Republic of Iraq
Ministry of Higher Education & scientific research
Department of Statistics
College of Administration and Economics
University of Al-Qadisiyah

Bayesian Reciprocal Lasso for Right Censored Data

A thesis submitted to the council of the college of Administration and Economics in partial fulfillment of the requirements for the degree of master of Statistics Sciences

By
Fatimah Kadhim Mohammed

Supervised by

Prof. Dr. Ahmad Naeem Flaih

1443 A.H **2022 A.D**

بسم الله الرحمن الرحيم

(وَيَسْأَلُونَكَ عَنِ الرُّوحِ ۖ قُلِ الرُّوحُ مِنْ أَمْرِ رَبِّي وَمَا أُوتِيتُم مِّنَ الْعِلْمِ إِلَّا قَلِيلًا)

صدق الله العظيم

(الاسراء الاية 85)

Acknowledgement

All praise is to ALLAH Almighty, who enabled me to complete this task successfully and our utmost respect To his last Prophet Mohammed and his holy family

My deepest gratitude is to my advisor, (Prof. Dr. Ahmad Naeem Flaih) for invaluable guidance, supervision and untiring efforts during this work.

I would like to thank all staff of the college of Administration and Economics.

Finally, I would like to thank all the kind, helpful and lovely people who helped me Directly or indirectly to complete this work

Table of Contents

Chapter One

Introduction and Literature Review

Chapter Two

Theoretical Part

Chapter Three

Simulation and Real Data Analysis

Chapter Four

Conclusions and Recommendations

List of Tables

List of Figures

Abstract

This thesis focuses on Bayesian reciprocal lasso (rlasso) regression in presence of a right censored limited dependent variable. Choosing the optimal subsets of predictor variables is the most common aim of regression analysis. The reciprocal lasso combined the reciprocal of L1-Norm in the penalty function. Nowadays many regression parameters estimation method including regularization methods have been developed to build a parsimonious model. Reciprocal lasso is a new regularization methods that provide a more parsimonious (variable selection with more interpretation) regression model. Few literature reviews about (rlasso) because of the new idea. We used the scale mixture of double Pareto (SMDP) and the scale mixture of truncated normal (SMTN) that proposed by Mallick et al. (2020) and we add modification for (SMTN) through new hierarchical prior model. We employed the (SMDP) and the modified (SMTN) in the structure model of the right censored dependent variable. Simulation examples have been conducted, as well as real data analysis to examine the behavior the posterior distributions. The results show that the employed scale mixture types outperform other common regularization methods in both the simulation and real data analysis. Overall the reciprocal lasso models provide an elegant foundation for a class of regularization methods that improves the sparse solution and closer to the true solution.

Chapter one

Introduction and Literature Review

1.1 Introduction

Statisticians have formulated statistical models to solve certain problems. The regression model demonstrates the relationship between the dependent (response) variable Y, and one or more independent (predictor) variables X. This relationship is defined as,

$$Y = f(X;\beta) + u$$

Where $u_i \sim N(0, \sigma^2)$, and $E(Y) = f(X;\beta)$. Then, the linear relationship between the predictor variables and the dependent (response) variable take the following function

$$f(X;\beta) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k \ldots\ldots\ldots\ldots (1.1)$$

Where k the number of predictor variables. It is very well known that the objective of the regression model (1.1) is to find the mean of Y. Many predictor variables included in the regression model affect Interpretation of the estimated model and may be inflated the variances of the parameter estimates which cause the poor prediction accuracy for the estimated model. So, variable selection procedure is essential to statistical modeling of many predictor variables problems which were recently found in many fields of scientific discoveries. Therefore, we can say that there is another objective of conducting the regression analysis which is called model selection. The quality of parameter estimate is measured by the bias and variance criteria of the estimators and then the prediction accuracy and interpretability of the regression model can be examined. See (Chatterjee and Hadi 2006), and (AlNasser 2014) for more details.

Usually the Ordinary least squares (OLS) used to solve (1.1) by the following minimizing problem of Residual Sum of Square (RSS),

$$\hat{\beta} = \underset{\beta}{argmin} \sum_{i=1}^{n} [y - f(X; \beta)]^2$$

It is known that OLS estimates are BLUE especially when $(n \geq k)$. But when k near the sample size n or $(k \geq n)$, the OLS estimate comes with high variances and biased estimates which leads to very poor prediction . To overcome this problem in using OLS, one can use the regularization method that depends on penalized methods which also tread the model selection problem. In this thesis, we will focus on the upper limit model. Also, the upper limit (upper censored) model will merge with the variable selection procedure by using the Bayesian regularization reciprocal lasso method. The Right (upper) censored regression model is more reliable if the variable selection procedure has been followed.

1.2 Literature Review

The analysis of limited dependent (response) variable is widely observed in many applications, where there is a boundary or limit on the response variable which means there are some of the values of $\boldsymbol{y}$ reach this limit or boundary. Limited dependent variable leads us to the censored sample which its observations are $(\mathbf{y_1}, \mathbf{y_2}, \dots, \mathbf{y_n})$ resulting from a latent variable $(\mathbf{y}^*)$ based on some structural function form. An awareness of this type of dependent variable is very important, because adopting the inappropriate statistical tool will yield an unsatisfied regression model. Hence, censored is only for the value of the dependent variable. In general, there are three types of censoring value (from below (left), from above (right), interval). In this thesis we are concerning in right censoring data.

In the analysis of the regression model, the number of covariates included in regression model brings the researcher to develop the mechanism of variable selection procedure. So, the variable selection procedure is treated with the regression form specification. The residual mean squares (**RMS**) criterion is a model selection tool defined by: (Chatterjee and Hadi 2006)

$$RMS = \frac{\sum_{i=1}^{n} u_i^2}{n-p}$$

Where $\boldsymbol{p}$ is the number of independent variables and $\boldsymbol{u}$ is the error, the smallest the ***RM***S the regression model is preferred.

Efroymson (1960) produced the stepwise method that utilized the model selection Forward Selection (FS) and Backward Elimination (BS) methods. The stepwise method calculation mechanism depends on the inclusion and deletion of predictor variables. Stepwise method basically is a modification of (FS and BE) methods.

Hoerl and Kennard (1970) introduced the ridge regression as procedure to overcome the problem of using the OLS in case of multicollinearty that present in the design matrix and/or when p is near n. Ridge regression produced biased estimators with small variances. The ridge regression model including parameters estimates that shrunk toward zero but not exactly equal to zero and then no variable selection is achieved. The ridge estimator defined by:

$$\hat{\beta} = \underset{\beta}{argmin}\, RSS + \lambda\, \|\beta\|^2$$

Where $\lambda \geq 0$ the shrinkage parameter and $\|\beta\|^2$ is L_2 – norm. Note that if $\lambda = 0$ then $\hat{\beta} = $ OLS estimates.

Mallows (1973) defined the following criterion that is called Mallows $\mathbf{C_p}$ criterion to assess the performance of the regression model:

$$C_p = \frac{\sum_{i=1}^{n} u_i^2}{\hat{\sigma}^2} + (2p - n),$$

where $\hat{\boldsymbol{\sigma}}^2$ is the estimated value of variance.

Akaike (1973) defined Akaike information criterion (**AIC**), which is a model selection tool by:

$$AIC_p = n\ln\left(\frac{\sum_{i=1}^{n} u_i^2}{n}\right) + 2p,$$

The smallest value of **AIC** the better model.

Hocking (1976) introduced an evaluating regression tool which is called *all possible* equations method that provides $\boldsymbol{2^p}$ equations ($\boldsymbol{p}$ is the number of covariates), here $\boldsymbol{RMS}, \boldsymbol{C_p},$ and $\boldsymbol{R^2}$ are used to select the best fit model. The drawback of all possible equations method is when the number of equations getting larger.

Schwarz (1978) defined the modified AIC criterion that is called Bayes Information criterion (BIC) by:

$$\mathrm{BIC_p} = \mathrm{n}\ln\left(\frac{\sum_{i=1}^{n} u_i^2}{n}\right) + \mathrm{p}(\ln \mathrm{n}),$$

The smallest value of BIC the better model.

Tibshirani (1996) developed new as regularization method named lasso which gives sparse solution for the linear regression coefficients. Lasso adds penalty function that include L1-norm function which is controlled by the shrinkage parameter. The parameter estimates for some predictor variables reach the zero value and the solution regards as sparse solution.

Fan and Li (2001) proposed a new regularization method called smoothly clipped absolute deviation (SCAD). The SCAD estimator has oracle properties. SCAD solutions are sparse by considering that the penalty function is discontinuous at zero and continue to ensure the stability of the solution.

All the above regularization methods are frequent methods. So, since we are interested in Bayesian estimation, the following studies are important to mention:

Group lasso was proposed by Bakin (1999) and then developed by Yuan and lin (2002). Group lasso as regularization method founded as a natural extension for lasso method to treats the effects at group level, also this method gives sparse solution between groups.

Seber and Lee (2003) introduced algorithms for the computations of the all possible subsets. These algorithms are reduced the computation by 50% through using the matrices of cross – product and sums of squares.

Tibshirani et al. (2005) proposed fused lasso as regularization method to treat the groups of predictor variables in meaningful way order. Where in many application fields, there are sets of correlated predictor variables, in this case lasso usually select randomly one predictor variable from the group.

Zou and Hastie (2005) proposed the elastic net method which provides sparse solution. Elastic net method combined the ridge and lasso penalty functions to handle the grouping effects of correlated predictor variables. They showed that the elastic net has the sparse solution.

Zou (2006) introduced the adaptive lasso method which is a new regularization method. This new penalized method scaling each parameter in the adaptive penalty functions with different weight. The adaptive lasso estimator has oracle properties.

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

Meinshausen (2007) introduced new regularization method that controls the bias of the lasso parameter. This new method called the relaxed lasso method, which is works under two shrinkage parameters, the first one controls the shrinkage of the regression parameter and then performs the variable selection procedure, and the second shrinkage parameter controls the amount of bias. The values of shrinkage parameters were obtained by cross validation method.

Zou and Zhang (2009) introduced adaptive elastic net that combined the adaptive lasso and elastic net methods. The adaptive elastic net enjoyed the good properties of the both combined penalized functions.

Park and Casella (2008) introduced the Bayesian analysis for the regularization method hased on lasso linear regression that developed the posterior distribution through new scale mixture for the prior distribution.

Mallick and Yi (2014) developed new scale mixture that mixed uniform distribution with particular gamma distribution $(2, \lambda)$ as the prior representation of the Laplace distribution. Therefore, based on the proposed scale mixture a new lasso solution has developed for the linear regression model, as well as, new hierarchical prior model and new Gibbs sampler algorithm have proposed. The new proposed model examined by simulation study and the results outperforms the new method over some exists regularization methods.

Song (2014) the first work that concerned the reciprocal lasso estimators that have the oracle property. This work was discussed the Bayesian variable selection procedure for ultra-high dimensional linear regression through the strategy of split-and-merge. The estimators are consistent and have asymptotic properties that give better results than the elastic net and lasso methods. Song and Liang (2015), and Song (2018) discussed the reciprocal L1-norm Bayesian variable selection in lasso methods. The reciprocal lasso (rlasso) proposed by song (2014) introduced the following penalty function:

$$P(\beta, \lambda) = \lambda \sum_{j=1}^{p} \frac{1}{|\beta_j|} I(\beta_j \neq 0) \cdots\cdots\cdots\cdots (1.1)$$

where λ is shrinkage parameter penalty function gives sparse solutions with infinity penalties , in contract of lasso that gives spares solution with nearly zero penalty funds. The function (1.1) is decreasing in the interval $(0, \infty)$, discontinuous at zero.

Alhamzawi (2016) proposed new Bayesian elastic net in Tobit quantile regression model. The proposed method is sparsity. He employed the gamma priors to develop the hierarchical prior model. New Gibbs sampler algorithm was introduced for the MCMC algorithm. Simulation studies have been conducted to examine the proposed model in terms of variable selection procedure, also the proposed has been applied to real data and the results show outperforms of the proposed model compared with some penalized method.

Alhamzawi (2017) proposed a new hierarchal prior model for the Tobit regression with lasso method. The prior distribution of the regression parameters is presented as Laplace distribution. The Laplace distribution presented as a scale mixture mixing uniform distribution with particular gamma distribution. Based on the proposed hierarchical prior model new Gibbs sampler algorithm has been proposed. Simulation studies have been conducted for Parameter estimation and variable selection, as well real data analysis have examined the behavior of the proposed model which showed that the outperforms compared with some other regularization methods.

