Anadolu Üniversitesi Bilim ve Teknoloji Dergisi A- Uygulamalı Bilimler ve Mühendislik
Anadolu University Journal of Science and Technology A- Applied Sciences and Engineering

ANADOLU UNIVERSITY

2017 - Volume: 18 Number: 2
Page: 403 - 418
DOI: 10.18038/aubtda.289280
Received: 02 February 2017 Revised: 23 April 2017 Accepted: 02 May 2017

MODELING HUMAN DEVELOPMENT INDEX USING FINITE MIXTURES OF DISTRIBUTIONS

Fatma Zehra DOĞRU *

Department of Econometrics, Faculty of Economics and Administrative Sciences, Giresun University, Giresun, Turkey

ABSTRACT

The Human Development Index (HDI) measures the development of a country which was designed by the United Nations Development Programme (UNDP). Since the values of HDI for different countries show differences according to the development of a country, the distribution of HDI may have one more mode, thick tail or skewness. Therefore, we can use mixtures of distributions to model the HDI data set to handle modality, heavy-tailedness and/or skewness. In this study, we propose to model the data set from the HDI report 2015 for 188 countries with finite mixtures of distributions. We give the basic scheme of the maximum likelihood (ML) estimation using Expectation-Maximization (EM) algorithm for the finite mixture model. To obtain the best model for HDI data set, we first find the appropriate cluster number using model-based clustering. Then, we use the finite mixture models obtained from some symmetric and/or heavy-tailed and skew and/or heavy-tailed distributions to find the best model for HDI data set.

Keywords: EM algorithm, Mixture model, ML, HDI

1. INTRODUCTION

The Human Development Index (HDI) is calculated for each country which was offered by United Nations Development Programme (UNDP) in UNDP reports 1990 [1]. This index is based on three essential dimensions of human development. The first one is to live a long and healthy life which was measured by average life at birth. The second one is the ability of being knowledgeable which was measured by mean years of schooling and expected years of schooling. The last one is the ability to have a properly standard of living which was measured by gross national income per capita. One can see the papers; [2-12] for more detailed explanations about the HDI. This index can be computed by taking the geometric mean of normalized indices for each of the three dimensions

$$HDI = (I_{Health} . I_{Education} . I_{Income})^{1/3}, \quad (1)$$

where $I_{Health}, I_{Education}$ and I_{Income} show the health, education and income indices. For the details about computation of the HDI, see the Technical notes of HDI report 2015 [13]. Also, the range of HDI according to the development of countries is given in Table 1. The value of HDI is decreasing when the human development level is decreasing.

Table 1. The range of HDI according to the development of countries

The Development Level	Range
Very high human development	0.800 and above
High human development	0.700-0.799
Medium human development	0.550-0.699
Low human development	Below 0.550

*Corresponding Author: fatma.dogru@giresun.edu.tr

Finite mixture models are dependent on a convex linear composition of a finite number of densities. These models are very popular for modeling and analyzing heterogeneous data sets in the presence of multimodality, skewness and heavy tails, concurrently. It is known that these models are a very powerful tool for supervised classification, unsupervised clustering, density estimation, pattern recognition, data mining, image analysis, machine learning etc. (see, [14-18]).

Since the normal distribution has wide applicability and computational tractability, in general, the components in a mixture model have normal distribution. However, in many real-world applications, the component densities may have asymmetric behavior and heavy tails. To overcome these problems, there are some studies which include mixture modeling based on heavy-tailed and/or skew distributions. To deal with heavy-tailedness, mixture model based on t distribution was proposed by [19]. Then, [20] studied the mixture model using the skew normal (SN) distribution [21,22] to model the data sets including asymmetric observations. Furthermore, [23] introduced a robust mixture modeling based on the skew t (ST) distribution [24] to handle both skewness and heavy-tailedness in the data, [25] proposed finite mixture modeling using the scale mixtures of SN distributions, [26] introduced mixtures of the skew Student-t normal distributions and [27] proposed finite mixture modeling based on the skew Laplace normal distribution and applied in mixture regression modeling.

In this paper, since the distribution of HDI may be multimode, heavy-tailed and/or skew, we consider to model the HDI data set using the finite mixtures of distributions which are some symmetric and/or heavy-tailed and skew and/or heavy-tailed distributions. We compare finite mixture model using the normal distribution, finite mixture model using the t distribution [16], finite mixture model using the skew normal distribution [20] and finite mixture model using the skew t distribution [23] for modeling the HDI data set.

The paper is designed as follows. In Section 2, we define the finite mixture model and give the maximum likelihood (ML) estimation using the Expectation-Maximization (EM) algorithm of this model. In Section 3, we select the number of clusters using model-based clustering for HDI data set. In Section 4, we obtain the best model for HDI data set by comparing the mixture models based on normal, t, skew normal and skew t distributions in terms of the values of information criteria. We give some conclusions in Section 5. Finally, we also give some details about estimation procedures that used in this paper in Appendix Section.

2. FINITE MIXTURES OF DISTRUBUTIONS

Let $y_1, y_2, \dots, y_n$ be a random sample come from a g-component finite mixtures of distributions. Then, the probability density function (pdf) of this mixture model is

$$f(y_j|\boldsymbol{\Theta}) = \sum_{i=1}^{g} w_i f_i(y_j;\theta_i), \qquad x \in \mathbb{R}, \tag{2}$$

where w_i shows the mixing probability with $\sum_{i=1}^{g} w_i = 1$, $0 \le w_i \le 1$, $f_i(y_j;\theta_i)$ is the pdf of the ith component parameterized with θ_i and $\boldsymbol{\Theta} = (w_1, \dots, w_{g-1}, \theta_1, \dots, \theta_g)'$ is the unknown parameter vector.

2.1. The ML Estimation of Finite Mixture Model

In general, the ML estimation method is used to find the parameter estimators of the finite mixture model. The ML estimator of $\boldsymbol{\Theta}$ can be found by maximization of the following log-likelihood function with regard to the unknown parameter

$$\ell(\mathbf{\Theta}) = \sum_{j=1}^{n} \log\left(\sum_{i=1}^{g} w_i\, f_i(y_j;\theta_i)\right). \tag{3}$$

However, since there is not an explicit maximizer of the log-likelihood function, a numerical algorithm, for instance, the EM ([28]) algorithm should be used to obtain the estimators for the parameters of interest. Now, we will use the following EM algorithm to get the ML estimators.

Let $\boldsymbol{Z}_j = (Z_{1j}, \dots, Z_{gj})'$ be the latent variables with

$$Z_{ij} = \begin{cases} 1, & if\ jth\ observation\ is\ coming\ from\ ith\ component \\ 0, & otherwise \end{cases} \tag{4}$$

where $j = 1, \dots, n$ and $i = 1, \dots, g$.

