Journal of Mathematics and Modeling in Finance (JMMF)
Vol. 2, No. 2, Summer & Autumn 2022
Research paper

Ridge Shrinkage Estimators in Finite Mixture of Generalized Estimating Equations

Sajad Nezamdoust[1], **Farzad Eskandari**[2]

[1] Department of Statistics, Faculty of Mathematical Sciences and Computer, Allameh Tabatabai University, Tehran, Iran

sa.nezamdoust@gmail.com

[2] Department of Statistics, Faculty of Statistics, Mathematics and Computer, Allameh Tabatabai University, Tehran, Iran

askandari@atu.ac.ir

Abstract:
The paper considers the problem of estimation of the parameters in finite mixture models.In this article, a new method is proposed for of estimation of the parameters in finite mixture models. Traditionally, the parameter estimation in finite mixture models is performed from a likelihood point of view by exploiting the expectation maximization (EM) method and the Least Square Principle. Ridge regression is an alternative to the ordinary least squares method when multicollinearity presents among the regressor variables in multiple linear regression analysis. Accordingly, we propose a new shrinkage ridge estimation approach. Based on this principle, we propose an iterative algorithm called Ridge-Iterative Weighted least Square (RI-WLS) to estimate the parameters. Monte-Carlo simulation studies are conducted to appraise the performance of our method. The results show that the Proposed estimator perform better than the IWLS method.

Keywords: Finite Mixture Model, Least Square Principle, Weighted Least Square, Ridge Estimation
MSC Classification 34K11

1 Introduction

Shrinkage estimation has become a important method for data modeling and has been considered by many researchers in various fields. Shrinkage estimation strategy attempts to incorporate prior uncertain information in the estimation procedure. Prior information may be available on some of the parameters, which are usually incorporated in the model as a constraint, resulted to have restricted models. Such prior information may be useful in increasing the performance of estimators. These are for example Saleh and Kibria (2011), Siray et al. (2015) and later Asar (2017). Finite mixture models (FMM) provide a flexible tool for modelling data when they are coming from more than one population. Applications of mixture distributions can be found in various fields of statistical applications such as biology, genetics,

[2] Corresponding author

Received: 31/08/2022 Accepted: 02/11/2022
https://doi.org/10.22054/jmmf.2022.69854.1073

engineering, marketing, and so on. In the field FMM, mixture of linear regressions has been studied widely, especially when no information about membership of the points assigned to each line was available. These models were introduced by Quandt and Ramsey (1978) as a very general form of switching regression. They used a technique based on moment generating functions to estimate the parameters. However, it has mainly been studied from a likelihood point of view. De Veaux (1989) developed an expectation maximization (EM) approach to fit the two regression situations. Jacobs, Jordan, Nolan, and Hinton (1991) and also Jiang and Tanner (1999) in machine learning programs used finite mixture of regression models. Jones and McLachlan (1992) applied mixtures of regressions to analyze real data and used the EM algorithm to fit these models. McLaughlin and Peel (2000) conducted a comprehensive review of finite mixture models in their book. Hawkins, Allen, and Stromberg (2001) used the score equation to estimate the number of components in a mixture of linear regression models. Zhou and Zhang (2004) developed asymptotic theory for maximum likelihood estimators in mixture regression models. Dias and Wedel (2004) have compared EM and Stochastic EM algorithms to estimate the parameters of Gaussian mixture model. Faria and Soromenho (2010, 2012) compared EM, SEM, and Classication EM to compute the maximum likelihood estimates of the parameters. Xu et al. (2012) proposed mix-GEE estimator based on a finite mixture model for the working correlation to analyze longitudinal data. Schepers (2015) and Eskandari and Ormoz (2016) improved random-starting method for the EM algorithm in finite mixture regression models. In 2020, Rezazadeh et al. developed an algorithm named Iterative Weighted least Square (IWLS) based on Generalized Estimating Equations concept to estimate the parameters.
In this paper, we propose RIWLS estimator. We show that our estimators in finite mixture regression models outperform the Iterative Weighted least Square (IWLS) estimators. We conduct a detailed Monte Carlo simulation to study the performance of the estimators in terms of their GMSE. The article is organized as follows: In Section 2, We discuss the finite mixture regression models and the IWLS method for estimating model parameters. In Section 3, we introduce shrinkage ridge of the parameters. In Section 4, we conduct Monte Carlo simulations to study the performance of the proposed estimators.

2 Finite Mixture of Linear Regressions

Let $y_1, \dots, y_n$ be random samples of size n from a population with density $f(y_i)$. The finite mixture of linear regression model for the i-th subject is given as follows:

$$y_i = \begin{cases} \boldsymbol{x}_i'\beta_1 + \epsilon_{i1} & ; \textit{with probability} \quad \pi_1 \\ \boldsymbol{x}_i'\beta_2 + \epsilon_{i2} & ; \textit{with probability} \quad \pi_2 \\ \vdots & \\ \boldsymbol{x}_i'\beta_K + \epsilon_{iK} & ; \textit{with probability} \quad \pi_K \end{cases}$$

