{1}{n}\sum_{j=1}^{n} \tau_{ij}{}^{(k)}, \quad (11)$$

$$\beta_i^{(k+1)} = \left(\sum_{j=1}^{h} \tau_{ij}{}^{(k+1)} \delta_{ij}{}^{(k+1)} \mathbf{X}_j \mathbf{X}_j^T\right) \quad (12)$$

$$\sigma_i^{2(k+1)} = \frac{2\sum_{j=1}^{h} \tau_{ij}{}^{(k+1)} \delta_{ij}{}^{(k+1)} \left|Y_j - \mathbf{x}_j^T \beta_i{}^{(k+1)}\right|}{\sum_{j=1}^{h} \tau_{ij}{}^{(k+1)}} \quad (13)$$

4. Repeat steps (2), (3) until the convergence is obtained.

We also assume that all σ_i^2 are equal, and the above EM algorithm, a common initial value for σ_i^2 are used, but σ^2 can be updated in M-step by

$$\sigma^{2(k+1)} = \frac{2\sum_{j=1}^{h}\sum_{i=1}^{g} \tau_{ij}{}^{(k+1)} \delta_{ij}{}^{(k+1)} \left|Y_j - \mathbf{x}_j^T \beta_i{}^{(k+1)}\right|}{n} \quad (14)$$

The robustness of the above EM procedure is resulted from the adoption of LTA regression; it is also obvious from the formulae of the updated β_is' in each iteration.
Note that the factor $\delta_{ij}{}^{(k+1)}$ is reversely related to the term $\left|Y_j - \mathbf{X}_j^T \beta_i{}^{(k)}\right|$, meaning that larger residuals give smaller values of $\delta_{ij}{}^{(k+1)}$, hence down weight the corresponding observations when calculating the estimates.

3. LTA Method by using t Distribution

In this section we assume that as Wei (2012), the error density $f_i(\epsilon)$ is a t-distribution with degrees of freedom v_i and scale parameter σ_i. Hence, given x_j , density function of y_j is:

$$f(y_j; x_j^T, \theta) = \sum_{i=1}^{g} \pi_i f(y_j; x_j^T \beta_i, \sigma_i^2, v_i), \tag{15}$$

where

$$f(y_j; x_j^T, \beta_i, \sigma_i^2, v_i) = \frac{\Gamma(\frac{v_i+1}{2})|\sigma_i|^{-1}}{(\pi_i v_i)^{\frac{1}{2}} \Gamma(\frac{v_i}{2})\{1+\delta(y_j, x_j^T \beta_i; \sigma_i^2)/v_i\}^{\frac{1}{2}(v_i+1)}} \ ,$$

(16)

and $\delta(y_j, x_j^T \beta_i; \sigma_i^2) = (y_j - x_j^T \beta_i)^2 / \sigma_i^2$

Let's assume that v_i are known. The unknown parameter θ can be estimated by maximizing the log likelihood

$$\sum_{j=1}^{n} \log\{\sum_{i=1}^{g} \pi_i f(y_j; x_j^T \beta_i, \sigma_i^2, v_i)\} \tag{17}$$

Note that the complete log likelihood function for (X, y, z) is

$$\log L_C(\theta; X, y, z) = \sum_{j=1}^{n} \sum_{i=1}^{g} z_{ij} \log\{\pi_i f(y_j; x_j^T \beta_i, \sigma_i^2, v_i)\}, \tag{18}$$

where $X = (X_1, X_2, \ldots, X_n)^T$; $y = (y_1, y_2, \ldots, y_n)$;$z = (z_{11}, \ldots, z_{ng})$. Based on the theory of EM algorithm, in E-step, given the current estimate θ^k at k^{th}iterative M-step, we calculate conditional expectation of the complete log likelihood$E(\log L_c(\theta; X, y, z)|X, y, \theta^k)$, which is simplified to the calculation of $E(z_{ij}|X, y, \theta^k)$ [See Wei (2012)]. In addition, at M-step, we compute the parameters which maximize

$$E(\log L_c(\theta; X, y, z)|X, y, \theta^k) = \sum_{j=1}^{n} \sum_{i=1}^{g} E(z_{ij}|X, y, \theta^k) \log\{\pi_i f(y_j; x_j^T, \beta_i, \sigma_i^2, v_i)\} \tag{19}$$

We note that there is no explicit solution for $\beta_i$and σ_i^2 .

Because the t-distribution can be considered as a scale mixture of normal distributions, we use the method of EM algorithm so that we can estimate unknown parameters and follow the following steps:

Let u be the latent variable such that

$$\epsilon|u \sim N\left(0, \frac{\sigma^2}{u}\right), u \sim gamma\left(\frac{1}{2}\nu, \frac{1}{2}\nu\right),$$

where gamma (α, γ) has density

$$f(u; \alpha, \gamma) = \frac{1}{\Gamma\alpha}\gamma^{\alpha}u^{\alpha-1}e^{-\gamma u}, \qquad u > 0$$

Then, marginally ϵ has a t-distribution with degrees of freedom ν and scale parameter σ. Therefore, Wei (2012) introduced another latent variable u to simplify the computation of M-step of EM algorithm.

Note that the complete likelihood for $(\boldsymbol{X}; \boldsymbol{y}; \boldsymbol{u}; \boldsymbol{z})$ is

$$\log L_c(\boldsymbol{\theta}; \boldsymbol{X}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{u})$$

$$= \sum_{j=1}^{n}\sum_{i=1}^{g} z_{ij} \log\{\pi_i\, \phi\left(y_i;\ \boldsymbol{x}_j^T\beta_i, \frac{\sigma_i^2}{u_i}\right) f(u_i; \frac{1}{2}\nu_i, \frac{1}{2}\nu_i)\}$$

$$= \sum_{j=1}^{n}\sum_{i=1}^{g} z_{ij}\log(\pi_i) + \sum_{j=1}^{n}\sum_{i=1}^{g} z_{ij}\log\left\{f\left(u_i; \frac{1}{2}\nu_i, \frac{1}{2}\nu_i\right)\right\}$$

$$+ \sum_{j=1}^{n}\sum_{i=1}^{g} z_{ij}\left\{-\frac{1}{2}\log(2\pi\sigma_i^2) + \frac{1}{2}\log(u_i) - \frac{u_i}{2\sigma_i^2}(y_i - \boldsymbol{x}_j^T\beta_i)^2\right\} \quad (21)$$

where $\boldsymbol{u} = (u_1, u_2, \dots, u_n)$ is independent of z.