Here, $\boldsymbol{z}$ is regarded as missing data and $\boldsymbol{y}$ is considered as observed data, where $\boldsymbol{y} = (y_1, \dots, y_n)$ and $\boldsymbol{z} = (z_1, \dots, z_n)$. Let $(\boldsymbol{y}, \boldsymbol{z})$ be the complete data. Then, the complete data log-likelihood function of $\mathbf{\Theta}$ can be obtained as

$$\ell_c(\mathbf{\Theta}; \boldsymbol{y}, \boldsymbol{z}) = \sum_{j=1}^{n}\sum_{i=1}^{g} z_{ij} \log\left(w_i f_i(y_j;\theta_i)\right). \tag{5}$$

However, since $\boldsymbol{z}_j$ is regarded as missing data, this function cannot be directly maximized to get the ML estimator of $\mathbf{\Theta}$. Thus, to handle the latency problem, we have to take the conditional expectation of the complete data log-likelihood function given the observed data y_j. This function will be as follows

$$E\left(\ell_c(\mathbf{\Theta}; \boldsymbol{y}, \boldsymbol{z})|y_j\right) = \sum_{j=1}^{n}\sum_{i=1}^{g} E\left(Z_{ij}|y_j\right) \log\left(w_i f_i(y_j;\theta_i)\right). \tag{6}$$

Here, the conditional expectation $E(Z_{ij}|y_j)$ can be calculated using the classical theory of mixture modeling.

EM algorithm:

1. Set starting point as $\mathbf{\Theta}^{(0)}$ and fix a stopping rule Δ.
2. E-step: Calculate the conditional expectation $\hat{z}_{ij}^{(k)}$ using the following equation for $k = 0,1,2,\dots$ iteration

$$\hat{z}_{ij}^{(k)} = E\left(Z_{ij}|y_j, \widehat{\mathbf{\Theta}}^{(k)}\right) = \frac{\widehat{w}_i^{(k)} f_i\left(y_j; \hat{\theta}_i^{(k)}\right)}{f\left(y_j|\widehat{\mathbf{\Theta}}^{(k)}\right)}\ . \tag{7}$$

Then, we can form the objective function $Q\left(\mathbf{\Theta}; \widehat{\mathbf{\Theta}}^{(k)}\right)$ as

$$Q\left(\mathbf{\Theta}; \widehat{\mathbf{\Theta}}^{(k)}\right) = \sum_{j=1}^{n}\sum_{i=1}^{g} \hat{z}_{ij}^{(k)} \log\left(w_i f_i(y_j;\theta_i)\right) \tag{8}$$

3. M-step: To obtain the $(k+1)th$ parameter estimate, maximize the $Q\left(\mathbf{\Theta}; \widehat{\mathbf{\Theta}}^{(k)}\right)$ with respect to $\mathbf{\Theta}$:

$$\hat{\theta}^{(k+1)} = \arg\max_{\theta} Q\left(\mathbf{\Theta}; \widehat{\mathbf{\Theta}}^{(k)}\right). \tag{9}$$

4. Repeat these E and M steps until the convergence criterion $\|\widehat{\mathbf{\Theta}}^{(k+1)} - \widehat{\mathbf{\Theta}}^{(k)}\| < \Delta$ is obtained. Alternatively, $\|\ell(\widehat{\mathbf{\Theta}}^{(k+1)}) - \ell(\widehat{\mathbf{\Theta}}^{(k)})\| < \Delta$ or $\|\ell(\widehat{\mathbf{\Theta}}^{(k+1)})/\ell(\widehat{\mathbf{\Theta}}^{(k)}) - 1\| < \Delta$ can also be used as a stopping rule (see [29] for more detailed explanations).

3. DETERMINING NUMBER OF COMPONENTS

To find the number of components g in a mixture model is a major problem. There are two commonly used techniques for choosing number of components which are information criteria and parametric bootstrapping of the likelihood ratio test statistic values for testing the following hypothesis

$H_0: g = g_0$
$H_1: g = g_0 + 1$
where $g_0 \in Z^+$ [30].

In this paper, we use package **mclust** [31-33] in R to select the number of components for the HDI data set. The best model can be obtained by using a statistical criteria for model selection after fitting models to the data set by the ML estimation method. The Bayesian Information Criterion (BIC) ([34]) is a model selection tool based on the maximized log-likelihood and a penalty term on the number of parameters in the model (see, [16,35] for more detailed explanations about model-based clustering).

In Figure 1, we give the BIC values provided in the **mclust** for the equal and unequal variance model parameterization and up to 9 clusters for the HDI data set. Also, we display the classification plot from **mclust** in this figure. We can observe from Figure 1 that the best model is equal variance with two components in terms of BIC. Also, the other two model unequal variance with 2 and 3 components follows this model.

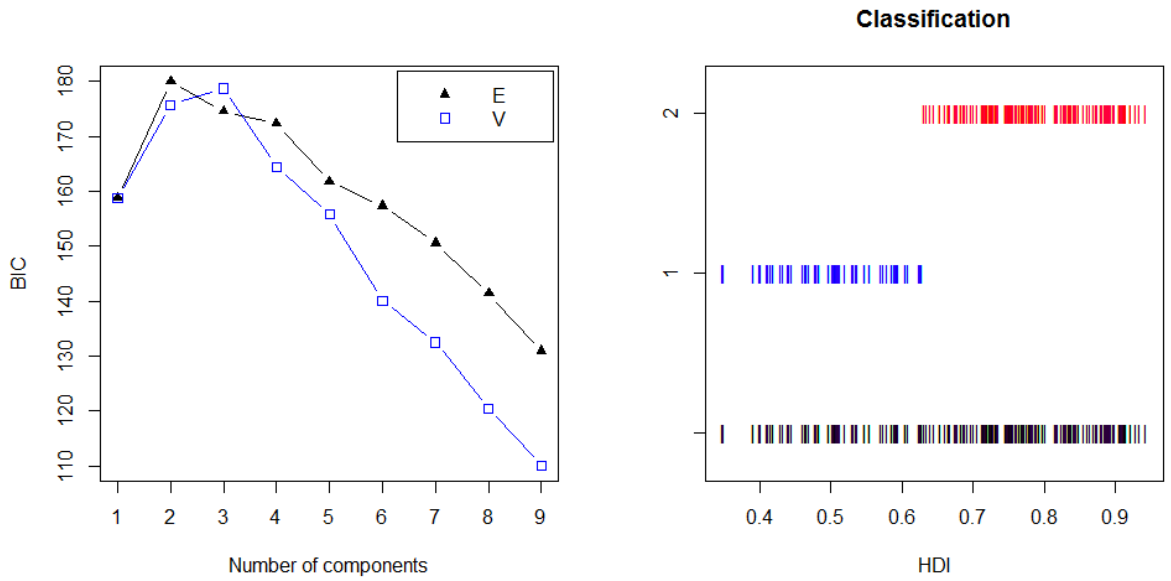

Figure 1. Left: BIC for the equal (E) and unequal (V) variance model parameterization and up to 9 clusters for the HDI data set. Right: The classification plot, all of the data is demonstrated at the bottom of the plot and the disjointed clusters displayed with different levels.

The silhouette statistic proposed by [36] is another way to find the number of groups in the data. The average silhouette width is used to estimate the number of clusters by using the separation with two or more clusters that produce the maximum value of average silhouette width. In Figure 2, we display the mean silhouette value with 9 clusters for the HDI data set which is obtained from **silhouette** function in

MATLAB R2013a. We have also similar observation from this figure that optimal silhouette value is achieved when the component number is equal 2.

