Periodicals of Engineering and Natural Sciences ISSN 2303-4521
Vol. 8, No. 1, March 2020, pp.345-359

Sparsity via new Bayesian Lasso

Ahmad Naeem Flaih, Taha Alshaybawee, Fadel Hamid Hadi Alhusseini
Statistics Department, College of Administration and Economics, University of Al-Qadisiyah,

ABSTRACT

Lasso estimate as the posterior mode assuming that the parameter β has prior density as double exponential distribution [1]. In this paper, we proposed Scale Mixture of Normals mixing with Rayleigh (SMNR) density on their variances to represent the double exponential distribution. Hierarchical model formulation presented with Gibbs sampler under SMNR as alternative Bayesian analysis of minimization problem of classical lasso. We conducted two simulation examples to explore path solution of the Ridge, Lasso, Bayesian Lasso, and New Bayesian Lasso (R, L, BL, NBL) regression methods through the prediction accuracy using the bias of the estimates with different sample sizes, bias indicates that the lasso regression perform well, followed by the NBL. The Median Mean Absolute Deviations (MMAD) used to compared the perform of the regression methods using real data, MMAD indicates that the proposed method (NBL) perform better than the others.

Keywords: Bayesian Lasso, mixing Rayleigh, shrinkage parameter ,Hierarchical Model, Gibbs sampling

Corresponding Author:

Ahmad Naeem Flaih
Statistics Department, University of Al-Qadisiyah
Address: Iraq, Al-Diwaniyah
E-mail: ahmad.flaih@qu.edu.iq

.

1. Introduction

The Bayes estimate of the parameters of linear regression through lasso method under constraint uses $\ell_1 - norm, \|\beta\|_1$. The lasso estimator $\hat{\beta}$ is the parameters estimate of the following linear regression model [1]:

$$y = X\underline{\beta} + \epsilon \qquad (1)$$

Where $y = (y_1, \dots, y_n,)'$ is the vector of the centered response variable, X *is* $n \times p$ matrix of standardized predictors (if they have different units) , and $\epsilon \sim N(0, \sigma^2)$ As minimization problem, the Lasso method to estimate the parameter β of model (1) by minimizing,

$$L(\beta, \lambda) = (\ell_2(y - X\underline{\beta}))^2 + \lambda\|\beta\|_1 \qquad (2)$$

Where $\ell_2(.)$ is the $\ell_2 -$ norm, and $\lambda \geq 0$ is the shrinkage parameter that decide the sparsity of $\hat{\beta}$. β. Lasso estimates in (2) is the posterior mode when the prior distribution of the regression parameter distributed according to double exponential density [1]. The solution path of lasso is multi-sub function(piecewise)linear that defined on a sequence of λ and suggest that the solution path of λ following algorithm called Least Angle Regression (LAR) which implies that the posterior distribution of regression parameters is linearly for sequence interval of $\lambda[\lambda_k, \lambda_{k+1}]$ [2]. Working of [1] motivate many authors to suggest new representations for the double exponential as prior density of regression coefficients. New Bayesian lasso considered through hierarchical model that represent the double exponential prior density as scale mixture of normal mixing with exponential distribution [3] which originally proposed in [4], then the full joint Bayesian posterior distribution of regression parameter under conditional Laplace density is

$$\pi(\beta, \sigma^2/y) \propto \pi(\sigma^2)\,(\sigma^2)^{\frac{-(n+p-1)}{2}} \exp\left[-\frac{1}{2\sigma^2}\left(y - X\underline{\beta}\right)'\left(y - X\underline{\beta}\right) - \lambda\sum_{j=1}^{n}|\beta_j|\Big/\sqrt{\sigma^2}\right] \qquad (3)$$

The joint posterior (3) have a lasso estimate that consider as the posterior mode estimate, appreciation to [3] they state that the prior distribution of regression parameter must be conditioning on σ2 to guarantees the unimodality of the full posterior distribution. New Bayesian lasso proposed [5] based on new hierarchical formulation that use the following representation of double exponential density as scale mixture of uniform distribution mixing with $Gamma(2, \lambda)$

$$\frac{\lambda}{2}\, e^{-\lambda|x|} = \int_{-u}^{u} \frac{1}{2u}\, \frac{\lambda^2}{\Gamma 2}\, u^{2-1}\, e^{\lambda u}\, du$$

Abbas [6] proposed new representation of the hierarchical model based on double exponential density as non-scale mixture of uniform mixing with standard exponential density,

$$\frac{\lambda}{2}\, e^{-\lambda|\beta_j|} = \int_{z_j > \lambda|\beta_j|}^{,} \frac{\lambda}{2}\, e^{-z_j}\, dz_j,$$

In this paper, along similar lines of [4], [3], [5], and [6], we proposed a new hierarchical formulation of Bayes lasso that conjugate normal prior for the regression coefficients and independent Rayleigh density on their variances ,i.e., we proposed using double exponential density as Scale Mixture of Normal mixing with Rayleigh density on their variances (SMNR). In section 2 we introduced the SMNR as the Laplace prior distribution. Also, in section 3 , we presented new hierarchical model formulation of the Laplace as SMNR, and Gibbs sampler algorithm presented in section 4. Two examples studied and real data analysis are presented in section 5 and section 6. In section conclusions have provided . Appendix A include the proof of the SMNR.

Following [3] our new full Bayesian analysis consider the conditional double exponential prior form as follows,

$$\pi(\beta/\sigma^2) = \prod_{j=1}^{p} \frac{\lambda}{2\sigma^2}\, exp\left[-\frac{\lambda|\beta_j|}{\sigma^2}\right]$$

2. Scale mixture of normal distribution

We construct new hierarchical model representation considering the double exponential prior density of the parameters as scale mixture of normal distribution mixing Rayleigh density. Following [7] and [8],generally the scale mixture of normals mixing with Rayleigh distribution g(σ)is ,

$$f(x) = \int_0^\infty \frac{1}{\sigma} \emptyset\left(\frac{x}{\sigma}\right) g(\sigma) d\sigma \qquad (4)$$

which is symmetric about zero and unimodal function. Mathematically it is well known that,
if $^{z}/_{s} \sim N(\mu = 0, s^2)$ with s∼Rayleigh(a), then z∼Laplace (μ = 0, a), and based on (4) we can write

$$\frac{1}{2a} \exp\left[-\frac{|z|}{a}\right] = \int_0^\infty \frac{1}{\sqrt{2\pi s^2}}\, e^{-z^2/2s^2}\, \frac{s}{a}\, e^{-s^2/2a}\, d\,s \qquad (5)$$

Appendix A contain the proof of (5). Let $a = \sigma^2/\lambda$, $z = \beta$, and $s = \sigma\sqrt{\tau}$ then (5) can be written as,

$$\frac{1}{2\sigma^2} \exp\left[-\frac{\lambda|\beta|}{\sigma^2}\right] = \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2\tau}} \exp\left\{-\frac{\beta^2}{2\sigma^2\tau}\right\} \frac{\lambda}{2}\, e^{-\lambda\tau/2}\, d\tau \qquad (6),$$

The prior density (6) is conditioning on the σ^2 which guarantees a unimodal posterior distribution, see [3] for more information.

3. New hierarchical model formulation

The new hierarchical model formulation of the full model under the new proposed scale mixture (6) defined as follows,

$$y^{n\times 1}|X,\beta,\sigma \sim N_n(X\beta,\sigma^2 I_n),$$
$$\beta^{p\times 1}|\sigma^2,\tau_1,\dots\tau_p \sim N_p(0,\sigma^2 D_\tau),$$
$$D_\tau = diag(\tau_1,\dots\tau_p),$$
$$\sigma^2,\tau_1,\dots\tau_{p\sim\pi(\sigma^2)d(\sigma^2)}\prod_{j=1}^{p}\frac{\lambda}{2}\,e^{-\lambda\tau_j/2}\;d\tau_j \qquad (7)$$

$\sigma^2,\tau_1,\dots,\tau_j > 0$. We can reach the conditional prior $\pi(\beta/\sigma^2)$ in (7) after integrating out $\tau_1,\dots,\tau_j$ in (7). As well as, we use the prior density $\pi(\sigma^2) = 1/\sigma^2$, or any inverse gamma to maintain the conjugacy in the proposed scale mixture, see [3].

