Al-Qadisiyah Journal for Administrative and Economic Sciences ISSNOnline : 2312-9883
QJAE, Volume 24, Issue 1 (2022) ISSNPrint : 1816-9171

Simulation Study for the Reciprocal Lasso Right Censored Response Variable

Fatima Kadhim Mohammed Dr. Ahmad Naeem Flaih

Abstract

This paper focuses on Bayesian reciprocal lasso regression with right censored response variable. Choosing the important variables that relevant on the response variable is very common goal of the regression analysis. The reciprocal lasso adds the reciprocal of L1-rorm in the penalty function. Reciprocal lasso is a regularization method that provides variable selection procedure with more interpretation regression model. We employed the scale mixture of double pareto (SMDP) and the scale mixture of truncated normal (SMTN) that proposed by Mallick et al. (2020) and we with some modification for (SMTN) in the right censored limited dependent variable. New hierarchical prior model and new Gibbs sampler algorithm have developed. Some simulation examples have conducted to analysis the behavior of the posterior distributions. The results show that the employed scale mixture types outperform other common regularization methods in both of the simulation.

Keywords: Reciprocal lasso, SMDP, SMTN, Gibbs sampler, Simulation.

1- Introduction

Statisticians are formulated the statistical models to solve certain problems. The regression model demonstrate the relationship between the dependent (response) variable Y, and one or more independent (predictor) variables X. This relationship 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 yellowing function

$$f(X;\beta) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k \dots\dots\dots\dots (1)$$

Where k the number of predictor variables. It is very well know that the objective of regression model (1) is to find the mean of Y. Many predictor variables included in regression model affect Interpretation of the estimated model and maybe inflated the variances of the parameter estimates which cause the poor prediction accuracy for the estimated model. So, variable selection procedure is essentially used for statistical modeling of many predictor variables problem which recently found in many fields of scientific fields. Therefore, we can say that there is another objective of conducting the regression analysis which is called model selection. The quality of parameter estimate 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) 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 precedence. To overcome this problem in using OLS, one can use the regularization method that depends on penalized methods which are also treads the model selection problem. In this paper 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. Right (upper) censored regression model is more reliable if the variable selection procedure has been followed. 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 y reach this limit or boundary. Limited dependent variable leads us to the censored sample which its observations are $(\mathrm{y}_1, \mathrm{y}_2, \dots, \mathrm{y}_\mathrm{n})$ resulting from a latent variable (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 yields 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 regression model, the number of independent variables included in regression model brings the researcher to develop the mechanism of variable selection procedure. So, the variable selection procedure treated with the regression form specification. The residual mean squares (**RMS**) criterion is a model selection tool, smallest the $\boldsymbol{RME}$ 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. Mallows (1973) defined the following criterion that is called Mallows $\mathbf{C_p}$ criterion to assess the performance of the regression model. Akaike (1973) defined Akaike information criterion (**AIC**), which is a model selection tool, 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 independent variables), here $\boldsymbol{RME}, \boldsymbol{C_p}$, and $\boldsymbol{R^2}$ are used to select the best fit model. Schwarz (1978) defined the modified **AIC** criterion that is called Bayes Information criterion (**BIC**), The smallest value of **BIC** the better model. The drawback of all possible equations method is when the number of equations getting larger.. One can see Draper and Smith (1998), Hastie et al. (2009), James et al. (2013), and Breaux (1967) for more details and information. Zou et al. (2007) discussed using BIC criterion in choosing the shrinkage parameter in lasso method.

But from the regularization function point of view, Hoerl and Kennard (1970a,1970b) 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 $\mathrm{L_2-norm}$. Note that if $\lambda = 0$ then $\hat{\beta} = \mathrm{OLS}$ estimates. Tibshirani (1996) developed new regularization method named lasso which gives sparse solution for the linear regression coefficients. Lasso adds penalty function that include L1-norm function which controlled by the shrinkage parameter. The parameter estimates for some predictor variables reach the zero value and the solution regards as sparse solution. So, since we interested in Bayesian estimation, the following studies are important to mention: Park and Casella (2008) introduced the Bayesian analysis for the regularization method based 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. 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 introduced for the MCMC algorithm. Simulation study have conducted to examine the proposed model in terms of variable selection procedure, also the proposed method have applied on real data and the results shows outperforms of the proposed model comparing with some penalized method. Alhamzawi (2017) proposed new hierarchal prior model for the Tobit regression with lasso method. The prior distribution of the regression parameters presented as Laplace distribution. The Laplace distribution presented as scale mixture mixing uniform distribution with particular gamma distribution. Based on the proposed hierarchical prior model new Gibbs sampler algorithm has proposed. Simulation study have 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 comparing with some other regularization 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 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 showed the outperforms of the proposed posterior distribution in part of variable selection and the efficiency of the proposed estimator. Song and Liang (2014) introduced the reciprocal Bayesian lasso for the high dimensional variable selection problem in the linear regression. Song (2015) mentioned that the reciprocal lasso estimators have the oracle property. 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