Alhusseini (2017) studied the Bayesian composite Tobit quantile regression model. In this work, asymmetric Laplace prior distribution is introduced as the prior distribution of the parameter regression. The scale mixture proposed by Mallick and Yi (2014) has been used to develop a new hierarchical prior model and new Gibbs sampler algorithm. The simulation examples and real data analysis showed that the proposed model gives better results than the other methods.

Rahim and Haithem (2018) introduced Bayesian elastic net in Tobit regression model. Variable selection procedure and coefficients estimates presented based on the proposed model. The new posterior distribution derived based on new hierarchical prior model as well as based on new Gibbs sampler algorithm. The scale mixture formula of truncated normal distribution mixing with exponential mixing distribution has used to represent the Laplace prior distribution. The proposed model outperforms the other existing method in simulation and real data analysis in terms of variable selection and parameter prediction error.

Soret et al. (2018) proposed new idea for parameter estimation of linear regression based on reversing the Buckley-James least squares algorithm to cope with left-censored data. Lasso penalized method have used to treats the situations of many predictor variables. The proposed method named Lasso-regularized Buckley-James least squares method which is works under non-parametric estimation with Kaplan-Meier and parametric estimation with normal distribution. Simulation analysis of the proposed model has examined, as well real data analysis has conducted based on clinical trials. The proposed model showed the less prediction error compared with other methods.

Hilali (2019) proposed a transformation for the scale mixture of double exponential prior distribution that developed by Mallick and Yi (2014). This new representation of the prior distribution employed into new hierarchical prior model and new Gibbs sampler algorithm. Bayesian adaptive lasso Tobit regression has used based on the new transformation. Variable selection procedure has examined under this proposed model with new posterior distribution. The results of simulation and real data analysis are comparable with some exists regularization methods.

Flaih et al. (2020) proposed using scale mixture that mixed Rayleigh with normal distribution in lasso and adaptive lasso regression. Moreover, the proposed scale mixture employed in deriving new hierarchical prior model as well as new Gibbs sampler algorithm. The results of simulation real data the analysis showed outperforms of the proposed posterior distribution in part of variable selection and the efficiency of the proposed estimator.

Mallick et al. (2020) introduced the reciprocal Bayesian lasso by employing scale mixture of double Pareto with truncated normal distribution. The liner reciprocal Bayesian lasso estimator is defined as follow

$$h(\beta) = \ argmin\, RSS + \lambda \sum_{j=1}^{p} \frac{1}{|\beta_j|}\, I(\beta_j \neq 0) \cdots\cdots\cdots\cdots (1.2)$$

Alhamzawi and Mallick (2021) introduced the Bayesian reciprocal lasso quantile regression by defined the following estimator:

$$Q(\beta) = \ argmin \sum_{i=1}^{n} \rho(y_i - x_i^T \beta) + \lambda \sum\nolimits_{j=1}^{p} \frac{1}{|\beta_j|} \ I(\beta_j \neq 0) \cdots \cdots \cdots (1.3)$$

Where $\rho(.)$ is the loss function.

1.3. Thesis Problem

The misspecification of the regression function that relates the response variable to predictor variables yields an inaccurate predicted model with un meaningful parameter estimates and with less ability of interpretability. This motivates many researchers in data analysis to develop variable selection methods. The variable selection procedure plays an important role in regression analysis. Most of the researchers in the regression analysis are looking for the parsimonious model (model selection) that have a less number of predictor variables with more ability of interpretability.

1.4. Thesis Objectives

In this thesis, we employed the reciprocal lasso method as a variable selection procedure in the right censored dependent variable in order to obtain a parsimonious model. Also, to illustrate how the reciprocal lasso works compared with other regularization methods. A simulation study is conducted with a view to examine the reciprocal lasso on the number of predictor variables. In addition to that, a practical analysis is used to illustrate the effect of reciprocal lasso on the variable selection procedure.

Chapter Two

Theoretical Part

2.1 Introduction

This chapter illustrates some important basic conceptions that are going to be used in the context of this thesis. Linear regression model, the right censored regression model, collinearity, variable selection procedure, lasso method, and reciprocal lasso method. Additionally, there will be a brief talk about the ordinary Least Squares **(OLS)**. Also, this chapter includes the theoretical part of the Bayesian reciprocal lasso right censored model.

2.2 Linear Regression Model

It is well known that the typical linear regression model is the usual tool of representing the $\boldsymbol{n}$ observations (x_1, y_1), … ,(x_1, y_n) selected randomly and independently from a particular population. The response (outcome) variable y_i; $i = 1, 2, \ldots, n$, is a function of the p predictor variables $x_{i1}, x_{i2}, \ldots, x_{ip}$, such that:

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

the matrix form of model (2.1) can be defines as follows,

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u}, \quad \ldots (2.2)$$

Where

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

$$\boldsymbol{\beta}_{(\boldsymbol{p}+\mathbf{1})\times\mathbf{1}} = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix}, \boldsymbol{u}_{\boldsymbol{n}\times\mathbf{1}} = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}.$$

$\boldsymbol{\beta}_j$ are the unknown parameters that we try to estimate and $\boldsymbol{u}$ is the random error. OLS method results are called BLUE if the error term satisfy; $E(u) = 0$, $Var(u) = \sigma^2 I$, $u \sim N(O, \sigma^2 I)$, and $Cov\,(u_i, u_j) = 0$, where $i \neq j$. (Chatterjee and Hadi 2006, AlNasser 2014). Following the above assumption of error term and taking the expected of (2.1), we have

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

Then, the error is as follows,

$$u = y_i - x_i^T \hat{\beta} \qquad \ldots (2.3)$$

The following minimization problem represents the solution of **OLS**:

$$argmin \sum_{i=1}^{n} (u_i)^2 = argmin \sum_{i=1}^{n} (y_i - x_i^T \hat{\beta})^2 \ldots (2.4)$$

So, based on the minimization problem (2.4), the OLS estimator defined by:

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad \ldots (2.5)$$

The **OLS** estimator in (2.5) is unbiased with smallest variance.

2.3 Censored Regression Models

Based on the type of the observations of response variable, the regression models are formulated. In many applications the observations of response variables are in some known ranges. So, the structural form of the regression observations (censored) are based on unobserved latent variable.

The censored model from the below limit (left constraint) is defined as follows: (Maddala, 1993)

$$y_i = \begin{cases} y_i^* & if \quad y_i^* \geq c \\ c & if \quad y_i^* < c \end{cases} \quad \cdots\cdots\cdots\cdots\cdots (2.6)$$

Where $y_i^* = x_i'\beta + u_i$, y_i^* is the latent variable (unobservable variable). The vector $(y_1, y_2, \dots, y_n)$ is the censured sample and β is the $(1 \times k)$ vector of unknown coefficients, x_i is the vector of known observations, u_i is the error term, $u_i \sim N(0, \sigma^2)$, and c is the known constant (censored point). Setting $c = 0$ in (2.6) yields Tobit regression model.

The Censored model from the upper limit (upper constraint) is defined as follows:

$$y_i = \begin{cases} y_i^* & if \quad y_i^* < c \\ c & if \quad y_i^* \geq c \end{cases} \quad \cdots\cdots\cdots\cdots\cdots (2.7)$$

We can equivalently write (2.7) as follows:

$$y_i = \min(y_i^*, c) \ \text{ or } \ y_i = \delta_i y_i^* + (1 - \delta_i)c,$$

Where $\delta_i = I_{(y_i^* < c)}$ is the censoring indicator.

2.4 Multicollinearity

The large number of predictor variables are the most difficult to interpret the regression model. The accurate parameter estimates the ordinary least squares and depends hugely on the sample size in relation to the number of predictor variables. When $\boldsymbol{k > n}$ the matrix ($\boldsymbol{X^T X}$) is singular, then the **OLS** estimation Suffers from high variances and cannot be uniquely**.** This problem presents because the design matrix $\boldsymbol{X}$ does not have full rank. The problem often happened when the columns of $\boldsymbol{X}$ being strongly correlated so that they are collinear (linear dependent). The collinear (mullticollinearity) problem causes inflated variances in the parameter estimates. Regularization (bias- variance tradeoff method) method such as ridge and lasso are used in case of presence of collinear problem.

2.5 Lasso Method

Lasso stands for least absolute shrinkage and selection operator, introduced by Tibshirani (1996), lasso is regularization method that add a penalty function with L1-norm to the residual sum of squares, lasso estimator is defined by

$$\hat{\beta} = argmin\left\|y_i - x_i^T \beta\right\|_2^2 + \lambda \sum_{j=1}^{k} \left|\beta_j\right|$$

Where the shrinkage parameter $\boldsymbol{\lambda \geq 0}$, and $L1 - norm = \sum_{j=1}^{k} \left|\beta_j\right|$.

Lasso provides variable selection procedure that sets some parameter estimations equals to zero. Lasso estimators do not hold oracle properties. Also lasso suffers from some limitations in case of correlated regressor in grouping.

2.6 Reciprocal Lasso Method

The reciprocal lasso (rlasso) is another regularization method that adds a particular penalty function to the residual sum of squares. In rlasso method Song (2014) proposed using the reciprocal lasso with new penalized function that has the following properties:

1- Decreasing in the range $(0, \infty)$.

2- Discontinuing at zero.

3- The function converges to ∞ when the parameters go to zero.

The rlasso gives more sparse solution by setting more parameter estimators equal to zero. The rlasso shows the oracle property. Also, the rlasso path solution for the parameter estimations are efficiently computed. The rlasso method produced better parsimonious model compared to lasso method with less prediction error. Obviously from minimization problem (1.2), large parameter estimate get small penalty, but small parameter estimate get infinite penalty.

The following figure (1) illustrates the lasso and reciprocal lasso functions behavior. (Alhamzawi, 2020).

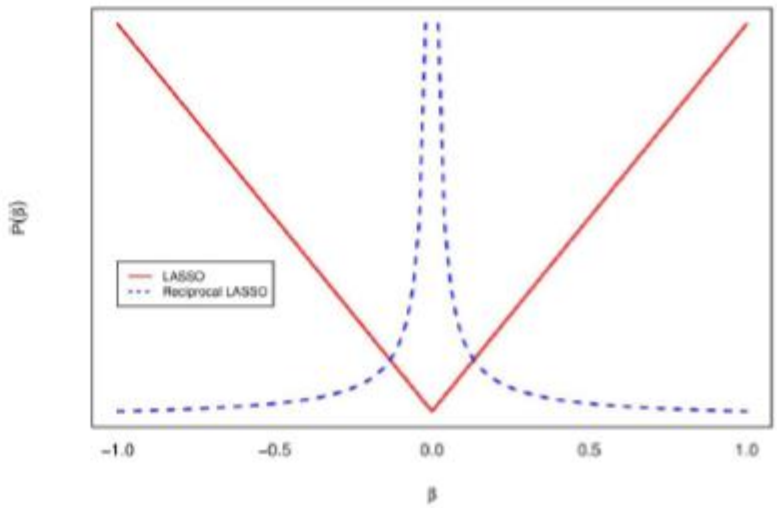

Figure (2.1) lasso and reciprocal lasso functions

2.7 Variable Selection Procedure

The large number of the covariates in the regression model yields a complex model in sense of interpretability and prediction accuracy. There are some practical cases when theoretical conditions or other assumptions determine the covariates to be included in the regression model. Such these cases the variable selection problem does not appear. Nevertheless in some cases the selecting variables for a regression model becomes an important procedure especially when there is no clear-out theory. Parsimonious model that implies the variable selection procedure (Sparsity) set some parameters estimates equal to zero. The variable selection procedure excluding the irrelevant covariates that do not impacts on the response variables by setting the para meters estimators are exactly equal to zero and including the relevant covariates that impact on the response variable. There are many selection procedures; such as, lasso-types, backward elimination, forward selection, stepwise selection, and all possible subset selection. The variable selection procedure and the problem of the function form specification for the regression model are linked to each other. One can see Miller (2002), Draper and Smith (1998), Hastie et al. (2009), Clarke et al. (2009), James et al. (2013), and Breaux (1967) for more details and information.