In order to use our proposed method, we can replace the last part of Equation (21) with LTA's robust criterion. In addition, the above second term doesn't involve unknown parameters. Therefore, based on the theory of EM algorithm, in E step, given the current estimate θ^k at k^{th}step, the calculation of $E(\log L_c(\boldsymbol{\theta}; \boldsymbol{X}, \boldsymbol{y}, \boldsymbol{u}, \boldsymbol{z})|\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{\theta}^{(k)})$is simplifiedto the calculation of $E(z_{ij}|\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{\theta}^{(k)})$ and of $E(u_{ij}|\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{\theta}^{(k)}, z_{ij} = 1)$. Then in M-step, we find the maximizer of

$$E\left(\log L_c(\theta;\boldsymbol{X},\boldsymbol{y},\boldsymbol{u},\boldsymbol{z})|\boldsymbol{X},\boldsymbol{y},\boldsymbol{\theta}^{(k)}\right) \propto \sum_{j=1}^{h}\sum_{i=1}^{g} E(z_{ij}|\,x,\boldsymbol{\theta}^{(k)})[\log(\pi_i) - \frac{1}{2}\log(2\pi\sigma_i^2) - \frac{E(u_{ij}|x,\theta^{(k)},z_{ij}=1)}{2\sigma_i^2}|y_i - x_j'\beta_i|\Big] \quad (22)$$

which has explicit solution for $\boldsymbol{\theta}$, where h is defined before.

Wei (2012) proposed the following EM algorithm to maximize (17). The steps of EM algorithm as:

- **EM Algorithm:**

1- Input initial values: $\pi_i^{(0)}, \beta_i^{(0)}$ *and* $\sigma_i^{2(0)}$.

2- E- step: at the $(k+1)^{th}$ iteration

$$E(z_{ij}|\boldsymbol{X},\boldsymbol{y},\boldsymbol{\theta}^{(k)}) = \tau_{ij}^{(k+1)} = \frac{\pi_i^{(k)} f\left(y_j;\boldsymbol{x}_j^T,\beta_i^{(k)},\sigma_i^{2(k)},\nu_i^{(k)}\right)}{\sum_{i=1}^{g}\pi_i^{(k)} f\left(y_j;\boldsymbol{x}_j^T,\beta_i^{(k)},\sigma_i^{2(k)},\nu_i^{(k)}\right)}, \quad (23)$$

and

$$E(u_{ij}|\boldsymbol{X},\boldsymbol{y},\boldsymbol{\theta}^{(k)},Z_{ij}=1) = u_{ij}^{(k+1)} = \frac{\nu_i^{(k)}+1}{\nu_i^{(k)} + \delta\left(y_j;\boldsymbol{x}_j^T,\beta_i^{(k)},\sigma_i^{2(k)},\nu_i^{(k)}\right)}. \quad (24)$$

3- M- step: At the $(k+1)^{th}$ iteration, the estimator of parameterscan be computed $(\pi_i,\beta_{i,},\sigma_i^2,\nu_i)$can be computing which maximize the expected complete log likelihood

$$\pi_i^{(k+1)} = \sum_{j=1}^{n}\tau_{ij}^{(k+1)}/n, \quad (25)$$

$$\beta_i^{(k+1)} = (\sum_{j=1}^{h}\boldsymbol{x}_j\boldsymbol{x}_j^T w_{ij}^{(k+1)})^{-1}\sum_{j=1}^{h}\boldsymbol{x}_j y_j w_{ij}^{(k+1)}, \quad (26)$$

and

$$\sigma_i^{2(k+1)} = \frac{\sum_{j=1}^{h} \tau_{ij}^{(k+1)} u_{ij}^{(k+1)} (y_i - x_j^T \beta_i^{(k+1)})^2}{\sum_{j=1}^{h} \tau_{ij}^{(k+1)}}, \quad (27)$$

where $w_{ij}^{(k+1)} = \tau_{ij}^{(k+1)} u_{ij}^{(k+1)}$. If we further assume that all $\sigma_i^2$are equal, and the above EM algorithm, a common initial value for$\sigma_i^2$are used, but σ^2 can be updated in M-step by

$$\sigma^{2(k+1)} = \frac{\sum_{j=1}^{h} \sum_{i=1}^{g} \tau_{ij}^{(k+1)} u_{ij}^{(k+1)} |y_j - x_j^T \beta_i^{(k+1)}|}{n} \quad (28)$$

4- Repeat E-step and M-step until the result can pass certain criterion.

From (24) in E-step, the weights $u_{ij}^{(k+1)}$ decrease if the standardized residuals increase and thus decrease the effects of the outliers to generate the robust estimate for mixture regression parameters. In addition, from (27) in M-step, we can see that larger residuals also have smaller effects on $\sigma_i^{2(k+1)}$ due to the weights $u_{ij}^{(k+1)}$.

4. LTA Method by using Normal Distribution

Let $\{(x_1,y_1),(x_2,y_2),\dots,(x_n,y_n)\}$be a sample. If it is assumed that the error terms have the normal distribution with 0 mean and σ^2variance in the mixture regression model, the estimator of $\boldsymbol{\theta}$can be found by maximizing the following log-likelihood function:

$$l(\boldsymbol{\theta}) = \sum_{j=1}^{n} \log\left(\sum_{i=1}^{g} \pi_i \phi(y_j; \boldsymbol{x}_j^T \beta_i, \sigma_i^2)\right). \quad (29)$$

However, since the direct maximization of (29) cannot be usually possible, in general, the EM algorithm is used to find the ML estimate of θ. Let Z_{ij}be the latent variables with

$$Z_{ij} = \begin{cases} 1 & \text{if } j\text{th observation is coming from } i\text{th componen} \\ 0 & \text{otherwise} \end{cases} \quad (30)$$

where $j = 1,\dots,n$ and $i = 1,\dots,g$. Here, $Z_j = (Z_{1j},\dots,Z_{gj})$ will be regarded as missing observations because they cannot be observable. Then, the complete data log-likelihood function for (y, Z_j) given X is obtained as

$$L_c(\theta; X, y, Z_j) = \left(\sum_{j=1}^{n}\sum_{i=1}^{g} Z_{ij}(\log(\pi_i) - \frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_i^2)\right) - \sum_{j=1}^{n}\sum_{i=1}^{g} Z_{ij}\frac{(y_j - x_j^T\beta_i)^2}{2\sigma_i^2},$$

(31)

where $X = (X_1, X_2, \dots, X_n)^T$ and $y = (y_1, y_2, \dots, y_n)$

The estimator based on the normal distribution in the mixture regression model will not be robust against the outliers because of the second term of the complete data log-likelihood function given in (31). This term is basically the least-squares (LS) criterion and it is known that the LS method is sensitive to the outliers. To obtain robust estimators, this term should be robustified. Therefore, we will take the complete data log-likelihood function given in (29) and use the LTA criterion given in (1) in this equation (Dogru and Arslan 2017). This results the following adapted complete data log-likelihood function

$$L_c(\theta; X, y, Z_j) = \left(\sum_{j=1}^{n}\sum_{i=1}^{g} Z_{ij}(\log(\pi_i) - \frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_i^2)\right) - \sum_{j=1}^{h}\sum_{i=1}^{g} Z_{ij}\frac{|r_i|_{j:n}}{2\sigma_i^2}.$$