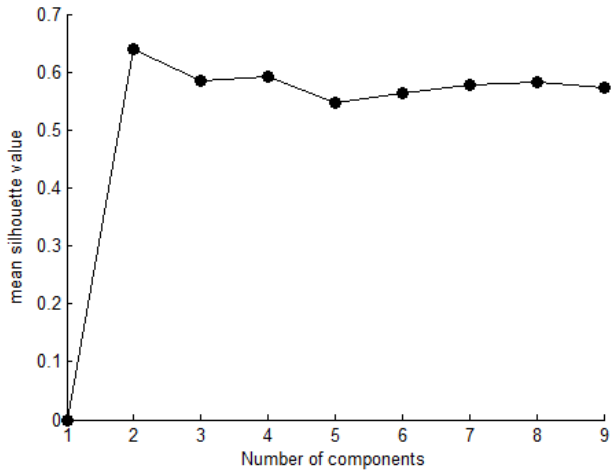

Figure 2. Mean silhouette value for the HDI data set

In summary, we draw a conclusion that there are two groups in the HDI data set according to the BIC plot and mean silhouette value. Now, in the Application Section, we take the component number as 2 and perform all computations using this assumption.

4. APPLICATION

In this section, we will use the HDI data set which includes the HDI values of 188 countries from the HDI report 2015 [13]. The data set is given in Table 2 which consists of HDI values of very high, high, medium and low human development countries. We note that according to the Technical Notes of [13] there are four groups of development countries, see Table 1. However, we use model-based clustering and obtain two groups. The first group includes low and more than half of medium development countries. The second group contains the rest of the countries which are high and very high development countries and a part of medium development countries.

Table 2. The HDI values of 188 countries

0.944	0.935	0.930	0.923	0.922	0.916	0.916	0.915	0.913	0.913	0.912	0.910
0.908	0.907	0.907	0.899	0.898	0.894	0.892	0.891	0.890	0.888	0.885	0.883
0.880	0.876	0.873	0.870	0.865	0.861	0.856	0.850	0.850	0.845	0.844	0.843
0.839	0.839	0.837	0.836	0.835	0.832	0.830	0.828	0.824	0.819	0.818	0.816
0.802	0.798	0.798	0.793	0.793	0.793	0.790	0.788	0.785	0.783	0.782	0.780
0.780	0.779	0.777	0.772	0.772	0.771	0.769	0.769	0.766	0.766	0.762	0.761
0.757	0.756	0.755	0.754	0.752	0.751	0.750	0.748	0.747	0.747	0.736	0.734
0.733	0.733	0.733	0.732	0.729	0.727	0.727	0.727	0.726	0.724	0.724	0.721
0.720	0.720	0.719	0.717	0.715	0.715	0.714	0.706	0.702	0.698	0.693	0.690
0.688	0.684	0.684	0.679	0.677	0.675	0.668	0.666	0.666	0.666	0.662	0.655
0.654	0.646	0.640	0.636	0.631	0.628	0.628	0.627	0.624	0.609	0.606	0.605
0.595	0.594	0.594	0.591	0.590	0.587	0.586	0.579	0.575	0.570	0.555	0.555
0.548	0.548	0.538	0.536	0.532	0.531	0.521	0.514	0.512	0.510	0.509	0.506
0.506	0.505	0.503	0.498	0.497	0.484	0.483	0.483	0.483	0.480	0.479	0.470
0.467	0.466	0.465	0.462	0.445	0.442	0.441	0.433	0.430	0.420	0.419	0.416
0.413	0.411	0.402	0.400	0.392	0.391	0.350	0.348				

We will attempt to determine the best model for the HDI data set by fitting mixtures of distributions. We will use finite mixture modeling using the normal distribution (MixN), finite mixture modeling using the t distribution (Mixt), finite mixture modeling using the skew normal distribution (MixSN) and finite mixture modeling using the skew t distribution (MixST) for modeling the HDI data set. The comparison will be done using the values of the Akaike Information Criterion (AIC) [37], the Bayesian Information Criterion (BIC) [34], the Efficient Determination Criterion (EDC) [38] and the Integrated Completed Likelihood Criterion (ICL) [39]. AIC, BIC and EDC can be computed by using the following form

$$-2\ell(\widehat{\mathbf{\Theta}}) + mc_n ,$$

where $\ell(\cdot)$ represents the maximized log-likelihood, m is the number of free parameters to be estimated under the interested model and c_n is the penalty term. Here, we use $c_n = 2$ for AIC, $c_n = \log(n)$ for BIC and $c_n = 0.2\sqrt{n}$ for EDC. The ICL has the form

$$-2\ell^*(\widehat{\mathbf{\Theta}}) + m \log n ,$$

where $\ell^*(\cdot)$ shows the integrated log-likelihood (see for more detail [25,40]).

In Table 3, we give the estimation results for MixN, Mixt, MixSN and MixST. This table contains the estimates, standard errors (SE) of estimates, log-likelihood and values of AIC, BIC, EDC and ICL. We use package **mixsmsn** [41,42] in R to obtain these estimation results for MixN, Mixt, MixSN and MixST. We also display the histogram of the HDI data set with the fitted densities obtained from MixN, Mixt, MixSN and MixST in Figure 3. We can observe from these results that the weight of the first group is 0.2946 and location of the first group is 0.4958. Also, the weight of the second group is 0.7054 and location of the second group is 0.7744. We can consider the dispersion of low-medium and medium-high-very high development countries. We can see from the values of AIC, BIC, EDC and ICL that MixN has the smallest AIC, BIC, EDC and ICL values. Therefore, MixN gives the best fit for modeling the HDI data set. Since the data may not be skewed and heavy-tailed, the best-fitted model is MixN. This result may be changed after years when the HDI values of countries and development dispersion of countries are changed.

Table 3. ML estimation results of the HDI data set for MixN, Mixt, MixSN and MixST

	MixN		Mixt		MixSN		MixST	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
w_1	0.7054	0.0578	0.6879	0.0624	0.7098	0.1272	0.7109	0.1339
μ_1	0.7744	0.0126	0.7783	0.0129	0.7491	1.1689	0.7524	1.1486
μ_2	0.4958	0.0185	0.5027	0.0198	0.5130	0.9188	0.5095	0.9749
σ_1	0.0932	0.0129	0.0902	0.0485	0.0970	0.2370	0.0958	0.2132
σ_2	0.0722	0.0143	0.0763	0.0490	0.0735	0.1821	0.0722	0.1671
λ_1	-	-	-	-	0.3298	17.1100	0.2784	16.5003
λ_2	-	-	-	-	-0.3382	18.1117	-0.2780	18.8293
ν	-	-	100.0000	114.8168	-	-	100.0000	111.2107
$\ell(\widehat{\mathbf{\Theta}})$	100.9591		101.8567		100.9169		101.6632	
AIC	-191.9182		-191.7134		-187.8338		-187.3264	
BIC	-175.7360		-175.5313		-165.1787		-164.6713	
EDC	-188.2069		-188.0022		-182.6379		-182.1306	
ICL	-154.7095		-153.3507		-144.2606		-143.3912	