2.8. **Bayesian Reciprocal Lasso Right Censored Regression**

In many practical situations, the researchers are interested in recording only the results of dependent variable that values less than a constant c (upper constant). This type of variable called limited dependent variable. So, the limited dependent variable that boundary from above named censoring from above (right censored data). The upper constraint model can be defined as the following structure. In this section, Bayesian Reciprocal Lasso Right Censored Regression (BRLRCR) has been developed by employing the scale mature for the prior distribution of interested parameter (β) that was proposed by Mallick et al. (2020) to discover the effects of the new scale mixture on the reciprocal lasso penalty function from the variable selection point of view. The appropriate prior distribution that copes with the penalty function in (1.2) is the inverse double exponential (Laplace) distribution,

$$\pi(\beta) = \prod_{j=1}^{k} \frac{\lambda}{2\beta_j^2} e^{-\frac{\lambda}{|\beta_j|}} \quad I(\beta_j \neq 0) \cdots\cdots\cdots (2.8)$$

See Mallick et al.(2020) for more details. The right censored regression model that defined in (2.7) can be rewritten as follows,

$$y_i = \begin{cases} x_i'\beta + u_i & if \quad x_i'\beta + u_i < c \\ c & if \quad x_i'\beta + u_i \geq c \end{cases} \qquad \cdots\cdots\cdots (2.9)$$

Mallick et al (2020) mentioned that the Bayesian reciprocal lasso is efficient from the computation algorithms point of views which provides efficient convergence for the posterior distribution of the interested parameters. See Song (2014) and Shine et al. (2018) for more details. We will employ the scale mixture of truncated normal (SMTN) that was proposed by Mallick et la. (2020) in right censored regression, as well as the scale mixture of double pareto (SMDP). Following Al. Athari (2011), Mallick et al. (2020), Alhamzawi and Mallick (2021) the prior distribution for the interested parameter of right censored regression model in (2.9) can be explain by the following definition and proposition :

Definition (1): Suppose that the random variable β takes the following pdf:

$$g(\beta) = \frac{\theta\lambda^{\theta}}{2\beta^{\theta+1}} I\{|\beta| \geq \lambda\} \qquad \cdots (2.10)$$

Where $\lambda > 0$ is the scale parameter, and $\theta > 0$ is the shape parameter. Then pdf (2.10) is called double Pareto (type I).

Definition (2): The Generalized Double Pareto (GDP) random variable X can be defined by the following p.d.f:

$$f(\beta) = \frac{1}{2\xi}\left[1 + \frac{(|\beta| - \mu)}{\alpha\xi}\right]^{(-\alpha\ -1)} \quad ; \ |X| \geq \mu \qquad \cdots\cdots (2.11)$$

Where $\mu \in (-\infty, \infty)$ the locator parameter, $\alpha \in (0, \infty)$ the scale parameter, and $\xi \in (0, \infty)$ is the shape parameter. Scale Mixture of Truncated Normal (SMTN) formulation proposed by Mallick et al. (2020) which is state that the marginal distribution of β takes inverse Laplace with parameter (λ) if:

$\beta \sim N(0, \tau), \tau \sim \exp(\varsigma^2/2),\ \varsigma \sim \exp(\eta)$, and $\eta \sim Inverse\ Gamma\ (2, \lambda)$.

2.9. BRLRCR Hierarchical Priors Distributions Models

Referring to the formula (2.8), the structure model (2.9), and with some modification for the above proposition and based on the work of Park and Casella (2008), we propose the following hierarchical prior model:

$$y_i = \begin{cases} x_i'\beta + u_i & if \quad x_i'\beta + u_i < c \\ c & if \quad x_i'\beta + u_i \geq c \end{cases} ,$$

$$y_i^*|x_i'\beta,\sigma^2 \sim N(x_i'\beta, \sigma^2 I_n); \quad i = 1,2, \dots . n$$

$$y^* = X_i'\beta + e_i,$$

$$\beta|\sigma^2, \tau \sim \prod_{j=1}^{p} N(0, \sigma^2\, \tau^2) ,$$

$$\tau_1^2, \dots, \tau_p^2 \sim \prod_{j=1}^{p} \frac{\delta^2}{2} e^{-\delta^2\tau_j^2/2} d\,\tau_j^2,\ \ \tau_1^2, \dots, \tau_p^2 > 0, \qquad (2.12)$$

$$\delta^2|\eta \sim Gamma\ (k, \eta),$$

$$\eta|\lambda \sim Inverse\ Gamma\ (2, \lambda),$$

$$\sigma^2 \sim \pi(\sigma^2) \propto \frac{1}{\sigma^2}$$

2.9.1. The Gibbs Sampler for BRLRCR Model

Now we can implement the hierarchical model (2.12) with a Gibbs sampler algorithm. The Gibbs sampler algorithm is a Markov Chain Monte Carlo (**MCMC**) algorithm that generates samples from the conditional distribution of a specific parameter given all other parameters. The hierarchical model (2.12) constructed in such a way that we can formulate the full conditional distributions which provides easy simulation. Now we can write the full joint density as follows:

$$f(y^*|\,\beta,\sigma^2)\,\pi(\sigma^2)\prod_{j=1}^{p}\pi(\beta_j|\tau_j^2,\sigma^2)\,\pi\left(\tau_j^2\right)\pi(\delta^2)\,\pi(\eta)=$$

$$\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}(y^*-X\beta)'(y^*-X\beta)}\,\frac{1}{\sigma^2}\prod_{j=1}^{p}\frac{1}{(2\sigma^2\tau_j^2)^{1/2}}\,e^{-\frac{\beta_j^2}{2\sigma^2\tau_j^2}}\,\frac{\delta^2}{2}\,e^{-\delta^2\tau_j^2/2}$$

$$\frac{1}{\eta^k}\,(\delta^2)^{k-1}e^{-\frac{\delta^2}{\eta}}\,\frac{\lambda^2}{\Gamma 2}(\eta)^{-2-1}e^{-\frac{\lambda}{\eta}}\,. \qquad \ldots (2.13)$$

Based on the hierarchical model (2.13) and the full joint density (2.12) it is easy to sample $y^*,\beta,\sigma^2,\tau^2,\delta^2,\eta,\lambda$. The full conditional posterior distributions are as follows:

1. The full conditional distribution of y^* is :

$$y_i^*|y_i,\beta\sim N_n(X_i'\beta,\sigma^2 I_n) \qquad \ldots (2.14)$$

2. The full conditional distribution of β_j is :

$$\pi(\beta/y_i^*,X,\tau^2,\sigma^2)\propto\pi(y_i^*/X,\beta,\sigma^2).\pi(\beta/\tau^2)$$

$$\propto e^{-\frac{1}{2\sigma^2}(y^*-X\beta)'(y^*-X\beta)}.\,e^{-\frac{1}{2\sigma^2}\beta' D_\tau^{-1}\beta}$$

Where $D_\tau = diag\left(\tau_1^2, \dots, \tau_p^2\right)$,

$$= \exp\left\{-\frac{1}{2\sigma^2}[(\beta'(X'X)\beta - 2y^*X\beta + y^{*\prime}y^*) + \beta' D_\tau^{-1}\beta]\right\}$$

$$= \exp\left\{-\frac{1}{2\sigma^2}[\beta'(X'X + D_\tau^{-1})\beta - 2y^*X\beta + y^{*\prime}y^*)]\right\}$$

Now let $C = X'X + D_\tau^{-1}$, then we have

$$= \exp\left\{-\frac{1}{2\sigma^2}[\beta' C\beta - 2y^*X\beta + y^{*\prime}y^*)]\right\}$$

$$= \exp\left\{-\frac{1}{2\sigma^2}(\beta - C^{-1}X'y^*)'C(\beta - C^{-1}X'y^*)\right\} \quad \dots(2.15)$$

Which is the multivariate normal with mean $C^{-1}X'y^*$and variance $\sigma^2 C^{-1}$.

3. The full conditional posterior distribution of σ^2 is:

$$\pi(\sigma^2/y_i^*, X, \beta)$$

$$\propto \pi(y_i^*/X, \beta, \sigma^2).\pi(\beta/\sigma^2).\pi(\sigma^2)$$

$$\propto -\frac{1}{(\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2}(y^*-X\beta)'(y^*-X\beta)} . \frac{1}{(\sigma^2\tau^2)^{\frac{p}{2}}} e^{-\frac{1}{2\sigma^2}\beta' D_\tau^{-1}\beta} . \frac{1}{\sigma^2}$$

$$= (\sigma^2)^{-(\frac{n}{2}+\frac{p}{2})-1} \exp\left\{-\frac{1}{2\sigma^2}[(y^* - X\beta)'(y^* - X\beta) + \beta' D_\tau^{-1}\beta]\right\}$$

$$= (\sigma^2)^{-(\frac{n+p}{2})-1} \exp\left\{-\frac{1}{2\sigma^2}[(y^* - X\beta)'(y^* - X\beta) + \beta' D_\tau^{-1}\beta]\right\} \quad \dots(2.16)$$

which is the invers gamma with shape parameter $\frac{n+p}{2} - 1$ and scale parameter $(y^* - X\beta)'(y^* - X\beta)/2 + \beta' D_\tau^{-1}\beta/2$.

4. The part of (2.13) including $\tau^2$is:

$$\pi\left(\tau_j^2/\delta^2,\eta\right) \propto \pi\left(\beta/\tau_j^2\right).\pi(\delta^2/\eta)$$

$$\propto \left(\frac{1}{\tau_j^2}\right)^{\frac{1}{2}} e^{-\frac{1}{2\sigma^2}\frac{\beta_j^2}{\tau_j^2}} e^{-\frac{\delta^2\tau_j^2}{2}}$$

$$\propto \left(\tau_j^2\right)^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\left(\frac{\beta_j^2/\sigma^2}{\tau_j^2} + \delta^2\tau_j^2\right)\right] \quad \cdots\cdots (2.17)$$

The last formula can be treated by using the invers Gaussian distribution and its invers form. Suppose that the invers Gaussian is:

$$f(X;a,b) = \left(\frac{b}{2\pi X^3}\right)^{\frac{1}{2}} \exp\left[\frac{-b(X-a)^2}{2a^2X}\right]$$

The invers of $f(X;\,.)$ is $f'(\,.)$ defined by

$$f'(y\,;a,b) = \left(\frac{b}{2\pi y}\right)^{\frac{1}{2}} \exp\left[\frac{-b(1-ay)^2}{2a^2y}\right]$$

Where $= X^{-1}$, then a formula (2.17) Can be rewrite as the reciprocal inverse Gaussian distribution as follows:

$$\propto \left(\frac{1}{\tau_j^2}\right)^{\frac{-3}{2}} \exp\left[-\frac{1}{2}\left(\frac{\beta_j^2}{\sigma^2\,\tau_j^2} + \frac{\delta^2}{1/\tau_j^2}\right)\right]$$

$$\propto \left(\frac{1}{\tau_j^2}\right)^{\frac{-3}{2}} \exp\left[-\frac{\beta_j^2\left(\left(1/\tau_j^2\right)-\sqrt{\delta^2\sigma^2/\beta^2}\right)^2}{2\sigma^2\left(1/\tau_j^2\right)}\right] \ldots (2.18)$$

So, we can say that $\left(\frac{1}{\tau_j^2}\right) \sim invers\ Gaussain\ with\ mean\ \sqrt{\frac{\delta^2\,\sigma^2}{\beta_j^2}}$ and shape parameter $\delta^2 = b$.

5. By following Park and Casella (2008), we assigned the gamma prior for δ^2. Then full conditional posterior distribution of δ^2 is defined as in follows:

$$(\delta^2)^{k-1} e^{-\frac{\delta^2}{\eta}} \left(\prod_{j=1}^{p} \frac{\delta^2}{2} e^{-\delta^2 \tau_j^2/2}\right)$$

$$=(\delta^2)^{p+k-1} \exp[-\delta^2(\tfrac{1}{2}\textstyle\sum_{j=1}^{p} \tau_j^2 + \tfrac{1}{\eta})] \quad \ldots (2.19)$$

This is also a gamma distribution with shape parameter $p + k$ and rate parameter $\frac{1}{2}\sum_{j=1}^{p} \tau_j^2 + \frac{1}{\eta}$.

6. The full conditional posterior distribution of η is defined as follows:

$$\pi(\eta/\delta^2, \lambda) \propto \pi(\delta^2/\eta).\pi(\eta|\lambda)$$

$$\propto \frac{1}{\eta^k} (\delta^2)^{k-1} e^{-\frac{\delta^2}{\eta}} \frac{\lambda^2}{\Gamma(2)} \eta^{-2-1} e^{-\lambda/\eta}$$

$$\propto \eta^{-(k+2)-1} e^{-\frac{1}{\eta}(\delta^2+\lambda)} \quad \ldots(2.20)$$

Recall the invers gamma distribution, consequently we can conclude that η is distributed according to inverse gamma with shape parameter (k+2) and scale parameter $(\delta^2 + \lambda)$.