Where y_i is the value of the response variable in the i-th observation $(i = 1, \dots, n)$; $\boldsymbol{x}_i'$ denotes the transpose of the $(p+1)$-dimensional vector of independent variables for the i-th observation, β_j $(j = 1, ..., K)$ denotes the $(p+1)$-dimensional vector of parameters for the j-th component, π_j's are the mixing probabilities $(0 < \pi_j < 1$, for all $j = 1, \dots, K$ and $\sum_{j=1}^{K} \pi_j = 1)$. Finally, ϵ_{ij} are the random errors with the following assumptions:

(i) $E(\epsilon_{ij}) = 0 \quad ; \quad i = 1, \dots, n \quad ; \quad j = 1, \dots, K$

(ii) $Var(\epsilon_{ij}) = \sigma_{ij}^2 \quad ; \quad i = 1, \dots, n \quad ; \quad j = 1, \dots, K$

(iii) $Cov(\epsilon_{ij}, \epsilon_{lr}) = 0 \quad ; i, l = 1, \dots, n \quad ; \quad j, r = 1, \dots, K \quad ; \quad j \neq r$

Thus, the expectation on y_i given x_i can be expressed as:

$$E(y_i|\boldsymbol{x}) = \sum_{j=1}^{K} \pi_j E\left(\boldsymbol{x}_i'\beta_j + \epsilon_{ij}\right) \tag{1}$$

By applying the above assumptions, we can rewrite (1) as follows:

$$E(y_i|\boldsymbol{x}_i) = \sum_{j=1}^{K} \pi_j \left(\boldsymbol{x}_i'\beta_j\right) = \boldsymbol{x}_i^{*'} P\beta = \boldsymbol{x}_i^{*'} B\, \pi \qquad ; i = 1, \dots, n \tag{2}$$

where

$$\boldsymbol{x}_i^{*'} = (\boldsymbol{x}_i', \boldsymbol{x}_i', \dots, \boldsymbol{x}_i')_{1 \times (p+1)K}$$

$$P = \begin{bmatrix} \pi_1 I_{p+1} & 0 & \dots & 0 \\ 0 & \pi_2 I_{p+1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \pi_K I_{p+1} \end{bmatrix}_{(p+1)K \times (p+1)K}$$

$$\beta' = \left(\beta_1', \dots, \beta_K'\right)_{1 \times (p+1)K}$$

$$\pi' = (\pi_1, \dots, \pi_k)_{1 \times K}$$

and B define as below:

$$B = \begin{bmatrix} \beta_1 & 0 & \dots & 0 \\ 0 & \beta_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \beta_K \end{bmatrix}_{(p+1)K \times K}$$

It is good to mention that estimation of these parameters will be explained in section 2.2. Equivalently, (2) can be expressed in the matrix form as follows:

$$E\left(Y|X^{*}\right) = X^{*}P\beta = X^{*}B\pi \tag{3}$$

where $X^{*'} = (\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_n^*)_{(p+1)K\times n}$ is not a full column rank matrix. which is based on the idea of H. Rezazadeh et al. (2020).

2.1 Identifiability of Finite Mixture Model

The term identifiability is used for the family of distributions in which distinct distributions are determined with different parameters. Thus, it is a major concern of finite mixture models. We quote DasGupta (2008) where he mentioned "Mixture models are riddled with difficulties such as nonidentifiability".

Teicher (1961, 1963) has begun research about the parameter identifiability for mixture models that is continued by Yakowitz and Spragins (1968), Hall and Zhou (2003), Hall et al. (2005), Elmore et al. (2005), Allman et al.(2009) and others. In the location mixture model, study the behavior of mixing distributions is a natural approach to address parameter estimation rates (Carroll and Hall, 1988,Zhang, 1990, Fan, 1991). For a class of over-fitted finite mixtures, Chen (1995) proposed a notion of strong identifiability and established the convergence of the mixing distribution. However, his work limited to models with a single scalar parameter and Nguyen (2013) removed this limit for a number of finite and infinite mixture models with multi-dimensional parameters by establishing rates of convergence of mixing distributions. Over-fitted mixtures in a Bayesian estimation situation studied by Rousseau and Mengersen (2011). In this paper we suppose identifiability for finite mixture models and continue the approach for these type of models.

2.2 Estimation of Paramters by Iterative Weighted Least Square Approach

Given a set of independent observations $y_1, y_2, \ldots, y_n$, corresponding to values $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ of the predictor x, the complete parameter set of the finite mixture regression model is $\boldsymbol{\theta}' = (\pi_1, \ldots, \pi_K, \beta_1, \ldots, \beta_K)$. The normal equations corresponding to the model (3) can be derived by least square functions:

$$if\ \boldsymbol{\pi}\ is\ fixed: \qquad Q\left(\boldsymbol{\theta}\right) = \left(Y - X^{*}P\beta\right)'\left(Y - X^{*}P\beta\right) \tag{4}$$

$$if\ \beta\ is\ fixed: \qquad Q\left(\boldsymbol{\theta}\right) = \left(Y - X^{*}B\pi\right)'\left(Y - X^{*}B\pi\right) \tag{5}$$

