Ministry of Higher Education and

Scientific Research

University of Al-Qadisiyah

College of Administration and Economics

Sparsity via Bayesian Elastic Net and Elastic Net Regression Models with an Application

A thesis submitted to the council of the college of Administration and Economics for the degree of master of science in Statistics

By

Ali Ahmed Hussein

Supervised by

Dr. Bahr kadhim mohammed

Statistics Department

College of Administration and Economics

University of Al-Qadisiyah

2022A.D 1443A.H

بسم الله الرحمن الرحيم

﴿ وَلَقَدْ آتَيْنَا دَاوُودَ وَسُلَيْمَانَ عِلْمًا ۖ وَقَالَا الْحَمْدُ لِلَّهِ الَّذِي فَضَّلَنَا عَلَىٰ كَثِيرٍ مِّنْ عِبَادِهِ الْمُؤْمِنِينَ﴾

صدق الله العظيم

[النمل : 15]

Abstract

Sparsity procedure (variable selection) helps to determine all of the predictor variables that related to the response variable, which makes the model more accurate and more interpretable. We proposed new Bayesian elastic net by employing the scale mixture of normal distribution mixing with Rayleigh distribution as Laplace prior distribution for the regression coefficients. Based on some mathematical transformations for this scale mixture, we proposed a new scale mixture of normal mixing with truncated gamma distribution. Also, we derived new Bayesian hierarchical priors model based on the proposed scale mixture, and hence we a new Gibbs sample algorithm have been developed. The simulation and real data analysis demonstrated that the proposed Bayesian elastic net gives sparse solution for the regression parameters. So, adaptation of the new Bayesian elastic net allows incorporate the proposed scale mixture in a meaningful way with explanation of the regression model. Overall the proposed Bayesian scale mixture provides a new method which improves the sparse solution.

ACKNOWLEDMENTS

After praise be to God and praise be to him , his Majesty , our gratitude and thanks , for your guidance in choosing this topic, I extend my sincere thanks and appreciation to my esteemed supervisor , Dr. Bahr Kazem, with sincerity and appreciation for accepting to supervise my thesis .

Many thanks to the distinguished professors, members of the discussion committee, who kindly accepted the discussion of this thesis, bearing the trouble of reading the thesis and giving sound notes to evaluate what happened in it .

I would also like to extend my sincere thanks to all academic staff in the department of Statistics, University of Al-Qadisiyah .

I also thank the health institution (Athena Medical Laboratory) in Babylon Governorate for their assistance in providing me with health data.

My dear father

My dear mother

My beloved wife

Dear friends and loved ones

Ali Ahmed Hussain

DEDICATION

Sincerity

To the symbol of giving ….. My dear father, may God give him long life .

To the spirit of life and the source of tenderness ….. My dear mother, may God give her long life .

To my companion, my life partner, and her future ….. My dear wife .

To the apple of my eyes, the fruit of my heart, the joy of my existence ….. my dear son .

To my brothers, sisters, friends and everyone who helped .

Table of Contents

List of Tables

List of Figures

Chapter One

Introduction
and
Literature Review

1.1 Introduction

Regression analysis attempts to estimate the population average of the response variable by using the information that the predictor variables are provided. The parameter estimates of the regression model are reliable if they offers a balance between the variance and bias, in addition to the model explainability. It is well known that the **OLS** estimates are biased and inconsistent (inflated variance) when the multicollinearity problem appears in the design matrix $\boldsymbol{X}$, or when the number of predictor variable **p** exceeds or near the number of observations **n**. Therefore, in these circumstances, the **OLS** estimates are usually not unique and unstable with high variances. The high variance in the **OLS** estimates motivated the authors to explore the regularization methods that were used to overcome the limitations of least squares estimates quality, James et al. (2013). The ridge regression method adds a penalty function to residuals sum of squares (**RSS**) to address the problem multicollinearity, where the penalty function contains the L2-norm. The ridge parameter estimates cannot be set to zero, Hoerl and Kennard (1970). The Model selection procedure in regression analysis aims to select the best fit estimated regression model through selecting the relevant predictor variables that affect the response variable and removing the irrelevant variables. (Tibshirani, 1996) produced Lasso method which is essentially regarded as a penalized method that provides variable selection procedure. Consequently, many authors developed other shrinkage methods to provide variable selection procedures such as relaxed lasso, fused lasso, adaptive lasso, elastic net, etc. Ridge as penalized method does not gives sparse solution and lasso has some drawbacks: (a) lasso does not have oracle properties, (b) lasso lacks the ability to deals with the correlated grouping of predictors, (c) select n variables if p>n.

The elastic net is a penalized function that simultaneously performs variable selection and shrinkage, also it has the ability to select groups of correlated predictor variables .The elastic net is the flexible regularization and variable selection method that combined two of the penalties functions. Moreover, the Elastic Net (**EN**) is another penalized method that proposed by Zou and Hastie (2005) to address the limitations of lasso method. EN method combined the ridge and lasso to the RSS term, **EN** method deals with many relevant predictors that have highly pairwise correlation and **EN** usually works better than lasso, (Osborne et al.2000a).

1.2 Thesis problem

Regression model analysis is a statistical method that investigates the relationship between predictor variables and the response variable for modeling the data, prediction, and variable selection procedure. On account of this, usually, use the ordinary least square (OLS) method that minimizes the residual sum of squared (RSS) and then produces the least squared estimator (LSE) which is known a BLUE. But , in the case of the predictor variables being highly correlated with each other , which is known as the multicollinearity or in case of the number of predictor variables greater than the sample size (p> n) , the OLS solutions are not unique , because of the non-singularity of $X'X$ matrix. To overcome the problem of multicollinearity , a penalized regression method has been proposed . That is called the Ridge method have not sparsity (set some variables equal to zero), property. Lasso is sparsity method but has some limitations and because of that many authors developed different penalized methods, like an elastic net to overcome some of these limitations performs of the variable selection (Lasso) and shrinkage the regression coefficients (Ridge). After that, many authors studied the penalized method from the Bayesian perspective through developing the Bayesian variable selection.

In this thesis variable selection with Bayesian elastic net regression context can be considered as a model selection problem. This motivates us to study the model selection problem with the Bayesian elastic net under new Scale mixture for the Laplace prior distribution.

1.3 Thesis Objectives

In this thesis there are two ideas and one comparative study which are as follows:

1. To propose a new Bayesian hierarchical model that considers the Laplace prior distribution as a Scale mixture of normal mixing with Rayleigh distribution.

2. To combine the Bayesian model selection problem with the penalized elastic net linear regression model under the prior distribution mentioned in idea One.

3. To perform a comparative study between the Bayesian penalized elastic net that proposed in idea one and the classical elastic net.

1.4 Literature Review

Obviously, the regression analysis methods are very widely popular tools that investigated the relationship between the response variable and the independent variable(s). This motivated many authors and researchers to develop various regression analysis tools that cope with the practical underlying situation. The ordinary least squared (OLS) method is very common tool to find the regression coefficient estimates. Moreover, violated the assumptions of (OLS) was the key idea behind searching for substitution methods for regression coefficient estimates. In addition for that the investigation about the more explanation model developed along with the model selection and variable selection procedures. The Ordinary Least Squares provided unbiased and smallest variance parameters estimates through minimizing of the Residual Sum of Squares (RSS),

$$\mathrm{RSS}(\beta) = \sum(y^{true} - \mathrm{f}(\mathrm{X};\ \beta))^2$$

In the regression analysis, the set of the independent variable that should be included in regression equation brings the attention of the researcher, because it is the first part of the regression analysis and then examine to see whether the regression equation was correct. So, the variable selection problem is related to the regression form specification. The residual mean squares (RMS) is a criterion for model selection,

$$RME = \frac{SSE}{n - p}$$

Where p is the number of independent variables and SSE is the sum of squares error, the smallest the SSE between two regression equation is preferred.

Efroymson (1960) introduced the stepwise method as a variable selection procedure combined the mechanism of both Forward Selection (FS) Procedure and Backward Elimination (BE) procedure. The calculation of the stepwise method depends on the inclusion and deletion of independent variables, it is essentially a modification method for (FS and BE) methods. The AIC and BIC are used to select the best- fitted model in the stepwise method. It is recommended to obtain the variance inflation factors (VIF) test or the eigenvalues of the correlation matrix of the independent variables as the first step to variable selection procedure.

Hoerl and Kennard (1970) introduced a theory about ridge regression with penalized function to estimate the parameters of multiple regression model by adding a small positive quantity (λ) to the inverse of (X^tX) matrix to address the problem of linearly dependent (correlation) of the independent variable. The ridge estimator is biased but with the smallest variance. Also, ridge methods can be applied in the case of ($n \geq p$) and regards as regularization methods. But ridge regression is not a variable selection method. Ridge uses the L2-norm as penalty function. The response variable in ridge regression is centered (Draper and Smith, (1998).

Mallows (1973) developed the Mallows $\mathrm{C_k}$ criterion to judge the performance of the regression function by using the following form,

$$\mathrm{C_p} = \frac{SSE_k}{\mathrm{s}^2} + (2\mathrm{p} - \mathrm{n}),$$

Where s^2 is the estimated variance.

Akaike (1973) introduced the Akaike information criterion (AIC) as model selection criterion that combined the most fit equation and the smaller number of independent variables, the AIC defines as follows,

$$AIC_p = n\ln\left(\frac{SSE_p}{n}\right) + 2p,$$

The model which has the smallest AIC value is the better model.

Hocking (1976) list the evaluating regression method that is called all possible equations which is gives 2^k equations (k is the number of independent variables), where we can use the (RME, C_k, R^2) to select the best model. The limitation of all possible equations is the larger number of equations when k getting larger.

Schwarz (1978) proposed a modification of the AIC is called Bayes Information criterion (BIC) which is defines as follows,

$$BIC_p = n\ln\left(\frac{SSE_p}{n}\right) + p(\ln n),$$

The model which has the smallest BIC value is the better model.

George and McCulloch (1993) proposed another method for utilizing an information criterion for model selection; this method is called stochastic search variable selection ($SSVS$). This method can be used in the well-known Bayesian algorithm ($MCMC$), so it depends on the probabilistic considerations in selecting the subsets of independent variables.

Tibshirani (1996) proposed the new variable selection method that is called Lasso. Lasso method can be regarded as a regularization method that adds the L1-norm penalty function to the RSS. Due to the L1-norm, lasso provides variable selection procedure by setting the parameter estimates to zero. Also, in this paper, there is a remarkable note about Bayes estimation for the linear regression model based on assuming that the parameter β follows the double exponential distribution as prior density.

Efron et al. (2004) introduced an algorithm to compute the lasso estimate this algorithm is called LARS. LARS used for sake of model selection; they proved that it is takes short time for computational implementation in lasso.

Zou and Hastie (2005) introduced the so-called elastic net, which is regarded as a regularization method that combined the ridge and lasso methods. It can be considered as a variable selection method that works simultaneously as variable selection and shrinkage method. Furthermore, the elastic net deals well with a grouping effect of correlated independent variables in contrast to lasso.