2.9.2. BRLRCR Computation Algorithm

The following parameters and variables have sampled based on Gibbs sampling algorithm:

1- Sampling $\boldsymbol{y}^*$: In this step we generate the latent variable y^* from truncated normal distribution with mean $(\boldsymbol{x_i}^{\boldsymbol{T}}\boldsymbol{\beta})$ and variance $(\boldsymbol{\sigma^2 I_n})$**.**

2- Sampling $\boldsymbol{\beta}$: In this step we generate $\boldsymbol{\beta}$ from normal distribution $\boldsymbol{C^{-1}X'y^*}$and variance $\boldsymbol{\sigma^2 C^{-1}}$**.**

3- Sampling $\boldsymbol{\sigma^2}$: In this step we generate $\boldsymbol{\sigma^2}$ from invers gamma with shape parameter $\boldsymbol{\frac{n+p}{2}-1}$ and scale parameter $(\boldsymbol{y}^*-\boldsymbol{X\beta})'(\boldsymbol{y}^*-\boldsymbol{X\beta})/\boldsymbol{2}+\boldsymbol{\beta' D_\tau^{-1}\beta/2}$**.**

4- Sampling $\boldsymbol{\tau^2}$ **:** In this step we generate $\boldsymbol{\tau^2}$ from inverse Gaussian with mean $\sqrt{\frac{\boldsymbol{\delta^2\,\sigma^2}}{\boldsymbol{\beta_j^2}}}$ and shape parameter $\boldsymbol{\delta^2}$.

5- Sampling δ^2 **:** In this step we generate $\boldsymbol{\delta^2}$ from a gamma distribution with shape parameter $\boldsymbol{p+k}$ and rate parameter $\boldsymbol{\frac{1}{2}\sum_{j=1}^{p}\tau_j^2+\frac{1}{\eta}}$.

6- Sampling $\boldsymbol{\eta}$: In this step we generate $\boldsymbol{\eta}$ inverse gamma with shape parameter **(k+2)** and scale parameter $(\boldsymbol{\delta^2+\lambda})$**.**

2.10. Extension on BRLRCR Models

In this section we employed the proposition and the hierarchical model that developed by Mallick et al. (2020) in the Bayesian reciprocal Laplace right censored regression model. Scale Mixture of Double Pareto (SMDP) formulation proposed by Mallick et al. (2020) which is state that if the prior distribution of β is β~Double inverse pareto (η, 1) and η~ inverse gamma (2, λ), then β follows inverse Laplace (λ).

$$y_i = \begin{cases} x_i'\beta + u_i & if \quad x_i'\beta + u_i < c \\ c & if \quad x_i'\beta + u_i \geq c \end{cases} ,$$

$$y_i^* | x_i'\beta, \sigma^2 \sim N(x_i'\beta, \sigma^2 I_n); \quad i = 1,2, \dots . n$$

$$\beta|\eta \sim \prod_{j=1}^{p} \frac{1}{uniform\left(-\frac{1}{\eta_j}, \frac{1}{\eta_j}\right)}$$

$$\eta|\lambda \sim \prod_{j=1}^{p} Gamma(2, \lambda)$$

$$\sigma^2 \sim \pi(\sigma^2) \tag{2.21}$$

2.10.1 Gibbs Sampler Computations

Connection with Bayesian lasso and reciprocal lasso the full conditional posterior distribution for the parameters in hierarchical prior model (2.21) of the (SMDP) Bayesian reciprocal Laplace right censored regression model are as follows Mallick et al. (2020):

$$y_i^*/y_i, \beta \sim N_n(X_i'\beta, \sigma^2 I_n)$$

$$\beta|y^*, X, u, \lambda, \sigma^2 \sim N_p\left(\hat{\beta}_{mle}, \sigma^2\left(\acute{X}X\right)^{-1}\right)\prod_{j=1}^{p} I\left\{\left|\beta_j\right| > \frac{1}{\sigma^2 u_i}\right\},$$

$$u|y^*, X, \beta, \lambda, \sigma^2 \sim \prod_{j=1}^{p} exp(\lambda)\, I\left\{u_j > \frac{1}{\sigma^2\left|\beta_j\right|}\right\},$$

$$\sigma^2, y^*, X, \beta, u, \lambda \sim Inv.Gamma\left(\frac{n-1}{p}, \frac{1}{2}(y^* - X\beta)'(y^* - X\beta)\right)$$

$$\lambda|\beta \sim Gamma(a + 2p, b + \textstyle\sum_{j=1}^{p} \frac{1}{|\beta_j|}). \qquad (2.22)$$

Chapter Three

Simulation and Real Data Analysis

3.1 Introduction

As the number of variables (parameters) getting larger in our model, the more difficulty in evaluating and analyzing the posterior distribution. Here is where the Gibbs sample algorithm becomes quite useful. Gibbs sample is a special case of MCMC technique and hence we can use the results of MCMC algorithm to make inference about the model and its parameters. We conducted some simulation examples and real data analysis to test the efficiency of the Gibbs sampler algorithm that mention in the theoretical chapter. Comparison is the main goal with some other regularization methods. We run the algorithm **13000** iterations with **3000** iterations have burned-in for reaching the stationary of posterior distribution.

3.2 Simulation Scenarios

In this subsection we are trying to simulate some scenarios for checking the efficiency of the proposed posterior distributions by using the Gibbs sampler algorithm. For the comparing purpose we have used the R.C. (right censored) regression model, Bayesian lasso R.C. regression, SMTN-reciprocal Lasso R.C. regression, and SMDP-reciprocal Lasso R.C. regression. As well as, we employed three different values of standard deviations (to guarantee the unimodel posterior distribution) for the regression models. Also, the criterions of Median of Mean Absolute Error **(MMAE)** and the Standard Error **(S.D.)** have used for assessing the quality of the estimated model.

$$\boldsymbol{MMAE} = \boldsymbol{Median}\ (\boldsymbol{mean}\ |\hat{\boldsymbol{f}} - \boldsymbol{f}|)$$

$\hat{\boldsymbol{f}} = \boldsymbol{x}_i^T \boldsymbol{\beta}^{predicted}$ and $\boldsymbol{f} = \boldsymbol{x}_i^T \boldsymbol{\beta}^{true}$.

3.2.1 Simulation Scenario One

In this simulation example, we introduced the scenario of the data generating process as following: $Y = X\beta + u$ where $X \sim N(0,1)$, and $u \sim N(0, \sigma^2)$. The correlation between the X_i and X_j is defined by $\rho^{|i-j|}$. Since the predictor variables have $\sigma^2 = 1$, then the design matrix of the predictor variables follows the multivariate normal distribution with mean equals to zero and variance-covariance matrix equal to $\boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}_{ij} = \rho^{|i-j|}$. The regression model that describe the true relationship between the response variable and predictor variables is defined as follows:

$$\boldsymbol{y_i = 5x_{i1} + u_i \quad i = 1, 2, \dots \dots \dots \dots . n = 200},$$

Where $\boldsymbol{\beta} = (\mathbf{5}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0})^{\boldsymbol{T}}$.

The following table shows the values of the **MMAD** and its **S.E.** criterions for different sample sizes and different values of the error variances with. We use $\rho = 0.5$, and $\sigma^2 \in \{1,3,5\}$ to test the effect of the noise in the data.

Table (1). MMAD and S.E. values for simulation one

Sample size		The methods			
----------------	σ^2	R.C. regression model	Bayesian lasso R.C. regression	SMTN-reciprocal Lasso R.C. regression	SMDP-reciprocal Lasso R.C. regression
n=25	1	0.824(0.527)	0.718(0.594)	0.622(0.354)	**0.542(0.385)**
	3	0.864 (0.435)	0.874 (0.644)	0.764 (0.673)	**0.484 (0.264)**
	5	1.374(0.612)	1.222(0. 836)	0.993(0.622)	**0.492(0.256)**
n=50	1	0.934(0.582)	0.856(0.584)	0.793(0.487)	**0.442(0.201)**
	3	0.891 (0.687)	0.944 (0.537)	0.675 (0.464)	**0.464 (0.281)**
	5	0.902(0.542)	0.984(0.582)	0.882(0.621)	**0.392(0.381)**
n=100	1	0.752(0.422)	0.819(0.572)	0.784(0.543)	**0.458(0.256)**
	3	0.766 (0.483)	0.764(0.428)	0.643 (0.534)	**0.458 (0.238)**
	5	0.729(0.482)	0.718(0.557)	0.681(0.639)	**0.512(0.386)**
n=150	1	0.911(0.482)	0.824 (0.482)	0.735 (0.474)	**0.389 (0.159)**
	3	0.825 (0.567)	0.733 (0.441)	0.513 (0.397)	**0.457 (0.267)**
	5	0.661 (0.474)	0.630 (0.377)	0.743 (0.390)	**0.387 (0.169)**
n=200	1	0.732 (0.487)	0.826 (0.576)	0.858 (0.438)	**0.437 (0.482)**
	3	0.629 (0.514)	0.728 (0.452)	0.834 (0.442)	**0.501 (0.373)**
	5	0.694 (0.437)	0.785 (0.552)	0.866 (0.462)	**0.411 (0.295)**
n=250	1	0.735 (0.414)	0.835 (0.553)	0.725(0.371)	**0.543 (0.412)**
	3	0.817 (0.514)	0.749 (0.398)	0.747(0.519)	**0.512 (0.581)**
	5	0.834 (05.38)	0.978 (0.610)	0.726 (0.431)	**0.415 (0.372)**

From table (1), values of **MMAD** and its **S.E.** that calculated based on the proposed regression models **(**SMTN-reciprocal Lasso R.C. regression**)** and **(**SMDP-reciprocal Lasso R.C. regression**)** are less than the values of other different methods **(**R.C. regression model**)** and **(**Bayesian lasso R.C. regression**)**. Therefore, the proposed models are comparable in terms of estimation accuracy and variable selection point of views through all the values of error distribution and the sample sizes.

Trace plot is a convergence diagnose tool, usually is using to indicate if the MCMC samples from the posterior distribution of parameter convergence to stationary distribution. The following figure (1) shows the trace plots which illustrate no flat bits and that MCMC algorithm suffer no slow mixing.

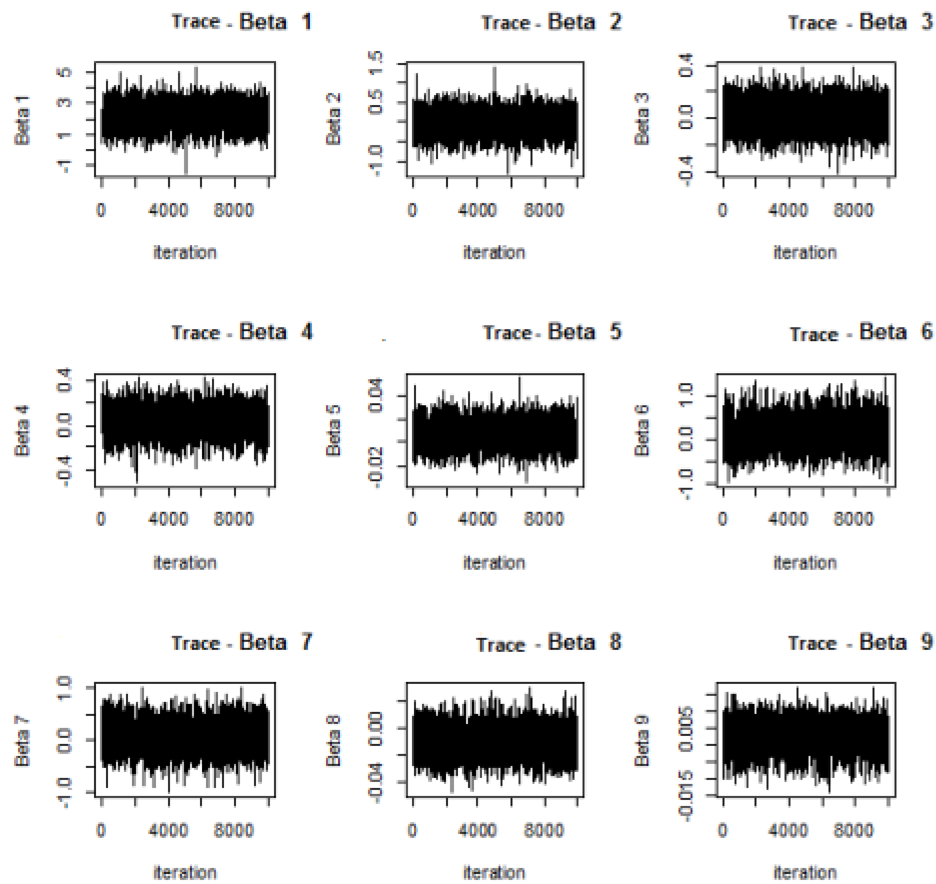

Figure 1. Trace plots of the parameter estimates $\beta_1 - \beta_9$.

