Periodicals of Engineering and Natural Sciences ISSN 2303-4521
Vol. 8, No. 2, June 2020, pp.1131-1140

Bayesian extensions on Lasso and adaptive Lasso Tobit regressions

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

ABSTRACT

Since lasso method launched, a lot of applications and extensions were run on it which made it to become deeply widely used in various discipline. In this paper, we proposed the Scale Mixture of Normals mixing with Rayleigh (SMNR) distribution on their variances to represent the double exponential distribution. Hierarchical model formula have derived with Gibbs sampler for SMNR. The proposed models; Bayesian Tobit Adaptive Lasso (BTAL) and Bayesian Tobit Lasso (BTL) models are illustrated using simulation example and a real data example through the prediction accuracy using the estimated relative efficiency with different sample. This is the first work that discussed regularization regression models under SMNR.

Keywords: Regularization regression, Rayleigh distribution, Tobit Model, Gibbs sampling, Hierarchical formulation

Corresponding Author:

Fadel Hamid Hadi Alhusseini
Department of Statistics, University of Al-Qadisiyah
Al-Qadisiyah, Iraq
E-mail: fadel.alhusiny@qu.edu.iq

1. Introduction

Regression models with predictors (k) exceeding the sample size ($k > n$) and/or when the matrix ($\acute{X}X$) be near singular produces less meaningful predicted model in the view of the interpretable and the prediction accuracy. Under this situation [17], introduced the Lasso penalized Residual Sum of Squares (RSS) as the following minimization problem,

$$_{argmin}\,(\ell_2(y - X\underline{\beta}))^2 + \lambda \sum_{j=1}^{k} |\beta_j|, \tag{1}$$

Where $\lambda > 0$, is the shrinkage parameter that controls the shrinkage amount in coefficients (β) which is the vector of unknown parameters. The second term in (1) is the penalty constraint that placed to reduce the number of the parameters in the following linear regression,

$$y = X\beta + \epsilon, \tag{2}$$

Where $y = (y_1, \dots \dots \dots . y_n)^{\prime}$ is vector of the centered response variable, $X\ is\ n \times k$ matrix of standardized predictors (if they have different units), and $\epsilon \sim N(0, \sigma^2)$.

The second term in (1) letting the lasso estimate of the irrelevant predictors shrunk exactly to zero [17], this feature called sparsity (variable selection) and then the regression model be more interpretable model. Lasso shrinkage produced estimator with a little bit bias and in the same time reducing the variance which improve the prediction accuracy. Consequently, the minimization problem in (1) can be viewed as the shrinkage estimation and sparsity in the same time. [15] introduced the focused lasso to deal with predictors that have parameters which are similar and ordered in some significate way. [8] suggest that the solution path of (λ) following algorithm called Least Angle Regression (LAR) which obtaining the lasso estimates of (β_j). [18] introduced the elastic net regularization method that combined the lasso and ridge penalty functions, as well works with collinearity predictors. [17] produced the group lasso variable selection method as a generalization of the classical lasso to select known groups of variables that have combined effect on the dependent variable. [19] proposed the following minimization problem to address the amount of bias in lasso estimates,

$$\hat{\beta}_{AL} = {}^{argmin}{}_{B}(\ell_2(y - X\underline{\beta}))^2 + \lambda \sum_{j=1}^{k} w_j|\beta_j| \qquad (3)$$

The estimator $\hat{\beta}_{AL}$ in (3) is called Adaptive Lasso (AL) estimator ,where $\lambda \geq 0$, is the shrinkage parameter , and w_j are the weights $w_j > 0$, which defined as $\widehat{w}_{j=1/|\alpha_j|^{\eta}}$, where α_j is an initial value of $\hat{\beta}_j$ and $\eta > 0$.

The adaptive lasso penalty function in Tobit quantile regression introduced by [3]. Newly, Bayesian lasso became more popular when [14], stated that the lasso estimate can be obtained as Bayesian posterior mode conjugated with independent Laplace priors for (β_j),

$$g(\beta_j) = \frac{1}{2v} e^{-\frac{\beta_j}{v}} \qquad (4)$$

Here $(v = {}^{1}/_{\lambda})$. The lasso estimate of (β) in (4) is the posterior mode estimate.

Gibbs sampler algorithm easily employed to use to obtain the Bayesian lasso estimates. Park and Casella [13] provided Bayesian analysis of the linear regression when the parameter (β) is distributed as double exponential density with Scale Mixture of Normals (SMN), mixing with exponential density on their variances. [5] compared the classical lasso results with the Bayesian Lasso results using certain Full Bayesian condition with hierarchical representation model. [12] proposed the following new minimization problem of the Bayesian lasso by assuming that the regularization parameter (λ) takes different (λ_j) for each parameter (β) instead of the same (λ) for every parameter as in lasso method.

$${}^{argmin}{}_{B}(\ell_2(y - X\underline{\beta}))^2 + \lambda \sum_{j=1}^{k} \lambda_j|\beta_j| \qquad (5)$$

[9] introduced a Bayesian regularization method that analogue to the adaptive lasso method whereby allowing to the scale parameter (λ) in the mixing density of the scale mixture of normals to vary from parameter to parameter.