4. The Gibbs sampler and the full conditional distributions

Gibbs sampler will implement the model (7) for Gibbs sampling the most useful algorithm of MCMC technique in Bayesian analysis which samples from the conditional distribution of a parameter given all the other parameters, see [9]. The construction of the hierarchical model is formulated such a way that there is full conditional distribution for each component of the estimate. Following [3], we can implement the model (7) and by using the following inverse gamma prior density on

$$\pi(\sigma^2) = \frac{\gamma^a}{\Gamma a}\,(\sigma^2)^{-a-1}\,e^{-\gamma/\sigma^2};\;\; \sigma^2, a, \gamma > 0$$

Where a and b are the hyper parameters , and as $a, b \to 0$ the prior $\pi(\sigma^2)$ will be proportional to $(1/\sigma^2)$ The full joint density can be written as follows:

$$f(y/\beta,\sigma^2)\pi(\sigma^2)\prod_{j=1}^{p}(\beta_j/\tau_j,\sigma^2)\,\pi(\tau_j) = \frac{1}{(2\pi\sigma^2)^{n/2}}\;e^{-\frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)}$$

$$\pi(\sigma^2)\frac{\gamma^a}{\Gamma a}\,(\sigma^2)^{-a-1}e^{-\gamma/\sigma^2}\;\prod_{j=1}^{p}\frac{1}{\sqrt{2\pi\sigma^2\tau_j}}\;e^{-\beta_j^2/2\sigma^2}\,\frac{\lambda}{2}\,e^{-\lambda\tau_j/2} \qquad (8)$$

Based on the full joint density, we can construct the following conditional distributions:

4.1. The full conditional distribution for β

Gibbs sampler technique need no more than an unnormalized posterior, so we have to eliminate all factors not involving the parameter β from (8) and the remaining part of the full joint density is proportional to that contains β, i.e., we are left with the following terms:

$$-\frac{1}{2\sigma^2}\,(y-X\beta)'(y-X\beta) - \frac{1}{2\sigma^2}\,\beta'\,D_\tau^{-1}\beta$$
$$= -\frac{1}{2\sigma^2}\,[\beta'(X'X)\beta - 2yX\beta + y'y + \beta'\,D_\tau^{-1}\beta] \qquad (9)$$
$$-\frac{1}{2\sigma^2}\,[\beta'(X'X - D_\tau^{-1})\beta - 2yX\beta + y'y]$$

Here, y is the centered response variable and let $C = X'X - D_\tau^{-1}$, ,then (9) can be rewrite as,

$$-\frac{1}{2\sigma^2}\,[\beta' C\beta - 2yX\beta + y'y],$$

Hence now we can rewrite the density of β as the exponent of

$$-\frac{1}{2\sigma^2}\left(\beta - C^{-1}X'y\right)' C\left(\beta - C^{-1}X'y\right),$$

Recalling the multivariate normal distribution $X\sim N(\mu,\Sigma)$, then we can say that β follows the multivariate normal density with mean $C^{-1}X'y$ and variance $\sigma^2 C^{-1}$,

4.2. The full conditional distribution for σ^2

The Gibbs sampler distribution for σ^2 is generating from the full joint distribution (8) and involving only σ^2

$$(\sigma^2)^{-\frac{n-1}{2}-\frac{p}{2}-(a+1)}\ e^{\left[-\frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)+\frac{1}{2\sigma^2}\beta' D_\tau^{-1}\beta+\frac{\gamma}{\sigma^2}\right]} \quad (10)$$

So, from (10) and the distribution pdf of inverse gamma, we can state that σ^2 is conditionally inverse gamma distribution,

$$\sigma^2/.\sim Inverse\_Gamma\ (\frac{n-1}{2}+\frac{p}{2}+a, (y-X\beta)'(y-X\beta)/2+\beta' D_\tau^{-1}\beta/2+\gamma.$$

4.3. The full conditional distribution for τ

The last variable that we have to sample in Gibbs sampler is the latent variable τ_j. the Gibbs sampler distribution for τ is the part of (8) that includes only τ_j is

$$(\tau_j)^{-\frac{1}{2}}\ exp\left[-\frac{1}{2}(\frac{\beta_j^2}{\sigma^2\,\tau_j}+\lambda\tau_j)\right] \quad (11)$$

The Inverse Gaussian distribution (IG) is,

$$f(x;\mu,\lambda)=\sqrt{\frac{\lambda}{2\pi x^3}}\ exp\left[-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right]; x>0,$$

Based on the distribution of the reciprocal of an inverse Gaussian variable introduced by [10], then we can rewrite (11) in the view of the reciprocal of the reciprocal of the inverse Gaussian distribution as follows:

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

Then, we can say that $1/\tau\sim IG(\sqrt{\frac{\lambda\sigma^2}{\beta^2}},\lambda)$, where λ is the shape parameter and $\sqrt{\frac{\lambda\sigma^2}{\beta^2}}$ is the mean (location) parameter.

4.4. Choosing the regularization parameter λ

The regularization parameter λ can be estimate by using empirical Bayes which is given by maximizing the data marginal likelihood and use the Monte Carlo EM algorithm to complement the Gibbs sampler [3]. As well, they state that the distribution of prior λ is the following Gamma form,

$$\pi(\lambda)=\frac{\delta^{\Upsilon}}{\Gamma\Upsilon}\ (\lambda)^{\Upsilon-1}\ e^{-\delta\lambda};\ \lambda,\delta\Upsilon>0, \quad (12)$$

From the full joint density (12), the terms involving (λ) together with gamma prior (8) are,

$$\prod_{j=1}^{p}\left(\frac{\lambda}{2}\ e^{-\frac{\lambda}{2}\tau_j}\right)(\lambda)^{\Upsilon-1}e^{-\delta\lambda}=\lambda^{p+\Upsilon-1}\exp\left[-\lambda\left(\frac{1}{2}\sum_{j=1}^{p}\tau_j+\delta\right)\right]$$

This is again the gamma distribution, i.e.,

$$\lambda\sim Gamma\left(p+\Upsilon,\frac{1}{2}\sum_{j=1}^{p}\tau_j+\delta\right),$$

Thus, we updated the regularization parameter λ via the above gamma distribution. Summarizing, we have the following hierarchical model for implement the Gibbs sampler:

$$\begin{aligned}
&y|X,\beta,\sigma^2\sim N_n(X\beta,\sigma^2 I_n),\\
&\beta|\tau\,\sigma^2,\sim N\left((X'X+D_\tau^{-1})^{-1}X'y,\sigma^2(X'X+D_\tau^{-1})^{-1}\right),\\
&\tau^{-1}/\sigma^2\sim IG(\sqrt{\frac{\lambda\sigma^2}{\beta^2}},\lambda),\\
&\sigma^2\sim Inverse-\ Gamma\ \left(\frac{n-1}{2}+\frac{p}{2}+a,(y-X\beta)'(y-X\beta)/2\right.\\
&\qquad\left.+\beta'\,D_\tau^{-1}\beta/2+\gamma\right), \qquad (13)\\
&\lambda\sim Gamma\ (p+\Upsilon,\tfrac{1}{2}\textstyle\sum_{j=1}^{p}\tau_j+\delta).
\end{aligned}$$

5. Simulation analysis

Simulation study is performed based on (13) to support the proposed scale mixture of normals and to identify many scenarios in which the New Bayesian Lasso (NBL) preforms well. For simulated examples, we study estimation accuracy is conducted to assess the accuracy of the estimation of Lasso parameters through shown how the bias of the estimators effects the quality of estimated regression model. We compared between the bias of parameter estimators of NBL, Ridge (R), Lasso (L), and Bayesian Lasso (BL) by using the following formula,

$$Bias\left[\hat{\beta}_j(\hat{\theta}_i)-\beta_j^{true}\right]$$

As well as, we use the statistic (MMAD) to compare the performance of different regression models (RR, LR, BLR, and NBLR) by using the following formula,

$$MMAD=median\left[mean\left(\left|X\hat{\beta}-X\beta^{true}\right|\right)\right]$$

Here, β^{true} is the vector of true parameter values. The generating process of data is as follows

$$y=X\underline{\beta}+e, \qquad (14)$$

Where X is distributed from normal with mean zero and variance one, $e\sim N(0,\sigma^2)$. The correlation between predictors $X_i\ and\ X_j$ is $\rho^{|i-j|}$, and the matrix of predictor variable observations are $X\sim N(0,\Sigma), here\ \Sigma_{ij}=\rho^{|i-j|}$ Before carry out any regression model, we standarized the predictors values and centered the response variable values. The Bayesian lasso and the new Bayesian lasso estimates are the posterior means, we use the Gibbs sampler with (100) samples and the burn out samples number is (4000). In R package lars for lasso, we used the LARS algorithm proposed by [2] to select the penalty parameter with (k = 10)-fold cross validation, see [11] who state that the best choice for k is 10.