We can estimate vector of parameters $\boldsymbol{\theta}$ by solving equations (4) and (5) iteratively. Since the variance of the errors is not fixed in this model, we should use $IWLS$ approach to obtain $\hat{\boldsymbol{\theta}}$. Let us define $\epsilon_i = \sum_{j=1}^{K} \pi_j \epsilon_{ij}$. If we apply the three random error assumptions explained in section 2, by defining $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)$, we can conclude that :

$$E(\boldsymbol{\epsilon}) = \mathbf{0}$$

and

$$Var(\boldsymbol{\epsilon}) = V$$

In this paper, we assume V is positive definite matrices. If we consider a transformation based on V^{-1}, the model (3) has become to:

$$E(V^{-\frac{1}{2}}Y|X^*) = V^{-\frac{1}{2}}X^*P\beta = V^{-\frac{1}{2}}X^*B\pi$$

By defining $Y_w = V^{-\frac{1}{2}}Y$, $X_w^* = V^{-\frac{1}{2}}X^*$, and $\boldsymbol{\epsilon}_w = V^{-\frac{1}{2}}\boldsymbol{\epsilon}$, we can rewrite this equation as follows:

$$E(Y_w|X_w^*) = X_w^*P\beta = X_w^*B\pi$$

The Weighted Least Square Principle minimizes the following two objective functions iteratively:

$$if\ \pi\ is\ fixed: \qquad Q_w(\boldsymbol{\theta}) = (Y_w - X_w^*P\beta)'(Y_w - X_w^*P\beta) \tag{6}$$

$$if\ \beta\ is\ fixed: \qquad Q_w(\boldsymbol{\theta}) = (Y_w - X_w^*B\pi)'(Y_w - X_w^*B\pi) \tag{7}$$

By taking derivative with respect to vectors β and π we have normal equations as below:

$$\hat{\beta}_{IWLS} = G_1P'X_w^{*'}Y_w \tag{8}$$

and

$$\hat{\pi} = G_2B'X_w^{*'}Y_w \tag{9}$$

Where G_1 and G_2 are generalized inverse of $P'X_w^{*'}X_w^*P$ and $B'X_w^{*'}X_w^*B$ respectively. In this paper, in the IWLS algorithm, instead of $\hat{\beta}_{IWLS}$ in equation (6), we can use the β penalty estimate by minimizing the following objective function.

$$\arg\min_{\beta}\left\{\|Y - X_w^*P\beta\|^2 + \sum_{l=1}^{K}\pi_l\left\{\sum_{j=1}^{p+1}p_{nl}(\beta_{lj})\right\}\right\} \tag{10}$$

Where $p_{nl}(\beta_{lj})$ s are non-negative and non-decreasing functions in $|\beta_{lj}|$. Here we used $p_{nl}(\beta_{lj}) = \gamma_{nl}\sum_{j=1}^{p+1}\beta_{lj}^2$ penalty function.

3 Ridge Regression Estimators

The multicollinearity phenomenon is one of the most serious cases Which we may encounter in a multiple regression model. The existence of multicollinearity may lead to the variance estimators for the regression coefficients be sensitive to small changes in the data and may often be enlarged. Because $rank(X'X) \leq rank(X)$, the $X'X$ matrix will be singular if there is an exact linear dependence in column vectors of X. These results can be extended to the finite mixture regression models in a straightforward way. Because $rank(X_w^{*}{'}X_w^*) \leq rank(X_w^*)$, the collinearity in X_w^* may cause problems in the $IWLS$ estimation as defined by equation (8). A popular numerical method to combat near multicollinearity is that of ridge regression estimator (RRE) due to Hoerl and Kennard (1970). Based on arguments similar to those that led to the ridge estimators in linear regression, we proposed the RRE in the finite mixture regression models as following:

$$\begin{aligned}\hat{\beta}_{RIWLS} &= \arg\min_{\beta}\,((Y_w - X_w^* P\beta)'(Y_w - X_w^* P\beta) + \sum_{l=1}^{K} \pi_l \gamma_{nl} \sum_{j=1}^{p+1} \beta_{lj}^2) \\ &= (P^{'} X_w^{*'} X_w^* P + \Gamma P)^{-1} P^{'} X_w^{*'} Y_w = G_3 P' X_w^{*'} Y_w \qquad (11)\end{aligned}$$

where $\gamma_{nl} > 0$ is the ridge parameter, $G_3 = (P^{'} X_w^{*'} X_w^* P + \Gamma P)^{-1}$ and

$$\Gamma = \begin{bmatrix} \gamma_1 I_{(p+1)} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \gamma_K I_{(p+1)} \end{bmatrix}_{(p+1)K \times (p+1)K}$$

3.1 Consequence of Estimations

In this subsection, we calculate the expected value, variance, estimate of expected value of response and mean of squared error (MSE).