(32)

where r_i and h are defined before. To run the EM algorithm, we will take the conditional expectation of the complete data log-likelihood function to get rid of the latency of z_{ij}

$$E(l_c(\theta; y, Z_j)|y_i) = \sum_{j=1}^{n}\sum_{i=1}^{g} E(Z_{ij}|y_i)(\log(\pi_i) - \frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_i^2)) - \sum_{j=1}^{h}\sum_{i=1}^{g} E(Z_{ij}|y_i)\frac{|r_i|_{j:n}}{2\sigma_i^2},$$

(33)

Note that the conditional expectation $E(Z_{ij}|y_i)$ can be calculated using the classical theory of the mixture modeling. Then, the steps of the EM algorithm for the mixture regression based on the LTA estimation method will be as follows:

- **EM Algorithm:**

1. Take initial estimate for the parameters, say $\boldsymbol{\theta}^{(0)}$ and fix a stopping rule Δ.
2. E-step: Compute the following conditional expectation when $\boldsymbol{y}$ and the current parameter value $\widehat{\boldsymbol{\theta}}^{(k)}$are given

$$\hat{z}_{ij}^{(k)} = E\left(Z_{ij}|y_j, \widehat{\boldsymbol{\theta}}^{(k)}\right) = \frac{\hat{\pi}_i^{(k)}\phi(y_j; \boldsymbol{x}_j{}^T\hat{\beta}_i, \hat{\sigma}_i{}^2)}{\sum_{i=1}^{g}\hat{\pi}_i^{(k)}\phi(y_j; \boldsymbol{x}_j{}^T\hat{\beta}_i, \hat{\sigma}_i{}^2)}. \quad (34)$$

3. M step: Compute the following estimates:

$$\hat{\pi}_i^{(k+1)} = \frac{1}{n}\sum_{j=1}^{n}\hat{z}_{ij}^{(k)}, \quad (35)$$

$$\hat{\beta}_i^{(k+1)} = (\sum_{j=1}^{h}\hat{z}_{ij}^{(k)}\boldsymbol{x}_j\boldsymbol{x}_j^T)^{-1}(\sum_{j=1}^{h}\hat{z}_{ij}^{(k)}\boldsymbol{x}_j y_j), \quad (36)$$

and

$$\hat{\sigma}_i^{2(k+1)} = c_\alpha \frac{\sum_{j=1}^{h}\hat{z}_{ij}^{(k)}\left|y_j - \boldsymbol{x}_j^T\hat{\beta}_i^{(k)}\right|}{\sum_{j=1}^{h}\hat{z}_{ij}^{(k)} - p}, \quad (37)$$

where $c_\alpha$is a consistency constant. For the normal errors, $c_\alpha$will be $(1-\alpha)/F_{x_3^2}(q_\alpha)$with
$q_\alpha = x_{1,1-\alpha}^2$[See Agulló *et al.*, (2008)]for the case multivariate normal errors. Here, $F_{x_3^2}$ shows the cumulative distribution function of the χ^2 distribution with 3 degrees of freedom and $q_\alpha$is the upper αpercent point of the χ^2 distribution with 1 degree of freedom.

4. Repeat E and M steps until the convergence criterion || $\widehat{\theta}^{(k+1)} - \widehat{\theta}^{(k)}$|| $< \Delta$is satisfied. Moreover, the absolute difference of the actual log-likelihood $||\ell(\widehat{\boldsymbol{\theta}}^{(k+1)}) - \ell(\widehat{\boldsymbol{\theta}}^{(k)})|| < \Delta$or $||\ell(\widehat{\boldsymbol{\theta}}^{(k+1)})/\ell(\widehat{\boldsymbol{\theta}}^{(k)}) - 1|| < \Delta$ can be used [See Dias and Wedel, (2004)] for the convergence.

5. Simulation Study

In this section, we will evaluate the performance of the proposed robust mixture regression method based on LTA method for the three different distributions through a simulation study. The comparison will also be made among different estimation methods using the estimated *mean squared errors* ($MSE's$) and bias of the parameter estimates for each estimation method.It is known that the estimate with the least sum of squares error and the least bias is the best. There have been great research efforts in dealing with the unbounded likelihood issue. See, for example, Hathaway (1985,1986), Chen *et al.* (2008), and Yao (2010).

The estimated*mean squared errors* ($MSE's$) is given by

$$\widetilde{MSE}(\hat{\theta}) = \sum_{j=1}^{M} \frac{(\hat{\theta}_j - \theta)^2}{M},$$

where M is the number of samples.

In our simulation study, we compare our method with some existing estimation procedures by generating sample data $(X_{j1}, X_{j2}, Y_j)_{j=1}^{n}$ from the following two-component mixture regression models which are also used in Wei (2012):

$$\mathbf{Y} = \begin{cases} 0 + X_1 + X_2 + \epsilon_1 & \text{if } Z = 1; \\ 0 - X_1 - X_2 + \epsilon_1 & \text{if } Z = 2. \end{cases}$$

where Z is a component indicator of Y with $P(Z = 1) = 0.25$. That is, the data are generated from a two-component mixture linear regression models with $\beta_1 = (\beta_{10}, \beta_{11}, \beta_{12})' = (0, 1, 1)$and

$\beta_2 = (\beta_{20}, \beta_{21}, \beta_{22})' = (0, -1, -1)$

The predictors$X_1 \sim N(0;\ 1)$, $X_2 \sim N(0;\ 1)$ are independent,the random errorϵ_1 and ϵ_2 are independent and have the same distribution as ϵ.

The following error distributions will be considered:

Case I: $\epsilon \sim N(0,1)$

Case II: $\epsilon \sim$ Laplacewith mean 0 and variance 1

Case III: $\epsilon \sim t_1$, t-distribution with degree of freedom 1

Case IV: $\epsilon \sim t_3$, t-distribution with degree of freedom 3

Case V: $\epsilon \sim$ lognormal (0,1)

Case VI: $\epsilon \sim 0{:}95N(0;\ 1)\ +\ 0{:}05N(0;\ 52)$.

Case VII: $\epsilon \sim N$ (0; 1) with 5% of high leverage outliers being $X1 = X2 = 20$ and $Y\ =\ 100$.