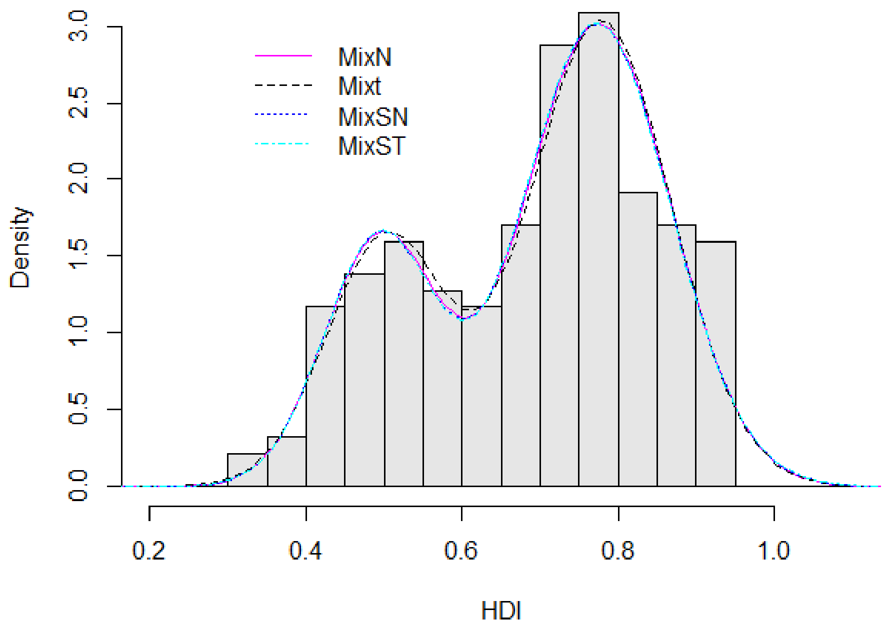

Figure 3. Histogram of the HDI data set with the fitted two component mixture densities obtained from MixN, Mixt, MixSN and MixST.

5. CONCLUSIONS

In this study, we have explored to model the HDI data set with the finite mixture models. We have given the general ML estimation concept of mixture model based on the EM algorithm. Then, we have chosen the cluster numbers using the model-based clustering. We have compared the finite mixture modeling using normal, t, skew normal and skew t distributions to find the best-fitted model for the HDI data set. We have observed from the estimation results that the finite mixture modeling using the normal distribution has the best fit for modeling the HDI data set. We also reason that the HDI data may not be skewed and heavy-tailed.

Further, in literature, there exist some studies which include multimodal distributions. For instance, some of them are [43-50] etc. These proposed distributions can be used to model data sets which have modality. In our study, we propose to model HDI data set using the mixtures of distributions. Therefore, as a future study, this study can be extended to compare modeling capability of mixture models over these multimodal distributions.

ACKNOWLEDGEMENTS

The author would like to thank two anonymous referees and the editor for their helpful comments and suggestions that have considerably improved the paper.

APPENDIX

In this part of the study, we consider the MixN, Mixt, MixSN and MixST for modeling the HDI data set. Therefore, we will give some details about these distributions and also the parameter estimations for these mixtures of distributions.

- Let Y be a random variable from normal distribution with the following pdf

$$f(y;\mu,\sigma^2)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(y-\mu)^2}{2\sigma^2}},\ \ y\in\mathbb{R}, \tag{10}$$

where $\mu\in\mathbb{R}$ and $\sigma^2>0$ show the mean and variance.

- Let Y be a random variable from Student's t distribution with the following pdf

$$f(y;\mu,\sigma^2,\nu)=\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{(y-\mu)^2}{\sigma^2\nu}\right)^{-\frac{\nu+1}{2}},\ \ y\in\mathbb{R}, \tag{11}$$

where $\mu\in\mathbb{R}$ is the location parameter, $\sigma^2>0$ is the scale parameter and $\nu>0$ is the degrees of freedom parameter.

- Let Y be a random variable from skew normal distribution [21,22] with the following pdf

$$f(y;\mu,\sigma^2,\lambda)=\frac{2}{\sigma}\phi\left(\frac{y-\mu}{\sigma}\right)\Phi\left(\lambda\left(\frac{y-\mu}{\sigma}\right)\right),y\in\mathbb{R}, \tag{12}$$

where $\mu\in\mathbb{R}$ is the location parameter, $\sigma^2>0$ is the scale parameter and $\lambda\in\mathbb{R}$ is the skewness parameter. Here, $\phi(\cdot)$ and $\Phi(\cdot)$ denote the pdf of the standard normal distribution and the cumulative distribution function (cdf) of the standard normal distribution.

- Let Y be a random variable from skew t distribution [24] with the following pdf

$$f(y;\mu,\sigma^2,\lambda,\nu)=\frac{2}{\sigma}t_\nu(\eta)T_{\nu+1}\left(\lambda\eta\sqrt{\frac{\nu+1}{\eta^2+\nu}}\right),\ \ \eta=\left(\frac{y-\mu}{\sigma}\right),\ \ y\in\mathbb{R}, \tag{13}$$

where $\mu\in\mathbb{R}$ is the location parameter, $\sigma^2>0$ is the scale parameter, $\lambda\in\mathbb{R}$ is the skewness parameter and $\nu>0$ is the degrees of freedom parameter. Here, $t_\nu(\cdot)$ and $T_\nu(\cdot)$ represent the pdf and the cdf of the Student's t distribution with degrees of freedom ν.

Now, we will give the estimates for the mixtures of distributions that used in this study.

Mixtures of normal distributions:

When the distribution of components is normal distribution in the mixture model given in (2) with the parameter vector $\mathbf{\Theta}=\left(w_1,\dots,w_{g-1},\mu_1,\dots,\mu_g,\sigma_1^2,\dots,\sigma_g^2\right)'$, the parameter estimation for the mixtures of normal distributions is given with the following EM algorithm.

EM algorithm:

1. Take initial parameter estimate $\mathbf{\Theta}^{(0)}$ and a stopping rule Δ.
2. E-step: Compute the following conditional expectation

$$\hat{z}_{ij}^{(k)} = E\left(Z_{ij} \middle| y_j, \widehat{\mathbf{\Theta}}^{(k)}\right) = \frac{\widehat{w}_i^{(k)} f_i\left(y_j; \hat{\mu}_i^{(k)}, \hat{\sigma}_i^{2(k)}\right)}{\sum_{i=1}^{g} \widehat{w}_i^{(k)} f_i\left(y_j; \hat{\mu}_i^{(k)}, \hat{\sigma}_i^{2(k)}\right)} \cdot \tag{14}$$

where $f_i\left(y_j; \mu_i, \sigma_i^2\right)$ is the pdf of the normal distribution given in (10).