We obtain expected value of $\hat{\beta}$ as follows:

$$E(\hat{\beta}_{RIWLS}) = G_3 P' X_w^{*'} X_w^* P\beta = H_1 \beta \qquad (12)$$

where $H_1 = G_3 P' X_w^{*'} X_w^* P$. Hence $\hat{\beta}_{RIWLS}$ is an biased estimator of β.
Also variance matrix is:

$$\begin{aligned}Var(\hat{\beta}_{RIWLS}) &= Var(G_3 P' X_w^{*'} Y_w) \\ &= G_3 P' X_w^{*'} \Sigma X_w^* P G_3' \qquad (13)\end{aligned}$$

where Σ is variance-covariance matrix of Y_w.
By using the estimation of $\hat{\boldsymbol{\theta}}$, vector of the estimated expected values $\widehat{E(Y_w)}$ corre-

sponding to the vector of observations $Y_w$is ,

$$\begin{aligned}\widehat{E(Y_w)} \equiv \hat{Y}_w &= X_w^* \hat{P}\hat{\beta} \\ &= X_w^* \hat{P} G_3 \hat{P}' X_w^{*'} Y_w \end{aligned} \tag{14}$$

The mean of square error of the Ridge estimator is as follow,

$$\begin{aligned} MSE(\hat{\beta}_{RIWLS}) &= E((\hat{\beta}_{RIWLS} - \hat{\beta})'\Phi^{-1}(\hat{\beta}_{RIWLS} - \hat{\beta})) \\ &= tr(H'\Phi^{-1}H) + \beta'(P'X_w^{*'}H'\Phi^{-1}HX_w^*P \\ &-P'X_w^{*'}H'\Phi^{-1} - \Phi^{-1}HX_w^*P + \Phi^{-1})\beta \end{aligned} \tag{15}$$

where Φ is $var(\hat{\beta})$ and $H = \left(P'X_w^{*'}X_w^*P + \Gamma P\right)^{-1} P'X_w^{*'}$.

3.2 Asymptotic Consistency of Ridge Estimates

For our purpose, we must have the following regular conditions.

$A_1)$ $\quad C_n = \frac{1}{n}\sum\limits_{i=1}^{n} x_i x_i' \to C$, n $\to \infty$

where C is a finite and positive definite matrix.

$A_2)$ $\quad \frac{1}{n}\max\limits_{1\leq i\leq n} x_i' x_i \to 0$, n $\to \infty$

We consider the asymptotic behavior of Ridge's objective function by first defining a random variable $Z_n(\phi)$:

$$Z_n(\phi) = \frac{1}{n}(\|Y_w - X_w^* P\phi\| + \sum_{l=1}^{K} \pi_l \gamma_{nl} \sum_{j=1}^{p+1} \phi_{lj}^2) \tag{16}$$

Which, it is minimized at $\phi = \hat{\beta}_n$. The following theorem tells us that $\phi = \hat{\beta}_n$ is a consistent estimator of β provided that $\gamma_{nl} = o(n)$.

Theorem : If C is nonsingular and $\frac{\gamma_{nl}}{n} \to \gamma_{0l} \geq 0$ then argmin$(Z_n(\phi)) \xrightarrow{p}$ argmin$(Z(\phi))$ where

$$Z(\phi) = \sum_{i=1}^{k} \pi_l^2 \sigma_l^2 + (P(\phi - \beta))'C(P(\phi-\beta)) + \sum_{l=1}^{K} \pi_l \gamma_{0l} \sum_{j=1}^{p+1} \phi_{lj}^2) \tag{17}$$

Thus, if $\gamma_{nl} = o(n)$ then $\gamma_{0l} = 0$ and argmin$(Z(\phi)) = \beta$ so that $\hat{\beta}_n$ is consistent.

Proof : We will show that $Z_n(\phi)$ defined above, converges in probability to $Z(\phi)$. The result will follow by applying established previous results from Pollard. To show convergence of Z_n:

$$\begin{aligned}
Z_n(\phi) &= \frac{1}{n}(\mathbb{Y}_w - X_w^* P\boldsymbol{\phi})'(\mathbb{Y}_w - X_w^* P\boldsymbol{\phi}) + \frac{1}{n}\sum_{l=1}^{K}\pi_l\gamma_{nl}\sum_{j=1}^{p+1}\phi_{lj}^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[\sum_{l=1}^{K}\pi_l(x_i'\beta_l + \varepsilon_{il}) - \sum_{l=1}^{K}\pi_l x_i'\phi_l\right]^2 + \frac{1}{n}\sum_{l=1}^{K}\pi_l\gamma_{nl}\sum_{j=1}^{p+1}\phi_{lj}^2 \\
&= \frac{1}{n}\left\{\sum_{i=1}^{n}\left(\sum_{l=1}^{K}\pi_l\varepsilon_{il}\right)^2 - 2\sum_{i=1}^{n}\left(\sum_{l=1}^{K}\pi_l\varepsilon_{il}\right)\left(\sum_{l=1}^{K}\pi_l x_i'(\phi_l - \beta_l)\right) + \sum_{i=1}^{n}\left(\sum_{l=1}^{K}\pi_l {x'}_i(\phi_l - \beta_l)\right)^2\right\} \\
&+ \frac{1}{n}\sum_{l=1}^{K}\pi_l\gamma_{nl}\sum_{j=1}^{p+1}\phi_{lj}^2 \qquad (18)
\end{aligned}$$

Assuming that $\varepsilon_i = \sum_{l=1}^{K}\pi_l\varepsilon_{il}$ will have:

$$E\left(\varepsilon_i\right) = E\left(\sum_{l=1}^{K}\pi_l\varepsilon_{il}\right) = 0 \qquad (19)$$

$$\operatorname{var}\left(\varepsilon_i\right) = \operatorname{var}\left(\sum_{l=1}^{K}\pi_l\varepsilon_{il}\right) = \sum_{l=1}^{K}\pi_l^2\sigma_l^2 \qquad (20)$$