In Case I, the error is exactly normally distributed and there are no outliers. It is often used to evaluate the efficiency of different estimation methods compared to the traditional MLE. For Case II, the estimation methods proposed by Song *et al.* 2014 which, as in the first Case, would serve a reference line to evaluate the performance of other estimation procedures. Both Case III and IV are heavy tailed distributions. In Case V we use lognormal distribution for random error as a skewed distribution to know whether the skewed distribution gave better results or not. In Case VI, the 5% data from $N(0; 25)$ are likely to be low leverage outliers. In Case VII, 5% observations are replicated high leverage outliers, which will be used to check the robustness of estimation procedures against the outliers in the x-direction. The following algorithm will be compared:

- MLE based on normality assumption
- Trimmed likelihood estimator (TLE) proposed by Neykov *et al.* (2007)
- The robust EM mixture regression based on t-distribution (Mixregt)
- MLE based on Laplace distribution (MixregL)
- The proposed robust EM mixture regression based on least trimmed sum of absolute deviations method (MixrigLTA) with using three cases of distribution for y. In case I we assume that error has Laplace distribution, case II; error has t distribution. Finally, case III; error has normal distribution.

From the simulation studies, we can see that if the true distribution of ϵ is normal or have Laplace distribution the MSEs of MLE procedure are slightly bigger than our proposed method MixregLTA_L and better than MixregLTA_t and MixregLTA_N when the sample size is 100.The larger the sample size, the better the results for all, but with the preference of the proposed method MixregLTA_L. While for other cases when the distribution of ϵ has a heavier

tail, contaminated by some outliers, or there are high leverage outliers in the data set or skewed, then MLE fails to provide reasonable estimates. It is clear that the performance of LTA_L method is the best in all cases, with some slight differences in preference between it and TLE, Mixregt and MixregL methods in some cases. For LTA_t we can say that it is in the fourth place after LTA_L, MixregL and TLE. But LTA_N is after them in preference. The overall performance of the Mixregt proposed by Wei (2012) is also satisfying when sample size gets bigger except for the Case VI when high leverage points present in the data set.
Tables (5.1), (5.2) and (5.3) show the estimated ($MSE's$) and bias (Bias) of the parameter estimates for each estimation method for the sample sizes $n = 100, 200\ and\ 400$ respectively. The number of replicates is 200. Based on Tables (5.1), (5.2) and (5.3), we can see that $MixrigLTA_L$has overall better or comparable performance than other six methods considered for Cases I to VII.

Table 5.1: The estimated ($MSE's$) and bias (Bias) of the parameter estimates for each estimation methods *MLE*, $TLE, Mixregt,\ MixregL, MixrigLTA_L, MixrigLTA_t$ and $MixrigLTA_N$ for n = 100

parameter		MLE		TLE		Mixrigt		MixregL		MixrigLTA		MixrigLTA		MixrigLTA	
		MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
β_{10}	c ~ N	0.093	0.019	0.139	0.033	0.109	-0.014	0.09	0.028	0.028	0.097	0.104	0.03	0.019	-0.011
β_{11}		0.074	0.015	0.212	-0.195	0.126	-0.023	0.095	-0.047	0.051	-0.084	0.146	-0.042	2.141	-1.448
β_{12}		0.134	-0.019	0.248	-0.195	0.133	-0.044	0.09	-0.023	0.042	-0.063	0.141	0.015	2.169	-1.46
β_{20}		0.018	-0.002	0.038	-0.004	0.019	-0.014	0.026	-0.017	0.012	-0.045	0.021	0.019	0.019	-0.011
β_{21}		0.02	-0.012	0.03	0.011	0.02	-0.004	0.026	0.007	0.014	0.012	0.025	-0.019	0.349	0.552
β_{22}		0.018	-0.012	0.024	0.034	0.02	-0.008	0.025	-0.001	0.012	0.005	0.02	-0.024	0.33	0.54
π_1		0.093	0.019	0.007	0.025	0.109	-0.014	0.004	0.011	0.006	0.031	0.006	0.015	0.062	0.25
β_{10}	Case II: c ~ Laplace	0.004	0.013	0.075	-0.007	0.005	0.009	0.052	0.019	0.006	0.031	0.125	-0.021	0.019	0
β_{11}		0.121	0.024	0.097	-0.107	0.103	0.022	0.054	-0.012	0.014	0.074	0.131	-0.004	2.137	-1.45
β_{12}		0.127	0.012	0.084	-0.077	0.135	-0.05	0.051	-0.012	0.021	-0.077	0.139	0.002	2.188	-1.467
β_{20}		0.15	0.01	0.013	0.004	0.102	-0.028	0.011	-0.006	0.019	-0.067	0.014	0.003	0.019	0
β_{21}		0.101	-0.025	0.013	0.007	0.013	-0.009	0.012	0.007	0.007	-0.061	0.013	-0.008	0.338	0.55
β_{22}		0.023	-0.002	0.013	0.019	0.013	0.002	0.012	0	0.005	0.033	0.014	0.004	0.322	0.533
π_1		0.043	-0.016	0.004	0.019	0.012	-0.006	0.052	0.019	0.008	0.04	0.005	0.008	0.062	0.25
β_{10}	Case III: c ~ t_1	0.006	0.009	3.2	-0.15	0.004	0.01	0.004	0.007	0.005	0.027	0.541	0.106	0.068	-0.004
β_{11}		13741.2	-7.49	1.886	-0.17	0.74	0.041	0.125	0.103	0.018	0.096	1.018	-0.451	2.315	-1.49
β_{12}		5478.51	5.183	1.797	-0.033	1.263	-0.65	0.416	-0.327	0.069	-0.186	0.916	-0.509	2.361	-1.505
β_{20}		9085.99	5.619	1.526	0.065	1.215	-0.525	0.472	-0.322	0.077	-0.192	0.185	-0.006	0.068	-0.004
β_{21}		13738.3	-7.14	0.774	-0.129	0.144	0.014	0.028	-0.022	0.011	-0.079	0.088	-0.008	0.355	0.51
β_{22}		5499.32	5.266	0.773	-0.065	0.111	-0.057	0.026	0.059	0.013	0.051	0.106	-0.043	0.341	0.495
π_1		9119.33	6.31	0.039	0.06	0.082	-0.023	0.042	0.069	0.01	0.046	0.022	0.096	0.062	0.25
β_{10}	Case IV: c ~ t_7	0.116	0.249	0.238	0.007	0.029	0.111	0.024	0.072	0.034	0.177	0.749	0.167	0.028	-0.006
β_{11}		1.542	-0.006	0.264	-0.135	0.451	-0.033	0.15	0.024	0.085	0.099	0.481	-0.034	2.203	-1.469
β_{12}		1.09	-0.076	0.239	-0.096	0.477	-0.136	0.155	-0.034	0.145	-0.077	0.413	-0.022	2.222	-1.475
β_{20}		1.526	-0.337	0.038	-0.008	0.599	-0.006	0.187	-0.013	0.185	0.01	0.065	-0.015	0.028	-0.006
β_{21}		0.446	-0.067	0.034	0.01	0.631	-0.068	0.032	0.008	0.022	-0.019	0.048	-0.007	0.328	0.531
β_{22}		0.159	0.071	0.026	0.01	0.139	0.022	0.038	0.004	0.04	-0.019	0.105	-0.041	0.321	0.525
π_1		0.229	0.026	0.007	0.037	0.256	-0.074	0.034	-0.022	0.026	-0.01	0.012	0.027	0.062	0.25