Example 1

In this example , we generate 50 datasets each with 20 observation from the true model (14), this example used by [1], here the true vector of parameters values is β=(3,1.5,0,0,2,0,0,0) with $\rho=0.5$, $\sigma=3$. We focus on the prediction accuracy aspect of the estimate quality through calculate the bias of the estimates for the different regression methods (RR, LR, BLR, and NBLR) under different sample sizes (20,50,100). The simulation results with the different scenarios summarized in Table (1), the effects of increasing the samples size shows that the NBL gives the less bias of estimate. As well, the results in Table (2) shows that the lasso regression performs better than the other methods based on the values of MMAD statistic, followed closely by NBL method, followed by BL method, and poorly perform of ridge method which has the highly bias in Table (1).

Table 1: The bias of parameter estimates along the path solution of each method with different sample sizes

N	Methods	Bias $\hat{\beta}_1$	Bias $\hat{\beta}_2$	Bias $\hat{\beta}_3$	Bias $\hat{\beta}_4$	Bias $\hat{\beta}_5$	Bias $\hat{\beta}_6$	Bias $\hat{\beta}_7$	Bias $\hat{\beta}_8$
20	RR	2.6060	1.1705	0.3306	0.4048	1.6632	0.3649	0.3355	0.3607
	LR	1.5183	0.3068	0.0959	0.2775	0.1647	0.2465	0.2232	0.3505
	BLR	2.4519	1.0145	0.3704	0.3498	1.3379	0.3773	0.4594	0.3783
	NBLR	1.0792	0.1989	0.0292	0.1879	0.1631	0.2458	0.1379	0.1912
50	RR	2.6141	1.1175	0.2883	0.2964	1.6344	0.2835	0.4198	0.3217
	LR	0.5100	0.6866	0.2170	0.1835	0.6387	0.2304	0.3230	0.3122
	BLR	1.4645	0.7517	0.3685	0.3815	0.9347	0.2892	0.4236	0.4594
	NBLR	0.3558	0.2412	0.1706	0.0973	0.3964	0.1064	0.1692	0.2409
100	RR	2.6048	1.1844	0.3673	0.3211	1.6308	0.3127	0.2995	0.3082
	LR	0.1854	0.2995	0.1845	0.2892	0.1939	0.2211	0.0767	0.1234
	BLR	0.7718	0.6267	0.5000	0.3084	0.6213	0.4109	0.3151	0.2490
	NBLR	0.1066	0.1146	0.1300	0.1577	0.1417	0.1853	0.0239	0.0814

Table 2: Median Mean Absolute Deviation along each method with different sample sizes

Methods	MMAD		
	N=20	N=50	N=100
RR	4.2308	3.0917	3.3232
LR	1.1963	0.7928	0.5108
BLR	2.9015	1.0753	0.6859
NBLR	2.3100	0.8905	0.5748

Trace plot as convergence diagnose tool, indicates from Figure (3) that the MCMC samples of the posterior distribution of regression coefficients convergence to stationary distribution (the vector of true parameter values). Also, the trace plots Figure shows no flat bits and does not suffer from slow mixing.

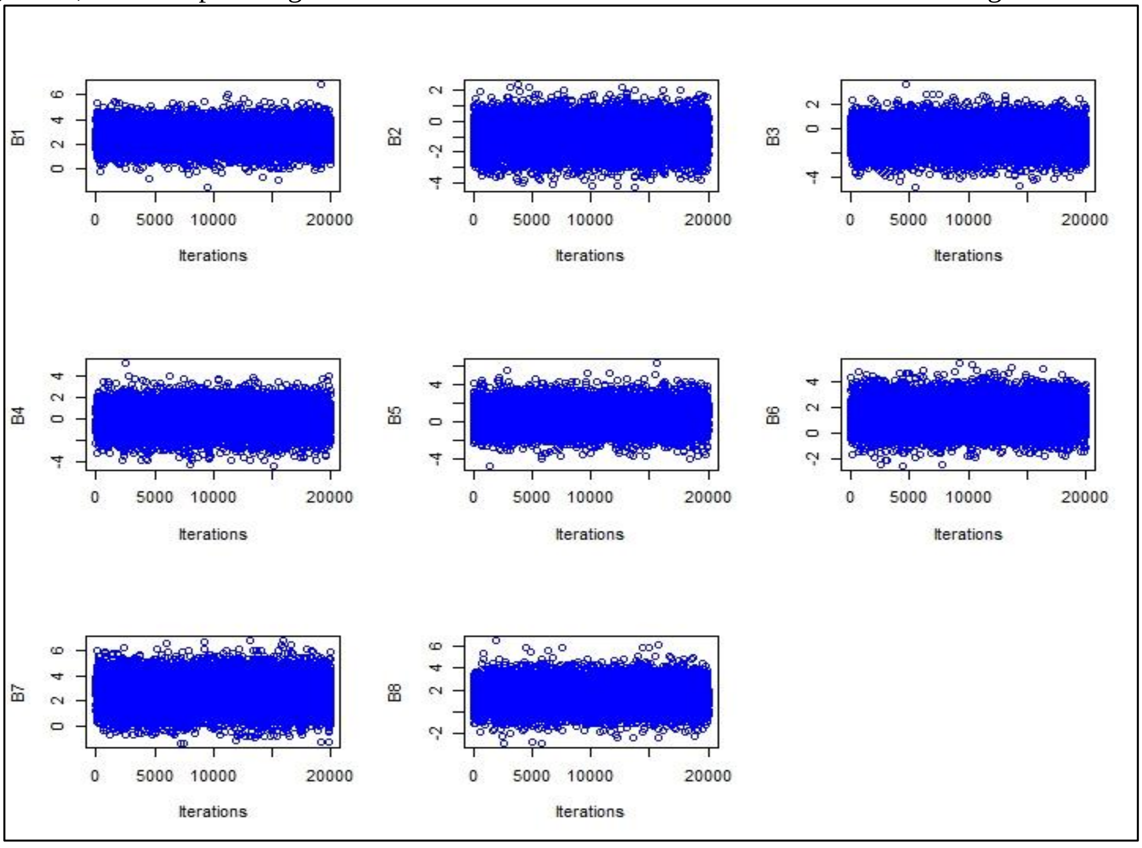

Figure 1. Trace plots of $\beta_1 - \beta_8$ of the Posterior parameter estimates

Figure 4 checks the distributions of the parameter estimates with sample sizes (20,50,100) respectively and it is clearly that the distribution of the parameter is fairly follow the normal distribution for each regression method.