Now, we let n → ∞ and note the following facts:

$$\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 \xrightarrow{p} \sum_{l=1}^{K}\pi_l^2\sigma_l^2 \text{ (by the law of large numbers)} \qquad (21)$$

Similarly since $E\left(\varepsilon_i x_i'\right) = 0$, it can be concluded

$$\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{l=1}^{K}\pi_l\varepsilon_{il}\right)\left(\sum_{l=1}^{K}\pi_l x_i'(\phi_l - \beta_l)\right) = \frac{1}{n}\sum_{i=1}^{n}\left(\varepsilon_i\right)\left(\sum_{l=1}^{K}\pi_l x_i'(\phi_l - \beta_l)\right) \xrightarrow{p} 0 \qquad (22)$$

Under the regularity conditions (A_1) have:

$$\sum_{i=1}^{n}\left(\sum_{l=1}^{K}\pi_l x_i'(\phi_l - \beta_l)\right)^2 = \left(P\left(\phi - \beta\right)\right)'\sum_{i=1}^{n}x_i x_i'\left(P\left(\phi - \beta\right)\right) \to \left(P\left(\phi - \beta\right)\right)' C\left(P\left(\phi - \beta\right)\right) \qquad (23)$$

On the other hand, according to the assumption $\frac{\gamma_{nl}}{n} \to \gamma_{0l} = 0$, therefore

$$\frac{1}{n}\sum_{l=1}^{K}\pi_l\gamma_{nl}\sum_{j=1}^{p+1}\phi_{lj}^2 \to \sum_{l=1}^{K}\pi_l\gamma_{0l}\sum_{j=1}^{p+1}\phi_{lj}^2 \qquad (24)$$

so that

$$Z_n(\phi) \to \sum_{i=1}^{K}\pi_l^2\sigma_l^2 + (P(\phi - \beta))'(P(\phi - \beta)) + \sum_{l=1}^{K}\pi_l\gamma_{0l}\sum_{j=1}^{p+1}\phi_{lj}^2) = Z(\phi) \qquad (25)$$

The pointwise convergence of $Z_n(\phi)$ to $Z(\phi)$ allows us to conclude that

$$\sup_{\phi \in E} |Z_n(\phi) - Z(\phi)| \xrightarrow{p} 0$$

, for any compact set E by Pollard. It follows that $\arg\min\left(Z_n(\phi)\right) \xrightarrow{p} \arg\min\left(Z(\phi)\right)$, This tells us that the ridge estimates are consistent.

4 Simulation Studies

We perform Monte Carlo simulations to test the performance of the finite sample of our proposed estimators. Computer code for generating the simulated data and manipulating it is written in R language. As considered in McDonald and Galarneau (1975), the explanatory variables are generated by

$$X_{ij} = \sqrt{1-\rho^2} z_{ij} + \rho z_{ip}, \; i = 1, 2, ..., n, \; j = 1, 2, ..., p$$

where z_{ij} are independent standard normal pseudo random numbers and ρ is the correlation between any two predictors. In this study, to investigate the effects of different degrees of collinearity on the estimators, we consider $\rho = 0.50$, 0.75 and 0.99 and $p = 2$ and 5.
According to p+1 and K, the β vector size dimension will be as Table 1. To simulate the response variable, we generate a random permutation of 5 normal random variables with the following distributions: $normal(-7, 1), normal(-3, 2)$, $normal(0, 1), normal(2, 1)$ and $normal(5, 3)$. See the density of y in Figure 1.

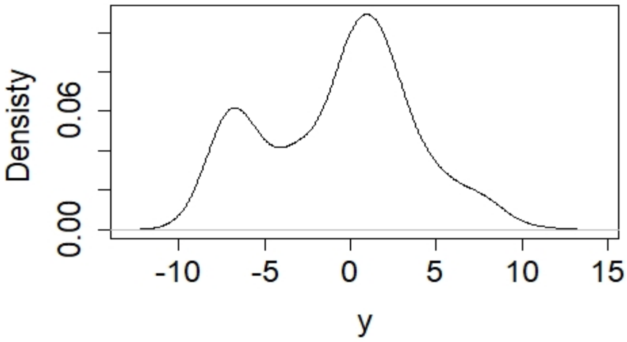

Figure 1: The density of y.

The behavior of $GMSE$ for first 500 iteration when $K = 4$ is shown in Figure 2. Table 2 shows the estimation of parameters for $K = 2, 3, 4$, $n = 500$, $p = 8$ and

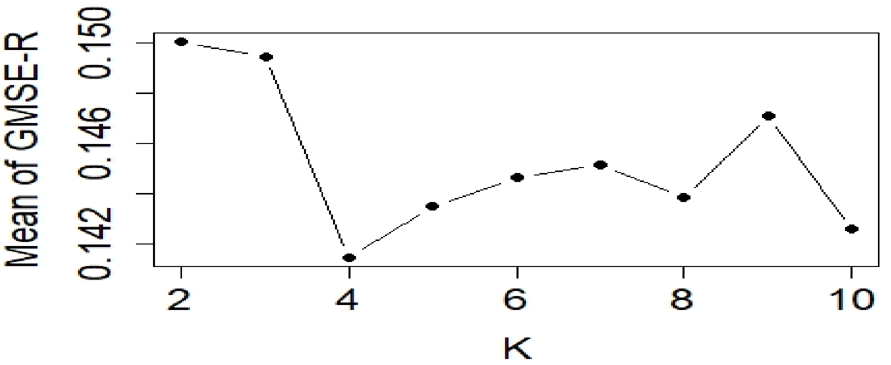

Figure 2: Mean value of the $GMSE$ in 10000 iteration for $K = 2, \ldots, 10$.