Zou (2006) introduced the adaptive lasso regularization method. The adaptive lasso is considered as two stages procedure. In this paper, the problem of biased estimator has been controlled through assigned different weights for each parameter in the penalty function.

Yuan and Lin (2006) introduced the so-called group lasso as new regularization method; the group lasso is a generalization for the lasso method. Group lasso method essential founded to deal with problem of selected grouped independent variables. Lasso selected individual independent variables but group lasso can select a set of small groups of independent variable .

Yaun and Lin. (2006) showed that lasso has not the ability to detect the effects of grouped variables. Also, they stated that the variable selection with Bayesian perception outperforms the variable selection with lasso and elastic net methods based on the efficiency criterion.

Zou et al. (2007) discussed the using of BIC criterion to choose the shrinkage parameter in lasso method.

Ghosh (2007) and Zou and Zhang (2009) introduced two adaptive elastic net regularization methods. These new regularization methods focused on the limitation of lasso in dealing with the presence of grouped independent variables and the inconsistent of estimators. The adaptive lasso overcomes the problem of an inconsistent estimator by imposing weights for the different parameters. Also, adaptive elastic net estimators have oracle properties (normality and consistent). We can say that this method is combining adaptive lasso and elastic net.

Park and Casella (2008) developed Gibbs sample algorithm based on new Bayesian hierarchal prior model. The scale mixture of normal mixing with exponential density have used as representation form for the double exponential prior distribution through the lasso linear regression. The results are very similar for the classic lasso results.

Hans (2009) introduced Bayesian estimation for lasso regression coefficients. New Gibbs sampling algorithms have been developed by imposing directly the Laplace prior on the lasso regression parameters and a gamma prior on the tuning parameter. The results emphases that the classical lasso results did not match the Bayesian results in terms of prediction.

Li and Lin (2010) introduced the parameter estimation of the elastic net model from the Bayesian perception. By using the Gibbs sampler algorithm based on considering that the prior density is a scale mixture of normal mixing with truncated Gamma. The linear regression model studied for variable selection and prediction accuracy, the proposed model outperforms in the variable selection procedure and is a comparable model in the terms of prediction accuracy.

Kyung et al. (2010) introduced the Bayesian estimation for the linear regression with proposed hierarchal models. The Gibbs sample algorithm has been developed for the lasso, elastic net, group lasso, and fused lasso methods. The results showed that the proposed hierarchal model outperforms the LARs algorithms from the Bayesian perception.

Hans (2011) introduced new Gibbs sampler algorithm to find the solution for the Bayesian estimates using the elastic net method. In this paper the values of the shrinkage parameters $(\lambda_1 \; and \; \lambda_2)$ are based on the 10-fold cross validation method. Also, the scale mixture of normal has used to make the computational of the Gibbs sampler algorithm easier. The proposed Gibbs sample algorithm considered as an alternative to SSVS method.

Mallick and Yi (2014) introduced new Bayesian lasso method that depends on new representation of the double exponential prior density as scale mixture of uniform mixing with special case of gamma distribution. Variable selection procedure has performed and parameter estimation explained based on the new lasso method.

Alhamzawi (2016) proposed the Tobit quantile Bayesian elastic net regression model. The variable selection procedure and coefficients estimation have developed through new Bayesian hierarchal prior model. The gamma priors have used in Gibbs sample algorithm. The results showed that from the simulation examples and real data analysis that the proposed model outperforms other methods.

Rahim and Haithem (2018) introduced new Bayesian elastic net regularization method for variable selection and parameter estimation in linear regression. New hierarchical form prior model have developed based on the location-scale mixture of normal mixing with gamma density. The simulation results and real data analysis results showed the outperforms of the proposed model.

Flaih et al. (2020) introduced new scale mixture of normal mixing Rayleigh density to represent the double exponential prior density. New hierarchal prior model has been developed and therefore new Gibbs sample algorithm have implement to calculate the mode of the posterior density of lasso regression model parameter. The proposed model is comparable in terms of variable selection and estimation accuracy.

Fadel et al. (2020) developed an extension for lasso Tobit and adaptive lasso Tobit regression models based on the proposed scale mixture in Flaih et al. (2020) .

Flaih et al. (2020). New Bayesian hierarchal model have developed and new Gibbs sampler algorithm have been implemented. Simulation and real data analysis have conducted to investigate the prediction accuracy.

Chapter two

Some

Basic concepts

2.1 Introduction

In typical linear regression model the standard way of representing the $\boldsymbol{n}$ observations (x_1, y_1), … ,(x_1, y_n) drawn randomly and independently from a specific population. Where y_i is the response variable for $i = 1, 2, \ldots, n$, as a function of the p covariates $x_{i1}, x_{i2}, \ldots, x_{ip}$ is:

$$y_i = \beta_0 + \sum_{i=1}^{n}\sum_{j=1}^{p}\beta_j x_{ij} + e_i \ , \qquad i = 1,2,3, \ldots, n \ldots (2.1)$$

In matrix form (2.1) can be rewritten as follows,

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{e}, \qquad \ldots (2.2)$$

Where

$$\boldsymbol{y}_{\boldsymbol{n}\times\boldsymbol{1}} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \boldsymbol{X}_{\boldsymbol{n}\times(\boldsymbol{p}+\boldsymbol{1})} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix},$$

$$\boldsymbol{\beta}_{(\boldsymbol{p}+\boldsymbol{1})\times\boldsymbol{1}} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \boldsymbol{e}_{\boldsymbol{n}\times\boldsymbol{1}} = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}.$$

and $\boldsymbol{\beta}_j$ are the unknown coefficients (parameters) which we want to be estimated and $\boldsymbol{e}$ is the random error term of the model. The least squares method usually used to find the parameter estimates and offer BLUE estimates under the following assumptions about the error term: (Chatterjee and Hadi 2006, AL Nasser 2014)

1. $E(e) = 0$

2. $V(e) = \sigma^2 I$

3. $e \sim N(0, \sigma^2 I)$

4. Cov (e_i, e_j) = 0, where $i \neq j$

By taking the expected value for the regression equation (2.1) and applying of the error term assumption, we have the following equation that represents the population average:

$$E(y_i|x_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} = x_i^T \hat{\beta}$$

Also, it is very well known that the estimated error is defines as follows

$$\hat{e} = y_i - x_i^T \hat{\beta} \text{ ... (2.3)}$$

Equation (2.3) represent the difference between the observed values of y_i and estimated values of $\hat{y} = x_i^T \hat{\beta}$.

The least squares parameter estimate can be obtained by the following minimization problem:

$$\sum_{i=1}^{n} g(y_i - x_i^T \hat{\beta}) = \sum_{i=1}^{n} (y_i - x_i^T \hat{\beta})^2 \text{ ... (2.4)}$$

Where $g(w) = w^2$ is representing the quadratic loss function, hence based on (2.4) the least squares estimator is defined as follows

$$\hat{\beta} = (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{y} \quad \text{... (2.5)}$$

The error term follows the normal distribution with mean zero and variance $\sigma^2 \boldsymbol{I_n}$. Under some assumptions about the error term in least squares method and taking the expected of equation (2.2), it easy to note that the model (2.2) can be rewritten as follows:

$$\boldsymbol{y} \sim N(\boldsymbol{X\beta}, \sigma^2 \boldsymbol{I_n}) \text{ ... (2.6)}$$

The model (2.6) describe the linearly relationship between the mean of y_i and predictor variables.

If the matrix $\boldsymbol{X}$ is a full rank this leads $X^{'}X$ to be also full rank, then the matrix $X^{'}X$ is nonsingular. If the predictor variables in $\boldsymbol{X}$ are linearly dependent or when $p > n$, then the matrix $X^{'}X$ will be singular (not invertible matrix) which leads to non-unique solution in (2.5) see James et al. (2013) and Hastie et al. (2015). The problem of linear dependent between the predictor variables is called the multicollinearity which the most well-known source of the high variances of the estimates.

2.2 Multicollinearity

Many data sets in different scientific areas have a functional form (model) with a large number of variables which leads to difficulty in the interpretation of the model. So, many predictor variables in the model that affects the response variable may weaken the model. It is well-known that the least squares method largely depends on the relationship between the sample size (n) and the number of predictor variables (p). So, if the matrix $X^{'}X$ is singular, then $(X^{'}X)^{-1}$ is not reached and the problem arise in using the least squares method because of the highly correlated columns in matrix X. This problem called multicollinearity which leads to high variances parameter estimates. So, to solve the problem of collinearity, Hoerl and Kennard (1970) proposed a penalized method called Ridge method. The variance inflation factor (VIF) has the ability to detect the multicollinearity, see (Seber and Lee 2003, Draper and Smith 1998) for more information about VIF.

2.3 Variable Selection Procedure

Variable selection procedure in regression analysis context can be viewed as one of the most goals in the analysis of the relationship between the variables that determined to be included in the model. If there is no clear cut theoretical determination of the variables included in the regression model, the variable selection procedure becomes very important. The model specification (model selection) and variable selection are linked together. The researcher always asks himself the following question: Which predictor variable(s) must be included in the regression model? Variable selection methods like lasso method can answer this question by removing the irrelevant predictor variable on the response variable via letting its coefficient equal to zero. Variable selection procedure yields parsimonious (more interpreted model with a small number of predictors) model. Based on the correlation matrix of the predictor variables, different ways proposed to the variable selection procedure are depends on the data analyzed if are not collinear or if collinear. The forward selection procedure (FS), the backward elimination procedure (BE), and stepwise method are the classical variable selection methods. (Chatterjee and Hadi 2006).

(Hesterberg et al.,2008) stated that the aims of model selection are as follows:

(1) Better prediction accuracy.

(2) Better interpretability of models – determining which predictor variables are relevant to the response variable, and

(3) Better model stability –addition data should not result in large variation in either using the subset of predictor variables, the associated parameters, and the prediction.

2.4 Ridge Regression model

Vastly growing and great importance for the regularization method noted recently. The ridge method which is a penalized method was proposed by Hoerl and Kennard (1970a, 1970b) to address the problem of multicollinearity that present in case of many correlated predictor variables where the matrix $X^{'}X$ is not invertible or nonsingular. The least squares estimator depends in its minimization function on the matrix $(X^{'}X)^{-1}$. The ill condition in $X^{'}X$ matrix can be address by ridge method which has the ability to address this problem by adding the penalty function to the residual sum of squares, that is,

$$\hat{\beta}_{ridge} = RSS(\beta) + \lambda||\beta||_2^2 \qquad ... (2.7)$$

Where λ is the shrinkage parameter and λ ≥ 0 , the second term in (2.7) is called the penalty function and $||\beta||_2^2 = \sum_{j=1}^{p}\beta_j^2$. When $\lambda = 0$, the function (2.7) becomes least squares estimator. Clearly, equation (2.7) applied the penalty function $\lambda\sum_{j=1}^{p}\beta_j^2$ for the coefficients $(\beta_{1,}\beta_{2,} \dots, \beta_p)$ but not for β_0.