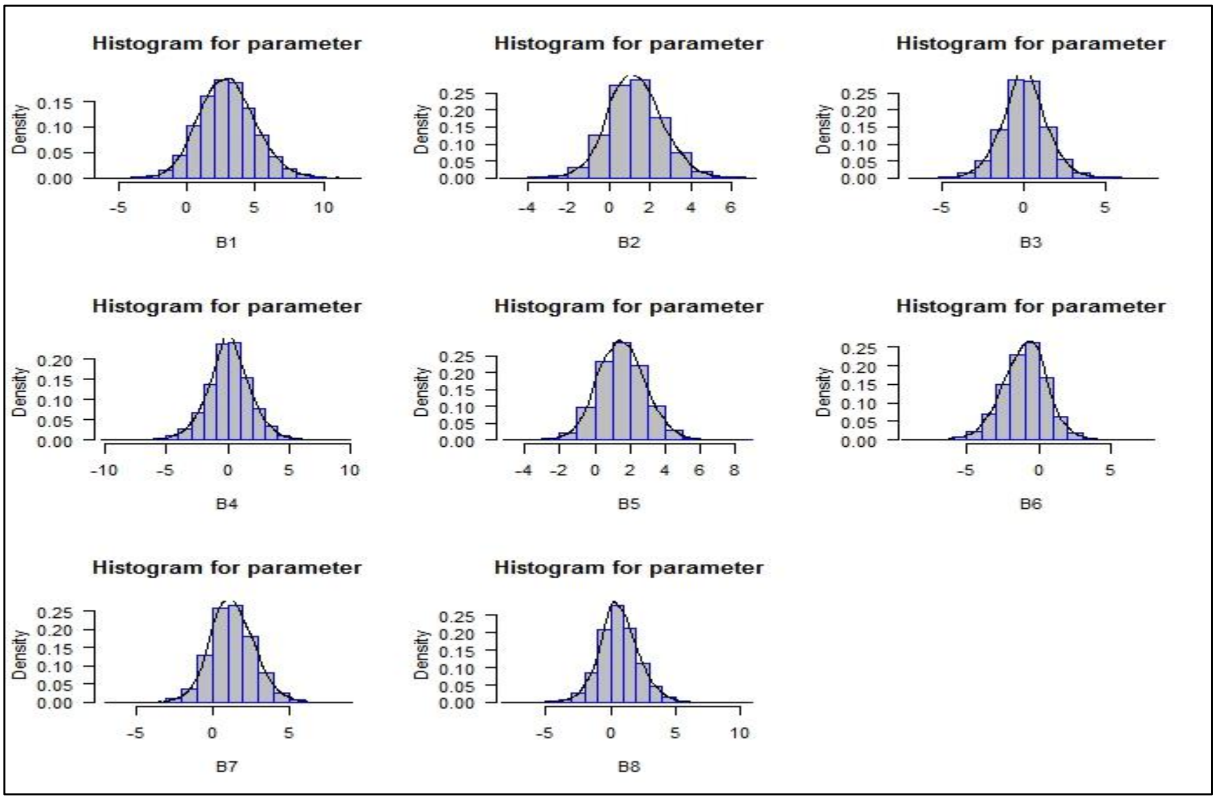

Figure 2. Histograms of parameter estimates $\beta_1 - \beta_8$ distributions

The boxplots from Figures 5,6,7 with different sample sizes (20,50,100) exhibits that the proposed NBL regression method does not suffer from the dispersion of the parameter estimates compared with the other methods, also we can see that the median is closely to the true parameter value dotted in the red horizontal line.

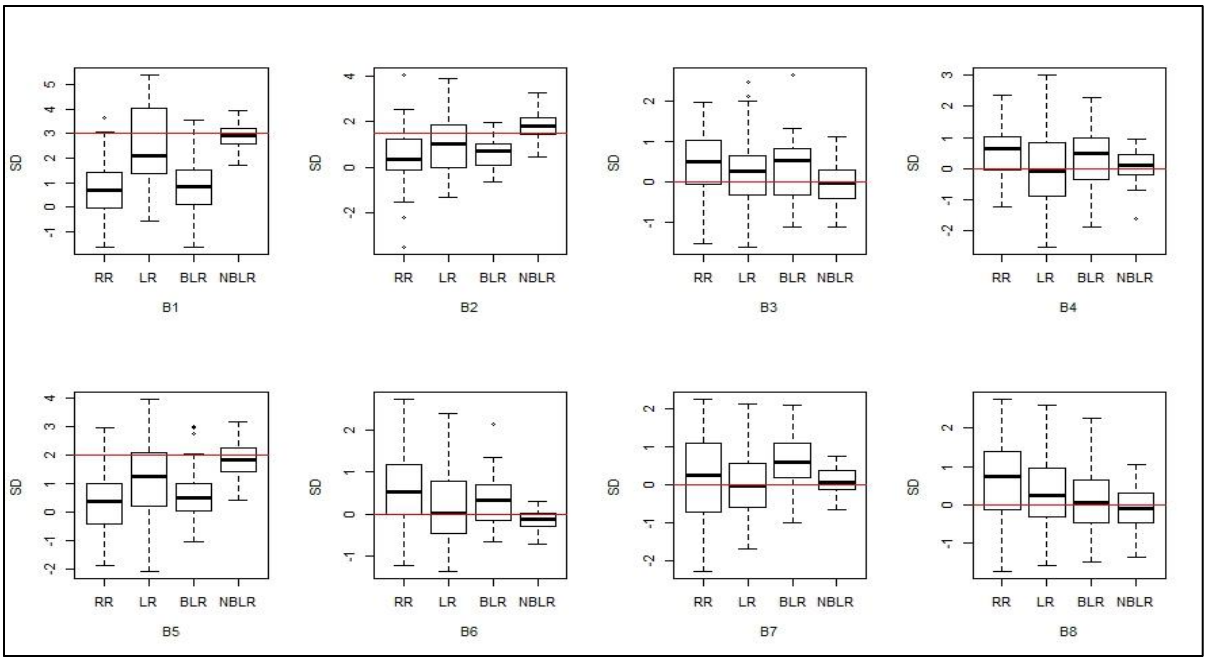

Figure 3. Comparison of performance between different method along with $\beta_1 - \beta_8$
and sample size 10

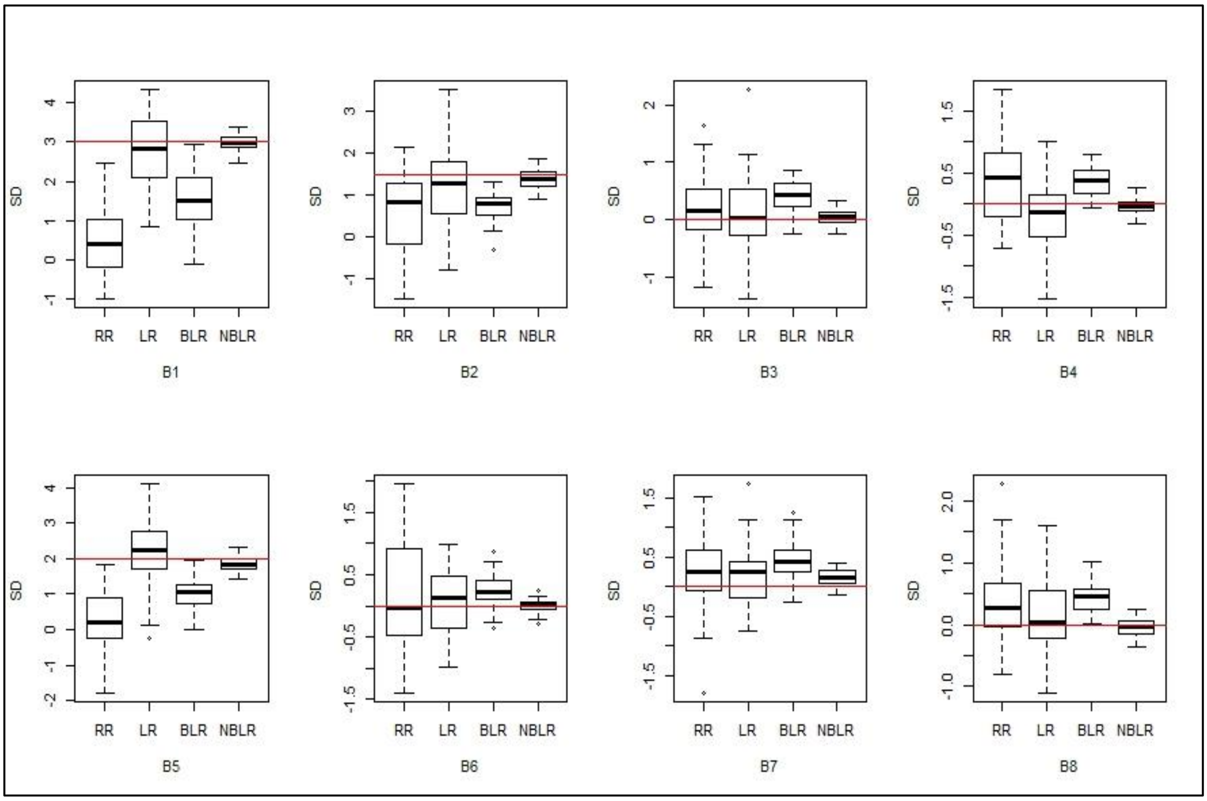

Figure 4. Comparison of performance between different method along with $\beta_1 - \beta_8$ and sample size 50