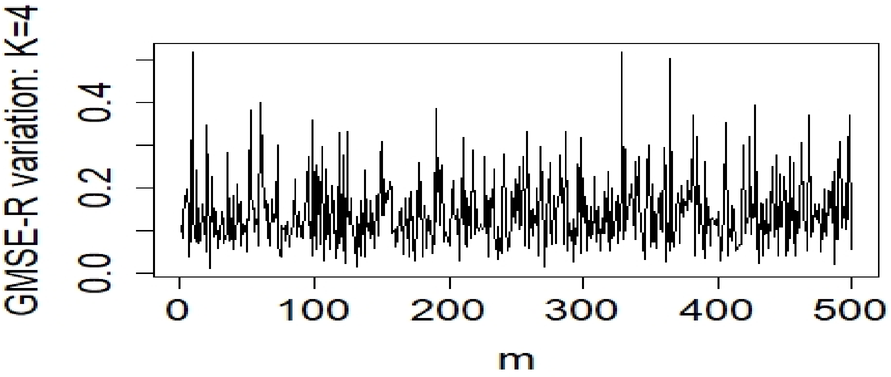

Figure 3: $GMSE$ for 500 iteration in the case $K = 4$

$\rho = 0.50$.

In step 0 in the $IWLS$ algorithm, suppose that $(\pi_1^{(0)}, \ldots, \pi_K^{(0)})$ are generated from the $uniform$ (0,1) distribution and calculate $P^{(0)}$. Continue $IWLS$ algorithm by calculating $\hat{\boldsymbol{\beta}}^{(0)}$ from equation (11) and go to step 1. Now recalculate $\hat{\boldsymbol{\pi}}$ by the estimated $\hat{\boldsymbol{\beta}}^{(0)}$ and repeat this step until the following condition is met:

$$\parallel \hat{\beta}_{t+1} - \bar{\hat{\beta}}_{1:t} \parallel < 0.001$$

We estimate the parameters of k = 2,3,,10 using the $IWLS$ and $RIWLS$ methods. To evaluate our method, we performed a comparative study between our proposed method and $IWLS$ method. In this comparison, we considered $GMSE$.

Table 1: The parametric size dimension for $p = 2$ and 5 and $K = 2, \dots, 10$.

K		2	3	4	5	6	7	8	9	10
p	2	6	9	12	15	18	21	24	27	30
	5	12	18	24	30	36	42	48	54	60

Table 2: Estimate of parameters for simulated data in the case $K = 2, 3, 4$.

K	$\hat{\pi}\pi$	$\hat{\beta}\beta_{\mathbf{RIWLS}}$	**SSE**($\hat{\beta}\beta_{\mathbf{RIWLS}}$)	$\overline{\mathbf{GMSE}}$	**var**(**GMSE**)
2	$\begin{pmatrix} 0.302 \\ 0.698 \end{pmatrix}$	$\begin{pmatrix} 0.101 \\ -0.871 \\ 0.223 \\ -0.012 \\ 0.462 \\ 1.055 \\ 0.191 \\ -0.408 \\ -0.137 \\ 0.233 \\ -2.017 \\ 0.518 \\ -0.027 \\ 1.071 \\ 2.443 \\ 0.442 \\ -0.945 \\ -0.316 \end{pmatrix}$	6153.408	0.282	0.0248
3	$\begin{pmatrix} 0.538 \\ 0.021 \\ 0.440 \end{pmatrix}$	$\begin{pmatrix} I \end{pmatrix}$	6153.408	0.227	0.0227

To assess the performance of our estimations $\hat{\beta}$, we used the following algorithm to obtain generalized mean square error ($GMSE$).

step1. Determine $\hat{Y}_w$ from equation (14) and generate a vector of observation from the normal distribution with mean $\hat{Y}_w$ and variance I.

step2. Estimate $\hat{\beta}_m$ by using the obtained vector of observations in step1 as the response variable in equation (11).

step3. Define $GMSE$ as follows:

$$GMSE(\hat{\beta}_m) = (\hat{\beta}_m - \hat{\beta})' \Phi^{-1} (\hat{\beta}_m - \hat{\beta})$$

We continue the algorithm for 10000 iterations and get a $GMSE$ on each iteration. As expected, the lower the mean of $GMSE$ is, the more accurate our parameter estimation would be. To obtain the best value for K, we applied the $GMSE$ algorithm with 10000 iterations for the range of $K = 2, \ldots, 10$, $n = 500$, $p = 8$ and $\rho = 0.50$. The results are presented in Figure 2 where mean of $GMSE$ after removing outliers is illustrated versus K.
where

$I = c(0.215, -1.859, 0.477, -0.025, 0.987, 2.252, 0.407, -0.871, -0.291, 0.009, -0.074, 0.019, -0.001,$
$0.039, 0.090, 0.016, -0.035, -0.012, 0.176, -1.519, 0.390, -0.020, 0.807, 1.841, 0.333, -0.712, -0.238).$

To evaluate our method, we performed a comparative study between our proposed method and $IWLS$ method. In this comparison, we considered $GMSE$ issue. The comparison is performed for $K = 4$ and for different values of n, p and ρ. The results are presented in Table 3. As can be seen, our method outperforms the $IWLS$ method.