Table 5.1: MSE and Bias of Point Estimates for n = 100

β_{10}	Case V: ε ~ lognormal (0; 1)	12.628	2.764	1.515	1.176	3.399	1.53	1.527	1.134	1.116	0.991	2.379	1.307	1.47	1.198
β_{11}		2.295	-0.4	0.117	-0.048	0.776	-0.304	0.19	-0.011	0.141	-0.028	0.537	-0.08	2.736	-1.635
β_{12}		2.946	-0.379	0.147	-0.056	0.627	-0.27	0.185	-0.026	0.146	-0.071	0.491	-0.084	2.664	-1.613
β_{20}		9.765	2.221	1.085	1.032	1.487	1.014	0.986	0.978	0.865	0.914	2.081	1.029	1.47	1.198
β_{21}		0.863	0.056	0.013	0.003	0.041	-0.004	0.026	0.008	0.032	-0.031	0.033	-0.018	0.196	0.365
β_{22}		2.051	-0.144	0.013	0.016	0.03	0.001	0.02	-0.002	0.03	-0.006	0.184	-0.037	0.214	0.387
π_1		0.075	0.075	0.005	0.033	0.017	0.045	0.005	0.002	0.006	0.024	0.013	0.041	1.47	1.198
	Case VI … 25)	0.025	0.022	0.124	0.046	0.013	0.025	0.006	0.01	0.007	0.024	0.195	0.091	0.022	0
		2.316	0.061	0.282	-0.209	0.342	0.008	0.156	0.07	0.046	0.095	0.431	-0.045	2.323	-1.509
		1.089	-0.096	0.221	-0.19	0.422	-0.111	0.119	-0.036	0.074	-0.071	0.336	-0.045	2.311	-1.505
		4.86	0.098	0.03	0.013	0.741	0.044	0.202	0.01	0.086	-0.081	0.033	0	0.022	0
		2.294	0.098	0.034	0.001	0.031	0.014	0.029	0.003	0.02	-0.033	0.027	-0.023	0.288	0.491
		1.267	-0.08	0.027	0.011	0.022	0.009	0.025	0.01	0.02	0.016	0.033	0.001	0.29	0.495
		1.534	-0.172	0.01	0.034	0.02	-0.01	0.028	0	0.025	0.008	0.008	0.019	0.062	0.25
	Case VII … 1) with 5% high leverage outliers	0.052	0.037	0.173	0.002	0.007	0.008	0.006	0	0.007	0.018	0.152	-0.057	0.024	0.006
		22.01	-3.086	0.248	-0.209	2.581	0.017	0.034	0.055	0.033	0.091	0.135	-0.016	2.473	-1.556
		6.353	1.398	0.219	-0.168	3.488	1.483	0.043	-0.033	0.051	-0.094	0.153	0.003	2.526	-1.574
		7.191	1.677	0.036	-0.002	3.61	1.514	0.041	-0.048	0.072	-0.101	0.024	0.003	0.024	0.006
		12.41	2.394	0.028	0	0.018	0.002	0.014	-0.05	0.016	-0.044	0.02	-0.004	0.251	0.444
		12.77	3.324	0.022	0.025	0.042	0.13	0.014	0.034	0.017	0.027	0.021	0.002	0.228	0.426
		11.91	3.206	0.007	0.017	0.04	0.128	0.013	0.029	0.019	0.025	0.005	0.003	0.062	0.25

Table 5.2: The estimated $(MSE's)$ and bias (Bias) of the parameter estimates for each estimation methods *MLE*, $TLE, Mixregt, MixregL, MixrigLTA_L, MixrigLTA_t$ and $MixrigLTA_N$ for n = 200

parameter		MLE		TLE		Mixrigt		MixregL		$MixrigLTA_L$		$MixrigLTA_t$		$MixrigLTA_N$	
		MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
β_{10}	Case I ε ~ N(	0.042	-0.003	0.073	-0.002	0.052	-0.002	0.063	0.037	0.018	0.079	0.046	0.017	0.01	-0.02
β_{11}		0.036	0.001	0.129	-0.162	0.048	-0.007	0.049	-0.03	0.021	-0.077	0.046	-0.022	2.12	-1.449
β_{12}		0.048	-0.011	0.174	-0.199	0.047	-0.04	0.062	-0.046	0.021	-0.081	0.05	0.009	2.144	-1.457
β_{20}		0.01	-0.011	0.018	-0.016	0.008	-0.008	0.013	-0.019	0.007	-0.058	0.01	-0.002	0.01	-0.02
β_{21}		0.009	-0.004	0.012	0.02	0.008	0.004	0.011	0.016	0.006	0.026	0.01	-0.01	0.323	0.551
β_{22}		0.009	0	0.017	0	0.009	0.001	0.013	0.006	0.007	0.04	0.01	-0.002	0.314	0.543
π_1		0.002	0.005	0.004	0.019	0.002	0.008	0.003	0.008	0.002	0.021	0.002	0.002	0.062	0.25
β_{10}	Case II: ε ~ Laplace	0.047	0.008	0.039	0.003	0.026	-0.006	0.022	0.025	0.009	0.075	0.041	0.001	0.008	-0.011
β_{11}		0.042	0.025	0.033	-0.064	0.027	-0.019	0.023	-0.018	0.009	-0.07	0.036	-0.01	2.141	-1.458
β_{12}		0.042	-0.004	0.028	-0.058	0.033	0.028	0.022	0.001	0.009	-0.061	0.037	0.002	2.14	-1.457
β_{20}		0.006	-0.002	0.007	-0.007	0.005	-0.003	0.005	-0.009	0.005	-0.062	0.005	0.003	0.008	-0.011
β_{21}		0.006	-0.011	0.007	0.007	0.005	-0.004	0.005	-0.001	0.003	0.043	0.007	-0.007	0.311	0.542
β_{22}		0.007	0.002	0.006	0.007	0.005	-0.004	0.005	-0.002	0.003	0.04	0.006	-0.004	0.311	0.543
π_1		0.002	0.002	0.002	0.019	0.026	-0.006	0.022	0.025	0.009	0.075	0.002	0.006	0.062	0.25