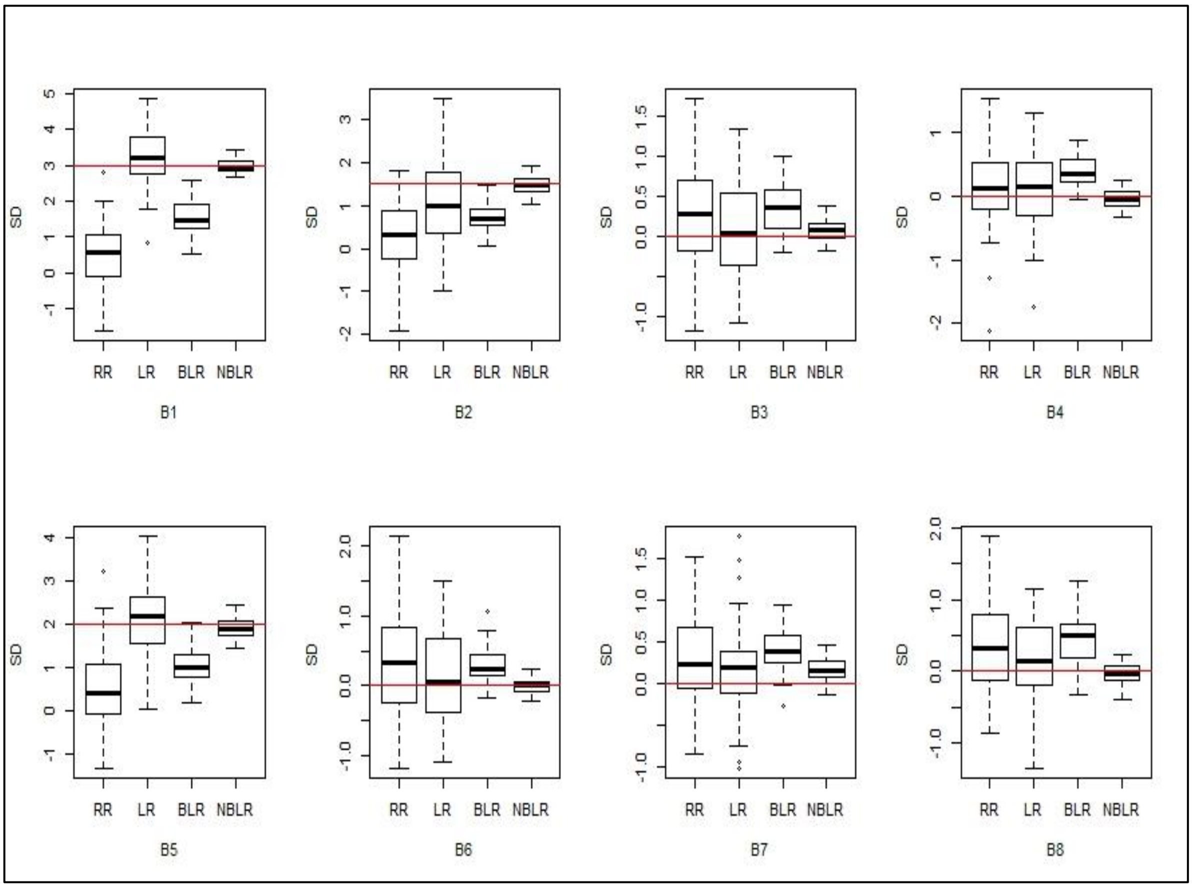

Figure 5. Comparison of performance between different method along with $\beta_1 - \beta_8$ and sample size 100

Example 2

Similar to example 1, we use β =(5,0,0,0,0,0,0,0), $\rho = 0.5$, $\sigma = 2$. Table 3 shows the bias of the estimates of the different methods with increasing of sample size, we can see the less bias attained in NBL. Consequently, the performance of the four methods improves. As well, Table 4 indicates that the less MMAD is in the proposed method NBL, followed by LR. Figure 3 shows that the distributions of the parameter estimates with different sample sizes (20,50,100) follows normal distribution for each regression method. Trace plots in Fig.6 shows that the MCMC samples convergence to stationarity, which means the well mixing of The MCMC samples and the target distribution. Figures (7,8,9) shows the boxplots with different sample sizes (20,50,100), clearly that the proposed NBL regression method does not suffer from the variability of the parameter estimates compared with the other methods, and clearly the median in the boxplot is closely to the true parameter value dotted in the red horizontal line.

Table 3. The bias of parameter estimates along the path solution of each method with different sample sizes

N	Methods	Bias $\hat{\beta}_1$	Bias $\hat{\beta}_2$	Bias $\hat{\beta}_3$	Bias $\hat{\beta}_4$	Bias $\hat{\beta}_5$	Bias $\hat{\beta}_6$	Bias $\hat{\beta}_7$	Bias $\hat{\beta}_8$
20	RR	4.6403	0.1934	0.1918	0.1876	0.2098	0.1514	0.1897	0.1584
	LR	0.1733	0.0406	0.2251	0.1176	0.1069	0.0302	0.1693	0.1210
	BLR	2.7769	0.1829	0.0004	0.1634	0.3170	0.1152	0.3429	0.1317
	NBLR	0.1021	0.0066	0.2570	0.0640	0.0834	0.0378	0.1380	0.0802
50	RR	4.6284	0.2140	0.2332	0.2768	0.1977	0.2117	0.2390	0.2523
	LR	0.1329	0.0633	0.1021	0.0926	0.0904	0.0626	0.0324	0.1701
	BLR	1.4222	0.1470	0.0934	0.2734	0.0829	0.1068	0.1723	0.2083
	NBLR	0.0444	0.0289	0.0839	0.0503	0.0535	0.0458	0.0134	0.1296
100	RR	4.6411	0.1848	0.1810	0.2356	0.2116	0.2088	0.1818	0.1991
	LR	0.1008	0.0370	0.0215	0.0398	0.0365	0.0531	0.0507	0.0614
	BLR	0.6309	0.0256	0.0309	0.1115	0.1212	0.1081	0.0901	0.0660
	NBLR	0.0265	0.0270	0.0082	0.0374	0.0070	0.0161	0.0449	0.0232

Table 4. Median Mean Absolute Deviation along each method with different sample sizes

Methods	MMAR		
	N=20	N=50	N=100
RR	2.5620	3.2542	2.8931
LR	0.5401	0.5310	0.3813
BLR	1.2383	0.7635	0.4009
NBLR	0.483	0.466	0.2706