Where β_0 measure of the average of the population response variable when $(x_{1,}x_2, \dots, x_p = 0)$. Standardization of the predictors leads to $(E(x) = 0,\ v(x) = 1)$. Also, centering the response variable value such that $(\frac{1}{n}\sum_{i=1}^{n} y_i = 0)$. Consequently, if the ridge parameter estimates are obtained, we can obtain the β_0 by the following equation,

$$\hat{\beta}_0 = \bar{y} - \sum_{j=1}^{p} \bar{x}_j\hat{\beta}_j,$$

Where $\bar{y}$ and $\bar{x}$ are the original mean values.

The ridge estimator can be defined as follows,

$$\hat{\beta}_{ridge} = (X^{'}X + \lambda \mathrm{I}_P)^{-1}X^Ty \qquad ... (2.8)$$

The ridge estimates cannot reach the zero and therefore it is not sparse method. The ridge estimators have the following properties:

1. $\mathrm{E}\left(\hat{\beta}\right) = (\mathrm{X}^{'}\mathrm{X} + \lambda\mathrm{I}_\mathrm{P})^{-1}\mathrm{X}^\mathrm{T}\mathrm{X}\beta$

So, we can say that $\hat{\beta}_{ridge}$ is biased estimator. The mean of ridge estimator becomes zero when λ goes to infinity.

2. $\mathrm{var}\left(\hat{\beta}\right) = \sigma^2(X^TX + \lambda I_P)^{-1}X^TX(X^TX + \lambda I_P)^{-1}$

Obviously the variance of ridge estimator becomes zero when λ goes to infinity. See Hoerl et al (1975) for more information…

2.5 Lasso Regression Model

Lasso is another penalized function which can deals with many predictor variables, unlike the ridge method, lasso can set some predictor variables parameters to zero, therefor lasso is a sparse method which can remove the irrelevant predictor variables by setting its estimates to zero, and include the relevant predictors variable in the estimated regression model. The lasso estimator defined as follows:

$$\hat{\beta}_{lasso} = RSS\ (\beta) + \lambda\|\beta\|_1 \quad \dots (2.9)$$

The second term in (2.9) is the penalty function, where $\lambda \geq 0$ is the shrinkage parameter and $\|\beta\|_1 = \lambda\sum_{j=1}^{P}\left|\beta_j\right|$ is $L_1 -$ norm. So, we can say that lasso method is a variable selection method. (Tibshirani,1996) . As in ridge method, also standardize the predictor variables and centered the response variable. The intercept term can be obtained after standardization and obtaining the lasso parameters estimates. The optimal values for the shrinkage parameter λ are in the interval $(0, \lambda_{max})$, where

$(\lambda_{max} = max_j\left|x_j^t y\right|)$ (Osborne et al. 2000b). For lasso computational purposes (Friedman et al. 2007 and Hastie et al. 2009) suggested the threshold value for the parameter estimates of lasso estimator.

2.6 Elastic Net Regression

The lasso has some limitations , such as , in the case of p > n lasso can select only at most n variables , lasso methods tend to select randomly only one predictor variable in the case of a group of predictor variables , and also for n > p case with high pairwise correlated predictor variables ridge method out performs lasso. Consequently, lasso is not an appropriate variable selection procedure in some cases. (Zou and Hastie, 2003) proposed new penalized method that combined ridge and lasso penalty functions namely called elastic net method to solve the drawbacks of lasso method.

Moreover, the lasso has some restrictions: (i) in the case of the number of predictor variables (P) is greater than sample size (n), the lasso select at most (n) variables out of (p), (ii) the lasso lacks the capacity to disclose the grouping information in case of correlated grouped (Celeux et al. 2012; Zou and Hastie 2005; Lee, 2016) and (iii) for usual n > p cases, if there are high pairwise correlated predictor variables observed that the ridge regression outperforms lasso in terms of prediction accuracy (Tibshirani, 1996, Zou & Hastie, 2005). The elastic net penalized method simultaneously performs shrinkage (ridge) and variable selection procedure (lasso) (Hans 2011, and Hans 2010). That is means, elastic net can selects groups of pairwise correlated predictor variables. The naïve elastic net estimator defined as follows:

$$\hat{\beta}_{\text{Naive elastic net}} = \text{argmin}\ [\text{RSS}\ (\beta)]$$

$$\text{subject to}\ \ (1-\kappa)\|\beta\|_1 + \kappa\ \|\beta\|_2^2 \leq \tau, \qquad \text{...(2.10)}$$

Where $\tau > 0$ is the shrinkage parameter and $K \in (0,1)$ which controlling the weighting of terms in the above condition. If K=1 the condition in (2.10) becomes ridge, and if K=0 the condition becomes lasso.

It is known that the condition in (2.10) has the ability of setting coefficients equal to zero, Zou and Hastie (2003). The problem in . (2.10) can be rewritten in term of penalized function as follows:

$$\hat{\beta}_{naive\ elastic\ net} = argmin\ [RSS\ (\beta)] + \lambda\{(1-\kappa)\|\beta\|_1 + \kappa\ \|\beta\|_2^2\}$$

$$= argmin\ [RSS\ (\beta)] + \lambda_1\|\beta\|_1 + \lambda_2\|\beta\|_2^2\ \ldots(2.11)$$

Where $\kappa = \frac{\lambda_1}{\lambda_1+\lambda_2}$ and $\lambda = \lambda_1 + \lambda_2$. So, the penalty function is defined by $\lambda_1\|\beta\|_1 + \lambda_2\|\beta\|_2^2$. If $\lambda_1 = 0$ ridge function is obtained and if $\lambda_2 = 0$ lasso function is obtained. Thus the elastic net estimator defined by

$$\hat{\beta}_{elastic\ net} = argmin(1+\lambda_2)\{[RSS\ (\beta)] + \lambda_1\|\beta\|_1 + \lambda_2\|\beta\|_2^2\} \ldots (2.12)$$

Chapter Three

Bayesian Elastic Net Regression Model

3.1 Introduction

The Elastic net penalized method is very commonly used in the regression model as a regularization method that combines the ridge and lasso penalty functions. Zou and Hastie (2005) introduced the elastic net method as sparse procedure that can deal with the effect of correlated variables in covariates , the elastic net estimator is defined as follows ,

$$\hat{\beta} = argmin \, \|y - X\beta\|^2{}_2 + \lambda_1\|\beta\|_1 + \lambda_2\|\beta\|^2 \ldots (3.1)$$

Where the elastic net penalized function is

$$h(\beta) = \lambda_1\|\beta\|_1 + \lambda_2\|\beta\|^2$$

here $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$ the penalties parameters .

Tibshirani (1996) suggested that Lasso estimates can be interpreted as posterior mode estimates when the regression parameters have independent and identical Laplace priors. Park and Casella (2008) introduced the Bayesian lasso method for the linear regression model by assuming that Laplace prior distribution of β can be represent as scale mixture of normal mixing with exponential distribution based on the Andrews and mallows (1974) work , where the prior of β is defined as follow :

$$\pi\,(\beta\,/\,\lambda\,,\sigma^2)=\prod_{j=1}^{p}\sum_{0}^{\infty}\frac{1}{\sqrt{2\pi\,\sigma^2 Z_j}}\exp\left[-\frac{\beta j^2}{2\sigma^2 Z_j}\right]\frac{\lambda^2}{2}\,e^{\frac{-\lambda^2 Z_j}{2}}dZ_j \ldots(3.2)$$

Mallick and Yi (2014) introduced new scale mixture for the Laplace distribution of uniforms mixing with gamma $(2\,,\lambda)$ and the results of Bayesian estimates are very promising. Flaih et al. (2020) introduced the prior of Laplace distribution as scale mixture of normal distribution mixing with Rayleigh distribution normal which is defined as the following representation:

$$\frac{\lambda}{2\sigma^2}\exp\left[\frac{-\lambda|\beta|}{\sigma^2}\right] = \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2 t_j}} exp[\frac{-\beta j^2}{2\sigma^2 t_j}]\frac{\lambda}{2}e^{-\frac{\lambda t_j}{2}}dt_j \text{ ... (3.3)}$$

In this thesis, following Flaih et al. (2020), I introduced a new hierarchy model of our proposed regression model.

3.2. Bayesian Elastic Net Hierarchal Model and Prior Distributions.

Flaih et al. (2020) introduced the Bayesian lasso regression model based on scale mixture representation (3.3). In this thesis I assumed the above formula (3.3) by considering the linear regression model:

$$E\ (\ y\ /\ X,\beta) = X\beta$$

Suppose that the scale mixture of Laplace distribution that mixing normal with Rayleigh distribution defined as follows,

If $x/y \sim N\ (\mu\ , y^2)$ with y~Ray (b), then x~Laplace (μ, b), that is:

$$\frac{1}{2b}e^{-\frac{|x-\mu|}{b}} = \int_0^\infty \frac{1}{\sqrt{2\pi y^2}}e^{-\frac{(x-\mu)^2}{2y^2}}\frac{y}{b}e^{-\frac{y^2}{2b}}dy \quad \text{... (3.4)}$$

by letting μ=0, X =β, and b = $\frac{\sigma^2}{\lambda_1}$, then (3.3) become as follows :

$$\frac{\lambda_1}{2\sigma^2}\mathrm{e}^{-\frac{\lambda_1|\beta|}{2\sigma^2}} = \int_0^\infty \frac{1}{\sqrt{2\pi y^2}}e^{-\frac{\beta_j^2}{2y^2}}\frac{\lambda y}{\sigma^2}e^{-\frac{\lambda_1 y^2}{2\sigma^2}}dy \quad \text{... (3.5)}$$

Zou and Hastie (2005) introduced the prior distribution of elastic net method $\pi(\beta)$ as:

$$\pi(\beta)\ \propto e^{-\lambda_1\|\beta\|^1 - \lambda_2\|\beta\|_2^2}\ , \qquad \text{... (3.6)}$$

Then by multiplying both sides of (3.5) with $e^{-\lambda_2\|\beta\|_2^2}$, we get the scale mixture that cope with the prior (3.6),

$$\frac{\lambda_1}{2\sigma^2}\mathrm{e}^{-\frac{\lambda 1|\beta j|}{2\sigma^2}-\frac{\lambda 2\beta_j^2}{2\sigma^2}} = \int_0^\infty \frac{1}{\sqrt{2\pi y^2}} e^{-\frac{\beta_j^2}{2\sigma^2}} e^{-\frac{\lambda_2\beta_j^2}{2\sigma^2}} \frac{\lambda_1}{\sigma^2} e^{-\frac{\lambda_1 y^2}{2\sigma^2}} \text{ dy}$$

$$= \int_0^\infty \frac{1}{\sqrt{2\pi y^2}} e^{-\frac{\beta_j^2}{2}\left(\frac{1}{y^2}+\frac{\lambda_2}{\sigma^2}\right)} \frac{\lambda_1 y}{\sigma^2} e^{-\frac{\lambda_1 y^2}{2\sigma^2}} \text{ dy}$$