parameter		MLE		TLE		Mixregt		MixregL		$MixrigLTA_L$		$MixrigLTA_t$		$MixrigLTA_N$	
β_{10}	Case III: $\epsilon \sim t_1$	49.98	0.073	1.026	-0.123	0.002	0.002	0.002	0	0.002	0.026	0.23	-0.009	0.03	-0.004
β_{11}		39.74	-1.465	0.906	0.103	0.264	-0.015	0.052	0.063	0.012	0.09	0.728	-0.381	2.324	-1.509
β_{12}		77.11	-1.162	0.904	-0.024	0.971	-0.502	0.361	-0.414	0.088	-0.243	0.706	-0.338	2.277	-1.496
β_{20}		43.17	0.118	0.774	0.128	0.903	-0.445	0.362	-0.408	0.079	-0.237	0.035	-0.006	0.03	-0.004
β_{21}		15.66	0.637	0.201	0.052	0.034	0.008	0.014	-0.043	0.008	-0.083	0.033	-0.033	0.29	0.491
β_{22}		11.57	0.253	0.253	-0.032	0.043	-0.017	0.015	0.077	0.007	0.066	0.029	0.018	0.294	0.504
π_1		0.058	0.231	0.02	0.031	0.039	-0.001	0.018	0.088	0.008	0.075	0.015	0.064	0.062	0.25
β_{10}	Case IV: $\epsilon \sim t_3$	0.296	-0.038	1.026	-0.123	0.019	0.078	0.029	0.093	0.04	0.198	0.108	0.016	0.016	0.004
β_{11}		0.579	-0.065	0.906	0.103	0.094	0.041	0.06	0.06	0.062	0.049	0.116	-0.04	2.251	-1.492
β_{12}		0.429	-0.109	0.904	-0.024	0.086	-0.031	0.054	-0.021	0.049	-0.04	0.095	0.04	2.215	-1.479
β_{20}		2.946	-0.124	0.774	0.128	0.129	-0.033	0.054	-0.029	0.05	-0.027	0.014	0.006	0.016	0.004
β_{21}		0.47	0.009	0.201	0.052	0.02	0	0.016	-0.002	0.013	-0.024	0.017	-0.003	0.285	0.508
β_{22}		0.324	0.004	0.253	-0.032	0.015	-0.009	0.014	-0.004	0.015	-0.009	0.019	-0.006	0.3	0.521
π_1		0.02	0.002	0.02	0.031	0.015	-0.005	0.016	0	0.017	-0.005	0.003	0.011	0.062	0.25
β_{10}	Case V: $\epsilon \sim$ lognormal (0; 1)	17.05	3.119	0.061	0.006	2.664	1.457	1.729	1.252	1.134	1.019	1.7	1.187	1.442	1.193
β_{11}		2.515	-0.488	0.078	-0.093	0.46	-0.268	0.087	-0.038	0.078	-0.011	0.157	-0.037	2.778	-1.656
β_{12}		2.491	-0.527	0.087	-0.132	0.384	-0.21	0.109	-0.01	0.077	-0.033	0.239	-0.062	2.742	-1.644
β_{20}		10.54	2.3	0.012	-0.013	0.969	0.912	0.956	0.97	0.85	0.915	1.111	0.952	1.442	1.193
β_{21}		0.72	0.123	0.01	0.015	0.007	0.003	0.012	0.003	0.013	-0.006	0.009	0.005	0.154	0.344
β_{22}		0.671	0.112	0.016	-0.004	0.017	-0.01	0.012	-0.008	0.014	-0.009	0.011	-0.002	0.167	0.356
π_1		0.076	0.049	0.003	0.013	0.005	0.02	0.003	-0.002	0.014	0.007	0.005	0.018	0.062	0.25
β_{10}	Case VI $\epsilon \sim 0$:	0.851	-0.008	0.08	0.02	0.004	0.008	0.003	0.001	0.003	0.021	0.06	0.004	0.011	-0.004
β_{11}		0.93	0.004	0.082	-0.082	0.076	-0.013	0.054	0.013	0.032	0.062	0.066	0.025	2.334	-1.519
β_{12}		1.369	0.097	0.082	-0.095	0.069	-0.008	0.055	-0.017	0.044	-0.024	0.075	0.005	2.331	-1.518
β_{20}		0.024	0.003	0.019	-0.008	0.072	-0.022	0.073	-0.028	0.044	-0.047	0.011	-0.006	0.011	-0.004
β_{21}		0.092	0.056	0.017	0.014	0.013	0.004	0.016	-0.007	0.011	-0.037	0.012	0.002	0.258	0.481
β_{22}		0.105	0.087	0.014	-0.009	0.011	0.006	0.014	0.008	0.014	0.01	0.015	0.001	0.259	0.482
π_1		0.01	-0.032	0.004	0.035	0.01	0.012	0.012	0.016	0.016	0.009	0.002	0.002	0.062	0.25
β_{10}	Case VII $\epsilon \sim N(0;$ with 5% high leverage outliers	12.42	-2.059	0.072	0.029	0.003	0.003	0.003	0	0.003	0.007	0.051	0.007	0.012	-0.013
β_{11}		4.272	1.347	0.095	-0.142	1.878	0.273	0.027	0.069	0.022	0.067	0.055	0.004	2.63	-1.613
β_{12}		5.16	1.656	0.097	-0.125	2.862	1.488	0.022	-0.041	0.034	-0.103	0.044	-0.005	2.636	-1.614
β_{20}		14.75	2.618	0.013	0.007	2.891	1.497	0.023	-0.056	0.035	-0.105	0.01	-0.014	0.012	-0.013
β_{21}		12.37	3.316	0.011	0.014	0.011	-0.005	0.009	-0.054	0.009	-0.05	0.012	-0.004	0.179	0.387
β_{22}		12.28	3.289	0.012	0.016	0.034	0.146	0.008	0.042	0.009	0.028	0.011	0	0.179	0.386
π_1		0.149	0.213	0.003	0.02	0.035	0.143	0.007	0.032	0.009	0.043	0.002	-0.004	0.062	0.25