In many practical situations, researcher adopting certain statistical techniques to deal with censored samples. [16] introduced the Tobit regression to deal with the data that experiencing left censored response variable (y) , which is defined as follows

$$y_i = \begin{cases} \acute{x}\beta + \epsilon_i & if \quad \acute{x}\beta + \epsilon_i > 0 \\ 0 & if\ otherwise \end{cases} \qquad (6)$$

Where $\epsilon \sim N(0, \sigma^2), y_i^* = \acute{x}\beta + \epsilon_i$ is the latent variable ,and $y_i = \max\{0, y_i^*\}$,

In [4] studied the Bayesian Tobit quantile regression when the regression parameter (β). have g-prioe density under ridge coefficient. Also, the Bayesian analysis of the Tobit quantile regression under elastic net penalty function was introduced in [5]. The Bayesian Tobit quantile regression with the new SMN was studied [6]. The Lasso and adaptive lasso Tobit regression models with new mixture of uniforms, mixing with standard exponential density as prior representation of the regression parameter (β), but the prior is not conditioning on (σ^2) was studied in [1].

In this paper, followed of [2], [13],[11], [1], we proposed a new hierarchical model formulation for Bayesian lasso Tobit and Bayesian adaptive lasso Tobit regressions considering that the prior distribution of the regression parameter (β) follows the Laplace distribution 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.

[13] addressed the problem of multiple modes in the posterior distribution ($\pi(\beta, \sigma^2)$ with the prior in (4) through conditioning on (σ^2) to assure the unimodal of the (β_j) posterior distribution, i.e.,

$$\pi(\beta/\sigma^2) = \prod_{j=1}^{k} \frac{\lambda}{2\sqrt{\sigma^2}} exp\left[-\frac{\lambda|\beta_j|}{\sqrt{\sigma^2}}\right] \tag{7}$$

Following [13] our new full Bayesian analysis consider the conditional Laplace prior form as follows,

$$\pi(\beta/\sigma^2) = \prod_{j=1}^{k} \frac{\lambda}{2\sigma^2} exp\left[-\frac{\lambda|\beta_j|}{\sigma^2}\right] \tag{8}$$

2. Scale mixture of normal mixing Rayleigh

Following [10] and [7], the scale mixture of normal s mixing the parameter σ on their variances is,

$$\mathrm{f(x)} = \int_0^\infty \frac{1}{\sigma} \emptyset \left[\frac{x}{\sigma}\right] h(\sigma) d\sigma \tag{9}$$

Hence, f(x) is symmetric about zero and unimodal function. So, based on (9) and

if $z/s \sim N(\mu = 0, s^2)$, $s \sim$ rayleigh (a) , then $z \sim$ Laplace ($\mu = 0, a$),

$$\frac{1}{2a} \exp\left[-\frac{|z|}{a}\right] = \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2}}\ e^{-z^2/2s^2} \frac{s}{a} e^{-s^2/2a} ds \tag{10}$$

Appendix A contain the proof of (10).

Let $a = {}^{\sigma^2}\!/_{\lambda}$, $z = \beta$, and $s = \sigma\sqrt{\tau}$, then (10) can be written as,

$$\frac{\lambda}{2\sigma^2} \exp\left[-\frac{\lambda|\beta|}{\sigma^2}\right] = \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2\tau}}\ e^{-\beta^2/2\sigma^2\tau} \frac{\lambda}{2} e^{-\lambda\tau/2} d\tau \tag{11}$$

The prior density (10) is conditioning on σ^2 to assure a unimodal posterior distribution.

3. Hierarchical Bayesian Lasso Tobit model

The new hierarchical full model formulation under the proposed scale mixture (10) defined as follows,

$$
\begin{aligned}
& y_{i=max\{0,y_i^*\}}, \qquad i=1,2,\dots\dots,n \\
& y_i^*|X,\beta,\sigma \sim N_n(X\beta,\sigma^2 I_n), \\
& \beta|\sigma^2,\tau_1,\dots\dots,\tau_k \sim N_K(0,\sigma^2 D_\tau), \\
& D_\tau = diag(\tau_1,\dots\dots,\tau_k), \\
& \sigma^2,\tau_1,\dots\dots,\tau_k \sim \pi(\sigma^2)d(\sigma^2)\prod_{j=1}^{k}\frac{\lambda}{2}exp\left\{-\frac{\lambda\tau_j}{2}\right\}d\tau_j \\
& \sigma^2,\tau_1,\dots\dots,\tau_k > 0,
\end{aligned}
\tag{12}
$$

The conditional prior $\pi(\beta/\sigma^2)$ in (8) can be obtained after integrating out $\tau_1,\dots\dots,\tau_k$ in (12) Also, we can use prior density $(1/\sigma^2)$, or any inverse gamma for $\pi(\sigma^2)$ to maintain the conjugacy in the SMNR. The Gibbs sampler algorithm can be implemented through the following full joint density:

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

$$
\pi(\sigma^2)\frac{\gamma^a}{\Gamma_a}(\sigma^2)^{-a-1}e^{-\gamma/\sigma^2}\prod_{j=1}^{k}\frac{1}{\sqrt{2\pi\sigma^2\tau_j}}\,e^{-\beta_j^2/2\sigma^2}\;\frac{\lambda}{2}e^{-\lambda\tau_j/2} \tag{13}
$$