3. M-step: Update the following parameter estimates

$$\widehat{w}_i^{(k+1)} = \frac{\sum_{j=1}^{n} \hat{z}_{ij}^{(k)}}{n}, \tag{15}$$

$$\hat{\mu}_i^{(k+1)} = \frac{\sum_{j=1}^{n} \hat{z}_{ij}^{(k)} y_j}{\sum_{j=1}^{n} \hat{z}_{ij}^{(k)}}, \tag{16}$$

$$\hat{\sigma}_i^{2(k+1)} = \frac{\sum_{j=1}^{n} \hat{z}_{ij}^{(k)} \left(y_j - \hat{\mu}_i^{(k)}\right)^2}{\sum_{j=1}^{n} \hat{z}_{ij}^{(k)}}. \tag{17}$$

4. Repeat E and M steps until the convergence criteria $\left\|\widehat{\mathbf{\Theta}}^{(k+1)} - \widehat{\mathbf{\Theta}}^{(k)}\right\| < \Delta$ is obtained.

<u>Mixtures of t distributions:</u>

When the distribution of components is t distribution in the mixture model given in (2) with the parameter vector $\mathbf{\Theta} = \left(w_1, \dots, w_{g-1}, \mu_1, \dots, \mu_g, \sigma_1^2, \dots, \sigma_g^2, \nu_1, \dots, \nu_g\right)'$, the parameter estimation for the mixtures of t distributions can be obtained as follows.

EM algorithm:

1. Set initial parameter estimate $\mathbf{\Theta}^{(0)}$ and a stopping rule Δ.
2. E-step: Calculate the following conditional expectations

$$\hat{z}_{ij}^{(k)} = E\left(Z_{ij} \middle| y_j, \widehat{\mathbf{\Theta}}^{(k)}\right) = \frac{\widehat{w}_i^{(k)} f_i\left(y_j; \hat{\mu}_i^{(k)}, \hat{\sigma}_i^{2(k)}, \hat{\nu}_i^{(k)}\right)}{\sum_{i=1}^{g} \widehat{w}_i^{(k)} f_i\left(y_j; \hat{\mu}_i^{(k)}, \hat{\sigma}_i^{2(k)}, \hat{\nu}_i^{(k)}\right)}, \tag{18}$$

$$\hat{u}_{1ij}^{(k)} = E\left(U_j \middle| y_j, \widehat{\mathbf{\Theta}}^{(k)}\right) = \frac{\hat{\nu}_i^{(k)} + 1}{\hat{\nu}_i^{(k)} + \left(\left(y_j - \hat{\mu}_i^{(k)}\right)/\hat{\sigma}_i^{(k)}\right)^2}, \tag{19}$$

$$\hat{u}_{2ij}^{(k)} = E\left(\log\left(U_j\right) \middle| y_j, \widehat{\mathbf{\Theta}}^{(k)}\right) = \psi\left(\frac{\hat{\nu}_i^{(k)} + 1}{2}\right) - \log\left(\frac{1}{2}\left(\hat{\nu}_i^{(k)} + \frac{\left(y_j - \hat{\mu}_i^{(k)}\right)^2}{\hat{\sigma}_i^2}\right)\right), \tag{20}$$

where $f_i\left(y_j; \mu_i, \sigma_i^2, \nu_i\right)$ is the pdf of the t distribution given in (11) and $\psi(\cdot)$ shows the digamma function.

3. M-step: The maximization yields the following parameter estimates

$$\widehat{w}_i^{(k+1)} = \frac{\sum_{j=1}^{n} \hat{z}_{ij}^{(k)}}{n}, \tag{21}$$

$$\hat{\mu}_i^{(k+1)} = \frac{\sum_{j=1}^{n} \hat{z}_{ij}^{(k)} \hat{u}_{1ij}^{(k)} y_j}{\sum_{j=1}^{n} \hat{z}_{ij}^{(k)} \hat{u}_{1ij}^{(k)}}, \tag{22}$$

$$\hat{\sigma}_i^{2(k+1)} = \frac{\sum_{j=1}^{n} \hat{z}_{ij}^{(k)} \hat{u}_{1ij}^{(k)} \left(y_j - \hat{\mu}_i^{(k)}\right)^2}{\sum_{j=1}^{n} \hat{z}_{ij}^{(k)}}. \tag{23}$$

Also, the parameter estimate for ν can be obtained by solving the following equation

$$\sum_{i=1}^{n} \hat{z}_{ij}^{(k)} \left(-\psi\left(\frac{\nu_i}{2}\right) + \log\left(\frac{\nu_i}{2}\right) + 1 + \hat{u}_{2i}^{(k)} - \hat{u}_{1i}^{(k)}\right) = 0. \tag{24}$$

4. Repeat E and M steps until the convergence rule $\left\|\widehat{\mathbf{\Theta}}^{(k+1)} - \widehat{\mathbf{\Theta}}^{(k)}\right\| < \Delta$ is satisfied.

Mixtures of skew normal distributions:

When the distribution of components is skew normal distribution in the mixture model given in (2) with the parameter vector $\mathbf{\Theta} = \left(w_1, \ldots, w_{g-1}, \mu_1, \ldots, \mu_g, \sigma_1^2, \ldots, \sigma_g^2, \lambda_1, \ldots, \lambda_g\right)'$, the results about parameter estimation for the mixtures of skew normal distributions is summarized as follows.

EM algorithm:

1. Choose initial parameter estimate $\mathbf{\Theta}^{(0)}$ and a stopping rule Δ.
2. E-step: Compute the following conditional expectations

$$\hat{z}_{ij}^{(k)} = E\left(Z_{ij} \middle| y_j, \widehat{\mathbf{\Theta}}^{(k)}\right) = \frac{\widehat{w}_i^{(k)} f_i\left(y_j; \hat{\mu}_i^{(k)}, \hat{\sigma}_i^{2(k)}, \hat{\lambda}_i^{(k)}\right)}{\sum_{i=1}^{g} \widehat{w}_i^{(k)} f_i\left(y_j; \hat{\mu}_i^{(k)}, \hat{\sigma}_i^{2(k)}, \hat{\lambda}_i^{(k)}\right)}, \tag{25}$$

$$\hat{t}_{1ij}^{(k)} = E\left(\gamma_j \middle| y_j, \widehat{\mathbf{\Theta}}^{(k)}\right) = \hat{\delta}_{\lambda_i}^{(k)} \hat{\eta}_{ij}^{(k)} + \sqrt{1 - \hat{\delta}_{\lambda_i}^{2(k)}} \frac{\phi\left(\hat{\lambda}_i^{(k)} \hat{\eta}_{ij}^{(k)}\right)}{\Phi\left(\hat{\lambda}_i^{(k)} \hat{\eta}_{ij}^{(k)}\right)}, \tag{26}$$

$$\hat{t}_{2ij}^{(k)} = E\left(\gamma_j^2 \middle| y_j, \widehat{\mathbf{\Theta}}^{(k)}\right) = 1 - \hat{\delta}_{\lambda_i}^{2(k)} + \hat{\delta}_{\lambda_i}^{(k)} \hat{\eta}_{ij}^{(k)} \hat{t}_{1ij}^{(k)}, \tag{27}$$

where $f_i\left(y_j; \mu_i, \sigma_i^2, \lambda_i\right)$ is the pdf of the skew normal distribution given in (12), $\hat{\delta}_{\lambda_i}^{(k)} = \hat{\lambda}_i^{(k)} \Big/ \sqrt{1 + \hat{\lambda}_i^{2(k)}}$ and $\hat{\eta}_{ij}^{(k)} = \frac{\left(y_j - \hat{\mu}_i^{(k)}\right)}{\hat{\sigma}_i^{(k)}}$.