Table 5.3: The estimated $(MSE's)$ and bias (Bias) of the parameter estimates for each estimation methods *MLE*, $TLE, Mixregt, MixregL, MixrigLTA_L, MixrigLTA_t$ and $MixrigLTA_N$ for n = 400

parameter		MLE		TLE		Mixrigt		MixregL		MixrigLTA		MixrigLTA		MixrigLTA	
		MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
β_{10}	Case I ε ~ N(	0.02	-0.001	0.041	0.012	0.019	0.004	0.028	0.014	0.009	0.082	0.02	0.007	0.006	-0.008
β_{11}		0.017	0.006	0.108	-0.178	0.016	0.002	0.022	-0.017	0.011	-0.088	0.021	-0.026	2.13	-1.457
β_{12}		0.015	0.01	0.096	-0.171	0.021	-0.009	0.026	-0.023	0.012	-0.081	0.023	-0.003	2.12	-1.454
β_{20}		0.004	-0.001	0.009	0.002	0.005	-0.004	0.006	-0.004	0.006	-0.068	0.004	-0.003	0.006	-0.008
β_{21}		0.005	-0.004	0.007	0.02	0.004	-0.001	0.006	0	0.004	0.041	0.005	-0.001	0.304	0.543
β_{22}		0.004	0.003	0.006	0.013	0.005	-0.004	0.007	0.003	0.004	0.041	0.006	-0.007	0.306	0.546
π_1		0.001	0	0.002	-0.001	0.001	0.004	0.001	0.003	0.002	0.025	0.001	0.003	0.062	0.25
β_{10}	Case II: ε ~ Laplace	0.019	-0.006	0.012	0.012	0.012	0.001	0.011	0.011	0.007	0.074	0.015	-0.009	0.006	-0.005
β_{11}		0.015	-0.002	0.013	-0.041	0.009	0.002	0.008	-0.001	0.007	-0.073	0.014	-0.007	2.131	-1.457
β_{12}		0.016	0.006	0.017	-0.05	0.01	-0.01	0.008	-0.012	0.007	-0.07	0.013	0.003	2.116	-1.452
β_{20}		0.004	-0.001	0.003	-0.003	0.002	0.002	0.002	0	0.004	-0.063	0.002	0	0.006	-0.005
β_{21}		0.003	-0.013	0.003	0.005	0.002	-0.004	0.002	-0.001	0.003	0.047	0.003	-0.006	0.302	0.543
β_{22}		0.005	-0.011	0.004	0.012	0.003	0.001	0.002	0.005	0.003	0.048	0.002	0	0.308	0.548
π_1		0.001	0.004	0.001	0.016	0.001	0.005	0.001	0.003	0.002	0.026	0.001	0.004	0.062	0.25
β_{10}	Case III: ε ~ t_1	221.3	0.942	0.735	-0.04	0.169	0	0.017	0.075	0.01	0.089	0.119	0.034	0.017	-0.008
β_{11}		346.1	-0.021	0.398	0.097	0.478	-0.244	0.205	-0.35	0.078	-0.247	0.37	-0.16	2.311	-1.514
β_{12}		362.7	-3.15	0.399	0.059	0.445	-0.259	0.193	-0.338	0.081	-0.252	0.315	-0.178	2.332	-1.52
β_{20}		221.3	0.942	0.021	-0.001	0.022	0	0.008	-0.062	0.008	-0.086	0.012	-0.011	0.017	-0.008
β_{21}		350.0	1.978	0.032	0.003	0.014	0.015	0.012	0.09	0.008	0.085	0.017	0.002	0.254	0.486
β_{22}		354.1	-1.15	0.093	-0.009	0.011	0.003	0.011	0.086	0.008	0.086	0.014	-0.002	0.251	0.48
π_1		0.061	0.246	0.008	0.003	0.009	0.035	0.034	0.145	0.044	0.207	0.008	0.034	0.062	0.25
β_{10}	Case IV: ε ~ t_3	0.525	0.082	0.037	-0.008	0.034	-0.008	0.033	0.02	0.022	0.06	0.038	0.011	0.007	-0.01
β_{11}		0.366	-0.081	0.039	-0.07	0.04	-0.008	0.034	-0.017	0.028	-0.043	0.042	0	2.279	-1.504
β_{12}		0.438	-0.091	0.037	-0.081	0.039	0.005	0.037	0.004	0.033	-0.039	0.04	0.003	2.291	-1.507
β_{20}		0.029	-0.004	0.008	-0.017	0.006	-0.007	0.007	-0.011	0.006	-0.022	0.006	0.003	0.007	-0.01
β_{21}		0.116	0.051	0.007	0.001	0.006	0.007	0.007	0.016	0.007	-0.007	0.006	-0.004	0.264	0.496
β_{22}		0.033	0.066	0.009	0.006	0.009	0.002	0.009	0.01	0.007	0	0.007	0.005	0.263	0.493
π_1		0.011	-0.008	0.002	0.023	0.002	0.004	0.002	-0.001	0.002	0.013	0.002	0.003	0.062	0.25

parameter		MLE		TLE		Mixrigt		MixregL		MixrigLTA		MixrigLTA		MixrigLTA	
β_{10}	Case V: ϵ ~ lognormal (0; 1)	21.68	3.468	1.235	1.102	1.384	1.115	1.429	1.166	1.036	0.99	1.192	1.057	1.42	1.186
β_{11}		1.761	-0.521	0.012	-0.022	0.067	-0.022	0.04	0.017	0.035	0	0.053	0.002	2.725	-1.642
β_{12}		2.36	-0.381	0.017	-0.023	0.067	-0.017	0.038	0.008	0.034	-0.004	0.043	-0.003	2.684	-1.628
β_{20}		12.96	2.499	1.07	1.032	0.888	0.937	0.967	0.979	0.845	0.914	0.866	0.924	1.42	1.186
β_{21}		0.373	0.212	0.003	-0.006	0.004	-0.006	0.006	-0.004	0.007	-0.012	0.004	-0.006	0.159	0.358
β_{22}		0.292	0.208	0.003	-0.001	0.003	0	0.006	0.005	0.007	-0.004	0.004	-0.004	0.171	0.372
π_1		0.089	0.053	0.002	0.027	0.001	0.005	0.001	-0.008	0.002	-0.004	0.001	0.005	0.062	0.25
β_{10}	Case VI ϵ ~ 0:	1.047	0.051	0.041	0.005	0.026	-0.017	0.031	0	0.017	0.058	0.029	-0.001	0.006	-0.002
β_{11}		0.387	0.036	0.048	-0.095	0.026	-0.003	0.028	-0.008	0.022	-0.021	0.027	0.023	2.401	-1.543
β_{12}		0.822	0.034	0.051	-0.119	0.02	0.011	0.024	0.001	0.027	-0.043	0.035	-0.007	2.394	-1.541
β_{20}		0.043	0.013	0.006	0.003	0.006	-0.005	0.007	-0.009	0.006	-0.029	0.005	0.003	0.006	-0.002
β_{21}		0.087	0.048	0.006	0.01	0.006	0.007	0.007	0.011	0.008	0.027	0.006	0.001	0.229	0.457
β_{22}		0.137	0.052	0.007	0.024	0.004	-0.01	0.006	-0.006	0.008	0.025	0.007	0.015	0.23	0.459
π_1		0.007	-0.032	0.001	0.003	0.001	-0.001	0.001	-0.004	0.002	-0.004	0.001	-0.006	0.062	0.25
β_{10}	Case VII ϵ ~ N(0: with 5% high leverage outliers	12.81	-2.173	0.033	0.01	1.206	0.206	0.013	0.059	0.012	0.071	0.026	-0.003	0.007	-0.005
β_{11}		4.656	1.49	0.049	-0.102	2.642	1.485	0.013	-0.037	0.02	-0.088	0.02	-0.012	2.76	-1.657
β_{12}		4.657	1.513	0.039	-0.098	2.699	1.503	0.015	-0.042	0.017	-0.09	0.022	-0.002	2.746	-1.653
β_{20}		10.79	2.284	0.005	0.002	0.006	0.003	0.006	-0.044	0.005	-0.052	0.005	0.005	0.007	-0.005
β_{21}		11.86	3.384	0.006	0.011	0.023	0.133	0.004	0.028	0.005	0.035	0.005	0.004	0.131	0.343
β_{22}		11.51	3.332	0.006	0.003	0.024	0.13	0.005	0.029	0.005	0.041	0.006	-0.006	0.134	0.347
π_1		0.133	0.191	0.002	0.004	0.008	-0.088	0.004	0.023	0.002	0.003	0.001	0.001	0.062	0.25