The full conditional for β can written as follows:

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

Recalling the multivariate normal distribution $X\sim(\mu,\Sigma)$,then we can say that β follows the multivariate normal density with mean $C^{-1}X^{'}y$ and variance σ^2C^{-1} . The full conditional distribution for σ^2 is the conditionally inverse gamma distribution,

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

The Gibbs sampler distribution for τ is the inverse Gaussian distribution,

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

Then, we can say that $(1/\tau_j)\sim IG\left(\sqrt{\frac{\lambda\sigma^2}{\beta_j^2}},\lambda\right)$, where λ is the shape parameter and $\sqrt{\frac{\lambda\sigma^2}{\beta_j^2}}$ is the mean (location) parameter. Choosing the regularization parameter λ, based on [13] the distribution of prior λ is the Gamma (γ,δ) , then the conditional posterior distribution of λ is

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

This is again the gamma distribution, i.e., $\lambda\sim\text{Gamma}(p+\gamma,\frac{1}{2}\sum_{j=1}^{k}\tau_j+\delta)$.

4. Hierarchical Bayesian adaptive Lasso Tobit model

Based on the adaptive lasso estimator of [12] in (5) the conditional Laplace prior which can be represented as the SMNR (11), the hierarchical formulation for the Bayesian adaptive lasso Tobit is defined as follows,

$$y_{i=max\{0,y_i^*\}}, \qquad i=1,2,\dots\dots,n$$

$$\begin{aligned}
&y_i^*|X,\beta,\sigma \sim N_n(X\beta,\sigma^2 I_n),\\
&\beta|\sigma^2,\tau_1,\dots\dots,\tau_k \sim N_K(0,\sigma^2\, D_\tau),\\
&D_\tau = diag(\tau_1,\dots\dots,\tau_k), \qquad\qquad (14)\\
&\sigma^2,\tau_1,\dots\dots,\tau_k \sim \pi(\sigma^2)d(\sigma^2)\prod_{j=1}^{k}\frac{\lambda_j}{2}exp\left\{-\frac{\lambda\tau_j}{2}\right\}d\tau_j\\
&\sigma^2,\tau_1,\dots\dots,\tau_k > 0,
\end{aligned}$$

The conditional prior $\pi(\beta/\sigma^2)$ in (8) can be obtained after integrating out$\tau_1,\dots\dots,\tau_k$ in (14). Also, we can use prior density $1/\sigma^2$,or any inverse gamma for $\pi(\sigma^2)$ to maintain the conjugacy in the SMNR. The Gibbs sampler algorithm is implementing with the following hierarchical model,

$$\begin{aligned}
&y|X,\beta,\sigma^2 \sim N_n(X\beta,\sigma^2 I_n),\\
&\beta|\tau\sigma^2 \sim N(\left(X'X+D_\tau^{-1}\right)^{-1}X'y,\sigma^2\left(X'X+D_\tau^{-1}\right)^{-1}), \qquad\qquad (15)\\
&\tau_j^{-1}\Big/\sigma^2 \sim IG\,(\frac{\sqrt{\lambda_j\sigma}}{|\beta_j|}\,,\lambda_j),\\
&\sigma^2 \sim Inverse-Gamma\,\left(\frac{n-1}{2}+\frac{p}{2}+a,(y-X\beta)'(y-X\beta)/2+\beta' D_\tau^{-1}\beta/2+\gamma\right),
\end{aligned}$$

Following [12] the full conditional posterior distribution of λ_j is defines as

$$\lambda_j \sim Gamma(1+\gamma,\tau_j+\delta) \qquad\qquad (16)$$

5. Simulation analysis

In this section, we have generated the observation of $(x_1,\dots\dots,x_p)$ predictors independently from $N_{n=8}(0,\Sigma)$. The matrix Σ is defined as $\Sigma_{ij}=0.5^{|i-j|}$. Also, we have generated 100 observations from the Tobit model $yi=\max\{0,y_i^*\}$, here $y_i^*=X\beta+e$. We conducted the simulation analysis based on sparse condition:

1- The true vector of parameter $(\beta=3,1.5\,,0,0\,,2,0,0,0)^T$ with error term has followed deterrents scenarios; $N(0,1)$, T-student distribution with 3 degrees of freedom, and Chi-squared distribution with 3 degrees of freedom. . In this paper we focused on the prediction accuracy of the parameter estimates for different regression models; Bayesian Tobit Model(BTM), Bayesian Median Tobit Model (BMTM) with our proposed models , Bayesian Tobit Adaptive Lasso Model (BTALM), and Bayesian Tobit Lasso Model (BTLM). The estimated relative efficiency (reff) statistics are used for comparing between the different models,

$$re\hat{f}f(\hat{\beta}_j)=\frac{S^2_{model(\hat{\beta}_j)}}{S^2_{proposed-model(\hat{\beta}_j)}}$$

$Here,\ S^2_{(\hat{\beta}_j)}$and $\hat{\beta}_j$ are defined as follows,

$$S^2_{(\hat{\beta}_j)} = \sum_{j=1}^{400} \frac{\left(\hat{\beta}_j^{\ k} - \bar{\beta}_j\right)}{400}$$

$$\bar{\beta}_j = \sum_{j=1}^{400} \frac{\hat{\beta}_j^{\ k}}{400}$$

Where, $\hat{\beta}_j^{\ k}$ is the target model parameter estimate with rth replications, and β_j is the true value of parameter. Table 1 and Table 2 show the values of the estimated relative efficiency of the proposed models relative to other models by using 100 and 200 sample sizes based on sparse conditions.