Let $\frac{1}{\sqrt{y^2}} = \frac{\sqrt{\frac{1}{y^2}+\frac{\lambda_2}{\sigma^2}}}{\sqrt{1+\frac{\lambda_2 y^2}{\sigma^2}}}$, then

$$\int_0^\infty \sqrt{\frac{1}{y^2}+\frac{\lambda_2}{\sigma^2}} \ \mathrm{e}^{-\frac{\beta_j^2}{2}(\frac{1}{y^2}+\frac{\lambda_2}{\sigma^2})} \cdot \frac{1}{\sqrt{1+\frac{\lambda_2 y^2}{\sigma^2}}} \frac{\lambda_1 y}{\sigma^2} \ e^{-\frac{\lambda_1 y^2}{2\sigma^2}} \ dy$$

Let t = $1 + \frac{\lambda_2 y^2}{\sigma^2} \quad \rightarrow \frac{1}{t-1} = \frac{\sigma^2}{\lambda_2 y^2}$

and

$$\frac{1}{y^2} + \frac{\lambda_2}{\sigma^2} = \frac{\lambda_2}{\sigma^2}\left(1 + \frac{\sigma^2}{\lambda_2 y^2}\right) = \frac{\lambda_2}{\sigma^2}\left(\frac{t}{t-1}\right)$$

From t = $1 + \frac{\lambda_2 y^2}{\sigma^2}$ if y=0 , and $y = \infty$, we get $t \in (1, \infty)$

$$\frac{\lambda_1}{2\sigma^2}\mathrm{e}^{-\frac{1}{2\sigma^2}(\lambda_1|\beta_j| + \lambda_2\beta_j^2)} \propto \int_1^\infty \sqrt{\frac{\lambda_2}{\sigma^2}\left(\frac{t}{t-1}\right)}$$

$$e^{-\frac{\beta_j^2}{2}\left(\frac{\lambda_2}{\sigma^2}\left(\frac{t}{t-1}\right)\right)} t^{-\frac{1}{2}} \frac{\lambda_1 y}{\sigma^2} e^{-\frac{\lambda_1}{2\sigma^2}\frac{t\sigma^2}{\lambda_2}} \frac{\sigma^2}{\lambda_2 2y} dt$$

$$\propto \int_1^\infty \sqrt{\frac{t}{t-1}} \sqrt{\frac{\lambda_2}{\sigma^2}} \ e^{-\frac{\beta_j^2}{2}\left(\frac{\lambda_2}{\sigma^2}\left(\frac{t}{t-1}\right)\right)} t^{-\frac{1}{2}} e^{-\frac{t\lambda_1}{2\lambda_2}} t \ dt$$

$$\propto \int_1^\infty \sqrt{\frac{\lambda_2}{\sigma^2}\frac{t}{t-1}}\, e^{-\frac{\beta_j^2}{2}\left(\frac{\lambda_2}{\sigma^2}\frac{t}{t-1}\right)}\, t^{-\frac{1}{2}}\, e^{-\frac{\lambda_1 t}{2\lambda_2}}\, dt \qquad \text{... (3.7)}$$

From (3.7), we can deal with β_j/σ^2 as Scale mixture of normal distributions N $(0,\frac{\sigma^2(t-1)}{\lambda_2\, t})$ mixing truncated gamma with shape parameter (1/2) and Scale parameter $(\frac{2\lambda_2}{\lambda_1})$, see Almusaedi and Flaih (2021a, 2021b), Alsafi and Flaih (2021) for more information. By formula (3.7), we have the following Bayesian elastic net linear regression (ENLR) hierarchical model,

$$y = X\beta + e\,,$$

$$y|X\,,\beta\,,\sigma^2 \sim N\,(X\beta\,,\sigma^2\, In\,)$$

$$\beta|\lambda_2\,,\sigma^2\,,t \sim \prod_{j=1}^{p} N\,\left(0,\left(\frac{\lambda_2}{\sigma^2}\frac{t_j}{t_j-1}\right)^{-1}\right)$$

$$t|\lambda_1,\lambda_2 \sim \prod_{j=1}^{p} truncated\ gamma\left(\frac{1}{2}\,,\frac{2\lambda_2}{\lambda_1}\right);t\ \in(1,\infty)$$

$$\sigma^2 \sim \text{ Inverse Gamma} \qquad \text{.... (3.8)}$$

3.3 Full Conditional Posterior Distributions of ENLR

By using the hierarchical model (3.8) the computational process through the Gibbs sampling become more easily and the full joint distribution is well defined as follows:

$$\pi(\beta|y,X\,,\sigma^2\,,t) \propto \pi(y/X\,,\beta\,,\sigma^2)$$

$$\pi\,(\sigma^2)\prod_{j=1}^{p}\pi\,(\beta_j|\ t_j\,,\sigma^2\,)\,\pi(t_j) \ =$$

$$\left(\frac{1}{\sqrt{2\,\pi\,\sigma^2}}\right)^n \; e^{-\frac{1}{2\,\sigma^2}\,(\,y-X\beta)'(y-X\beta)} \;\; . \; \frac{\tau^{\alpha}}{\sqrt{\alpha}} \;\; (\sigma^2)^{-\alpha-1} \;\; e^{-\frac{\tau}{\sigma^2}} \quad \prod_{j=1}^{p} \sqrt{\frac{\lambda_{\,2}\,t}{\sigma^2\,(t-1)}}$$

$$e^{-\frac{\beta_j^2}{2}\left(\frac{\lambda_{\,2}}{\sigma^2}\cdot\frac{t}{(t_j-1)}\right)} \;\; t^{-\frac{1}{2}} \;\; e^{\frac{-\lambda_1}{2\lambda_{\,2}}t} \quad ...(3.9)$$

Remark that, y variable is centered and x is standardized. Now the full conditional posterior distributions are as follows:

1- The parts that includes β, $\pi(\beta)$ in the joint distribution (3.9) is

$$e^{-\frac{1}{2\,\sigma^2}\,(\,y-X\beta)'(y-X\beta)-\frac{1}{2\sigma^2}\lambda_2\beta^{'}\,A\,\beta} \qquad , \text{where } \; A=\left(\frac{t}{t-1}\right)$$

$$= \exp\left[-\frac{1}{2\,\sigma^2}\;\{\left(\beta'\;(X'X)\;\beta-2yx\beta+y^{'}y\right)+\lambda_{\,2}\;\beta'\;A\;\beta\;\}\right]$$

$$= \exp\left[-\frac{1}{2\sigma^2}\{\beta^{'}\left(X^{'}X+\lambda_2 A\right)\beta-2yX\beta+y^{'}y\}\right]$$

$$= \exp\left[-\frac{1}{2\,\sigma^2}\;\{\left(\beta^{'}\;C\beta-2yx\beta+y^{'}y\right)\}\right]$$

Where C= $X^{'}X+\lambda_2 A$

$$\exp\left\{-\frac{1}{2\sigma^2}\left(\beta^{'}C\beta-2yX\beta+y^{'}y\right)\right\} \quad ...\;(3.10)$$

Let $\left(\beta-C^{-1}X^{'}y\right)^{'}C\left(\beta-C^{-1}X^{'}y\right)=\beta C^{'}\beta-2yX\beta+y^{'}(XC^{-1}X)^{'}y$

then (3.10) Can rewrite as follows:

$$\exp\left[-\frac{1}{2\,\sigma^2}\;\{\left(\beta-\;C^{-1}X^{'}y\right)^{'}C\;\left(\beta-C^{-1}X^{'}y\right)+\;y^{'}(In-XC^{-1}X^{'})y\;\;\right] \quad ...\;(3.11)$$

The second part of (3.11) does not involve β, so we can reduce (3.11) as follows

$$\exp\left[-\frac{1}{2\,\sigma^2}\;\{\left(\beta-\;C^{-1}X^{'}y\right)^{'}C\;\left(\beta-C^{-1}X^{'}y\right)\right] \;...\;(3.12)$$

We can say that (3.12) is the multivariable normal distribution with mean $C^{-1} X^{'} y$ and variance $\sigma^2 C^{-1}$.

The second Conditional posterior distribution is for σ^2, $\pi(\sigma^2)$. The terms that involve σ^2 in the full joint distribution (3.9) are as follows

$$(\sigma^2)^{-\frac{n}{2}} \quad (\sigma^2)^{-\alpha-1} \quad (\sigma^2)^{-\frac{p}{2}} - \frac{1}{e^{2\sigma^2}}(y - X\acute{\beta})(y - X\beta) - \frac{\tau}{\sigma^2} - \frac{\beta^{'}\lambda_2 A\beta}{2\sigma^2}$$

$$= (\sigma^2)^{-\frac{n}{2}-\frac{p}{2}-\alpha-1} \; -\frac{1}{e^{2\sigma^2}}\{(y-X\beta)^{'}(y-X\beta)+\tau+\beta^{'}\lambda 2A\beta\} \quad \text{... (3.13)}$$

The formula (3.13) is the inverse gamma distribution with shape parameter

$(\frac{n}{2} + \frac{p}{2} + \alpha)$ and Scale parameter $\frac{(y-X\beta)^{'}(y-X\beta)}{2} + \frac{\beta^{'}\lambda 2A\beta}{2} + \tau$.

The third part in the conditional posterior distribution of (t_j). The parts of (3-9) involve (t_j) are

$$\sqrt{\frac{\lambda_2}{\sigma^2}\frac{tj}{t_j - 1}} \; e^{-\frac{\beta_j^2}{2}(\frac{\lambda_2}{\sigma^2}\frac{t_j}{t_j-1})tj^{-\frac{1}{2}}} \; e^{-\frac{\lambda 1}{2\lambda 2}t_j}$$

Then based on the (Chhikara and Folks 1988) works, the distribution of $(t-1)$ is the generalized inverse Gaussian distribution and defined as follows,

$$(t-1)\sim GIG(\lambda = \frac{1}{2}, a = \frac{\lambda 1}{4\lambda_2\sigma^2}, \chi = \frac{\lambda_2\beta_j^2}{\sigma^2}), \text{...(3.14)}$$

Then, $(t-1)^{-1}$ variable follows the full conditional inverse Gaussian distribution with $\mu = \frac{\sqrt{\lambda_1}}{(2\lambda_2|\beta_j|)}$ and $\lambda = \frac{\lambda_1}{4\lambda_2\sigma^2}$. See (Chhikara and Folks 1988) for more details.

3.4 Choosing the Shrinkage parameters λ_1 and λ_2.

Following Li and Lin (2010) and Park and Casella (2008), we can take the log for the following functions and the maximization problem is solving as follows:

$$\frac{\partial R}{\partial \lambda_1} = \frac{p}{\lambda_1} + \frac{p\lambda_1}{4\lambda_2} E\left[\left\{\Gamma_U\left(\frac{1}{2}, \frac{\lambda_1^2}{8\sigma^2\lambda_2}\right)\right\}^{-1} \varphi\left(\frac{\lambda_1^2}{8\sigma^2\lambda_2}\right)\frac{1}{\sigma^2}\middle|\lambda^{(k-1)}, y\right] - \frac{\lambda_1}{4\lambda_2}\sum_{j=1}^{p} E\left[\frac{\tau_j}{\sigma^2}\middle|\lambda^{(k-1)}, y\right]$$

$$\frac{\partial R}{\partial \lambda_2} = -\frac{P\lambda_1^2}{8\lambda_2^2} E\left[\left\{\Gamma_U\left(\frac{1}{2}, \frac{\lambda_1^2}{8\sigma^2\lambda_2}\right)\right\}^{-1} \varphi\left(\frac{\lambda_1^2}{8\sigma^2\lambda_2}\right)\frac{1}{\sigma^2}\middle|\lambda^{(k-1)}, y\right] - \frac{1}{2}\sum_{j=1}^{p} E\left[\frac{\tau_j}{\tau_J - 1}\ \frac{\beta_j^2}{\sigma^2}\middle|\lambda^{(k-1)}, y\right] + \frac{\lambda_1^2}{8\lambda_2^2}\sum_{j=1}^{p} E\left[\frac{\tau_j}{\sigma^2}\middle|\lambda^{(k-1)}, y\right],$$

…(3.15)

Where $\varphi(t) = t^{-\frac{1}{2}} e^{-t}$.