Figure (2) shows the distributions of the parameter estimates $\beta_1 - \beta_9$ and it is very clear that the distribution of the parameter follows the normal distribution for all parameter estimates.

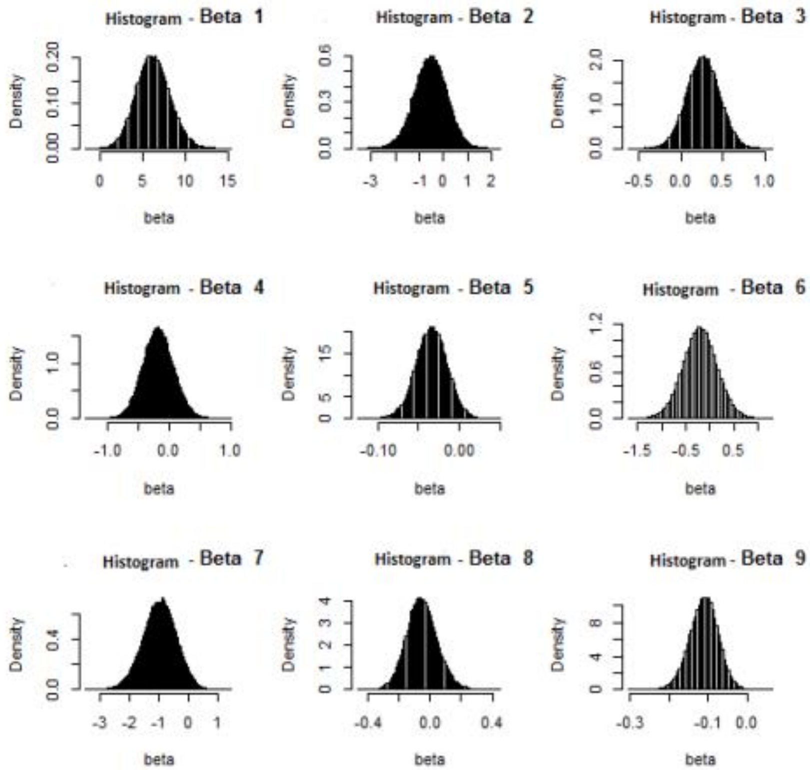

Figure (2) Histograms of parameter estimates $\beta_1 - \beta_9$.

3.2.2. Simulation Scenario Two

In this example and based on the same process in the sample one, we supposed the following dense true parameter vector, $\boldsymbol{\beta} = (\mathbf{0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85})^{\boldsymbol{T}}$ and then the true regression model is defined as follows:

$$\boldsymbol{y_i = 0.85x_{i1} + 0.85x_{i2} + 0.85x_{i3} + 0.85x_{i4} + 0.85x_{i5} + 0.85x_{i6} + 0.85x_{i7} + 0.85x_{i8} + 0.85x_{i9} + u_i \quad ;\ i = 1, 2, \dots \dots \dots \dots . n = 200}$$

The following table shows the values of the **MMAD** and its **S.E.** criterions.

Table (2) MMAD and S.E. values for simulation example Two

Sample size		The methods			
----------------	σ^2	R.C. regression model	Bayesian lasso R.C. regression	SMTN-reciprocal Lasso R.C. regression	SMDP-reciprocal Lasso R.C. regression
n=25	1	0.984(0.634)	0.816(0.621)	0.872(0.622)	**0.454(0.264)**
	3	0.754 (0.762)	0.758 (0.875)	0.824 (0.572)	**0.464 (0.264)**
	5	0.737(0.565)	0.862(0. 567)	0.806(0.661)	**0.401(0.308)**
n=50	1	0.762(0.351)	0.933(0.657)	0.831(0.530)	**0.536(0.364)**
	3	0.725 (0.506)	0.928 (0.634)	0.756 (0.368)	**0.585 (0.358)**
	5	0.831(0.604)	0.986(0.637)	0.952(0.764)	**0.465(0.288)**
n=100	1	0.864(0.534)	0.769(0.437)	0.837(0.375)	**0.504(0.359)**
	3	0.675 (0.487)	0.834(0.506)	0.935 (0.346)	**0.537 (0.325)**
	5	0.839(0.534)	0.864(0.638)	0.738(0.428)	**0.468(0.283)**
n=150	1	0.837(0.535)	0.953 (0.531)	0.836 (0.528)	**0.524 (0.385)**
	3	0.626 (0.635)	0.768 (0.375)	0.739 (0.425)	**0.573 (0.345)**
	5	0.768 (0.539)	0.752 (0.437)	0.734 (0.418)	**0.454 (0.209)**
n=200	1	0.853 (0.567)	0.952 (0.634)	0.769 (0.548)	**0.573 (0.306)**
	3	0.961 (0.428)	0.674 (0.534)	0.542 (0.392)	**0.561 (0.286)**
	5	0.724 (0.579)	0.865 (0.635)	0.767 (0.549)	**0.468 (0.372)**
n=250	1	0.926 (0.526)	0.758 (0.457)	0.824(0.586)	**0.589 (0.242)**
	3	0.864 (0.452)	0.861 (0.537)	0.735(0.426)	**0.453 (0.276)**
	5	0.736 (0.573)	0.647 (0.522)	0.655 (0.514)	**0.321 (0.216)**

From table (2), values of **MMAD** and its **S.E.** that calculated based on the proposed regression models **(**SMTN-reciprocal Lasso R.C. regression**)** and **(**SMDP-reciprocal Lasso R.C. regression**)** are less than the values of other different methods **(**R.C. regression model**)** and **(**Bayesian lasso R.C. regression**)**. Therefore, the proposed models are comparable in terms of estimation accuracy and variable selection point of views through all the values of error distribution and the sample sizes.

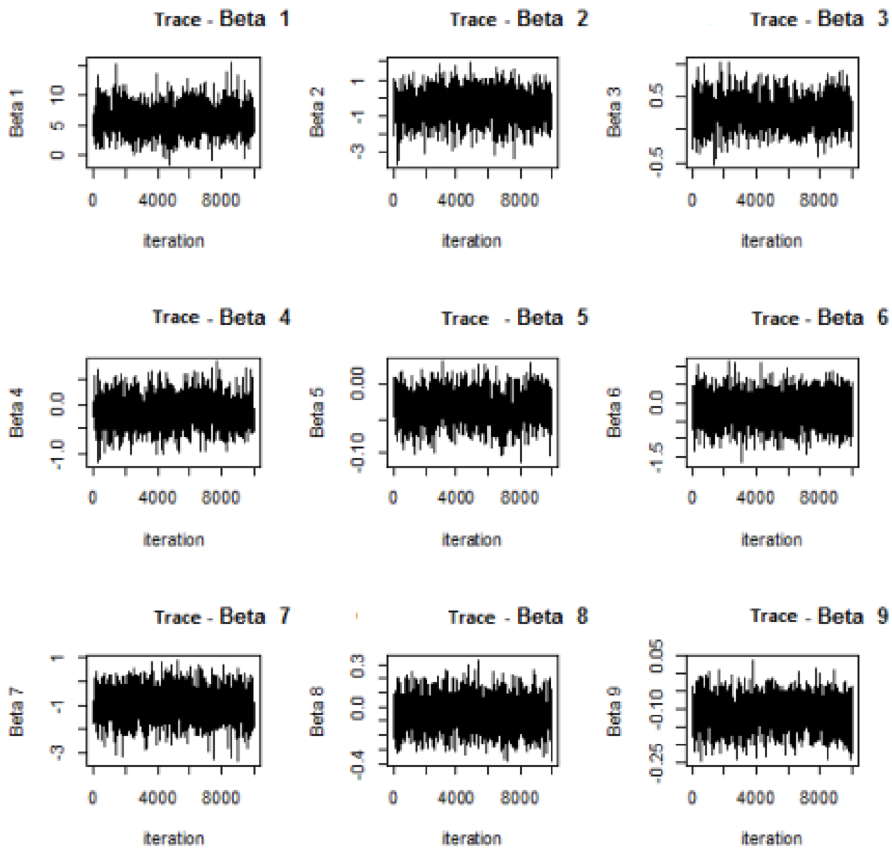

Figure 3. Trace plots of the parameter estimates $\beta_1 - \beta_9$.

The above figure (3) shows the trace plots which illustrate no flat bits and that MCMC algorithm suffer no slow mixing.

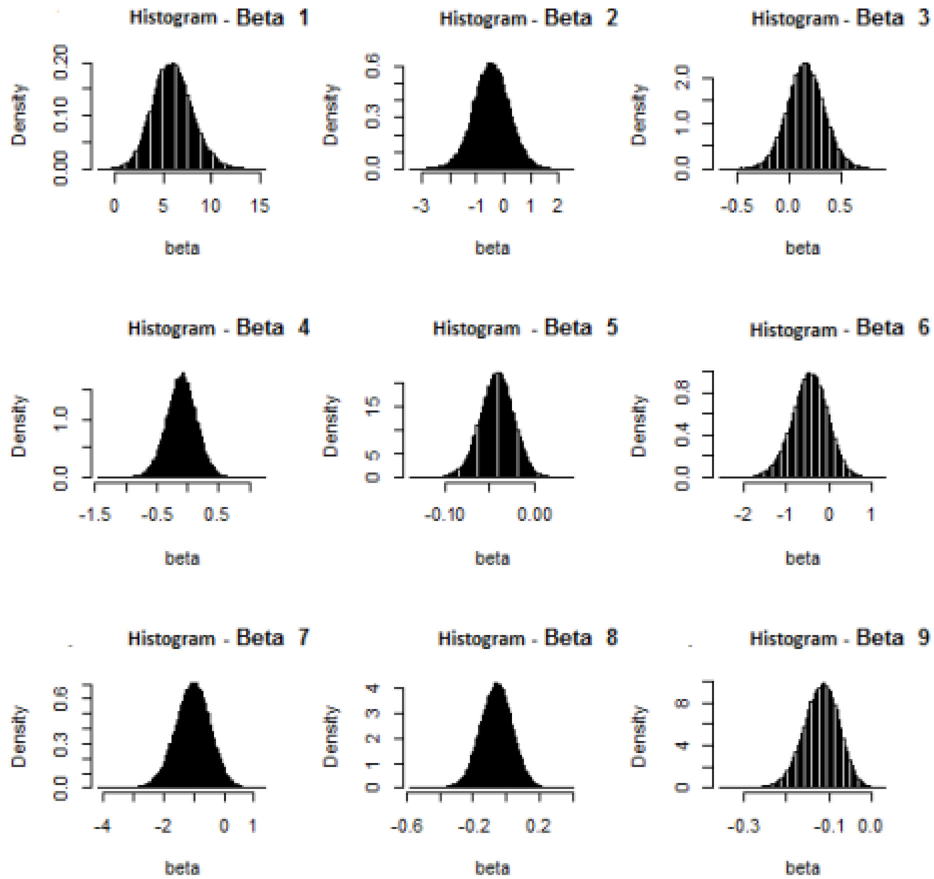

Figure (4) Histograms of parameter estimates $\beta_1 - \beta_9$.

Figure (4) shows the distributions of the parameter estimates $\beta_1 - \beta_9$ and it is very clear that the distribution of the parameter follows the normal distribution for all parameter estimates.

3.2.3 Simulation Scenario Three

In this simulation example, I supposed the sparse true vector of parameter $\boldsymbol{\beta} = (\mathbf{5}, \mathbf{0}, \mathbf{0}, \mathbf{1}, \mathbf{0}, \mathbf{2}, \mathbf{0}, \mathbf{1.5}, \mathbf{0})^{\boldsymbol{T}}$, hence the regression model is defined as follows

$$\boldsymbol{y_i = 5x_{i1} + x_{i4} + 2x_{i6} + 1.5x_{i8} + e_i \quad i = 1, 2, \ldots\ldots\ldots\ldots n = 100},$$

The explanatory variables of the above model are generated from the multivariate normal distribution. The following table shows the values of the **MMAD** and **S.E.** criterions.