$$h(\beta) = \ argmin\ RSS + \lambda \sum_{j=1}^{p} \frac{1}{|\beta_j|} I(\beta_j \neq 0) \cdots\cdots\cdots\cdots (2)$$

where $\lambda \geq 0$ 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 (2) is decreasing in the interval $(0, \infty)$, Discontinuous at zero. The 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 $\eta \sim Inverse\ Gamma\ (2, \lambda)$.

Also,

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

$$Q(\beta) = \ argmin \sum_{i=1}^{n} \rho(y_i - x_i^T \beta) + \lambda \sum\nolimits_{j=1}^{p} \frac{1}{|\beta_j|} I(\beta_j \neq 0) \cdots \cdots \cdots (3)$$

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

2- SMTN Hierarchical Priors for Bayesian models

Referring to the formula of proposition that Mallick et al. (2020) and based on the work of Park and Casella (2008), we propose the following hierarchical prior model scale mixture of truncated normal (SMTN):

$$y_i = \begin{cases} x_i'\beta + u_i & if \quad x_i'\beta + u_i < c \\ c & if \quad x_i'\beta + u_i \geq c \end{cases} \quad , \ where\ \mathbf{c}\ is\ a\ censored\ point \qquad \ldots (4)$$

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

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

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

$$\tau_1^2, \ldots, \tau_p^2 \sim \prod_{j=1}^{p} \frac{\delta^2}{2} e^{-\delta^2 \tau_j^2 / 2} d \, \tau_j^2 ; where \quad \tau_1^2, \ldots, \tau_p^2 > 0,$$

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

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

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

3- SMTN Full Conditional Posterior Distributions

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

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

3.2- Sampling $\boldsymbol{\beta}$: In this step we generate $\boldsymbol{\beta}$ from normal distribution $\boldsymbol{C^{-1} X' y^*}$ and variance $\boldsymbol{\sigma^2 C^{-1}}$. **Where** $C = X'X + D_\tau^{-1}$ **and** $D_\tau = ding\left(\tau_1^2, \ldots, \tau_p^2\right)$.

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

3.4- 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 $\boldsymbol{\delta^2}$.

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

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

4. Extension on Reciprocal lasso right censored 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).

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

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

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

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

$$\sigma^2 \sim \pi(\sigma^2) \tag{5}$$

Connection with Bayesian lasso and reciprocal lasso the full conditional posterior distribution for the parameters in hierarchical prior model (5) of the (SMDP) Bayesian reciprocal Laplace right censored regression model are as follows Mallick et al. (2020):

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

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

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

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

$$\lambda|\beta \sim Gamma(a + 2p, b + \textstyle\sum_{j=1}^{p} \frac{1}{|\beta_j|}). \tag{5}$$

4.1 Simulation Experimental

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 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 **12000** iterations with **2000** iterations have burned-in for reaching the stationary of posterior distribution by using R programming language.

4.2 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 Deviation **(S.D.)** have used for assessing the quality of the estimated model.

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

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

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:

4.3 Simulation Example

In this example and based on the same process in the sample, we supposed the following dense true parameter vector, $\boldsymbol{\beta} = (\boldsymbol{0.85}, \boldsymbol{0.85}, \boldsymbol{0.85}, \boldsymbol{0.85}, \boldsymbol{0.85}, \boldsymbol{0.85}, \boldsymbol{0.85}, \boldsymbol{0.85}, \boldsymbol{0.85})^{\boldsymbol{T}}$ 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 (1) MMAD and S.E. values for simulation

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)**
	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)**
n=250	5	0.736 0.573	0.647 (0.522)	0.655 (0.514)	**0.321 (0.216)**

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.