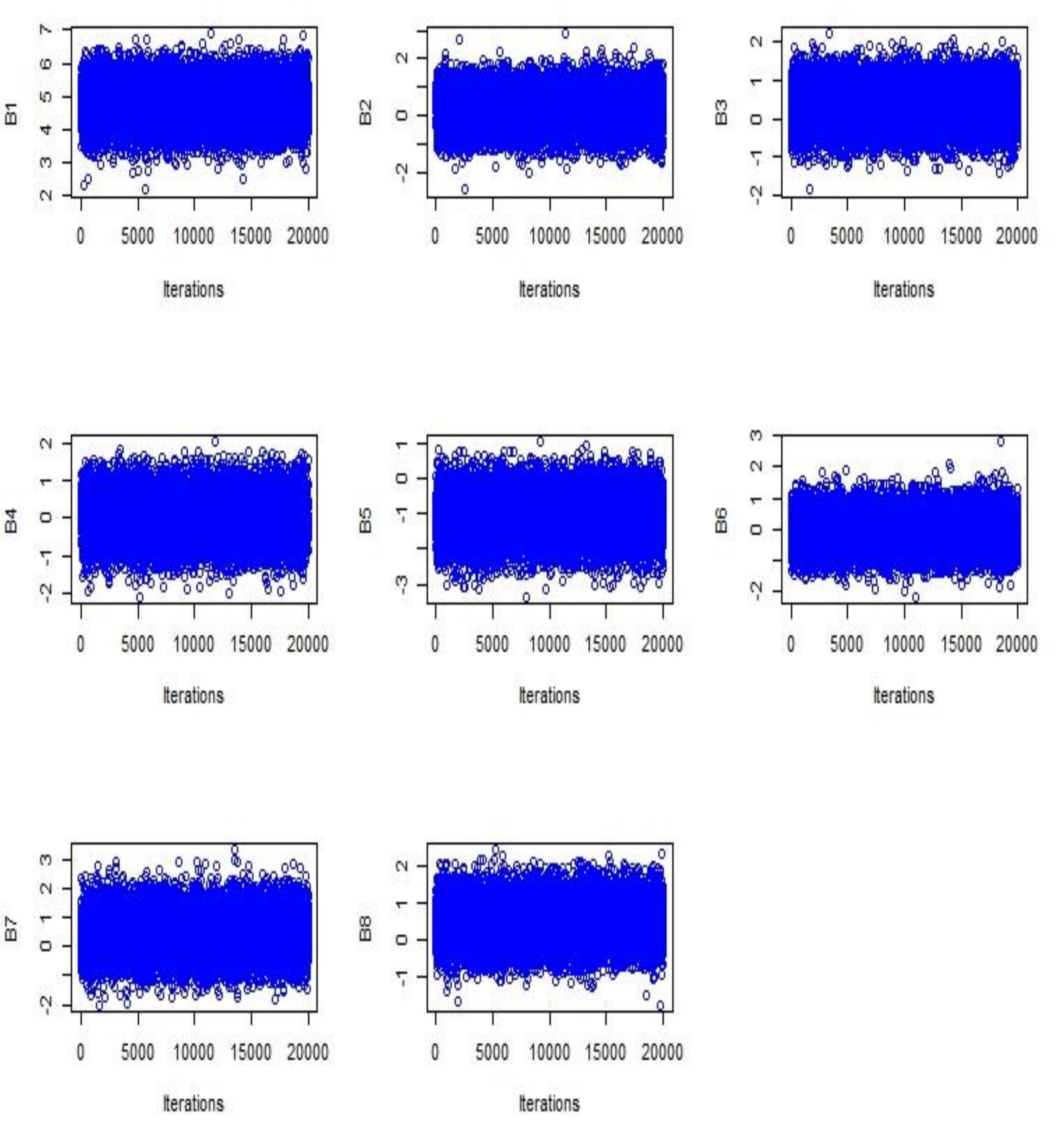

Figure 6. Trace Plots of $\beta_1 - \beta_8$ of the Posterior parameter estimates

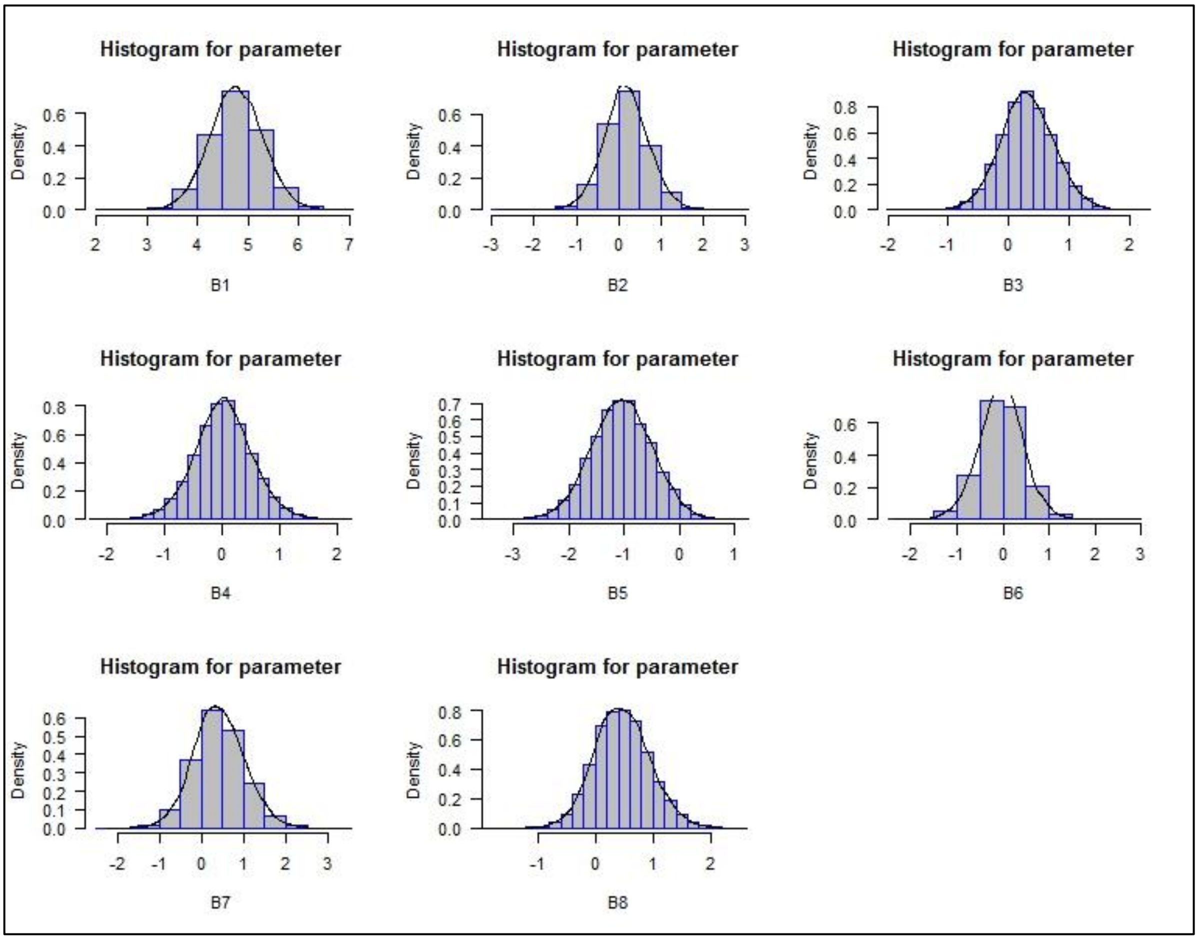

Figure 7. Histograms of parameter estimates $\beta_1 - \beta_8$ distributions

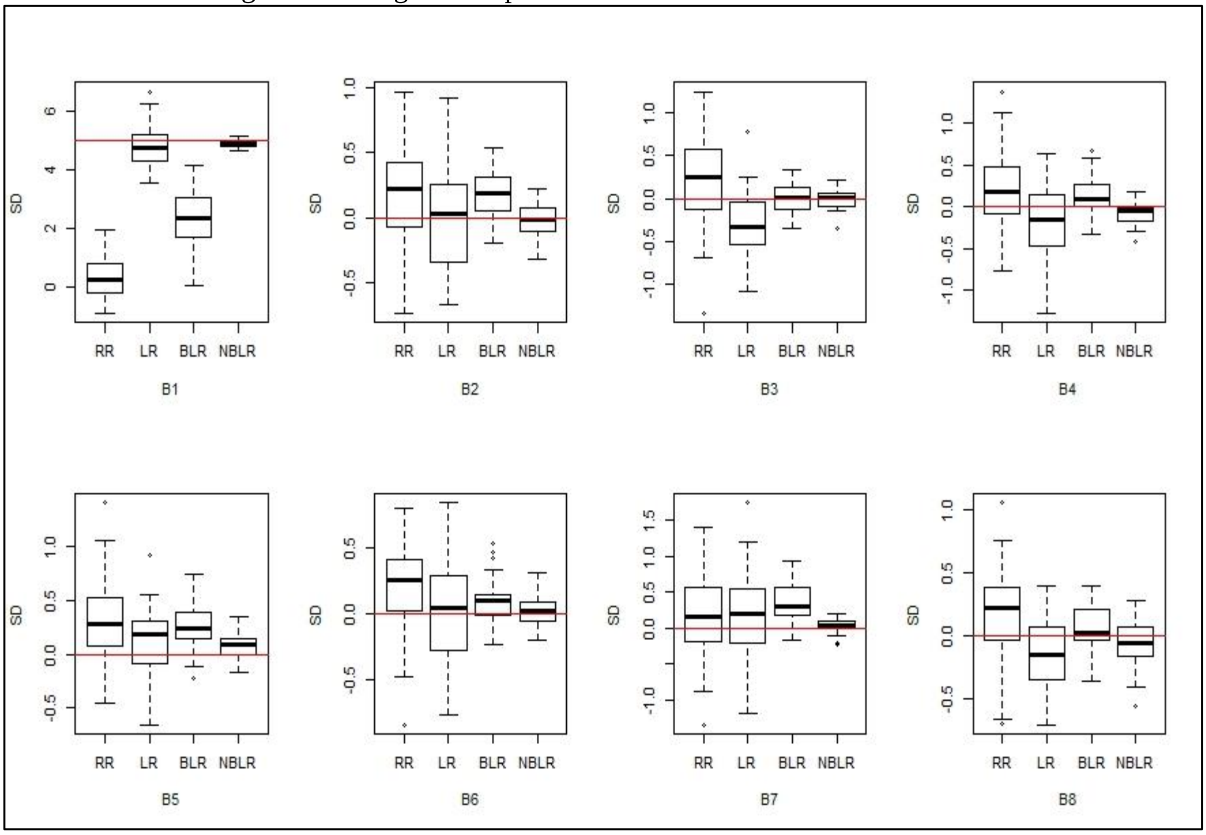

Figure 8. Comparison of performance between different method along with $\beta_1 - \beta_8$ and sample size 10