Table (3). MMAD and S.E. values for simulation example Three

Sample size		The methods			
----------------	σ^2	R.C. regression model	Bayesian lasso R.C. regression	SMTN-reciprocal Lasso R.C. regression	SMDP-reciprocal Lasso R.C. regression
n=25	1	0.745(0.546)	0.647(0.585)	0.654(0.325)	**0.451(0.215)**
	3	0.745 (0.545)	0.765 (0.455)	0.635 (0.364)	**0.575 (0.185)**
	5	0.765(0.451)	0.844(0. 564)	0.784(0.397)	**0.414(0.254)**
n=50	1	0.835(0.524)	0.724(0.565)	0.694(0.536)	**0.451(0.434)**
	3	0.754 (0.476)	0.865 (0.446)	0.435 (0.465)	**0.447 (0.371)**
	5	0.804(0.554)	0.875(0.574)	0.774(0.444)	**0.386(0.171)**
n=100	1	0.823(0.424)	0.726(0.381)	0.785(0.474)	**0.458(0.274)**
	3	0.644 (0.574)	0.645(0.547)	0.514 (0.644)	**0.467 (0.257)**
	5	0.648(0.574)	0.647(0.556)	0.734(0.398)	**0.453(0.214)**
n=150	1	0.816(0.464)	.762 (0.467)0	0.742 (0.415)	**0.478 (0.288)**
	3	0.867 (0.826)	.644 (0.454)0	0.614 (0.536)	**0.576 (0.346)**
	5	0.641 (0.365)	0.440 (0.466)	0.684 (0.380)	**0.436 (0.368)**
n=200	1	0.675 (0.423)	0.754 (0.373)	0.752 (0.347)	**0.386 (0.273)**
	3	0.728 (0.352)	0.792 (0.634)	0.745 (0.454)	**0.408 (0.264)**
	5	0.762 (0.396)	0.642 (0.434)	0.787 (0.374)	**0.584 (0.287)**
n=250	1	0.865 (0.436)	0.964 (0.362)	0.785(0.568)	**0.421 (0.237)**
	3	0.954 (0.675)	0.463 (0.251)	0.657(0.349)	**0.321 (0.276)**
	5	0.785 (0.432)	0.897 (0.470)	0.545 (0.337)	**0.405 (0.263)**

From table (3), values of **MMAD** and its **S.E.** that calculated based on the proposed regression models **(**SMTN-reciprocal Lasso R.C. regression**)** and **(**SMDP-reciprocal Lasso R.C. regression**)** are less than the values of other different methods **(**R.C. regression model**)** and **(**Bayesian lasso R.C.

regression). Therefore, the proposed models are comparable in terms of estimation accuracy and variable selection point of views through all the values of error distribution and the sample sizes.

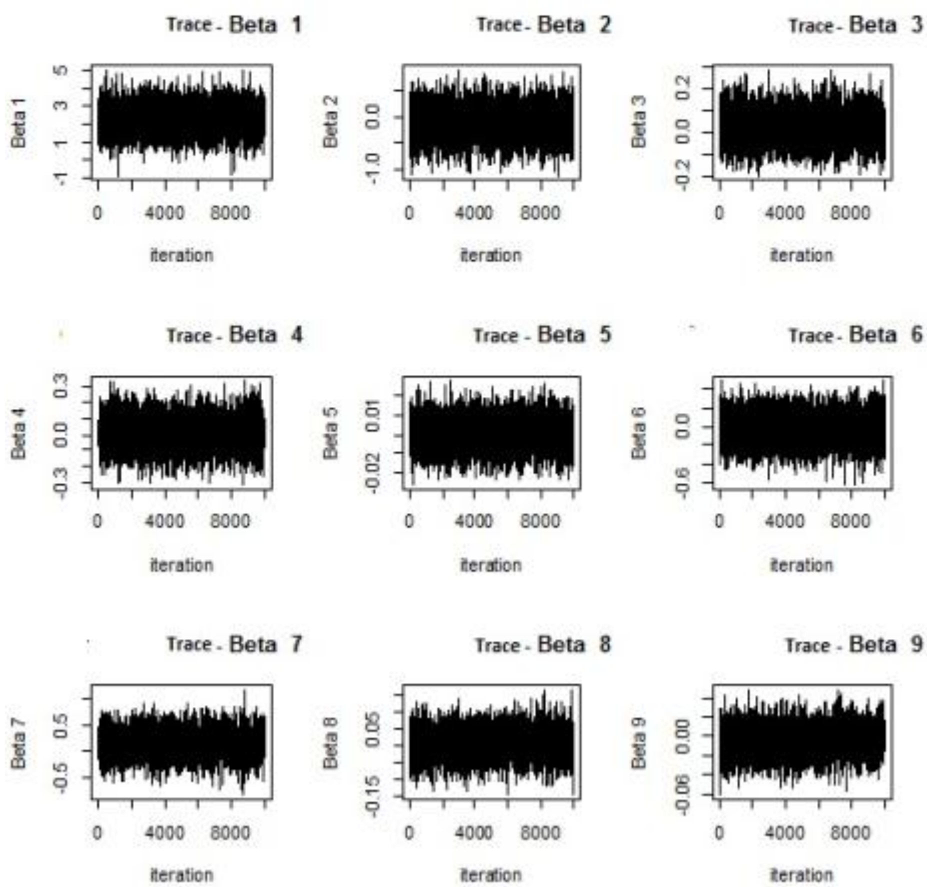

Figure 5. Trace plots of the parameter estimates $\beta_1 - \beta_9$.

The above figure (5) shows the trace plots which illustrate no flat bits and that MCMC algorithm suffer no slow mixing.

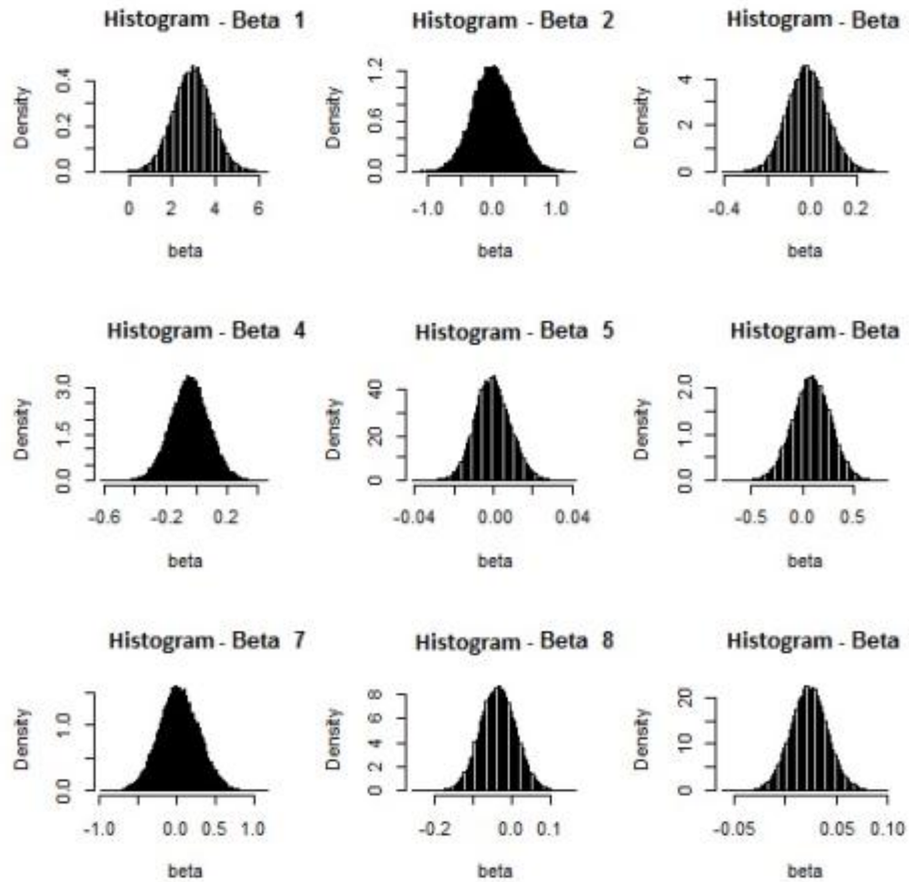

Figure (6) Histograms of parameter estimates $\beta_1 - \beta_9$.

Figure (6) shows the distributions of the parameter estimates $\beta_1 - \beta_9$ and it is very clear that the distribution of the parameter follows the normal distribution for all parameter estimates. In the next figures we illustrated the results of simulation scenario two with the true vector of dense $\beta = (0.85)_9^T$.

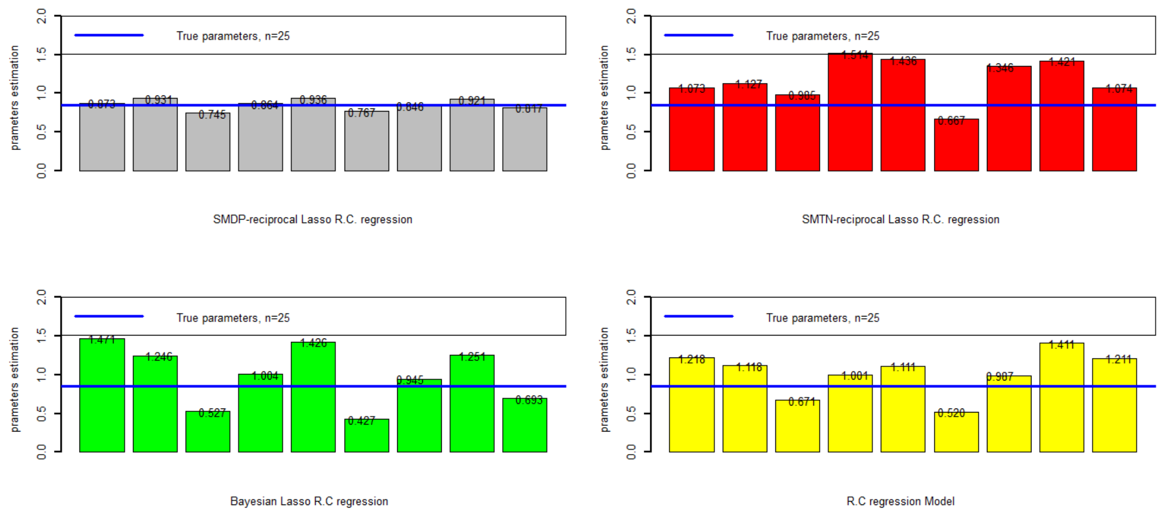

Figure (7) True vector and parameter estimates $\beta_1 - \beta_9$ with sample size=25

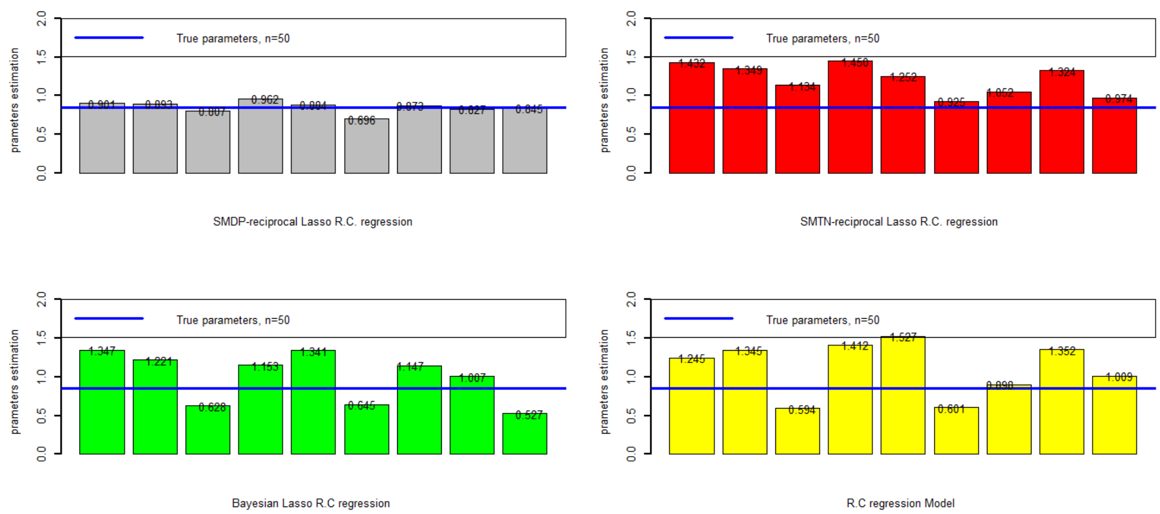

Figure (8) True vector and parameter estimates $\beta_1 - \beta_9$with sample size=50