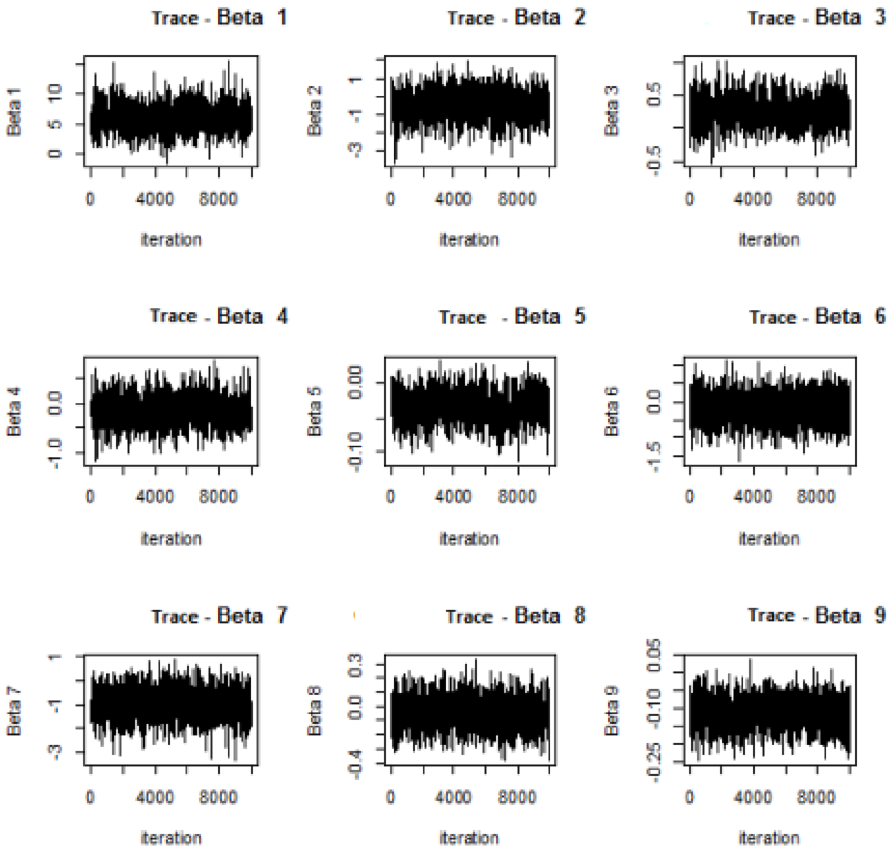

Figure 1. Trace plots of the parameter estimates $\beta_1 - \beta_8$.

The above figure (1) shows the trace plots which illustrate no flat bits and that MCMC algorithm suffer no slow mixing.