Table 1. Relative efficiency under dense condition and sample size=100

Error Dist.	Method	$Eff(\hat{\beta}_1)$	$Eff(\hat{\beta}_2)$	$Eff(\hat{\beta}_3)$	$Eff(\hat{\beta}_4)$	$Eff(\hat{\beta}_5)$	$Eff(\hat{\beta}_6)$	$Eff(\hat{\beta}_7)$	$Eff(\hat{\beta}_8)$
NORMAL	*BTM*	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	BMTM	0.5399	1.2762	1.0210	1.3080	0.9088	1.5207	1.1854	1.2299
	BTALM	0.4443	0.5942	0.4497	0.5427	0.3559	0.6244	0.6319	0.3770
	BTLM	0.4546	0.6018	0.4537	0.5250	0.3584	0.6212	0.6280	0.3706
T-STUDENT (3)	*BTM*	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	BMTM	0.1528	0.1695	0.1059	0.1089	0.1161	0.1697	0.1900	0.2053
	BTALM	0.0439	0.1017	0.1063	0.0595	0.0736	0.0834	0.0638	0.0777
	BTLM	0.0452	0.0997	0.1055	0.0572	0.0733	0.0846	0.0660	0.0775
CHI-SQUARE (3)	*BTM*	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	BMTM	0.2119	0.2463	0.3616	0.2842	0.2888	0.4633	0.3003	0.2662
	BTALM	0.2013	0.2083	0.3846	0.2840	0.2314	0.4060	0.3606	0.3558
	BTLM	0.1900	0.2178	0.3840	0.2644	0.2369	0.4078	0.3557	0.3506

Table 2. Relative efficiency under dense condition and sample size=200

Error Dist.	Method	$Eff(\hat{\beta}_1)$	$Eff(\hat{\beta}_2)$	$Eff(\hat{\beta}_3)$	$Eff(\hat{\beta}_4)$	$Eff(\hat{\beta}_5)$	$Eff(\hat{\beta}_6)$	$Eff(\hat{\beta}_7)$	$Eff(\hat{\beta}_8)$
NORMAL	*BTM*	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	BMTM	2.7695	1.9071	1.2953	3.1325	3.3123	2.6089	2.0824	1.5432
	BTALM	1.0798	0.7836	0.4083	1.1906	1.5063	1.1433	0.9543	1.1829
	BTLM	1.0995	0.7987	0.4200	1.1895	1.5411	1.1441	1.0046	1.1858
T-STUDENT (3)	*BTM*	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	BMTM	0.2207	0.3930	0.2967	0.4954	0.2677	0.3268	0.3457	0.3546
	BTALM	0.1323	0.2781	0.2316	0.2333	0.1641	0.2372	0.2560	0.1787
	BTLM	0.1314	0.2744	0.2363	0.2419	0.1646	0.2420	0.2623	0.1816
CHI-SQUARE (3)	*BTM*	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	BMTM	0.3879	0.6695	0.3177	0.4351	0.5069	0.6959	0.6095	0.5332
	BTALM	0.3428	0.6230	0.2379	0.3432	0.5109	0.6945	0.5152	0.6847
	BTLM	0.3395	0.6040	0.2593	0.3269	0.5076	0.6840	0.4917	0.6746

In general, it can be seen that the relative efficiency (reff) values in Table 1 and Table 2 obtained from the proposed models (BTALM), and (BTLM) are more efficiency than the other models (BTM, BMTM) as the sample size increasing from 100 to 200 especially when the error term followed the Normal distribution, which means that the variances of the different parameters in the sparse model decreases compared with the other models. Figures 1 and 2 show the distributions of the parameter estimates, we can see that the distribution of the parameter is following the normal distribution under BTAL and BTL models, respectively.