3. M-step: Update the following parameter estimates

$$\widehat{w}_i^{(k+1)} = \frac{\sum_{j=1}^{n} \hat{z}_{ij}^{(k)}}{n}, \tag{28}$$

$$\hat{\mu}_i^{(k+1)} = \frac{\sum_{j=1}^{n} \hat{z}_{ij}^{(k)} \left(y_j - \hat{\alpha}_i^{(k)} \hat{t}_{1ij}^{(k)} \right)}{\sum_{j=1}^{n} \hat{z}_{ij}^{(k)}}, \tag{29}$$

$$\hat{\alpha}_i^{(k+1)} = \frac{\sum_{j=1}^{n} \hat{z}_{ij}^{(k)} \hat{t}_{1ij}^{(k)} \left(y_j - \hat{\mu}_i^{(k)} \right)}{\sum_{j=1}^{n} \hat{z}_{ij}^{(k)} \hat{t}_{2ij}^{(k)}}, \tag{30}$$

$$\hat{\kappa}_i^{2(k+1)} = \frac{1}{\sum_{j=1}^{n} \hat{z}_{ij}^{(k)}} \sum_{j=1}^{n} \hat{z}_{ij}^{(k)} \left(\left(y_j - \hat{\mu}_i^{(k)} \right)^2 - 2\hat{\alpha}_i^{(k)} \hat{t}_{1ij}^{(k)} \left(y_j - \hat{\mu}_i^{(k)} \right) + \hat{\alpha}_i^{2(k)} \hat{t}_{2ij}^{(k)} \right) \tag{31}$$

$$\hat{\sigma}_i^{2(k+1)} = \hat{\kappa}_i^{2(k+1)} + \hat{\alpha}_i^{2(k+1)}, \tag{32}$$

$$\hat{\lambda}_i^{(k+1)} = \hat{\delta}_{\lambda_i}^{(k+1)} \left(1 - \hat{\delta}_{\lambda_i}^{2(k+1)} \right)^{-1/2}. \tag{33}$$

4. Repeat E and M steps until the convergence rule $\left\| \widehat{\mathbf{\Theta}}^{(k+1)} - \widehat{\mathbf{\Theta}}^{(k)} \right\| < \Delta$ is obtained.

<u>Mixtures of skew t distributions:</u>

If the distribution of components is skew t distribution in the mixture model given in (2) with the parameter vector $\mathbf{\Theta} = \left(w_1, \dots, w_{g-1}, \mu_1, \dots, \mu_g, \sigma_1^2, \dots, \sigma_g^2, \lambda_1, \dots, \lambda_g, \nu_1, \dots, \nu_g \right)'$, the parameter estimation for the mixtures of skew t distributions will be given with the following EM algorithm.

EM algorithm:

1. Choose initial parameter estimate $\mathbf{\Theta}^{(0)}$ and a stopping rule Δ.
2. E-step: Calculate the following conditional expectations

$$\hat{z}_{ij}^{(k)} = E\left(Z_{ij} \middle| y_j, \widehat{\mathbf{\Theta}}^{(k)} \right) = \frac{\widehat{w}_i^{(k)} f_i \left(y_j; \hat{\mu}_i^{(k)}, \hat{\sigma}_i^{2(k)}, \hat{\lambda}_i^{(k)}, \hat{\nu}_i^{(k)} \right)}{\sum_{i=1}^{g} \widehat{w}_i^{(k)} f_i \left(y_j; \hat{\mu}_i^{(k)}, \hat{\sigma}_i^{2(k)}, \hat{\lambda}_i^{(k)}, \hat{\nu}_i^{(k)} \right)}, \tag{34}$$

$$\hat{s}_{1ij}^{(k)} = E\left(Z_{ij} \tau_j \middle| y_j, \widehat{\mathbf{\Theta}}^{(k)} \right) = \hat{z}_{ij}^{(k)} \left(\frac{\hat{\nu}_i^{(k)} + 1}{\hat{\eta}_{ij}^{2(k)} + \hat{\nu}_i^{(k)}} \right) \frac{T_{\hat{\nu}_i^{(k)}+3} \left(\widehat{M}_{ij}^{(k)} \sqrt{\frac{\hat{\nu}_i^{(k)} + 3}{\hat{\nu}_i^{(k)} + 1}} \right)}{T_{\hat{\nu}_i^{(k)}+1} \left(\widehat{M}_{ij}^{(k)} \right)}, \tag{35}$$

$$\hat{s}_{2ij}^{(k)} = E\left(Z_{ij} \gamma_j \tau_j \middle| y_j, \widehat{\mathbf{\Theta}}^{(k)} \right) = \frac{1}{\hat{\sigma}_i^{(k)}} \hat{\delta}_{\lambda_i}^{(k)} \left(y_j - \hat{\mu}_i^{(k)} \right) \hat{s}_{1ij}^{(k)}$$

$$+ \hat{z}_{ij}^{(k)} \frac{\sqrt{1 - \hat{\delta}_{\lambda_i}^{2(k)}}}{\pi \hat{\sigma}_i^{(k)} \hat{f}\left(y_j \right)^{(k)}} \left(\frac{\hat{\eta}_{ij}^{2(k)}}{\hat{\nu}_i^{(k)} \left(1 - \hat{\delta}_{\lambda_i}^{2(k)} \right)} + 1 \right)^{-\left(\frac{\hat{\nu}_i^{(k)}}{2} + 1 \right)}, \tag{36}$$

$$\hat{s}_{3ij}^{(k)} = E\left(Z_{ij}\gamma_j^2\tau_j \middle| y_j, \widehat{\boldsymbol{\Theta}}^{(k)}\right) = \hat{\delta}_{\lambda_i}^{2(k)}\left(\frac{y_j - \hat{\mu}_i^{(k)}}{\hat{\sigma}_i^{(k)}}\right)^2 \hat{s}_{1ij}^{(k)} + \hat{z}_{ij}^{(k)}\left\{\left(1 - \hat{\delta}_{\lambda_i}^{2(k)}\right)\right.$$

$$\left. + \frac{\hat{\delta}_{\lambda_i}^{(k)}\left(y_j - \hat{\mu}_i^{(k)}\right)\sqrt{1 - \hat{\delta}_{\lambda_i}^{2(k)}}}{\pi\hat{\sigma}_i^{2(k)}\hat{f}(y_j)^{(k)}}\left(\frac{\hat{\eta}_{ij}^{2(k)}}{\hat{\nu}_i^{(k)}\left(1 - \hat{\delta}_{\lambda_i}^{2(k)}\right)} + 1\right)^{-\left(\frac{\hat{\nu}_i^{(k)}}{2}+1\right)}\right\}, \tag{37}$$

$$\hat{s}_{4ij}^{(k)} = E\left(Z_{ij}\log(\tau_j) \middle| y_j, \widehat{\boldsymbol{\Theta}}^{(k)}\right) = \hat{z}_{ij}^{(k)}\left\{\psi\left(\frac{\hat{\nu}_i^{(k)} + 1}{2}\right) - \log\left(\frac{\hat{\eta}_{ij}^{2(k)} + \hat{\nu}_i^{(k)}}{2}\right)\right.$$