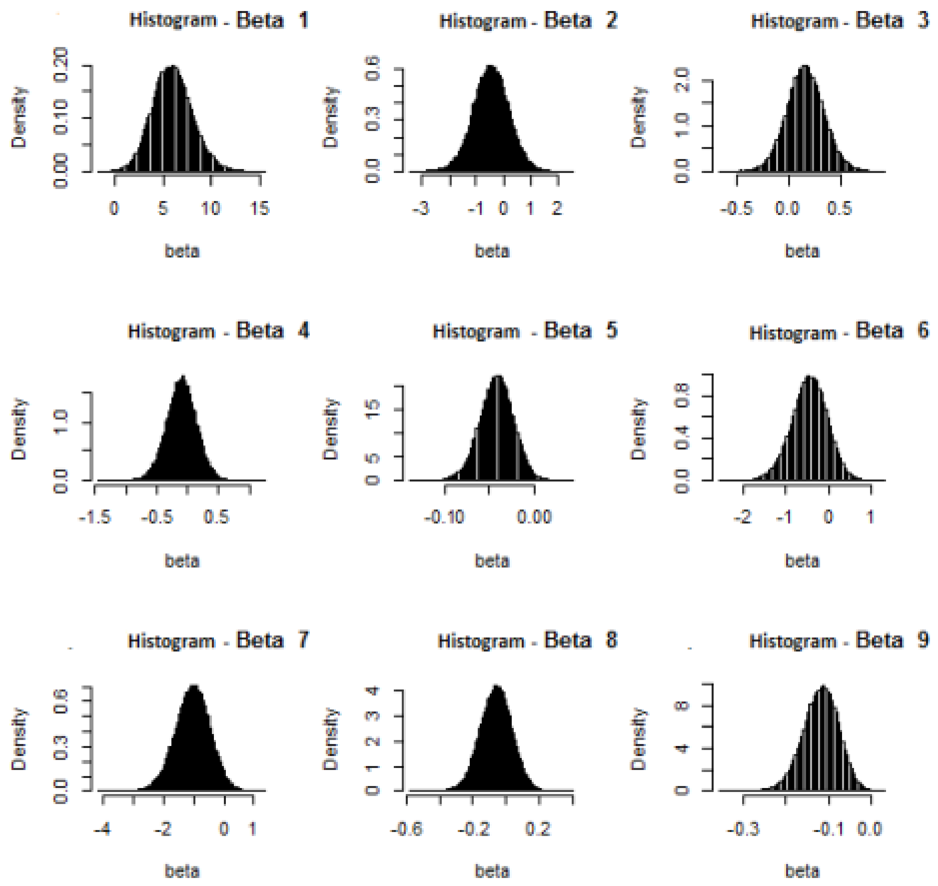

Figure (2) Histograms of parameter estimates $\beta_1 - \beta_8$.

Figure (2) shows the distributions of the parameter estimates $\beta_1 - \beta_8$ and it is very clear that the distribution of the parameter follows the normal distribution for all parameter estimates.

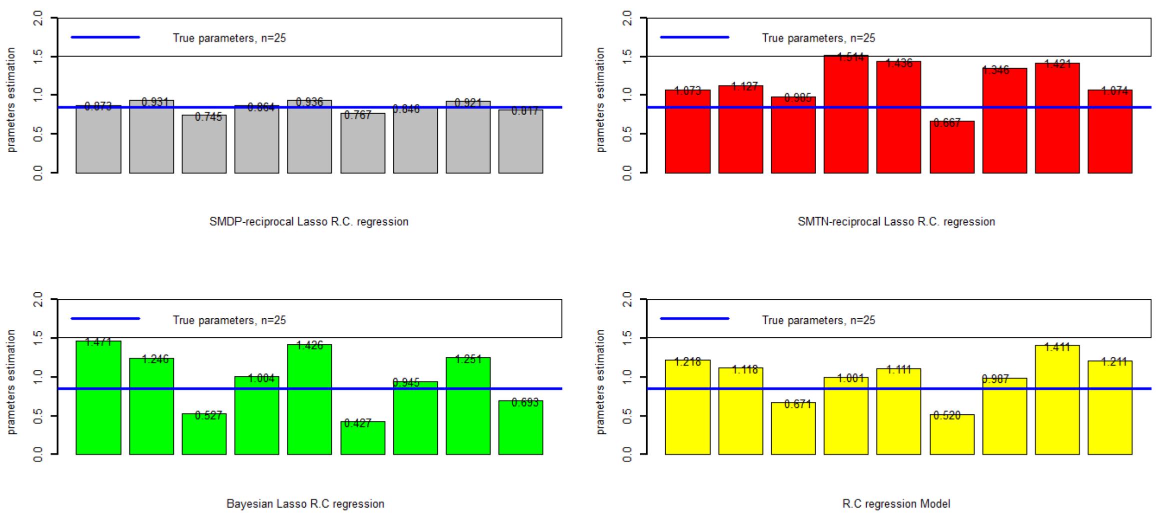

Figure (3) True vector and parameter estimates $\beta_1 - \beta_8$ with sample size=25