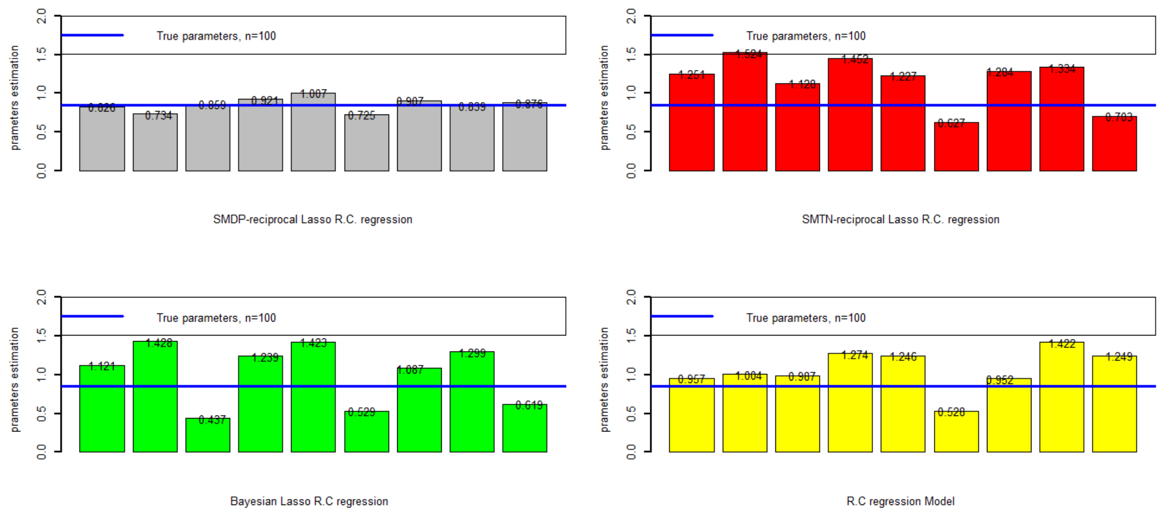

Figure (9) True vector and parameter estimates $\beta_1 - \beta_9$ with sample size=100

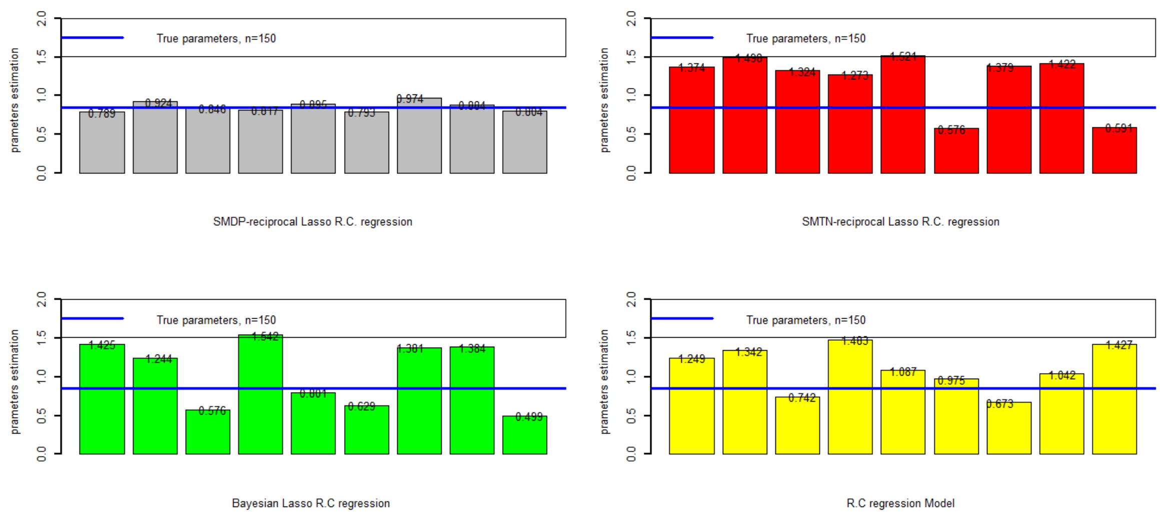

Figure (10) True vector and parameter estimates $\beta_1 - \beta_9$ with sample size=150

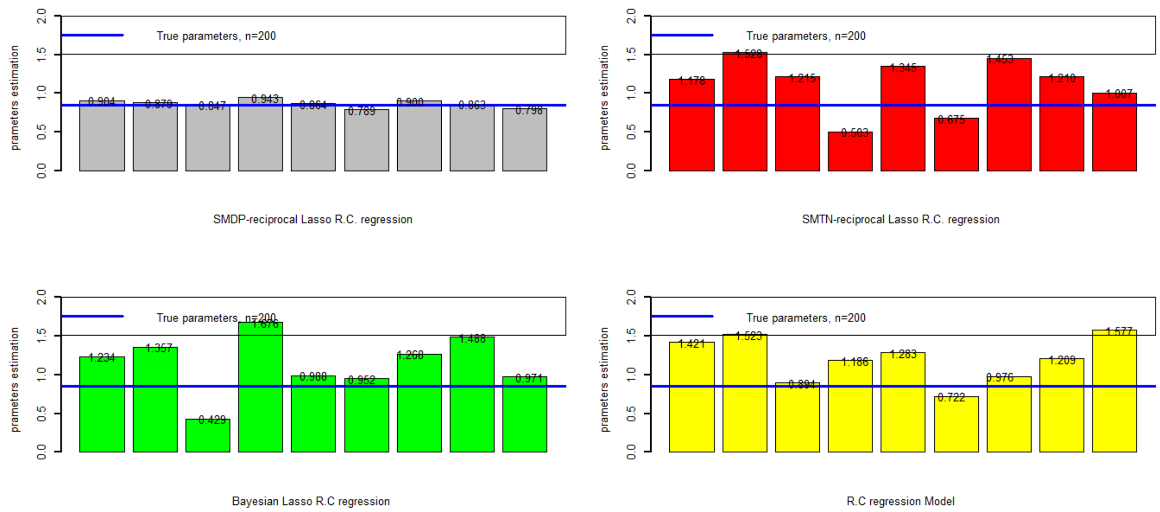

Figure (11) True vector and parameter estimates $\beta_1 - \beta_9$ with sample size=200, σ^2=1

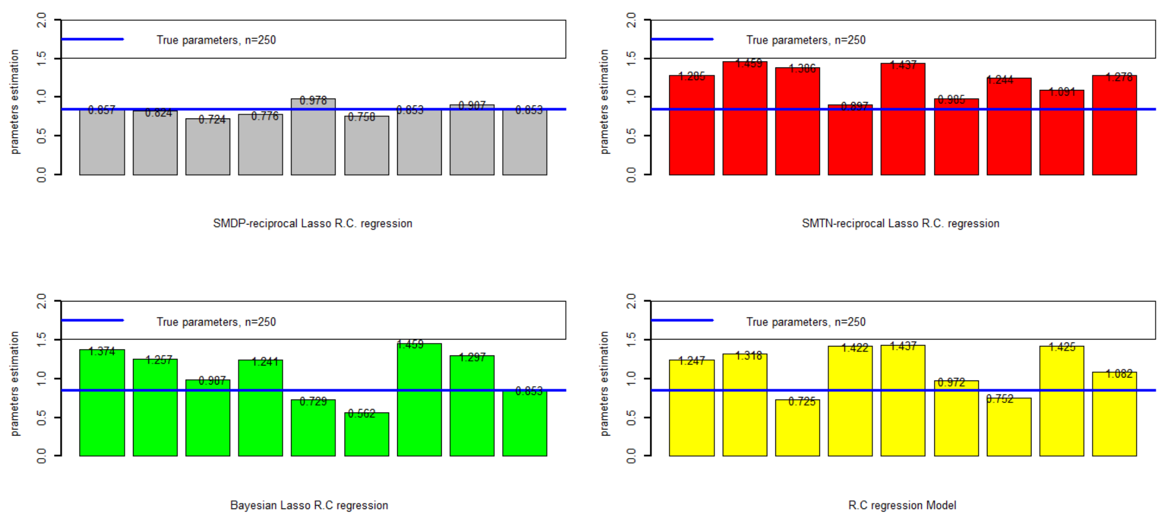

Figure (12) True vector and parameter estimates $\beta_1 - \beta_9$ with sample size=250, σ^2=1

Figures (7) to (12) represented the parameter estimates by the proposed models in addition to other two models, where the blue line (true vector) compared with the parameters estimates under different sample sizes and different estimation method. Therefore, clearly that the blue line is much closed to the parameter estimates for the proposed models (colored bars), so we can say that the proposed model (SMDP) gives the best fit and the (SMTN) model is comparable with the other methods under the different sample sizes.

3.3 Real Data analysis:

In this section we summarized the described of the real data that we collected from the central laboratory and Rafidain's Valley in the province of Babylon. To cope with the objectives of this thesis we focused on limited dependent response variable (right censored). After employing the simulation method to show the preference of our proposed method in estimating parameters and selecting variables compared to a group of previous methods. We will test the behavior of our method with real data, which also focuses on a medical phenomenon that includes the response variable that represents the normal blood sugar level within the range (80-180) for 55 patients. In this study, we focus on the normal limits of blood sugar, so the censored point is 180 and when the values are above the censored point then it will be set to 180. The independent variables represented are as follows:

X1: the patient's weight (in kilograms).

X2: the patient's age.

X3: the number of meals for the patient per day

X4: Are there genetic factors?

X5: Is the patient under psychological pressure?

X6: Does the patient have pancreatic disease?

X7: Does the patient have covid19?

X8: the patient's monthly income

X9: The number of hours of exercise per day

For comparison purpose we employed the two proposed regularization methods (SMTN-reciprocal Lasso R.C. regression and SMDP-reciprocal Lasso R.C. regression) with two other methods (R.C. regression model and Bayesian lasso R.C. regression) by using the median mean absolute deviation (MMAD) and the mean absolute error (MAE) criterion. These criteria are used to assess the prediction accuracy of the different models.

Table (4). Values of MSE and MAE with its Standard errors

	R.C. regression model	**Bayesian lasso R.C. regression**	**SMDP-reciprocal Lasso R.C. regression**	**SMTN-reciprocal Lasso R.C. regression**
MSE	0.852(0.493)	0.847(0.506)	0.848(0.495)	0.482(0.328)
MAE	0.848(0.507)	0.757(0.416)	0.712(0.392)	0.573(0.377)

Table (4) shows the values of the MSE and MAE, where the proposed models give the less values comparing with the other two methods. This result supports the simulation results and indicated the high prediction accuracy for the proposed models.