Chapter Four

Simulation and Real Data Analysis

4.1 Simulation Study

In this section ,a simulation study is conducted to show the behavior of our proposed method , Bayesian elastic net linear regression (BENLR) using R package (Lasso u) and compared, and the elastic net linear (ENLR) regression model (enr) by implementing the (rn) R package, by implementing the R package . Our comparison is based on the parameters estimates of the different models different elastic net . Also , we used the median mean absolute deviation (MMAD) criterion .

$$MMAD = median\left[mean\,|x^T\hat{\beta} - x^T\beta^{true}|\right] \quad \ldots (4.1)$$

MMAD and the standard deviation (SD) are used to measure the performance of prediction accuracy for different model . The Gibbs sample algorithm have been used with 10000 iterations to generate the stability of the posterior distribution of the interested parameter assuming the number of observation is n = 400 , the first 1000 iterations have burned in . We generated the observations of predictor variables from $X \sim N(0, \Sigma)$, where the matrix $\sum_{ij} = \rho^{|i-j|}$, with three distributions of (**i.i.d.**) error terms .

1- Simulation example one

In this example , we assumed that the true vector of the true parameter (very sparse), $\boldsymbol{\beta} = (\mathbf{1}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0})$ with error distributed according to standard normal $\boldsymbol{e_i \sim N}\,(\mathbf{0}, \mathbf{1})$**,** $\boldsymbol{e_i \sim N}\,(\mathbf{1}, \mathbf{1}), \boldsymbol{e_i \sim N}\,(\mathbf{2}, \mathbf{2}) + \boldsymbol{N}\,(\mathbf{2}, \mathbf{2}), \boldsymbol{e_i \sim LAP}\,(\mathbf{1}, \mathbf{0}), \boldsymbol{and\ e_i \sim \chi^2_{(4)}}$. I generated the observations of the predictors $\mathbf{X_1}, \ldots\ldots, \mathbf{X_9}$ from the multivariate normal $\mathbf{N_{n=9}}\,(\mathbf{0}, \sum)$, here $\sum$ is the variance-covariance matrix which is defined as $\sum_{\mathbf{ij}} = \mathbf{0.7}^{|\mathbf{i-j}|}$. The true relationship between the predictor variables and response variable base on the above true vector is $f(X) = \sum_{j=1}^{9} X_j\beta_j$, So the correct model is defined by $f(X) = X_1\beta_1$,

Table (4.1) values of MMAD and SD in example one

Sample Size	Comparison Methods		$e_i \sim N(0,1)$	$e_i \sim N(1,1)$	$e_i \sim N(2,2) + N(2,2)$	$e_i \sim Lap(1,0)$	$e_i \sim x^2_{(4)}$
Small Sample	n=15	BENLR	1.235(0.672)	1.130(0.845)	1.521 (0.832)	1.303 (0.672)	1.612(0.792)
		ENLR	1.543(0.830)	1.317(0.992)	1.834 (0.970)	1.452 (0.757)	1.704 (0.822)
	n=25	BENLR	1.373(0.238)	1.240(0.152)	1.234 (0.632)	1.546(0.499)	1.703(0.643)
		ENLR	1.645(0.536)	1.325(0.273)	1.564 (0.874)	1.769 (0.682)	1.892(0.782)
	n=35	BENLR	1.245(0.563)	1.547(0.482)	1.529 (0.353)	1.346 (0.482)	1.446(0.583)
		ENLR	1.482(0.834)	1.618(0.834)	1.865 (0.932)	1.634 (.0542)	1.782 (0.671)
Meddle Sample	n=45	BENLR	1.282(0.451)	1.417(0.683)	1.220 (0.493)	1.030 (.0534)	1.106 (0.585)
		ENLR	1.597(0.780)	1.632(0.745)	1.836 (0.698)	1.573 (.0840)	1.839 (0.732)
	n=55	BENLR	1.256(0.245)	1.361(0.391)	1.435 (0.634)	1.310 (0.427)	1.420 (0.370)
		ENLR	1.562(0.792)	1.620(0.407)	1.834 (0.803)	1.478 (0.896)	1.838 (0.550)
	n=65	BENLR	1.069(0.327)	1.452(0.075)	1.520 (0.311)	1.564 (0.183)	1.305 (0.183)
		ENLR	1.623(0.832)	1.971(0.621)	1.733 (0.504)	1.826 (0.202)	1.352(0.420)
Large Sample	n=100	BENLR	1.855(0.358)	1.746(0.352)	1.107 (0.432)	1.523(0.453)	1.352(0. 534)
		ENLR	(1.964)(0.563)	(1.832)(0.832)	(1.543)(0.678)	(1.854) (0.704)	(1.676)(0.828)
	n=200	BENLR	1.241(0.332)	1.230(0.282)	1.781 (0.405)	1.530 (0.204)	1.682 (0.387)
		ENLR	(1.537)(0.564)	(1.676)(0.564)	(1970)(0.653)	(1.675)(0.734)	(1.754)(0.673)
	n=300	BENLR	1.604(0.450)	1.638(0.564)	1.754 (0.563)	1.039 (0.356)	1.651(0.432)
		ENLR	(1.722)(0.792)	(1.927)(0.671)	(1.934)(0.892)	(1.527)(0.643)	1.643)(0.854)

Table (**4.1**) provided the values of the MMAD and SD quality measures of the estimated regression models for the proposed method (BENLR) and the (ENLR) based on three types of sample sizes , small samples **(n=15 , n=25 , n=35)** , middle samples **(n=45 , n=55 , n=65)** , and large samples **(n=100 , n=200 , n=300)** . Clearly the values of MMAD criterion are the smallest in the proposed methods compared with the other methods under all different type of error distributions. Also, the SD criterion shows the preference of the proposed method under different types of sample sizes and under different error distributions .Consequently, the proposed method is a promising regularization method.

2- Simulation example two

In this example, we supposed that the true vector of parameters (sparse) $\boldsymbol{\beta} = (\ \mathbf{1}, \mathbf{0}, \mathbf{0}, \mathbf{1}, \mathbf{0}, \mathbf{1}, \mathbf{0}, \mathbf{1}, \mathbf{0}\ )$ with error distributed according to standard normal $\boldsymbol{e_i} \sim \boldsymbol{N}\ (\ \mathbf{0}, \mathbf{1}\ )$ **,** $\boldsymbol{e_i} \sim \boldsymbol{N}\ (\ \mathbf{1}, \mathbf{1}\ ), \boldsymbol{e_i} \sim \boldsymbol{N}\ (\ \mathbf{2}, \mathbf{2}\ ) + \boldsymbol{N}\ (\ \mathbf{2}, \mathbf{2}\ ), \boldsymbol{e_i} \sim \boldsymbol{LAP}\ (\ \mathbf{1}, \mathbf{0}\ ), \boldsymbol{and}\ \boldsymbol{e_i} \sim \boldsymbol{\chi}^{\mathbf{2}}_{(\mathbf{4})}$

I generated the observations of the covariates $\mathbf{X_1}, \dots\dots, \mathbf{X_9}$ from the multivariate normal $\mathbf{N_{n=9}}\ (\ \mathbf{0}, \sum)$, here $\sum$ is the variance-covariance matrix which is defined as $\sum_{\mathbf{ij}} = \mathbf{0.7}^{|\mathbf{i-j}|}$. The true relationship between the predictor variables and response variable base on the above true vector is

$$f(X) = \sum_{j=1}^{9} X_j \beta_j,$$

So the correct model is defined by

$$f(X) = X_1\beta_1 + X_4\beta_4 + X_6\beta_6 + X_8\beta_8,$$

Table (4.2). values of MMAD and SD of example Two

Sample Size	Comparison Methods		$e_i \sim N(0,1)$	$e_i \sim N(1,1)$	$e_i \sim N(2,2) + N(2,2)$	$e_i \sim Lap(1,0)$	$e_i \sim x^2_{(4)}$
Small Sample	n=15	BENLR	1.364(0.453)	1.223 (0.573)	1.332 (0.353)	1.165 (0.758)	1.232 (0.563)
		ENLR	1.573 (0.748)	1.473 (0.736)	1.637 (0.572)	1.342 (0.394)	1.640 (0.662)
	n=25	BENLR	1.443 (0.283)	1.234 (0.263)	1.323 (0.157)	1.439(0.231)	1.006(0.346)
		ENLR	1.647 (0.834)	1.634 (0.463)	1.839 (0.453)	1.854 (0.537)	1.538(0.782)
	n=35	BENLR	1.362 (0.334)	1.433 (0.273)	1.245 (0.434)	1.245 (0.245)	1.234(0.456)
		ENLR	1.563 (0.673)	1.823 (0.439)	1.734 (0.664)	1.547 (0.465)	1.482(0.706)
Meddle Sample	n=45	BENLR	1.282 (0.436)	1.275 (0.055)	1.493 (0.565)	1.265 (0.346)	1.464(0.161)
		ENLR	1.453 (0.764)	1.453 (0.673)	1.745 (0.856)	1.733 (0.546)	1.845(0.456)
	n=55	BENLR	1.645 (0.453)	1.238 (0.459)	1.334 (0.264)	1.464 (0.354)	1.365(0.579)
		ENLR	1.934 (0.854)	1.852 (0.673)	1.652 (0.345)	1.655 (0.566)	1.934(0.935)
	n=65	BENLR	1.178 (0.327)	1.005 (0.075)	1.563 (0.352)	1.045 (0.164)	1.156(0.254)
		ENLR	1.465 (0.845)	1.454 (0.756)	1.745 (0.564)	1.456 (0.303)	1.458(0.846)
Large Sample	n=100	BENLR	1.045 (0.045)	1.273 (0.245)	1.354 (0.322)	1. 007(0.322)	1.256(0.233)
		ENLR	(1.212)(0.845)	(1.565)(0.372)	(1.783)(0.173)	(1.222)(0.433)	(1.475)(0.452)
	n=200	BENLR	1.435 (0.332)	1.543 (0.346)	1.697 (0.435)	1.157 (0.253)	1.235(0.633)
		ENLR	1.635 (0.732)	(1.812)(0.845)	(1.845)(0.674)	(1.676)(0.445)	(1.875)(0.343)
	n=300	BENLR	1.665 (0.553)	1.453 (0.334)	1.534 (0. 389)	(1.274)(0.341)	(1.452)(0.323)
		ENLR	(1.912)(0.675)	(1.756)(0.311)	(1.781)(0.564)	(1.771)(0.422)	(1.881)(0.706)

Table (**4.2**) displays MMAD and SD values as measurement for testing the quality of the estimated regression models based on the proposed methods (BENLR) and the (ENLR) under three types of sample sizes , small samples **(n=15 , n=25 , n=35)** , middle samples **(n=45 , n=55 , n=65)** , and large samples **(n=100 , n=200 , n=300)** . Obviously, the values of MMAD criterion are the smallest in the proposed methods compared with the other methods under all different type of error distributions .In addition , the SD criterion show the preference of the proposed methods under different type of sample sizes and error terms distributions .Hence , the proposed method is a promising regularization method.