5 Discussion

In this paper, we have combined the idea of the Iterative Weighted least Square ($IWLS$) and the ridge shrinkage estimators, for estimation parameters of finite mixture model. We studied the asymptotic performance of the proposed estimator. A Monte Carlo simulation study conducted to compare the estimators using $GMSE$ for various configurations of parameter size (p), degree of multicollinearity (ρ) and different sample size (n). Simulation results show that the $RIWLS$ outperforms $IWLS$.

Bibliography

[1] Allman, E. S., C. Matias, and J. A. Rhodes. 2009. Identifiability of parameters in latent structure models with many observed variables. *The Annals of Statistics* 37 (6A):3099132. :10.1214/09-AOS689.

[2] Carroll, R. J., and P. Hall. 1988. Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association* 83 (404):11846. :10.2307/2290153.

[3] DasGupta, A. 2008. Mixture Models and Nonparametric Deconvolution. *In Asymptotic Theory of Statistics and Probability*, 57191. New York: Springer.

[4] De Veaux, R. D. 1989. Mixtures of linear regressions. *Computational Statistics & Data Analysis*, 8(3), 227-45. :10.1016/0167-9473(89)90043-1.

[5] Dias, J. G., and M. Wedel. 2004. An empirical comparison of EM, SEM and MCMC performance for problematic Gaussian mixture likelihoods. *Statistics and Computing*, 14(4), 323-32. :10.1023/B:STCO.0000039481.32211.5a.

[6] Elmore, R., P. Hall, and A. Neeman. 2005. An application of classical invariant theory to identifiability in nonparametric mixtures. *In Annales de l'institut Fourier* 55 (1):128. :10.5802/aif.2087.

[7] Eskandari, F., and E. Ormoz. 2016. Finite Mixture of Generalized Semiparametric Models: Variable Selection via Penalized Estimation. *Communications in Statistics-Simulation and Computation*, 45(10), 3744-59. :10.1080/03610918.2014.953687.

[8] Fan, J. 1991. On the optimal rates of convergence for nonparametric deconvolution problems. *The Annals of Statistics*, 1257-72. :10.1214/aos/1176348248.

[9] Faria, S., and G. Soromenho. 2010. Fitting mixtures of linear regressions. *Journal of Statistical Computation and Simulation*, 80 (2):20125. :10.1080/00949650802590261.

[10] Faria, S., and G. Soromenho. 2012. Comparison of EM and SEM algorithms in Poisson regression models: A simulation study. *Communications in Statistics-Simulation and Computation*, 41(4): 497509. :10.1080/03610918.2011.594534.

[11] Frank, I.E., and J.H Friedman. 1993. An Statistical View of Some Chemometrics Regression Tools. *Technometrics*, 35, 109-135. : 10.1080/00401706.1993.10485033.

[12] Hall, P., A. Neeman, R. Pakyari, and R. Elmore. 2005. Nonparametric inference in multivariate mixtures. *Biometrika* 92(3), 667-78. :10.1093/biomet/92.3.667.

[13] Hall, P., and X. H. Zhou. 2003. Nonparametric estimation of component distributions in a multivariate mixture. *The Annals of Statistics* 31 (1):20124. :10.1214/aos/1046294462.

[14] Hawkins, D. S., D. M. Allen, and A. J. Stromberg. 2001. Determining the number of components in mixtures of linear models. *Computational Statistics & Data Analysis* 38(1), 15-48. :10.1016/S0167-9473(01)00017-2.

[15] Hoerl, A. E, and Kennard R. W.1970. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* 12, 55-67. : 10.1080/00401706.1970.10488634.

[16] Jacobs, R. A., Jordan, M. I., Nowlan, S. J. and Hinton, G. E. 1991. Adaptive mixture of local experts. *Neural Computation* 3, 79-87. : 10.1162/neco.1991.3.1.79.

[17] Jiang, W. and Tanner, M. A. 1999. Hierarchical mixtures-of-experts for exponential family regression models: Approximation and maximum likelihood estimation. *The Annals of Statistics* 27, 987-1011. : 10.1214/aos/1018031265.

[18] Jones, P. N., and G. J. McLachlan. 1992. Fitting finite mixture models in a regression context. *Australian Journal of Statistics* 34 (2):23340. :10.1111/j.1467-842X.1992.tb01356.x.

[19] McDonald, G. C., and D.I. Galarneau. 1975. A monte carlo evaluation of ridge-type estimators. *Journal of the American Statistical Association* 70 (350):40716. : 10.1080/01621459.1975.10479882.

[20] McLachlan, G. J. and Peel, D. (2000), Finite Mixture Models, New York: Wiley.