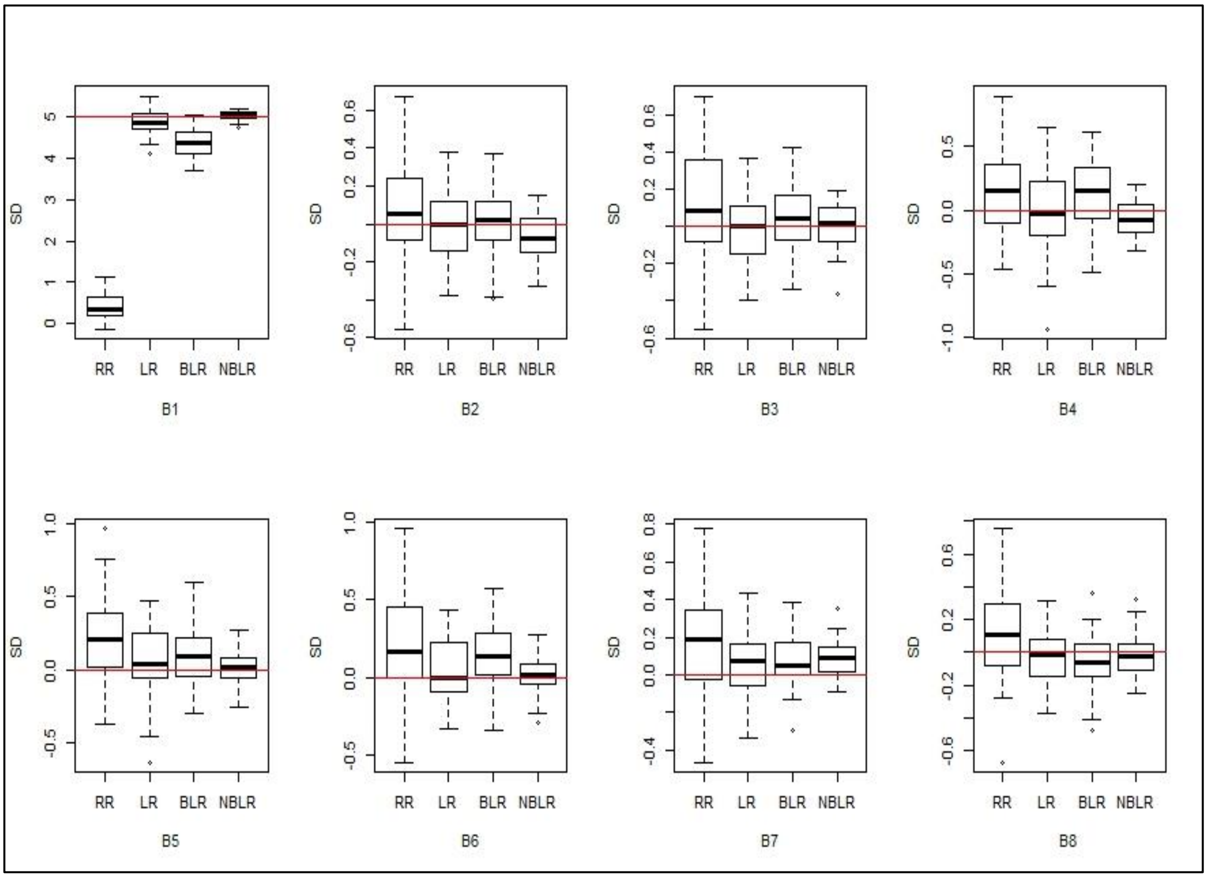

Figure 9. Comparison of performance between different method along with $\beta_1 - \beta_8$ and sample size 50

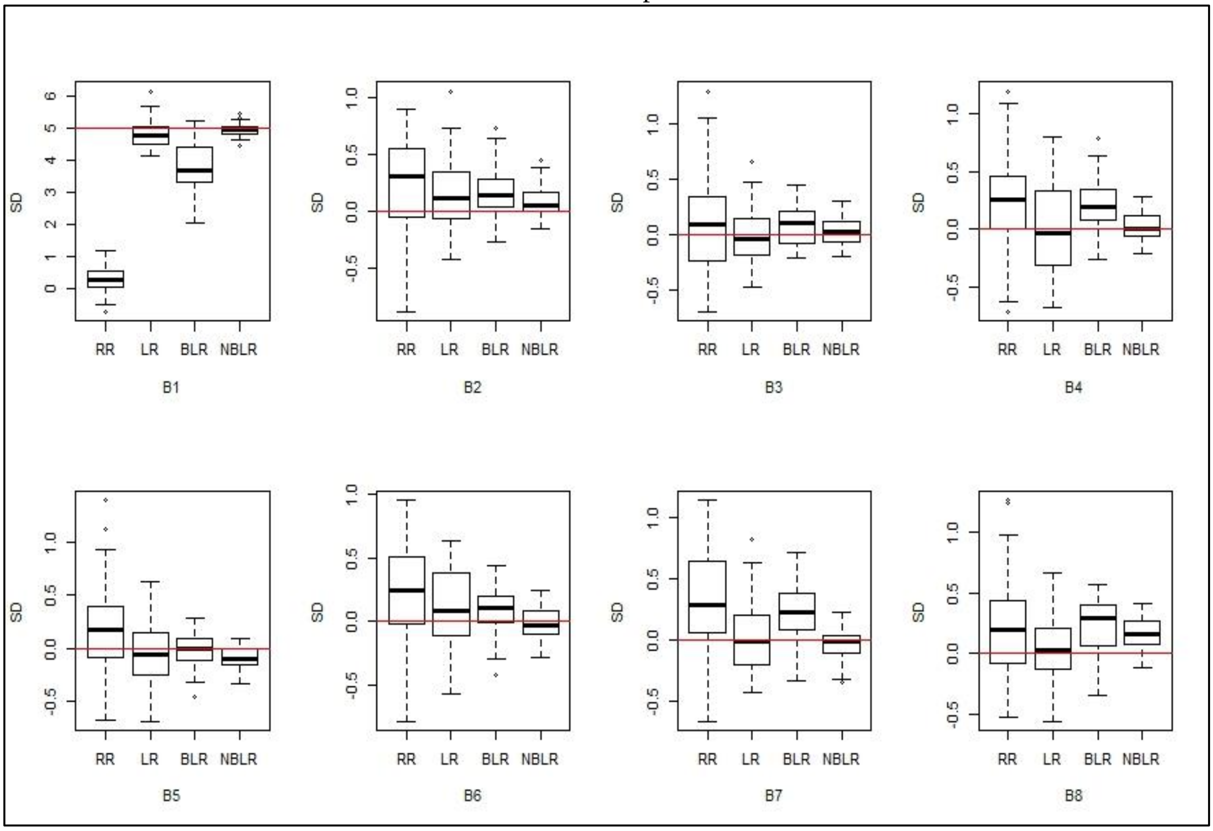

Figure 10. Comparison of performance between different method along with $\beta_1 - \beta_8$ and sample size 100

6. Analysis with Boston housing data

In 1978 Harrison and Rubinfeld study the relationship between the demand on clean air as response variable and 13 predictor variables; per capita crime rate by town (crim), proportion of residential land zoned for lots over 25,000 sq.ft(zn), proportion of non-retail business acres per town(indus), Charles River dummy variable (= 1 if tract bounds river; = 0 otherwise) (chas), nitric oxides concentration (parts per 110 million) (nox), average number of rooms per dwelling (rm), proportion of owner-occupied units built prior to 1940 (age), weighted distances to five Boston employment centers (dis), index of accessibility to radial highways (rad), full-value property-tax rate per USD 10,000 (tax), pupil-teacher ratio by town(ptratio, $1000\left(B-0.63\right)^2$ where B is the proportion of blacks by town (b), lower status of the population (1stat).These data collected in Boston state which contains 506 census area (observation).