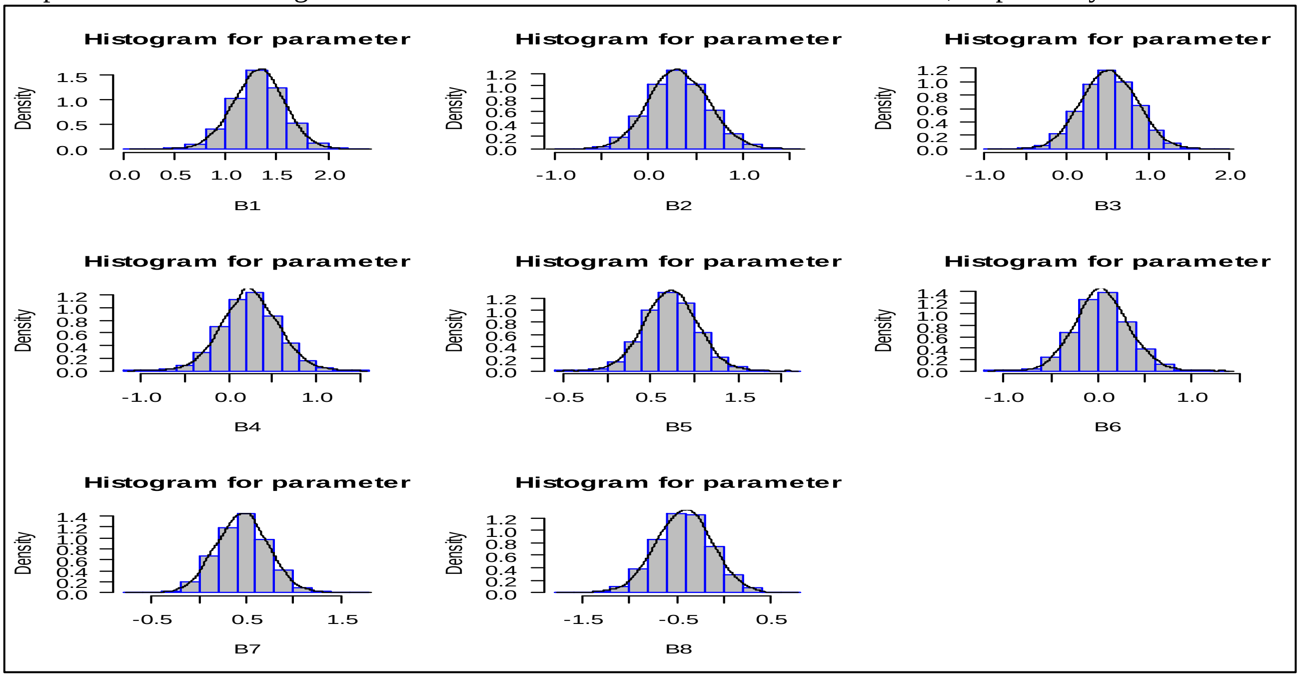

Figure 1. Histograms of parameter estimates of the BTALM

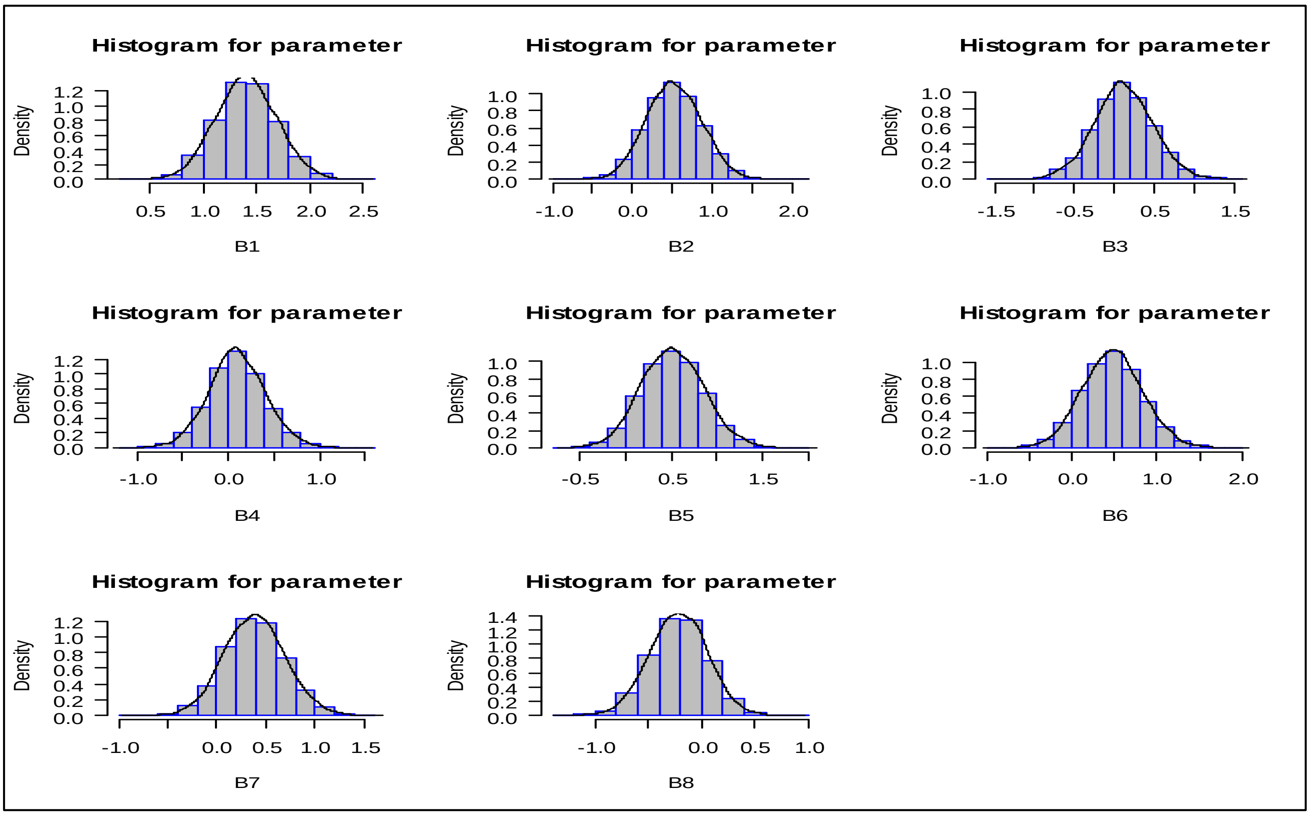

Figure 2. Histograms of parameter estimates of the BTLM

The diagnostic convergence tool, trace plot show in the following Figures (3) and (4) explain that the MCMC samples of the posterior pdf of regression parameters convergence to target distribution. Also, the trace plots show that the MCMC does not suffer from slow mixing.