[21] Nguyen, X. 2013. Convergence of latent mixing measures in finite and infinite mixture models. *The Annals of Statistics* 41 (1):370400. :10.1214/12-AOS1065.

[22] Pearson, K. 1894. Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London.* A 185:71110. :10.1098/rsta.1894.0003.

[23] Pollard, D. 1991. Asymptotics for Least Absolute Deviation Regression Estimators. *Econometric Theory* 18699. : 10.1017/S0266466600004394.

[24] Quandt, R. E., and J. B. Ramsey. 1978. Estimating mixtures of normal distributions and switching regressions. *Journal of the American statistical Association* 73 (364):7308. :10.2307/2286266.

[25] Rezazadeh, H., F. Eskandari, M. Bameni Moghadam and E. Ormoz. 2020. Variable selection in finite mixture of generalized estimating equations.*Communications in Statistics - Simulation and Computation.* :10.1080/03610918.2019.1711406.

[26] Rousseau, J., and K. Mengersen. 2011. Asymptotic behaviour of the posterior distribution in overfitted mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73 (5):689710. :10.1111/j.1467-9868.2011.00781.x.

[27] Schepers, J. 2015. Improved random-starting method for the EM algorithm for finite mixtures of regressions. *Behavior research methods* 47 (1):13446.

[28] Searle, S. R. 1997. Linear models. Hoboken, New Jersey: John Wiley & Sons.

[29] Teicher, H. 1961. Identifiability of mixtures. *The Annals of Mathematical Statistics* 32 (1):2448.

[30] Teicher, H. 1963. Identifiability of finite mixtures. *The Annals of Mathematical Statistics* 34 (4): 12659. :10.1214/aoms/1177703862.

[31] Tibshirani, R. 1996. Regression shrinkage and selection via the Lasso. *Journ al of the Royal Statistical Society, Series B* 58, 267-88. : 10.1111/j.2517-6161.1996.tb02080.x.

[32] Xu, L., N. Lin, B. Zhang, and N. Shi. 2012. A Finite mixture model for working correlation matrices in generalized estimating equations. *Statistica Sinica* 22 (2):75576. :10.5705/ss.2010.090.

[33] Yakowitz, S. J., and J. D. Spragins. 1968. On the identifiability of finite mixtures. *The Annals of Mathematical Statistics* 39 (1):20914. :10.1214/aoms/1177698520.

[34] Zhang, C. H. 1990. Fourier methods for estimating mixing densities and distributions. *The Annals of Statistics* 18 (2):80631. :10.1214/aos/1176347627.

[35] Zhu, H. T., and H. Zhang. 2004. Hypothesis testing in mixture regression models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 66 (1):316. :10.1046/j.1369-7412.

How to Cite: Sajad Nezamdoust[1], Farzad Eskandari[2], *Ridge Shrinkage Estimators in Finite Mixture of Generalized Estimating Equations*, Journal of Mathematics and Modeling in Finance (JMMF), Vol. 2, No. 2, Pages:91–104, (2022).

The Journal of Mathematics and Modeling in Finance (JMMF) is licensed under a Creative Commons Attribution NonCommercial 4.0 International License.

Table 2. Continued.

K	$\hat{\boldsymbol{\pi}}$	$\hat{\beta}_{\mathbf{RIWLS}}$	$\mathbf{SSE}(\hat{\beta}_{\mathbf{RIWLS}})$	$\overline{\mathbf{GMSE}}$	$\mathbf{var}(\mathbf{GMSE})$
4	$\begin{pmatrix} 0.118 \\ 0.171 \\ 0.130 \\ 0.581 \end{pmatrix}$	$\begin{pmatrix} 0.057 \\ -0.496 \\ 0.127 \\ -0.007 \\ 0.264 \\ 0.601 \\ 0.109 \\ -0.233 \\ -0.078 \\ 0.083 \\ -0.717 \\ 0.184 \\ -0.009 \\ 0.381 \\ 0.869 \\ 0.157 \\ -0.336 \\ -0.112 \\ 0.063 \\ -0.544 \\ 0.140 \\ -0.007 \\ 0.289 \\ 0.659 \\ 0.119 \\ -0.255 \\ -0.085 \\ 0.283 \\ -2.441 \\ 0.627 \\ -0.032 \\ 1.297 \\ 2.958 \\ 0.535 \\ -1.144 \\ -0.383 \end{pmatrix}$	6153.408	0.279	0.0240

Table 3: Comparison of the proposed method and $IWLS$ for K=4

ρ	method	p=2			p=5		
		n			n		
		100	**200**	**500**	**100**	**200**	**500**
0.50	**RIWLS**	**0.063**	**0.025**	**0.013**	**0.271**	**0.202**	**0.069**
	IWLS	**0.071**	**0.027**	**0.014**	**0.285**	**0.222**	**0.076**
0.75	**RIWLS**	**0.033**	**0.047**	**0.0117**	**0.216**	**0.184**	**0.092**
	IWLS	**0.037**	**0.048**	**0.0129**	**0.245**	**0.187**	**0.100**
0.99	**RIWLS**	**0.103**	**0.068**	**0.034**	**0.117**	**0.147**	**0.066**
	IWLS	**0.129**	**0.076**	**0.035**	**0.118**	**0.158**	**0.068**