3- Simulation example three

In this example, we assumed that the true vector of the true parameter (dense) $\boldsymbol{\beta} = (\mathbf{0.85}, \mathbf{0.85}, \mathbf{0.85}, \mathbf{0.85}, \mathbf{0.85}, \mathbf{0.85}, \mathbf{0.85}, \mathbf{0.85}, \mathbf{0.85})$ with error distributed according to standard normal $\boldsymbol{e_i} \sim \boldsymbol{N}\ (\mathbf{0}, \mathbf{1})$, $\boldsymbol{e_i} \sim \boldsymbol{N}\ (\mathbf{2}, \mathbf{2}) + \boldsymbol{N}\ (\mathbf{2}, \mathbf{2}), \boldsymbol{e_i} \sim \boldsymbol{LAP}\ (\mathbf{1}, \mathbf{0}), \boldsymbol{and}\ \boldsymbol{e_i} \sim \boldsymbol{\chi}^{\mathbf{2}}_{(\mathbf{4})}$

I generated the observations of the covariates $\mathbf{X_1}, \ldots\ldots, \mathbf{X_9}$ from the multivariate normal $\mathbf{N_{n=9}}\ (\mathbf{0}, \sum)$, here $\sum$ is the variance-covariance matrix which is defined as $\sum_{\mathbf{ij}} = \mathbf{0.7}^{|\mathbf{i-j}|}$. The true relationship between the predictor variables and response variable base on the above true vector is $f(X) = \sum_{j=1}^{9} X_j \beta_j$, So the correct model is defined by

$$f(X) = \sum_{j=1}^{9} 0.85 X_j,$$

Table (4.3). values of MMAD and SD of example Three

Sample Size	Comparison Methods		$e_i \sim N(0,1)$	$e_i \sim N(1,1)$	$e_i \sim N(2,2) + N(2,2)$	$e_i \sim Lap(1,0)$	$e_i \sim x^2_{(4)}$
Small Sample	n=15	BENLR	1.675 (0.344)	1.386 (0. 776)	1.285 (0.937)	1.930 (0.393)	1.383 (0.362)
		ENLR	1.283 (0.385)	1.495 (0.350)	1.364 (0.296)	1.696 (0.535)	1.672 (0.582)
	n=25	BENLR	1.898 (0.483)	1.292 (0.272)	1.352 (0.317)	1.782(0.562)	1.231 (0.452)
		ENLR	2.021 (0.536)	2.200 (0.652)	2.564 (0.674)	2.069 (0.682)	2.003 (0.563)
	n=35	BENLR	1.565 (0.346)	1.654 (0.452)	1.845 (0.453)	1.456 (0.456)	1.674 (0.450)
		ENLR	2.464 (0.834)	2.065 (0.834)	2.078 (0.732)	2.004 (.0786)	2.454 (0.780)
Meddle Sample	n=45	BENLR	2.002 (0.051)	1.417 (0.683)	1.672 (0.493)	1.653 (0.385)	1.452 (0.230)
		ENLR	2.597 (0.433)	2.003 (0.792)	2.112 (0.562)	2.573 (0.840)	2.021(0.3657)
	n=55	BENLR	1.564 (0.674)	1.673 (0.564)	1.562 (0.634)	1.423 (0.427)	1.008 (0.543)
		ENLR	2.456 (0.857)	2.620 (0.407)	2.005 (0.460)	2.478 (0.096)	1.845 (0.670)
	n=65	BENLR	1.956 (0.463)	1.563 (0.194)	1.206 (0.435)	1.546 (0.354)	1.563 (0.322)
		ENLR	2.071 (0.544)	2.534 (0.342)	2.116 (0.537)	2.399 (0.653)	2.054 (0.745)
Large Sample	n=100	BENLR	1.782 (0.452)	1.765 (0.653)	1.435 (0.175)	1.534 (0.264)	1.845 (0.343)
		ENLR	2.071 (0.544)	2.564 (0.934)	2.005 (0.341)	2.071 (0.544)	2.563 (0.544)
	n=200	BENLR	1.673 (0.364)	1.830 (0.282)	1.781 (0.405)	1.530 (0.204)	1.807 (0.432)
		ENLR	2.342 (0.649)	2.023 (0.450)	2.217 (0.671)	2.115 (0.620)	2.316 (0.782)
	n=300	BENLR	1.759 (0.423)	1.673 (0.337)	1.673 (0.452)	1.867 (0.340)	1.684 (0.632)
		ENLR	2.342 (0.685)	2.125 (0.644)	2.233 (0.6754)	2.43 (0.564)	2.231 (0.875)

Table(**4.3**) illustrate the MMAD and SD which are the measures of quality for the estimated regression models of the proposed methods (BENLR) and the (ENLR) based on three types of sample sizes , small samples **(n=15 , n=25 , n=35)** , middle samples **(n=45 , n=55 , n=65)** , and large samples **(n=100 , n=200 , n=300)** . It is very clear that the values of MMAD criterion are the smallest values in the proposed methods compared with the other methods under all different type of error distributions. As well as, the SD criterion shows the preference of the proposed method under different type, of sample sizes and error terms distributions. Eventually, the proposed method is a promising regularization method. Figure (4.1) shows different plots for $e \sim N$ (0,1) error term distributions and different sample sizes, three lines of the parameter estimates based the proposed model (BENLR) ,(ENLR) model , and the true vector of the coefficients.

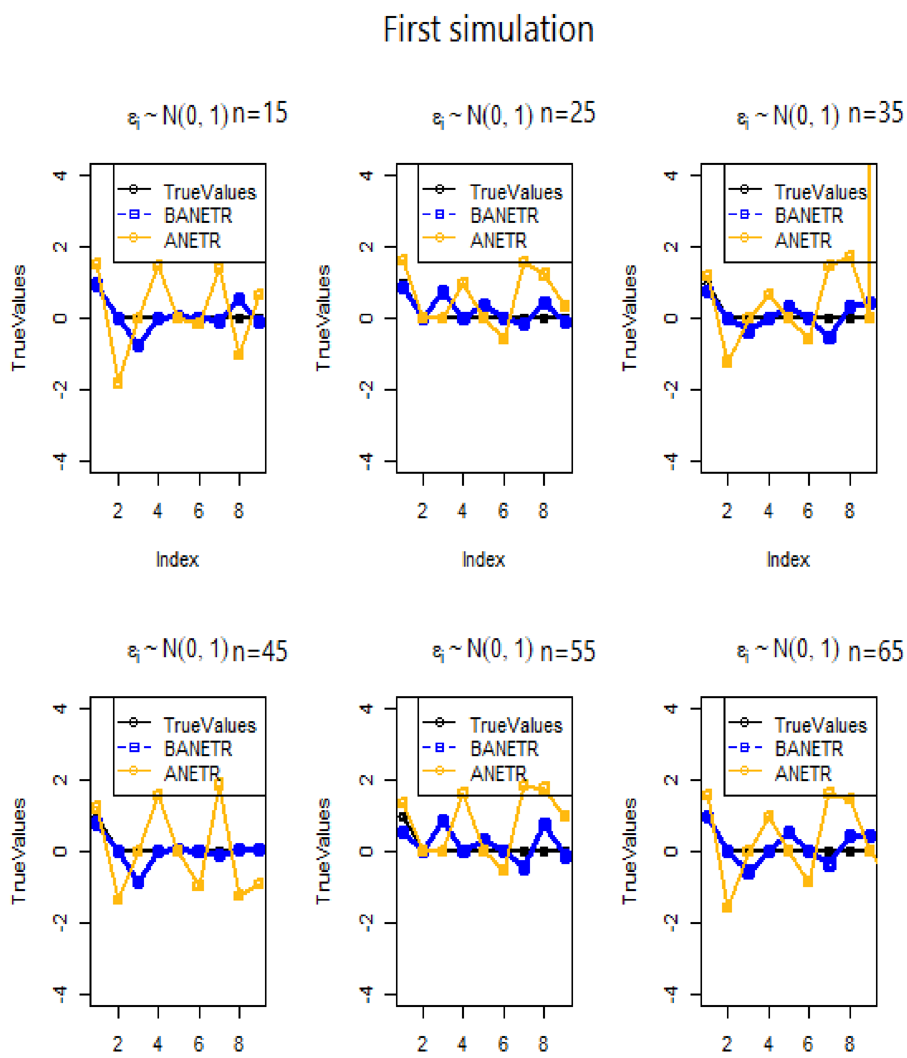

Figure (4.1) parameter estimates fitted lines of example one

Clearly, the proposed model (BENLR) is a comparable and gives best fit. Where the first simulation assumed the very sparse vector $\beta = (1,0,0,0,0,0,0,0,0)$ with black lined, the proposed model parameter estimates with blue line, and (ENLR) model parameter estimates with orange line. Hence, the blue line fits the true vector in all different plots.

Also, figure (4..2) shows different plots for $e \sim N\ (0,1)$ error term distributions and different sample sizes for the second simulation example (sparse model) $\boldsymbol{\beta} = (\mathbf{1}, \mathbf{0}, \mathbf{0}, \mathbf{1}, \mathbf{0}, \mathbf{1}, \mathbf{0}, \mathbf{1}, \mathbf{0})$, three lines of the parameter estimates based the proposed model (BENLR) , (ENLR) model , and the true vector of the coefficients.