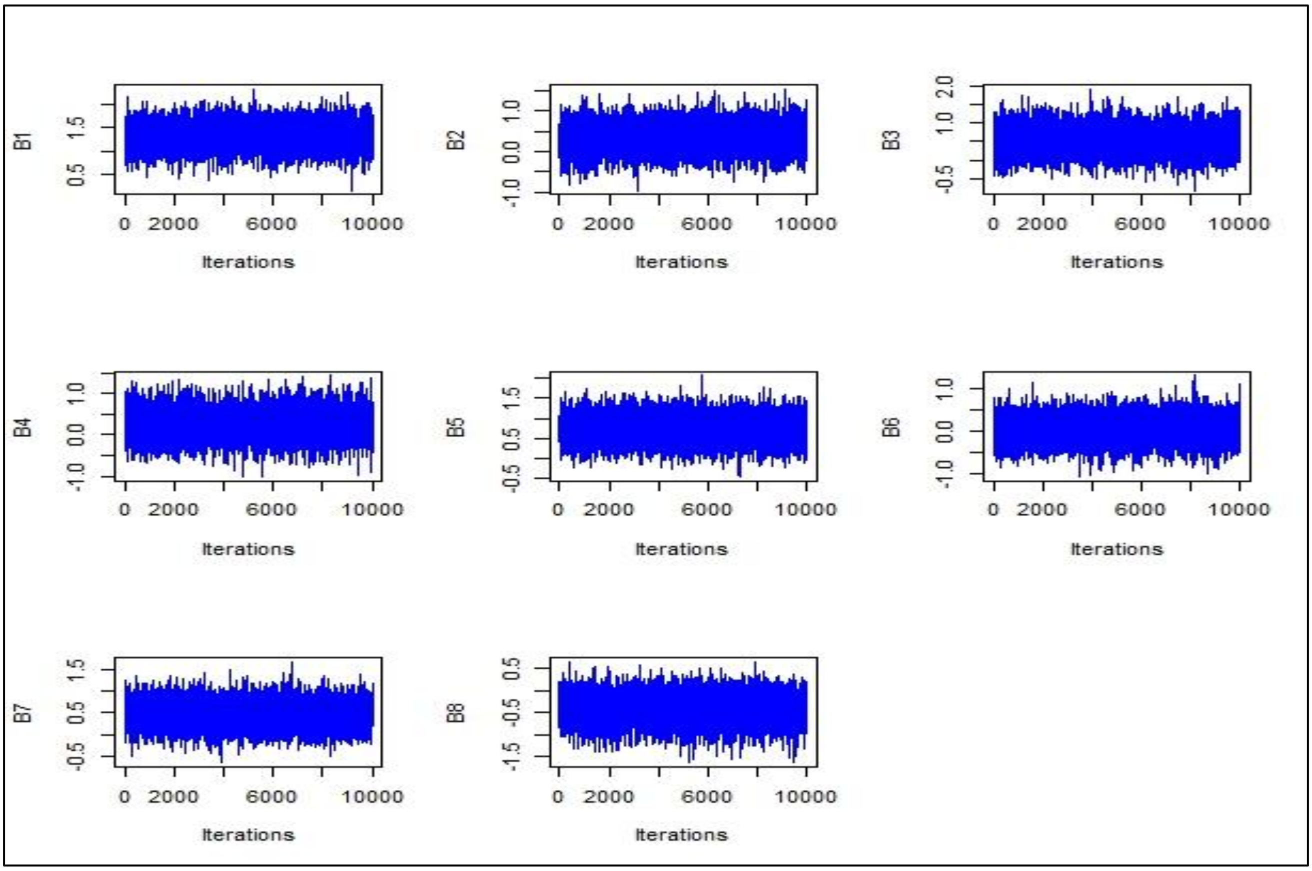

Figure 3. Trace Plots of $\beta_1 - \beta_8$ of parameter estimates of BTALM

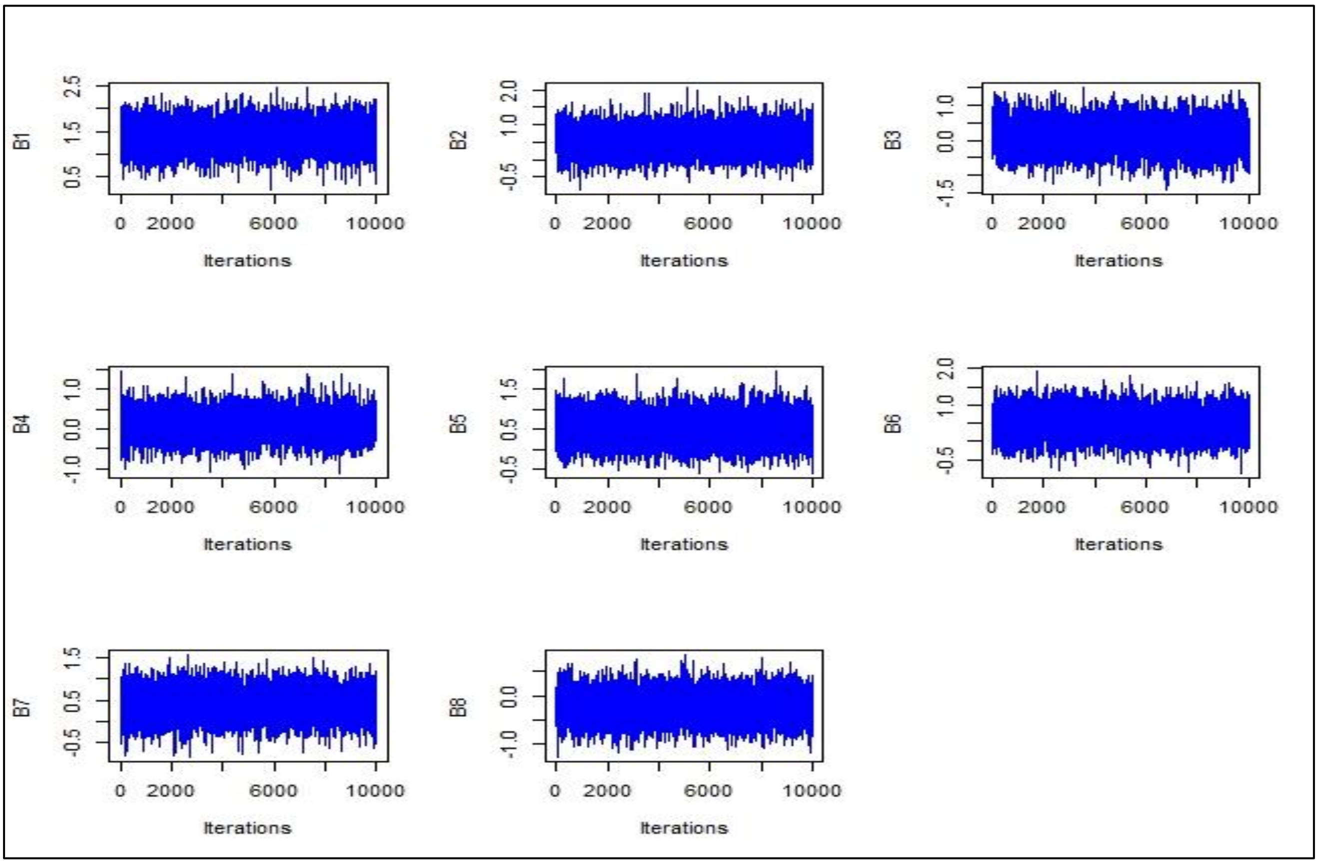

Figure 4. Trace Plots of $\beta_1 - \beta_8$ of parameter estimates of BTLM

6. Analysis of real data

We have collected the data from AL-Shamiya General Hospital that located in the south of Iraq. The sample size is 250 observations. The response variable represent the number of active sperm (censored about zero) , as well as six predictors variables ; (testosterone) the normal size of testosterone in blood 8.2-34 n.mol/l, (Prolactin) the normal size of Prolactin it is 1.5-19 ng/ml less than 19, (pH semen) the normal of PH semen 7.1-8 , (Viscosity semen) the normal od Viscosity semen its 20-30 minutes, smoking (=1 if yes, 0 otherwise) and Sperm antibodies (=1 if blood has sperm antibodies , 0 otherwise). Since the predictor variables have different measure units, we standardized their values and then centered the response variable values.

We followed [11] to estimate the penalty parameter λ value which is the posterior mean when the prior distribution is gamma (16) while implement the MCMC/Gibbs sampler algorithm.

We implement the proposed Bayesian conditional posterior distributions based on BTLAM and BTLM models in (2) and compared with (BTM, and BMTM) regressions. Table (3) contains the estimates of 6 parameters of the predictor variable of (BTM, BMTM, BTALM, and BTLM) regressions, clearly the BTAL and BTL models are outperform better than BTM and BMTM in terms of the sparsity. Table 4 contains the values of the Mean Squares Error (MSE) for regression models (BTM, BMTM, BTALM, and BTLM) and exhibits that BTALM and BTLM performs better than other regression models, followed by BTM and BMTM respectively.

Table 4. The MSE values for BTM, BMTM, BTALM, and BTLM

Methods	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$
BTM	1.2405	-1.3445	-2.5405	1.3002	10.9993	-3.1399
BMTM	0.8806	-1.4892	-1.1832	1.1445	10.8787	-6.7115
BTALM	1.2793	0.0000	1.8417	1.4461	8.6933	-1.8745
BTLM	1.2812	0.0000	1.0081	1.4489	8.7936	-2.0152

The boxplots in Figure 7 explains that the proposed models BTALM and BTLM does not suffer from the dispersion of the parameter estimates compared with the other models.