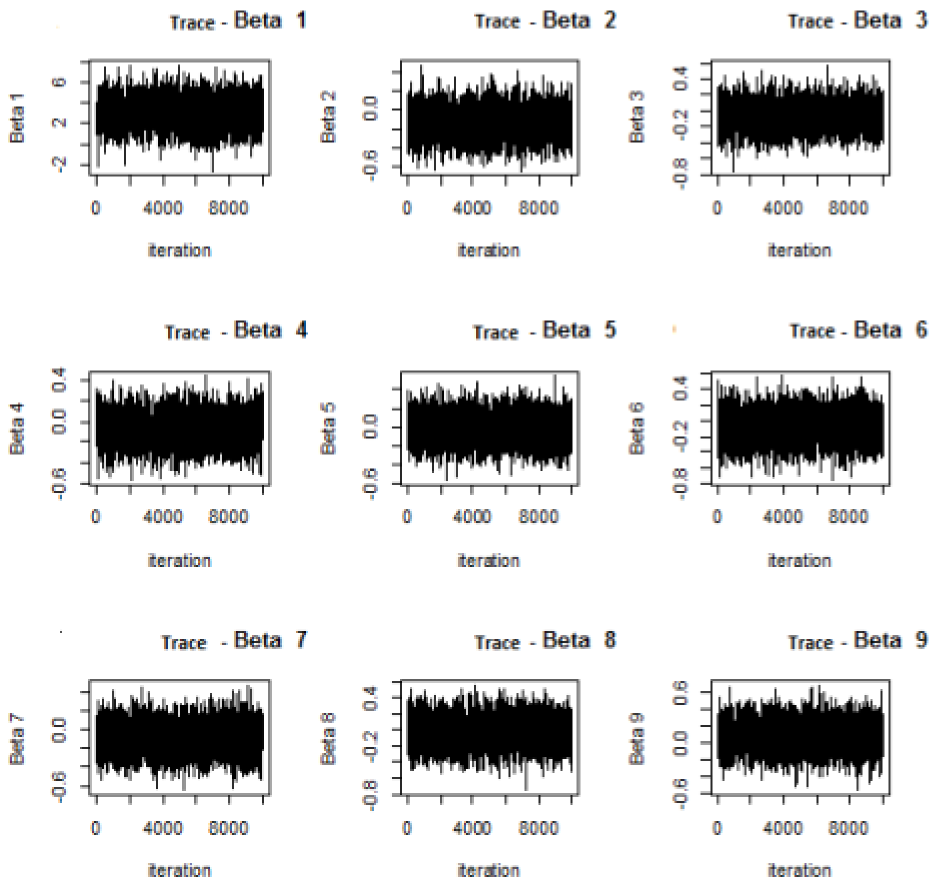

Figure (13). Trace Plot of Real Parameters

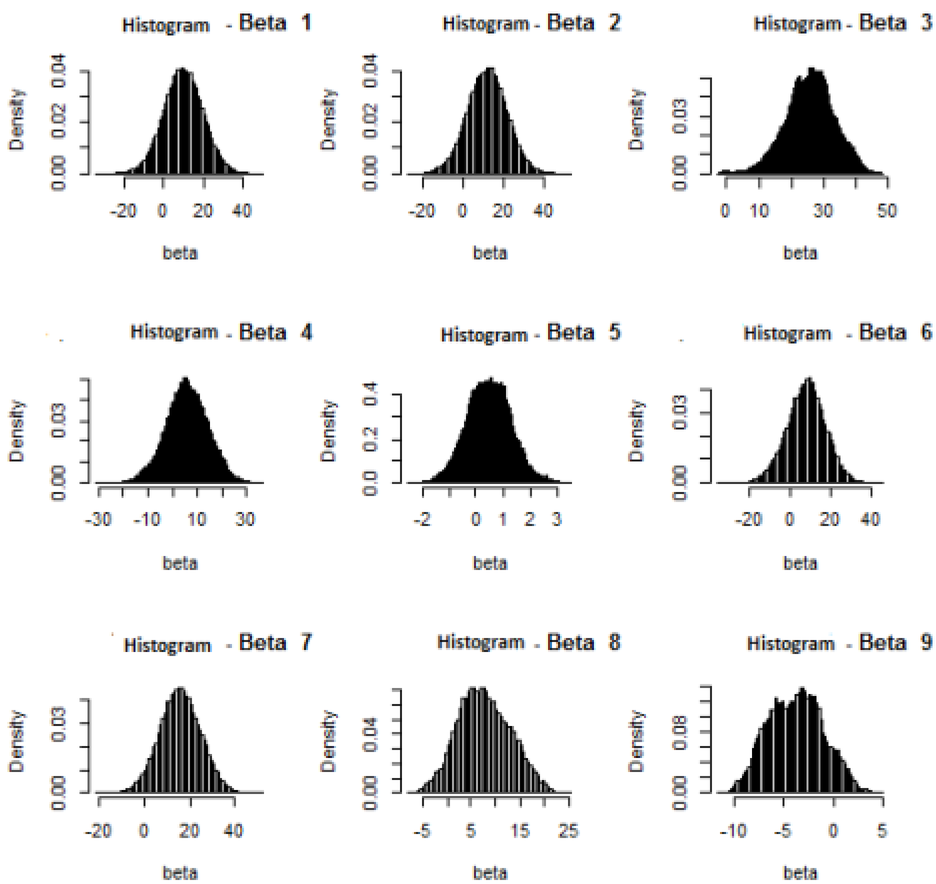

Figure (14) Histograms of real parameters

The above figure (13) shows the trace plots for the parameter estimates with the predictor variable observations of the (SMDP) model, which indicates the stationary of the proposed MCMC algorithm. Also, figure (14) illustrated that the proposed regularization methods gives parameter estimates follows the normal distribution under the (SMDP) model. The following table give the parameter estimates of the predictor variables.

Table (5) Parameter estimates

Variables name	**variables symbol**	**R.C. regression model**	**Bayesian lasso R.C. regression**	**SMDP-reciprocal Lasso R.C. regression**	**SMTN-reciprocal Lasso R.C. regression**
The patient's weight (in kilograms)	X1	1.576	0.923	1.325	0.425
The patient's age.	X2	0.773	0.733	1.221	0.841
The number of meals for the patient per day	X3	0.952	1.658	1.006	0.625
Are there genetic factors	X4	1.262	1.239	1.894	0.872
Is the patient under psychological pressure	X5	0.00	0.00	0.00	0.00
Does the patient have pancreatic disease	X6	0.285	0.474	0.039	0.059
Does the patient have covid19	X7	1.496	1.357	1.735	0.983
The patient's monthly income	X8	0.239	0.452	0.102	0.00
The number of hours of exercise per day	X9	0.961	0.806	0.722	0.492

Table (5) shows the parameter estimates of the predictor variables, where there are two irrelevant predictor variables (X5 and X8) excluded from the proposed model by setting its parameter values equal to zero. Consequently, the two proposed models are comparable to other model

Chapter Four

Conclusions and Recommendation

4.1 Conclusions:

In this thesis we have introduced the Bayesian reciprocal lasso regularization method in right censored dependent variable regression model. We write down the most important conclusions based on the theoretical and practical sides of thesis:

1- The few literature reviews about the Bayesian reciprocal lasso motivates us to open the door on this types of regularization method.

2- This thesis has introduced the Bayesian reciprocal lasso right censored regression model under two scale mixture, the first one is the scale mixture of truncated normal, and the second one is the scale mixture of uniform mixing with gamma that was proposed by Mallick and Yi (2014) but with using the reciprocal variable of gamma distribution to cope with the idea of the reciprocal lasso.

3- We have proposed a new Bayesian hierarchical model for the right censored regression model based on the mentioned scale mixtures. We have employed the scale mixtures to examine the performance of the introduced Bayesian reciprocal lasso in right censored regression model according to the suggested hierarchical model.

4- In addition to that we have focused on the comparison of the quality of the coefficient estimates and variable selection problem in simulation study and in real data. Therefore, we have used to criterion to test the performance of coefficient estimation methods; the median mean absolute deviation **(MMAD)** and standard deviation **(S.D)**. The simulation study and real data analysis shows that the proposed models give comparable results and outperform the other methods.

4.2 Recommendations

Based on the theoretical and data analysis aspects:

1- we have recommend the interested researchers in the field of Bayesian regularization method to develop the scale mixture of the reciprocal lasso since its results gives parsimonious model.

2- More development is required to hierarchical model and the scale mixture to other regression models; such as, Tobit model, Binary model, and Elastic net model. Moreover, we recommend using the reciprocal lasso model in other fields of data, such as, economic research, social, and health research.

3- Finally, the ability of reciprocal lasso in providing (parsimonious) sparse models gives more benefit for researchers interested in the field of variable selection problem.

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.

Al-Athari Faris M, (2011).Parameter Estimation for the Double Pareto • Distribution. Journal of Mathematics and Statistics 7(4).

Alhamzawi, R., & Mallick, H. (2020). Bayesian reciprocal LASSO quantile regression. Communications in Statistics-Simulation and Computation, 1-16.

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

Alhamzawi, R. and Haithem Taha Mohammad Ali. (2018). The Bayesian Elastic Net Regression. Communication in Statistics- Simulation and Computation.

Alhusseini, F. H. H. (2017). New Bayesian Lasso in Tobit Quantile Regression. Romanian Statistical Review Supplement, 65 (6), 213-229.

AlNasser, Hassan (2014). On Ridge Regression and Least Absolute Shrinkage and Selection Operator. B.Sc., University of Victoria.

Bakin, Sergey.(1999). Adaptive regression and model selection in data mining problems. Doctor of Philosophly of The Australian National University.

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.

Chatterjee, S. & Hadi, A.S (2013). Regression Analysis by example. Wiley series in probability and statistics, Fifth edition.

Clarke, B., Fokou, E., Zhang, H. H.(2009).Principles and theory for data mining and machine learning Springer series in statistics (pp. xv, 781 p.).

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

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

Fan, J. and R. Z. Li (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association 96, 1348-1360.

Flaih, A.N., Alsaadony, T., EIsalboukh, M. (2020). New Scale Mixture for Bayesian Adaptive Lasso Tobit Regression. Al-Rafidain University College For Sciences 2020, Volume , Issue 46, Pages 493-505

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.

Hastie, T., Tibshirani, R., and Wainwright, M. (2009). Statistical Learning with Sparsity: The Lasso and Generalizations. Taylor Francis Group, Forida, USA.

Hilali, H., K., A. (2019). Bayesian adaptive Lasso Tobit regression with a practical application. MSc. Thesis. Statistics Department College of Administration and Economics University of Al-Qadisiyah.

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.

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.

Maddala,G.S.(1993). Limited-Dependent and Qualitative Variables in Econometrics. Cambridge University Press.

Miller, Alan. (2002). Subset Selection in Regression. 2nd ed. Chapman and Hall/CRC.

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

Mallick, H., R. Alhamzawi, and V. Svetnik (2020). The reciprocal Bayesian lasso. arXiv preprint arXiv:2001.08327.

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

Meinshausen, Nicolai. (2007). Relaxed Lasso. Computational Statistics & Data Analysis.Volume 52, Issue 1, Pages 374-393.

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.

Shin,M., Bhattacharya,A. and Johnson,V.E. (2018).Scalable Bayesian variable selection using nonlocal prior densities in ultrahigh-dimensional settings. Statistica Sinica, 28(2):1053.

Song,Qifan.(2014).Variable selection for ultra-high dimensional data. Phd. Dissertation,Texas A&M University.

Song, Qifan. (2018). An overview of reciprocal L1-regularization for high dimensional regression data. Wiley Interdisciplinary Reviews: Computational Statistics, 10(1):e1416, 2018.

Song, Qifan and Liang,Faming. (2015). High-dimensional variable selection with reciprocal L1- regularization. Journal of the American Statistical Association, 110(512):1607{1620, 2015.

References

Soret, P, Avalos, M., Wittkop, L., Commenges, D., and Rodolphe, T. (2018). Lasso regularization for left-censored Gaussian outcome and high-dimensional predictor. BMC Medical Research Methodology volume 18, Article number: 159.

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

Tibshirani, R., Saunders, M., Rosset,S., and Knight, K. (2005). Sparsity and smoothness via the fused lasso. *J. R. Statist. Soc.* B.**67**, *Part* 1, *pp.* 91–108.

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.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, *101*(476), 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., & Zhang, H. H. (2009). On the adaptive elastic-net with a diverging number of parameters. Annals of statistics, 37(4), 1733.

الخلاصة

تركز هذه الرسالة على الانحدار البايزي لطريقة معكوس لاسو (rlasso) في وجود متغير تابع محدود خاضع للرقابة. يعد اختيار المجموعات الفرعية المثلى للمتغيرات المتنبأ بها هو الهدف الأكثر شيوعًا في موضوع تحليل الانحدار. حيث تجمع طريقة معكوس لاسو على مقلوب L1-Norm في دالة الجزاء. في الوقت الحاضر ، تم تطوير العديد من طرق تقدير معاملات الانحدار بما في ذلك طرق التنظيم لبناء النموذج الشحيح. ان طريقة معكوس لاسوهو أسلوب تنظيم جديد يوفر نموذج انحدار أكثر شحًا (اختيار متغير مع مزيد من التفسير). ان قلة الدراسات السابقة حول (rlasso) يعود لسبب الفكرة الجديدة لهذه الطريقة. فقد استخدمنا خليط مقياس لتوزيع الباريتو المزدوج (SMDP) وكذلك خليط المقياس للتوزيع الطبيعي المبتور (SMTN) الذي اقترحه Mallick وآخرون سنة (2020). حيث قمنا باجراء تعديلًا لـ (SMTN) من خلال اقتراح نموذج هرمي مسبق جديد. وقد استخدمنا الخليط (SMDP) والخليط (SMTN) المعدل في النموذج الهيكلي للمتغير التابع الخاضع للرقابة من اليمين. تم إجراء تجارب محاكاة ، بالإضافة إلى تحليل بيانات حقيقية لفحص سلوك التوزيعات اللاحقة. أظهرت النتائج أن أنواع خليط المقياس المستخدم تتفوق على طرق التنظيم الشائعة الأخرى في كل من المحاكاة وتحليل البيانات الحقيقية. بهذا يمكن القول انه بشكل عام ان نماذج معكوس لاسو يوفر أساسًا رائع لفئة من طرق التنظيم التي تعمل على تحسين الحل الصفري الذي يتميز بالتقارب إلى الحل الحقيقي.

جمهورية العراق
وزارة التعليم العالي والبحث العلمي
جـــــامـعــــة القـــادســــية
كــلية الادارة والاقتصــاد
قسم الاحصـــاء

معكوس لاسو البيزية لبيانات المراقبة من اليمين

رسالة مقدمـــة الى مجلس كلية الادارة والاقتصاد في جامعة القادسية
و هي جزء من متطلبات نيل درجة ماجستير في علوم الاحصاء

تقدمت بهـا

فاطمة كاظم محمد

باشراف

أ.د. احمد نعيم فليح

٢٠٢٢ م ١٤٤٣ هـ