$$+ \left(\frac{\hat{\nu}_i^{(k)} + 1}{\hat{\eta}_{ij}^{2(k)} + \hat{\nu}_i^{(k)}}\right)\left(\frac{T_{\hat{\nu}_i^{(k)}+3}\left(\hat{\lambda}_i^{(k)}\hat{\eta}_{ij}^{(k)}\sqrt{\frac{\hat{\nu}_i^{(k)} + 3}{\hat{\eta}_{ij}^{2(k)} + \hat{\nu}_i^{(k)}}}\right)}{T_{\hat{\nu}_i^{(k)}+1}\left(\hat{\lambda}_i^{(k)}\hat{\eta}_{ij}^{(k)}\sqrt{\frac{\hat{\nu}_i^{(k)} + 1}{\hat{\eta}_{ij}^{2(k)} + \hat{\nu}_i^{(k)}}}\right)} - 1\right)$$

$$+ \frac{\hat{\lambda}_i^{(k)}\hat{\eta}_{ij}^{(k)}\left(\hat{\eta}_{ij}^{2(k)} - 1\right)}{\sqrt{\left(\hat{\nu}_i^{(k)} + 1\right)\left(\hat{\eta}_{ij}^{2(k)} + \hat{\nu}_i^{(k)}\right)^3}} \frac{t_{\hat{\nu}_i^{(k)}+1}\left(\hat{\lambda}_i^{(k)}\hat{\eta}_{ij}^{(k)}\sqrt{\frac{\hat{\nu}_i^{(k)} + 1}{\hat{\eta}_{ij}^{2(k)} + \hat{\nu}_i^{(k)}}}\right)}{T_{\hat{\nu}_i^{(k)}+1}\left(\hat{\lambda}_i^{(k)}\hat{\eta}_{ij}^{(k)}\sqrt{\frac{\hat{\nu}_i^{(k)} + 1}{\hat{\eta}_{ij}^{2(k)} + \hat{\nu}_i^{(k)}}}\right)}$$

$$\left. + \frac{1}{T_{\hat{\nu}_i^{(k)}+1}\left(\hat{\lambda}_i^{(k)}\hat{\eta}_{ij}^{(k)}\sqrt{\frac{\hat{\nu}_i^{(k)} + 1}{\hat{\eta}_{ij}^{2(k)} + \hat{\nu}_i^{(k)}}}\right)} \int_{-\infty}^{\widehat{M}_{ij}^{(k)}} g_{\hat{\nu}_i^{(k)}}(x)\, t_{\hat{\nu}_i^{(k)}+1}(x)\,dx\right\}, \tag{38}$$

where $f_i\left(y_j; \mu_i, \sigma_i^2, \lambda_i, \nu_i\right)$ is the pdf of the skew t distribution given in (13) and

$$\hat{\delta}_{\lambda_i}^{(k)} = \hat{\lambda}_i^{(k)} \Big/ \sqrt{1 + \hat{\lambda}_i^{2(k)}},\ \hat{\eta}_{ij}^{(k)} = \frac{\left(y_j - \hat{\mu}_i^{(k)}\right)}{\hat{\sigma}_i^{(k)}},\ \widehat{M}_{ij}^{(k)} = \hat{\lambda}_i^{(k)}\hat{\eta}_{ij}^{(k)}\sqrt{\frac{\hat{\nu}_i^{(k)}}{\hat{\eta}_{ij}^{2(k)} + \hat{\nu}_i^{(k)}}},$$

$$\hat{f}(y_j)^{(k)} = \sum_{i=1}^{g} \widehat{w}_i^{(k)} \frac{2}{\hat{\sigma}_i^{(k)}} t_{\hat{\nu}_i^{(k)}}\left(\hat{\eta}_{ij}^{(k)}\right) T_{\hat{\nu}_i^{(k)}+1}\left(\widehat{M}_{ij}^{(k)}\right),$$

$$g_{\hat{\nu}_i^{(k)}}(x) = \psi\left(\frac{\hat{\nu}_i^{(k)}+2}{2}\right) - \psi\left(\frac{\hat{\nu}_i^{(k)}+1}{2}\right) - \log\left(1 + \frac{x^2}{\hat{\nu}_i^{(k)}+1}\right) + \frac{x^2\left(\hat{\nu}_i^{(k)}+1\right) - \hat{\nu}_i^{(k)} - 1}{\left(\hat{\nu}_i^{(k)}+1\right)\left(\hat{\nu}_i^{(k)}+1+x^2\right)}.$$

3. M-step: Update the following parameter estimates

$$\widehat{w}_i^{(k+1)} = \frac{\sum_{j=1}^{n} \hat{z}_{ij}^{(k)}}{n}, \tag{39}$$

$$\hat{\mu}_i^{(k+1)} = \frac{\sum_{j=1}^{n} \left(\hat{s}_{1ij}^{(k)} y_j - \hat{\alpha}_i^{(k)} \hat{s}_{2ij}^{(k)} \right)}{\sum_{j=1}^{n} \hat{s}_{1ij}^{(k)}}, \tag{40}$$

$$\hat{\alpha}_i^{(k+1)} = \frac{\sum_{j=1}^{n} \hat{s}_{2ij}^{(k)} \left(y_j - \hat{\mu}_i^{(k)} \right)}{\sum_{j=1}^{n} \hat{s}_{3ij}^{(k)}}, \tag{41}$$

$$\hat{\kappa}_i^{2(k+1)} = \frac{1}{\sum_{j=1}^{n} \hat{z}_{ij}^{(k)}} \sum_{j=1}^{n} \hat{s}_{1ij}^{(k)} \left(y_j - \hat{\mu}_i^{(k)} \right)^2 - 2\hat{\alpha}_i^{(k)} \hat{s}_{2ij}^{(k)} \left(y_j - \hat{\mu}_i^{(k)} \right) + \hat{\alpha}_i^{2(k)} \hat{s}_{3ij}^{(k)}, \tag{42}$$

$$\hat{\sigma}_i^{2(k+1)} = \hat{\kappa}_i^{2(k+1)} + \hat{\alpha}_i^{2(k+1)}, \tag{43}$$

$$\hat{\lambda}_i^{(k+1)} = \hat{\delta}_{\lambda_i}^{(k+1)} \left(1 - \hat{\delta}_{\lambda_i}^{2(k+1)} \right)^{-1/2}. \tag{44}$$

Also, the parameter estimate for ν_i can be found by solving the following equation

$$-\psi\left(\frac{\nu_i}{2}\right) + \log\left(\frac{\nu_i}{2}\right) + 1 + \frac{\sum_{j=1}^{n} \left(\hat{s}_{4ij}^{(k)} - \hat{s}_{1ij}^{(k)} \right)}{\sum_{j=1}^{n} \hat{z}_{ij}^{(k)}} = 0. \tag{45}$$

4. Repeat E and M steps until the convergence criterion $\left\| \widehat{\boldsymbol{\Theta}}^{(k+1)} - \widehat{\boldsymbol{\Theta}}^{(k)} \right\| < \Delta$ is obtained.