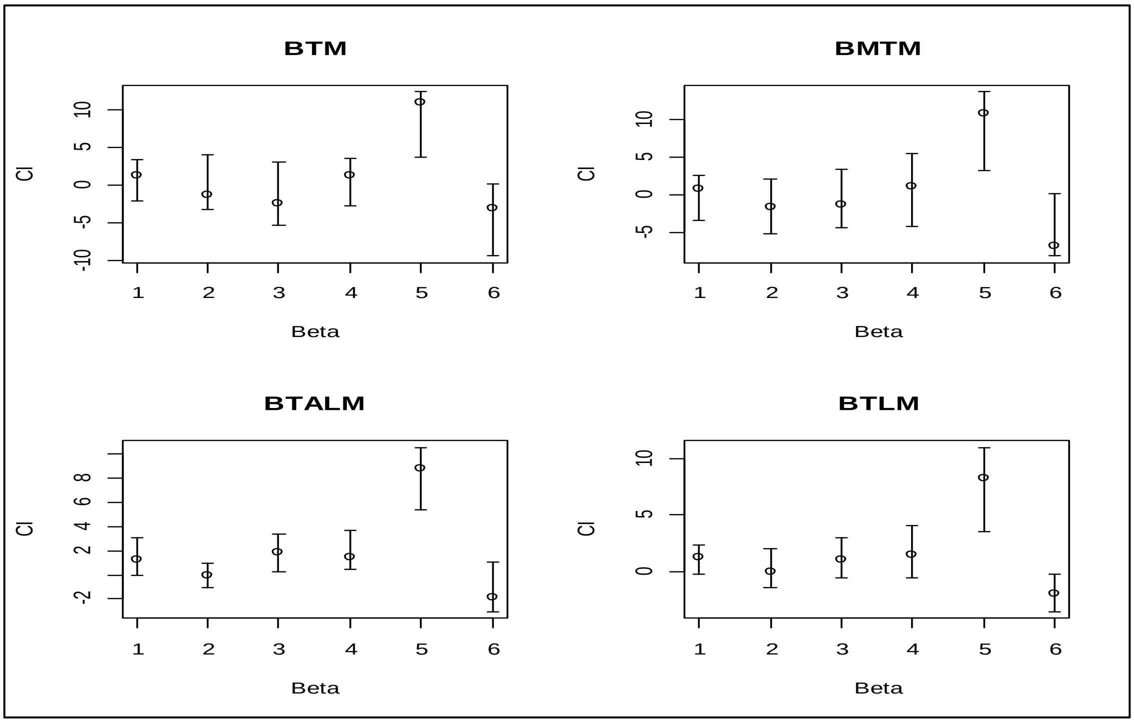

Figure 5. Comparison of performance between different method along with $\beta_1 - \beta_6$

7. Conclusions

New regularization regressions; Bayesian Tobit Adaptive Lasso (BTAL) and Bayesian Tobit Lasso (BTL) models 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. We have conducted simulation analysis to explore the path solution of the proposed regularization regressions, also we performed real data analysis. The results of simulation presented some evidence of competitable of the proposed regression models to the others exists models, but with outperform of the new Bayesian regression models in the real data in views of variable reduction.

References

[1] H.Abbas," Bayesian Lasso Tobit regression,"*Journal of AL-Qadisiyah for computer science and mathematics*, 11(2): 2504-2521.2019.

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

[3] R. Alhamzawi, "Tobit quantile regression with adaptive lasso penalty", The 4th International Scientific Conference of Arab Statistics 450 ISSN, pp. 1681-6870,2013.

[4] R, Alhamzawi and K. Yu," Bayesian tobit quantile regression using g-prior distribution with ridge parameter," *Journal of Statistical Computation and Simulation*, vol. 85 (14), 2903-2918,2015.

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

[6] F. H. H, Alhusseini , "New Bayesian Lasso in Tobit Quantile Regression," *Romanian Statistical Review Supplement* , vol. 65.6 pp 213-229,2017.

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

[8] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani R.,"Least angle regression,"*The Annals of statistics,* vol. 32(2), 407-499,2004.

[9] J. E. Griffin, and P. J. Brown," Bayesian adaptive Lassos with non-convex penalization.," *Technical report. Institute of Mathematics, Statistics and Actuarial Science*, University of Kent, 2010.

[10] S. Kotz, T. Kozubowski, and K. Podgorski, "The Laplace Distribution and Generalizations," Springer *ISBN*, vol. 0-8176-4166-1,2001.

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

[12] C.M. N. Leng, Tran and D. Nott,"Bayesian adaptive Lasso," arXiv*:*1009.2300v1,2010.

[13] T. Park, and G. Casella, "The Bayesian lasso," *Journal of the American Statistical Association,* vol.103(482), 681-686,2008.

[14] R. Tibshirani, "Regression shrinkage and selection via the lasso,"*Journal of the Royal Statistical Society. Series B (Methodological),* vol.267-288,1996.

[15], R. Tibshirani, M. Saunders, S. Rosset, J. Zhu and K. Knight, "Sparsity and smoothness via the fused lasso," *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* vol. 67(1), 91-108,2005.

[16] J. Tobin, "Estimation of relationships for limited dependent variables, "*Econometrica: journal of the Econometric Society,* vol. 24-36,1958.

[17] M. Yuan, and Y. Lin, "Model selection and estimation in regression with grouped variables," Journal of the Royal Statistical Society, Series B 68, 49–67,2006.

[18] H. Zou, and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* vol. 67(2), 301-320,2005.

[19] H. Zou, " The adaptive lasso and its oracle properties," *Journal of the American statistical association*, vol. 101(476), 1418-1429, 2006.