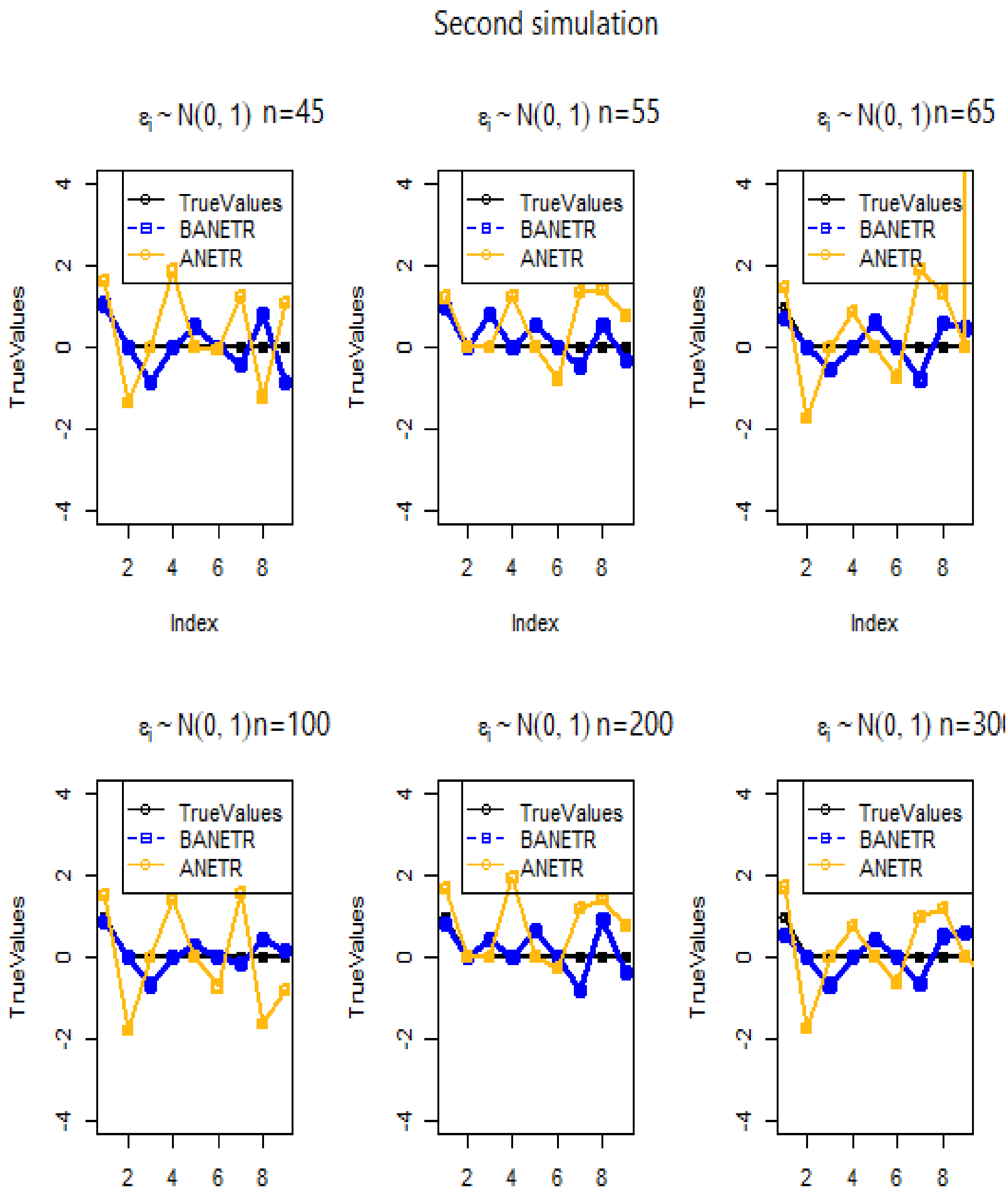

Figure (4.2) parameter estimates fitted lines of example two

Obviously, the proposed model (BENLR) is a comparable and gives best fit. Where the second simulation assumed the sparse vector with black lined, the proposed model parameter estimates with blue line, and (ENLR) model parameter estimates with orange line. Hence, the blue line fits the true vector in all different plots.

Figure (4.3) shows different plots for $e \sim N\,(0,1)$ error term distributions and different sample sizes for the second simulation example (dense model)

$\boldsymbol{\beta} = (\mathbf{0.85}, \mathbf{0.85}, \mathbf{0.85}, \mathbf{0.85}, \mathbf{0.85}, \mathbf{0.85}, \mathbf{0.85}, \mathbf{0.85}, \mathbf{0.85})$ three lines of the parameter estimates based the proposed model (BENLR) , (ENLR) model , and the true vector of the coefficients

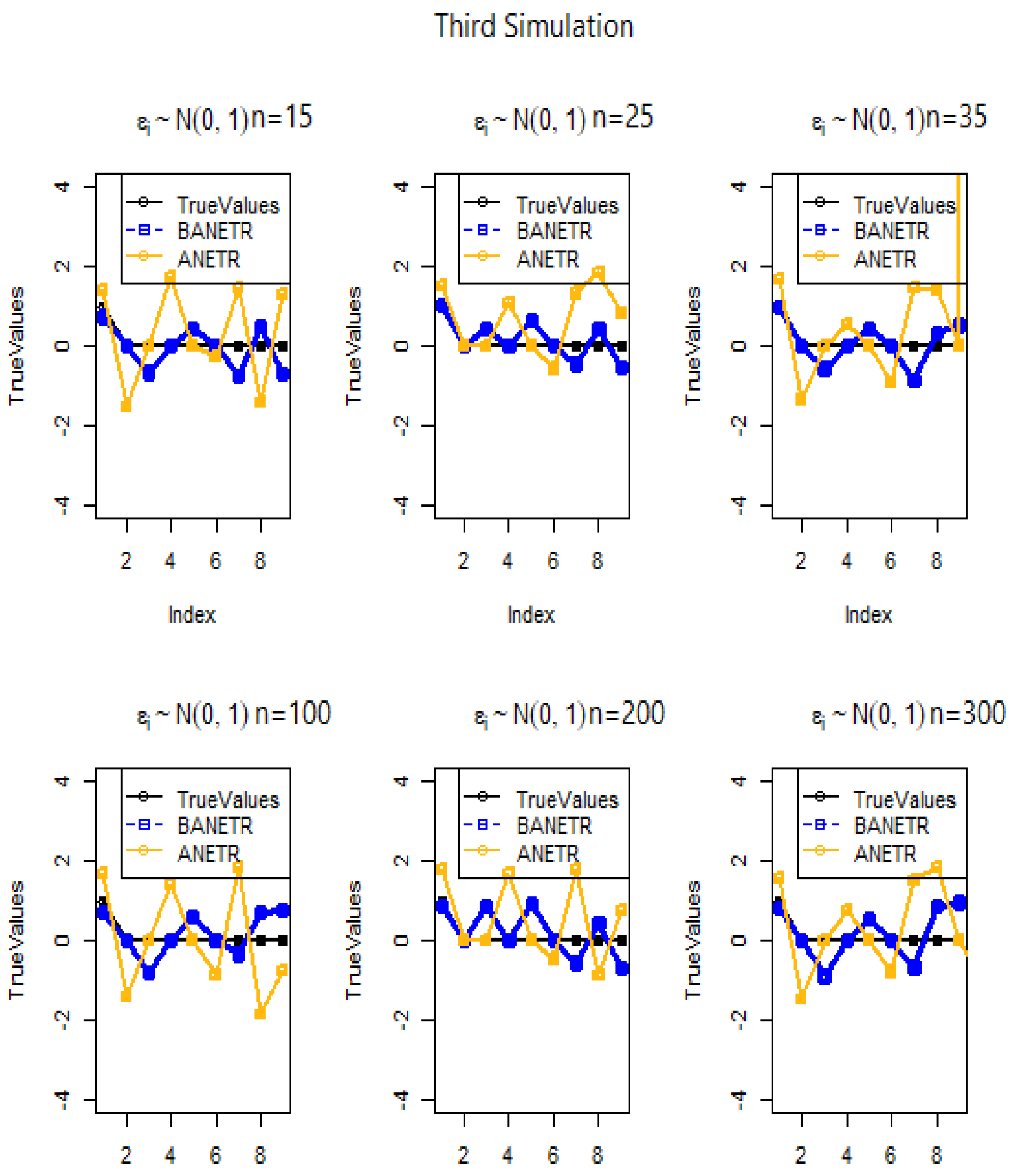

Figure (4.3) parameter estimates fitted lines of example three

Obviously, the proposed model (BENLR) is a comparable and gives best fit. Where the second simulation assumed the sparse vector $\boldsymbol{\beta} = (\mathbf{1}, \mathbf{0}, \mathbf{0}, \mathbf{1}, \mathbf{0}, \mathbf{1}, \mathbf{0}, \mathbf{1}, \mathbf{0})$, with black lined, the proposed model parameter estimates with blue line, and (ENLR) model parameter estimates with orange line. Hence, the blue line fits the true vector in all different plots.

4.2 Real Data Analysis

We will examine the proposed method and compare it with other models. Real-life case have studied based on the blood viscosity syndrome disease data by considering the blood viscosity syndrome as response variable (**y**), and the explanatory variables (**X**) The data collected from pathological analyzes of patients visiting the Advancing Surgical Care disease in the province of Babylon Centre . In addition to a set of questions posed by the researcher to affected Persons, this work was conducted on a sample that included (n=97) Person. The following data contains information that records visits of blood viscosity patients to Marjan Teaching Hospital in Babil Governorate Moreover , we used (97) models of different people , that is we took a simple the random sample , patients were drawn to study the factors affecting patients' blood viscosity (response variable), while the predictor variables are as follows :

$\boldsymbol{y_i}$	Blood viscosity sy ndrome
$\boldsymbol{X_1}$	Person gender
$\boldsymbol{X_2}$	Age of person
$\boldsymbol{X_3}$	Environment / elevated / flat
$\boldsymbol{X_4}$	Occupation
$\boldsymbol{X_5}$	Anemia
$\boldsymbol{X_6}$	Temperature

X_7	Genetics factor
X_8	Person weight
X_9	Blood pressure
X_{10}	Mental state
X_{11}	Kidney disease
X_{12}	Drink water and fluids
X_{13}	Congenital heart defects
X_{14}	Decreased plasma levels in the blood
X_{15}	Lung disease
X_{16}	Dietary pattern \ fats
X_{17}	Drinking alcoholic beverages
X_{18}	Playing sports
X_{19}	Smoking
X_{20}	Medicines and drugs
X_{21}	Increasing the amount of proteins in the blood

Table (4.4) Value of mean square error (MSE) and mean absolute error (MAE)

Methods	MSE	MAE
BENLR	19.077	9.236
ENLR	23.112	13.243

From table (4.4), the value of the quality measurement MSE of the proposed method gives the less value comparing (MSE=19.077) with the MSE of the ENLR (MSE=23.112). Also, the other measurement of quality criteria MAE of the proposed method gives the less (MAE=9.236) .Consequently, it is clearly that the proposed model outperforms the other method in terms of prediction accuracy.

Table (4.5) Parameter estimates of the predictor variables

Descriptive variables	Variables	$\hat{\beta}$ BENLR	$\hat{\beta}$ENLR
Person gender	x_1	1.466	2.875
Age of person	x_2	0.0044	0.000
Environment / elevated / flat	x_3	0.000	0.122
Occupation	x_4	0.287	0.000
Anemia	x_5	0.000	0.000
Temperature	x_6	0.370	0.000
Genetics factor	x_7	0.716	0.000
Person weight	x_8	0.000	0.000
Blood pressure	x_9	0.107	0.000
Mental state	x_{10}	0.000	-0.0506
Kidney disease	x_{11}	0.081	0.000
Drink water and fluids	x_{12}	0.189	0.000
Congenital heart defects	x_{13}	-0.020	0.000
Decreased plasma levels in the blood	x_{14}	-0.271	-0.224
Lung disease	x_{15}	0.364	0.000
Dietary pattern \ fats	x_{16}	0.000	-0.243
Drinking alcoholic beverages	x_{17}	0.650	0.000
Playing sports	x_{18}	0.000	0.000
Smoking	x_{19}	-0.182	0.000
Medicines and drugs	x_{20}	-0.112	0.000
Increasing the amount of proteins in the blood	x_{21}	0.000	0.000

From table (4.5) the proposed model removed the irrelevant predictor variable that does not influence the response variable, So we can say that the proposed model provide variable selection procedure. For example, the parameter estimate of the (Environment / elevated / flat) variable takes zero value, and so on for the other variables (Anemia , Person weight, Mental state, Dietary pattern \ fats, Increasing the amount of proteins in the blood). Eventually, the relevant predictor variables that effects the response variable (Blood viscosity) are (Person gender, Age of person, Occupation, Genetics factor, Blood pressure, Kidney disease, Drink water and fluids, Congenital heart defects, Decreased plasma levels in the blood, Lung disease, Drinking alcoholic beverages, Smoking, Medicines and drugs).