REFERENCES

[1] UNDP. Human Development Report 1990. New York: Oxford University Press, 1990.

[2] Bhanojirao VV. Human development report 1990: review and assessment. World Development 1991; 19(10): 1451-1460.

[3] Desai M. Human development: Concept and measurement. Eur Econ Rev 1991; 35: 350-357.

[4] Doessel DP, Gounder R. International Comparisons of the Standards of Living and the Human Development Index. Discussion Papers in Economics No. 72, Department of Economics, University of Queensland, Brisbane, 1991.

[5] Hopkins M. Human development revisited: A new UNDP report. World Development 1991; 19(10): 1469-1473.

[6] Kelley AC. The Human Development Index:" Handle with Care". Popul Dev Rev 1991; 315-324.

[7] McGillivray M. The human development index: yet another redundant composite development indicator? World Development, 1991; 19(10): 1461-1468.

[8] McGillivray M, White H. Measuring development?: a statistical critique of the UNDP's human development index. ISS Working Paper Series/General Series, 1992; 135: 1-25.

[9] Pyatt G. Poverty: a wasted decade. Eur Econ Rev 1991; 35(2-3): 358-365.

[10] Trabold-Nübler H. The human development index-a new development indicator?. Intereconomics, 1991; 26(5): 236-243.

[11] Anand S, Sen A. Human Development Index: Methodology and Measurement. Human Development Report Office Occasional Paper No. 12, UNDP, New York, 1992.

[12] Lind NC. Some thoughts on the human development index. Soc Indic Res 1992; 27(1): 89-101.

[13] UNDP. Human Development Report 2015. New York: Oxford University Press, 2015.

[14] Titterington DM, Smith AFM, Markov UE. Statistical Analysis of Finite Mixture Distributions. New York: Wiley, 1985.

[15] McLachlan GJ, Basford KE. Mixture Models: Inference and Application to Clustering. New York: Marcel Dekker, 1988.

[16] McLachlan GJ, Peel D. Finite Mixture Models. New York: Wiley, 2000.
[17] Bishop CM. Pattern Recognition and Machine Learning. Singapore: Springer, 2006.

[18] Frühwirth-Schnatter S. Finite Mixture and Markov Switching Models. New York: Springer, 2006.

[19] Peel D, McLachlan GJ. Robust mixture modeling using the t distribution. Stat Comput 2000; 10: 339-348.

[20] Lin TI, Lee JC, Yen SY. Finite mixture modelling using the skew normal distribution. Stat Sinica 2007b; 17(3): 909-927.

[21] Azzalini A. A class of distributions which includes the normal ones. Scand J Stat 1985; 12(2): 171-178.

[22] Azzalini A. Further results on a class of distributions which includes the normal ones. Statistica 1986; 46(2): 199-208.

[23] Lin TI, Lee JC, Hsieh WJ. Robust mixture modeling using the skew t distribution. Stat Comput 2007a; 17: 81–92.

[24] Azzalini A, Capitanio A. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution. J Roy Stat Soc B Met 2003; 65(2): 367-389.

[25] Basso RM, Lachos VH, Cabral CRB, Ghosh P. Robust mixture modeling based on scale mixtures of skew-normal distributions. Comput Stat Data An 2010; 54(12): 2926-2941.

[26] Ho H, Pyne S, Lin T. Maximum likelihood inference for mixtures of skew Student-t-normal distributions through practical EM-type algorithms. Stat Comput 2012; 22(1): 287-299.

[27] Doğru FZ, Arslan O. Parameter estimation for mixtures of skew Laplace normal distributions and application in mixture regression modeling. Commun Stat Theory DOI: 10.1080/03610926.2016.1252400 (accepted) 2017.

[28] Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. J Roy Stat Soc B Met 1977; 39: 1-38.

[29] Dias JG, Wedel M. An empirical comparison of EM, SEM and MCMC performance for problematic gaussian mixture likelihoods. Stat Comput 2004; 14: 323-332.

[30] McLachlan GJ. On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. J Roy Stat Soc C-App 1987; 36: 318-324.

[31] Fraley C, Raftery AE. mclust: Software for model-based cluster analysis. J Classif 1999; 16: 297-306.

[32] Fraley C, Raftery AE. Model-based clustering, discriminant analysis and density estimation. J Am Stat Assoc 2002; 97: 611-631.

[33] Fraley C, Raftery AE. Enhanced software for model-based clustering, density estimation, and discriminant analysis: mclust. J Classif 2003; 20: 263-286.

[34] Schwarz G. Estimating the dimension of a model. Ann Stat 1978; 6(2): 461-464.

[35] Fraley C, Raftery AE. Model-based Microarray Image Analysis. R News 2006; 6: 60-63.

[36] Kaufman L, Rousseeuw PJ. Finding Groups in Data: An Introduction to Cluster Analysis, New York: John Wiley & Sons, 1990.

[37] Akaike H. 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, Budapest: Akademiai Kiado, 1973.

[38] Bai ZD, Krishnaiah PR, Zhao LC. On rates of convergence of efficient detection criteria in signal processing with white noise. IEEE T Inform Theory 1989; 35: 380-388.

[39] Biernacki C, Celeux G, Govaert G. Assessing a mixture model for clustering with the integrated completed likelihood. IEEE T Pattern Anal 2000; 22: 719-725.

[40] Cabral CRB, Lachos VH, Prates MO. Multivariate mixture modeling using skew-normal independent distributions. Comput Stat Data An 2012; 56(1): 126-142.

[41] Prates MO, Lachos VH, Cabral C. mixsmsn: Fitting finite mixture of scale mixture of skew-normal distributions. R package version 0.3-2, 2011.

[42] Prates MO, Lachos VH, Cabral C. mixsmsn: Fitting finite mixture of scale mixture of skew-normal distributions. J Stat Softw 2013; 54(12): 1-20.

[43] Elal-Olivero D. Alpha-skew-normal distribution. Proyecciones (Antofagasta), 2010; 29(3): 224-240.

[44] Jamalizadeh A, Arabpour AR, Balakrishnan N. A generalized skew two-piece skew-normal distribution. Stat Pap 2011; 52(2): 431-446.

[45] Gui W. A generalization of the slashed distribution via alpha skew normal distribution. Stat Method Appl 2014; 23(4): 547-563.

[46] Çankaya MN, Bulut YM, Doğru FZ, Arslan O. A bimodal extension of the generalized gamma distribution. Rev Colomb Estad 2015; 38(2): 371-384.

[47] Acitas S, Senoglu B, Arslan, O. Alpha-skew generalized t distribution. Rev Colomb Estad 2015; 38(2): 353-370.

[48] Hassan MY, El-Bassiouni MY. Bimodal skew-symmetric normal distribution. Commun Stat Theory 2016; 45(5): 1527-1541.

[49] Ramires TG, Ortega EM, Cordeiro GM, Hens N. A bimodal flexible distribution for lifetime data.J Stat Comput Sim 2016; 86(12): 2450-2470.

[50] Rasekhi M, Hamedani GG, Chinipardaz R. A flexible extension of skew generalized normal distribution. METRON, 2017; 75(1): 87-107.