Since the predictor variables have different measure units, we standardized their values and then centered the response variable values. The estimated penalty parameter λ value is the posterior mean when the prior distribution is gamma with parameters (a=1, b=0.1) in Gibbs sampler algorithm, also we use (k=10) fold cross validation in classical lasso to select the regularization parameter, see [5]. We implement our proposed Bayesian conditional posterior distributions to estimate Lasso parameter in (1) and compared with (Ridge, Lasso, and Bayesian Lasso) regressions. Table (6) contains the estimates of 13 coefficients of the predictor variable and compared it with (Ridge, Lasso, and Bayesian Lasso) regressions, clearly the NBL is outperform better than RR,LR, and BL in terms of the sparsity, i.e., the NBL picks approximately (%53) of zeros coefficients (7) followed by the classical lasso. As well as, Table (7) contain the values of the Residual Mean Squares Error (RMSE) for regression methods (RR, LR, BL, and NBL) and exhibits than NBL performs the better than other regression models ,followed by LR. The data is available in R package MASS.

Table 5. Standardized parameter estimates for RR, LR, BL, and NBL

Methods	crim	zn	indus	chas	nox	rm	age	dis	rad	tax	ptratio	black	lstat
RR	-0.030	0.010	-0.045	0.684	-2.219	0.867	-0.008	0.035	-0.024	-0.002	-0.187	0.003	-0.083
LR	-0.083	0.035	0.000	2.630	-14.686	3.963	0.000	-1.250	0.183	-0.007	-0.905	0.009	-0.522
BLR	-0.083	0.050	-0.060	0.294	-0.133	1.439	0.005	-0.975	0.252	-0.013	-0.776	0.009	-0.703
NBLR	0.000	0.047	0.000	2.602	-12.060	4.524	0.000	0.000	0.255	0.000	0.000	0.011	0.000

Table 6. Comparison performance of RR, LR, BL, and NBL methods based on MMAD

Methods	RMSE
RR	24.8303
LR	31.9082
BLR	40.3455
NBLR	15.1699

7. Conclusions

In this paper, new Bayesian lasso method for variable selection have proposed based on the Laplace prior distribution as scale mixture of normals mixing with Rayleigh distribution on their variances. New hierarchical model representation and new Gibbs sampler algorithm have developed. Two simulation examples conducted to explore the path solution of the proposed method, as well as we performed real data analysis. The results of simulation presented some evidence of Comparable of the proposed method to the others methods, but with outperform of the new Bayesian method in the real data in views of sparsity.

References

[1] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological),* Vol. 58, No. 1, pp. 267-288, 1996.

[2] B. Efron, T. Hastie, I. Johnstone, and, R. Tibshirani," Least angle regression," *The Annals of statistics*, Vol. 32, No. 2, 407–499, 2004.

[3] T. Park, and, G. Casella , " The Bayesian lasso," *Journal of the American Statistical Association*, Vol. 103, No. 482, 2008

[4] D.F. Andrews, and C.L. Mallows ,"Scale mixtures of normal distributions," *Journal of the Royal Statistical Society. Series B (Methodological),*36, 99-102, 1974.

[5] H. Mallick, and N. Mallick, "A new Bayesian lasso," *Statistics and its interface*, vol.7(4), 571, 7(4): 571–582,2014.

[6] H. K. Abbas, " Bayesian Lasso Tobit regression," *Journal of AL-Qadisiyah for computer science and mathematics*, vol.11(2): 2521-3504, 2019.

[7] S. Kotz, , T., Kozubowski, and K. Podgorski , "The Laplace Distribution and Generalizations.," *Springer* I*SBN*: 0-8176-4166-1, 2001.

[8] K. C. Chu, "Estimation and detection in linear systems with elliptical errors, " *IEEE Trans. Auto. Control*, 18, 499(505), 1973.

[9] R. Heide , "the safe-Bayesian lasso. Master Thesis mathematical institute," *Leiden University,* 52(1) :374-393, 2007.

[10] R. S. Chhikara and J. L. Folks, “The Inverse Gaussian Distribution: Theory: Methodology, and Applications," *Marcel Dekker, Inc*. 1989.

[11] R. Kohavi , "A study of cross-validation and bootstrap for accuracy estimation and model selection.," *Morgan Kaufmann Publishers Inc. San Francisco,* CA, USA -21137-1143,1995.

Appendix A

$$\frac{1}{2a}\exp\left[-\frac{|z|}{a}\right]=\int_0^\infty \frac{1}{\sqrt{2\pi s^2}}\ e^{-z^2/2s^2}\ \frac{s}{a}\ e^{-s^2/2a}\quad ds$$

The RHS of above equation can be written as,

$$\frac{1}{a}\int_0^\infty \frac{s}{\sqrt{2\pi s^2}}\exp\left[-\frac{(s^2)^2+\sqrt{a^2}\ z^2}{2as^2}\right]\ ds$$

$Let \sqrt{a}=b, then$

$$\frac{1}{b^2}\int_0^\infty \frac{s}{\sqrt{2\pi s^2}}\exp\left[-\frac{(s^2-b|z|)^2}{2b^2s^2}-\frac{2s^2b|z|}{2b^2s^2}\right]\ ds$$

$$\frac{1}{b^2}\exp\left[-\frac{|z|}{b}\right]\int_0^\infty \frac{s}{\sqrt{2\pi s^2}}\,exp\left[-\frac{(s^2-b|z|)^2}{2b^2s^2}\right]\ ds$$

$$Let\ \mu=b,|z|, and\ \ \lambda=|z|^2$$

$$\frac{1}{b^2}\exp\left[-\frac{|z|}{b}\right]\int_0^\infty \frac{1}{\sqrt{2\pi}}\,exp\left[-\frac{(s^2-\mu)^2}{2b^2s^2}\right]\ ds$$

$$\frac{1}{b^2}\exp\left[-\frac{|z|}{b}\right]\int_0^\infty \frac{1}{\sqrt{2\pi}}\,exp\left[-\frac{\lambda(s^2-\mu)^2}{2\mu^2s^2}\right]\ ds$$

$$Let\ s^2=z, then$$

$$\frac{1}{2b^2}\exp\left[-\frac{|z|}{b}\right]\int_0^{\infty}\frac{1}{\sqrt{2\pi z}}\exp\left[-\frac{\lambda(z-\mu)^2}{2\mu^2 z}\right]\ dz$$

$$\frac{1}{2b^2}\exp\left[-\frac{|z|}{b}\right]\int_0^{\infty}\frac{z\sqrt{\lambda}}{\sqrt{\lambda}\ \ \sqrt{2\pi}z^3}\exp\left[-\frac{\lambda(z-\mu)^2}{2\mu^2 z}\right]\ dz$$

$$\frac{1}{\sqrt{\lambda}\ \ 2b^2}\exp\left[-\frac{|z|}{b}\right]\int_0^{\infty} z\ \sqrt{\frac{\lambda}{2\pi z^3}}\exp\left[-\frac{\lambda(z-\mu)^2}{2\mu^2 z}\right]\ dz$$

Where $\int_0^{\infty} z\ \sqrt{\frac{\lambda}{2\pi z^3}}\exp\left[-\frac{\lambda(z-\mu)^2}{2\mu^2 z}\right]dz = \mathrm{E}(z), and\ z \sim IG(\mu, \mu^3/\lambda),$

$$\frac{1}{\sqrt{\lambda}\ \ 2b^2}\exp\left[-\frac{|z|}{b}\right]\mu$$

$$\frac{1}{|z|\ \ 2b^2}\exp\left[-\frac{|z|}{b}\right]|z|b$$

$$\frac{1}{2b}\exp\left[-\frac{|z|}{b}\right]$$

Hence the proof (5).