6.Conclusion

In this paper, we propose a new robust estimation procedure for the mixture linear regression models by assuming the random error has a Laplace distribution or t distribution or normal distribution. The robustness is achieved essentially by LTA procedure, and implemented by the EM algorithm. The efficiency and effectiveness of the proposed EM algorithm depends upon the fact that the Laplace or t distribution is a scale mixture of a normal distribution and a distribution of a function of exponentially distributed random variable. The simulation study shows that the proposed method LTA_L is superior to and comparable to existing robust estimation procedures in all simulation conditions, but LTA_t and LTA_N are less preferred, and some other methods are preferred over them, such as Mixrigt and MixrigL in some cases.Specifically, we have the following findings:

1. The MLE works the best for Case I $\epsilon \sim N(0,1)$ and Case II: $\epsilon \sim$ Laplace with mean 0 and variance 1, but fails to provide reasonable estimates for Case III, VI and VII. And in the rest of the cases its results are not bad.
2. TLE has relatively better performance than Mixregt for Case IV, V and VII when n = 100, but have close performance to Mixregt in Case II and VI. But for Case I and III we find that Mixrigt is preferred. The results we obtained did not differ at the sample size of 200 except in the fourth Case; in this Case we find that Mixrigt is better than TLE. At n = 400, TLE has close performance to Mixregt, except for the last Case where Mixregt works better than TLE.
3. MixregL has better performance than TLE and Mixregt in most Cases whether n = 100 or 200 or 400. Otherwise, their performance is comparable.
4. MixrigLTA$_L$ have overall better performance than others.

.

References

1. Agulló, J., Croux, C., and Van Aelst, S. (2008). “The multivariate Least-Trimmed Squares Estimator,” *Journal of Multivariate Analysis*, Vol. 99, pp. 311–338.

2. Andrews, D.F. and Mallows, C.L. (1974). “Scale Mixtures of Normal Distributions”, Journal of the Royal Statistical Society. *(B)*, Vol. 36, pp. 99-102.

3. Bassett, G.W. (1991). “Equivariant, Monotonic, 50% Breakdown Estimators,” *The American Statistician*, Vol. 45, pp. 135-137.

4. Chen, J., Tan, X., and Zhang, R. (2008). “Inference for Normal Mixture in Mean and Variance,” *Statistica Sincia*, Vol. 18, pp. 443-465.

5. Croux, C., Rousseeuw, P.J., and Van Bael, A. (1996). “Positive-Breakdown Regression by Minimizing Nested Scale Estimators,” *Journal of Statistical Planning and Inference*, Vol. 53, pp. 197-235.

6. Dias, J. G.,Wedel, M. (2004). An empirical comparison of EM, SEM and MCMC performance for problematic Gaussian mixture likelihoods. *Statistics and Computing* 14:323–332.

7. Doğru, F.Z. and Arslan, O. (2017). “Robust Mixture Regression Modeling using the Least Trimmed Squares (LTS)-Estimation Method,” *Article in Communication in Statistics- Simulation and Computation*, Vol. 0, pp. 1-13.

8. Hathaway, R. J. (1985). “A Constrained Formulation of Maximum-Likelihood Estimation for Normal Mixture Distributions,” *Annals of Statistics*, Vol. 13, pp. 795-800.

9. Hathaway, R. J. (1986). “A Constrained EM Algorithm for Univariate Mixtures,” *Journal of Statistical Computation and Simulation*, Vol. 23, pp. 211-230.

10. Hawkins, M.D. and Olive, D. (1999). “Applications and Algorithms for Least Trimmed Sum of Absolute Deviations Regression,” Southern Illinois University Carbondale, Articles and Preprints, Department of Mathematics, *Computational Statistics and Data Analysis*, Vol. 32, pp. 119-134

11. H¨ossjer, O. (1991). “Rank-Based Estimates in the Linear Model with High Breakdown Point,” Ph.D. Thesis, Report 1991:5, Department of Mathematics, Uppsala University, Uppsala, Sweden.

12. H¨ossjer, O. (1994). “Rank-Based Estimates in the Linear Model with High Breakdown Point,” *Journal of the American Statistical Association*, Vol. 89, pp. 149-158.

13. Neykov, N., Filzmoser, P., Dimova, R., and Neytchev, P. (2007). “Robust fitting of mixtures using the trimmed likelihood estimator,”. Computational Statistics and Data Analysis, 52, 299-308.

14. Phillips, R.F. (2002). “Least Absolute Deviations Estimation Via the EM Algorithm,” *Statistics and Computing*, Vol. 12, pp. 281-285.

15. *Song, W., Yao, W. and Xing, Y. (2014). “Robust Mixture Regression Model Fitting By Laplace Distribution,”* Computational Statistics and Data Analysis,*Vol. 71, pp. 128-137.*

16. Tableman, M. (1994a). “The Influence Functions for the Least Trimmed Squares and the Least Trimmed Absolute Deviations Estimators,” *Statistics and Probability Letters*, Vol. 19, pp. 329-337.

17. Tableman, M. (1994b). “The Asymptotics of the Least Trimmed Absolute Deviations (LTAD) Estimator,” *Statistics and Probability Letters*, Vol. 19, pp. 387-398.

18. Wei, Y. (2012). “Robust Mixture Regression Models Using T-Distribution” Master of Science Department of Statistics College of Arts and Sciences, Kansas State University Manhattan, Kansas.

19. Yao, W. (2010). “A profile Likelihood Method for Normal Mixture with Unequal Variance,” *Journal of Statistical Planning and Inference*, Vol.140, pp. 2089-2098.