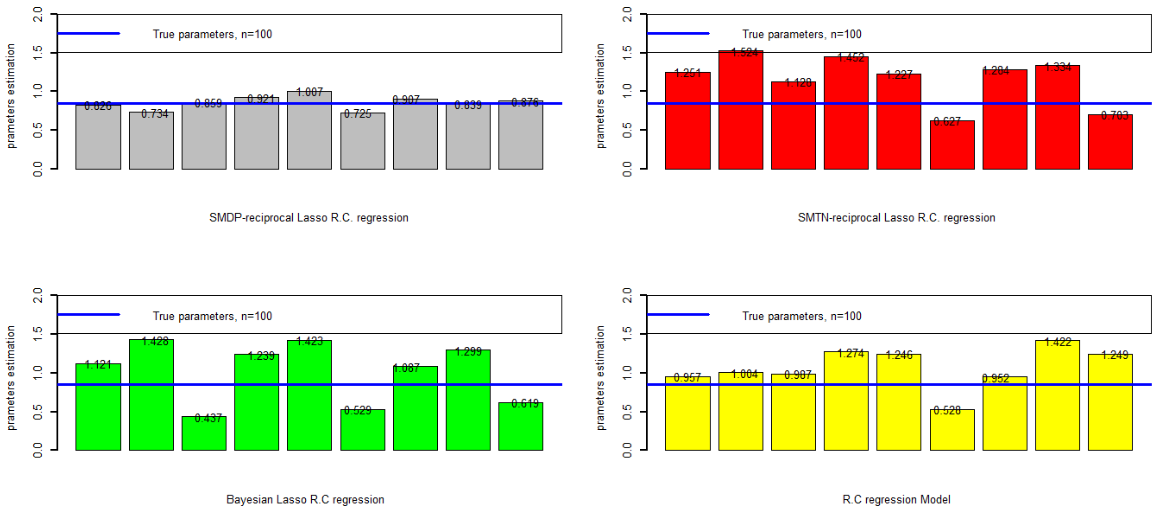

Figure (4) True vector and parameter estimates $\beta_1 - \beta_8$ with sample size=100

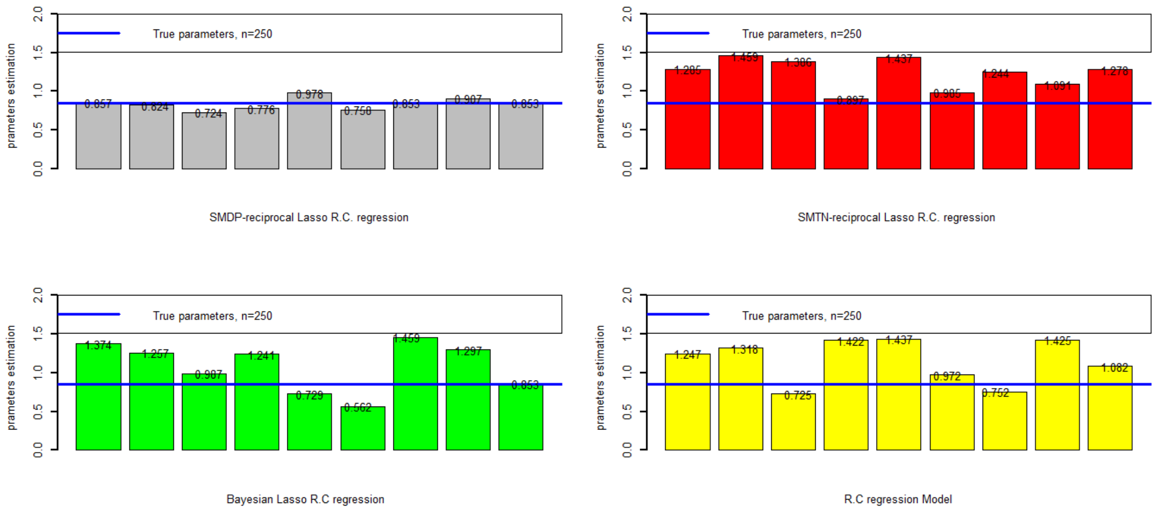

Figure (5) True vector and parameter estimates $\beta_1 - \beta_8$ with sample size=250

Figures (3) to (5) 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.

Conclusions

New scale mixture of Rayleigh distribution mixing with normal distribution have developed as the prior distribution of the Laplace distribution. Consequently, we produced new Bayesian hierarchical model for elastic net in linear regression. Gibbs sampler algorithm have developed have developed to examine the convergence of the proposed posterior distributions. Some simulation scenarios have implemented based on the proposed model. The result of simulation shows that the proposed method clearly outperforms the other method from the variable selection procedure view.

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

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

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.

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.

Flaih, A.N, Alshaybawee, and Al husseini F.H, (2020). sparsity via new Bayesian Lasso . Periodicals of Engineering and Natural Sciences , vol.8 , issue 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.

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.

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.

Qifan Song. (2018). An overview of reciprocal L1-regularization for high dimensional regression data. Wiley Interdisciplinary Reviews: Computational Statistics, 10(1): 1416.

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

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

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

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