Chapter Five

Conclusions and Recommendations

5.1 Conclusions

In this thesis, we proposed a new scale mixture of normal distribution mixing with truncated gamma distribution for the Laplace prior distribution. So, based on that new scale mixture, we developed new Bayesian hierarchical model for elastic net linear regression, there for, new Gibbs sampler algorithm have developed. The simulation study have implemented for the proposed model and compared the results with some exist estimation method. The outperforms of the proposed method clearly in results of simulation example as well in the real data analysis results which explained that the penalized proposed method outperformed the other method in terms of variable selection procedure.

5.2 Recommendations

The Bayesian penalized method that we proposed motivates us to recommend the researchers who are interested in the study of the Bayesian regularization models to develop different regression models by following the same methodology in this thesis through studying the convergence of Gibbs sample algorithm for the posterior distribution. Hence, by using the same scale mixture that is explained in the context of this thesis, we can develop the Bayesian adaptive elastic net in Binary regression, quantile regression and logistic regression models.

References

Akaike, H. (1973), “Information Theory and an Extension of Maximum Likelihood Principle,“ in Second International Symposium on Information Theory (B. N. Petrov and F. Caski, Eds.) Akademia Kiado, Budapest, 267-28 1.

Alhamzawi, R. (2016). Bayesian elastic net Tobit quantile regression. Communications in Statistics-Simulation and Computation, 45(7), 2409-2427.

Almusaedi, M.H.M. and Flaih, A.N.(2021b). Employing the Bayesian Elastic Net in Quantile Regression with an Application. AL-Qadisiyah Journal For Administrative and Economic sciences. (Accepted for publishing).

Almusaedi, M.H.M. and Flaih, A.N. (2021a). Simulation Study for Penalized Bayesian Elastic Net Quantile Regression. Al-Qadisiyah journal of pure science. P- ISSN 1997-2490 , E- ISSN 2411-3514.

Al-Nasser, A. D. (2014). Two steps generalized maximum entropy estimation procedure for fitting linear regression when both covariates are subject to error. Journal of Applied Statistics, 41(8), 1708-1720.

ALsafi, M.R. and Flaih, A.N. (2021). Sparsity in Bayesian Elastic Net in Tobit Regression with an Application. AL-Qadisiyah Journal For Administrative and Economic sciences. (Accepted for publishing).

Andrews, D. F., & Mallows, C. L. (1974). Scale mixtures of normal distributions. Journal of the Royal Statistical Society: Series B (Methodological), 36(1), 99-102.

Alhamzawi, R., & Ali, H. T. M. (2018). The Bayesian elastic net regression. Communications in Statistics-Simulation and Computation, 47(4), 1168-1178.

Billor, N., Chatterjee, S., & Hadi, A. S. (2006). A re-weighted least squares method for robust regression estimation. American journal of mathematical and management sciences, 26(3-4), 229-252.

Celeux, G., Anbari, M. E., Marin, J.-M., & Robert, C. P. (2012). Regularization in regression: Comparing Bayesian and frequentist methods in a poorly informative situation. Bayesian Analysis, 7(2), 477-502.

Draper, N. R., & Smith, H. (1998). Applied regression analysis (3rd ed.). New York: John Wiley & Sons.

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. The Annals of statistics, 32(2), 407-499.

Efroymson, M.A. (1960). "Multiple Regression Analysis," In: A. Ralston and H. S. Wilf, Eds., Mathematical Methods for Digital Computers, John Wiley, New York.

Flaih, A.N, Alshaybawee, and Al husseini F.H, (2020). sparsity via new Bayesian Lasso . Periodicals of Engineering and Natural Sciences , vol.8 , is suc 1, 345-359.

Friedman, J., Hastie, T., Hoefling, H. and Tibshirani, R. (2007). Pathwise coordinate optimization. Annals of Applied Statistics 2.

George, E.I. and McCulloch, R.E..(1993).Variable Selection Via Gibbs Sampling. Journal of the American Statistical Association, Vol. 88, No. 423. (Sep., 1993), pp. 881-889.

Ghosh, S. (2007). Adaptive elastic net: An improvement of elastic net to achieve oracle properties. Preprint, 1.

Hans, C. (2009). Bayesian lasso regression. Biometrika, 96, 835-845.

Hans, C. (2010). Model uncertainty and variable selection in Bayesian lasso regression. Statistics and Computing, 20, 221-229.

Hans, C. (2011). Elastic net regression modeling with the orthant normal prior. Journal of the American Statistical Association, 106(496), 1383-1393.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). Unsupervised learning. In The elements of statistical learning (pp. 485-585). Springer, New York, NY.

Hastie, T., Tibshirani, R., and Wainwright, M. (2015). Statistical Learning with Sparsity: The Lasso and Generalizations. Taylor Francis Group, Forida, USA.

Hesterberg, T., Choi, N. H., Meier, L., & Fraley, C. (2008). Least angle and L1 penalized regression: A review. Statistics Surveys, 2, 61-93.

Hocking, R.R. (1976) The Analysis and Selection of Variables in Linear Regression. Biometrics, 32, 1-50.

Hoerl, A. E., and Kennard, R. W. (1970a). Ridge regression: Applications to nonorthogonal problems. Technometrics, 12(1), 69-82.

Hoerl, A. E., and Kennard, R. W. (1970b). Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1), 55-67.

Hoerl, A. E., Kennard, R. W., and Baldwin, K. F. (1975). Ridge regression, some simulations. Communications in Statistics, A4, 105-123.

James, G.,Witten, D., Hastie, T., and Tibshirani, R. (2013). An Introduction to Statistical Learning: With Applications in R. Springer Publishing Company, Incorporated, New York.

Kyung, G., Ghosh, M., Gill, J., &, Casella M. (2010). Penalized regression, standard errors, and Bayesian lassos. *Bayesian analysis*, *5*(2), 369-411.

Li, Q., & Lin, N. (2010). The Bayesian elastic net. Bayesian analysis, 5(1), 151-170.

Mallick, H., & Yi, N. (2014). A new Bayesian lasso. Statistics and its interface, 7(4), 571-582.

Mallows, C.L. (1973). Some Comments on C_p. Technometrics, Volume 15, 1973 - Issue 4.

Osborne, M.R., Presnell,B. and Turlach,B.A.(2000a). On the LASSO and Its Dual. Journal of Computational and Graphical Statistics Vol. 9, No. 2 (Jun., 2000), pp. 319-337. Published By: Taylor & Francis, Ltd.

Osborne, M.R., Presnell,B. and Turlach,B.A.(2000b).A new approach to variable selection in least squares problems. IMA Journal of Numerical Analysis, Volume 20, Issue 3, July 2000, Pages 389–403.

Park, T., & Casella, G. (2008). The bayesian lasso. Journal of the American Statistical Association, 103(482), 681-686. Choi, W. W., Weisenburger, D. D., Greiner, T. C., Piris, M.

Schwarz, G. (1978), “Estimating the Dimensions of a Model,” Annals of Statistics, 121,461-464.

Seber, G. A. F. and Lee, A.J. (2003). Linear Regression Analysis, Second Edition. John Wiley & Sons, Inc. Book Series: Wiley Series in Probability and Statistics.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), 267-288.

Tibshirani, R. (1997). The lasso method for variable selection in the Cox model. Statistics in medicine, 16(4), 385-395.

Yuan, M., & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 68(1), 49-67.

Zhu, J., Rosset, S., Tibshirani, R., & Hastie, T. J. (2003). 1-norm support vector machines. In Advances in neural information processing systems (p. None).

Zou, H. (2006). The adaptive lasso and its oracle properties. Journal of the American Statistical Association, 101, 1418-1429. 161

Zou, H. (2006). The adaptive lasso and its oracle properties. Journal of the American Statistical Association, 101, 1418-1429.

Zou, H., & Hastie, T. (2005). Regression shrinkage and selection via the elastic net, with applications to microarrays. JR Stat Soc Ser B, 67.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society, Series B, 67, 301

Zou, H., & Zhang, H. H. (2009). On the adaptive elastic-net with a diverging number of parameters. Annals of statistics, 37(4), 1733.

Zou, H., Hastie, T., & Tibshirani, R. (2007). On the degrees of freedom of the lasso. The Annals of Statistics, 35, 2173-2192.

الخلاصة

ركزت الرسالة على اسلوب تصفير معالم المتغيرات المتنبأ بها الغير مهمة اتي ليس لها تاثير على المتغير التابع (إجراء اختيار المتغيرات) جنبًا إلى جنب مع الشبكة المرنة البايزية وطرق التنظيم التقليدية للشبكة المرنة في نموذج الانحدار الخطي. تم التركيز ايضا على توضيح طرق التنظيم مثل طريقة ريدج وطريقة اللاسو وطريقة الشبكة المرنة. قمنا أيضًا بدراسة شبكة بايز المرنة من خلال استخدام الخليط القياسي لخلط التوزيع الطبيعي مع توزيع رايلي كتوزيع لابلاس السابق لمعامل الانحدار ومن خلال إجراء بعض التحويلات الرياضية لمزيج المقياس هذا ، اقترحنا مزيجًا جديدًا من التوزيع الطبيعي مع توزيع كاما المقطوع . بالإضافة إلى ذلك ، اقترحنا نموذج بيز الهرمي جديد بناءً على مزيج المقياس المقترح ، ومن ثم اقترحنا خوارزمية عينة جبس جديدة. إن طريقة الشبكة المرنة البيزية المقترحة قادرة على التوصل إلى حلول صفرية لبعض المعالم ، عن طريق تعيين بعض المعلمات المقدرة بدقة تصل إلى الصفر وهذا واضح من نتائج المحاكاة وكذلك من نتائج تحليل البيانات الحقيقية.

جمهورية العراق
وزارة التعليم العالي والبحث العلمي
جامعة القادسية
كلية الادارة والاقتصاد
قسم الاحصاء
الدراسات العليا

الحل الصفري باستخدام نموذج انحدار الشبكة المرنة البيزية والشبكة المرنة مع تطبيق

رسالة قدمها الطالب

علي احمد حسين

الى
مجلس كلية الادارة والاقتصاد في جامعة القادسية وهي جزء من متطلبات نيل شهادة الماجستير في علوم الاحصاء

باشراف

م . د. بحر كاظم محمد

1443هـ **2022م**