New Scale Mixture for Bayesian Adaptive Lasso Tobit Regression

Dr. Ahmad N. Flaih[1], Dr. Muhannad F. Al-Saadony[2], Dr. Hassan Elsalloukh[3]
[1,2]Department of Statistics, University of Al-Qadisiyah
[3]Department of Statistics / University of Arkansas at Little Rock, USA

Abstract

Abbas [1] proposed new hierarchical representation of the adaptive Bayesian lasso model as uniform density , mixing with standard exponential distribution based on a transformation of the mixture of uniform density and a particular gamma distribution formulation provided by Mallick & Yi [2] They consider the new proposed hierarchical formulation model and prior distributions, as well as the full Conditional posterior distributions structural under non conditioning on σ^2 which makes the uncertainty, of a unimodal full posterior, Conditioning on σ^2 is important, because it guarantees a unimodal full posterior Park and Casella [3]. So, we can conclude that [1] proposed new hierarchical representation utilizing a Non- scale mixture distributions, which needs to deal with this problem . To address this problem we consider new hierarchical representation of the adaptive Bayesian lasso for Tobit model based on scale mixture of Uniform density, mixing with standard exponential distribution.

Keywords: Bayesian Adaptive, Left censored, Hierarchical Model, MCMC, Posterior distribution.

Corresponding Author: Dr. Hassan Elsalloukh

Department of Mathematics and Statistics
University of Arkansas at Little Rock
2801 S. University,
Little Rock, AR 72204
E-mail: hxelsalloukh@ualr.edu

مقياس خليط جديد لنموذج انحدار توبت البيزي لاسو المكيف

د. أحمد نعيم فليح[1] د. مهند فائز[2] د. حسن الصلبوخ[3]
[1،2]جامعة القادسية- كلية الادارة والاقتصاد – قسم الاحصاء
[3]جامعة أركنساس – قسم الإحصاء

المستخلص: أقترح عباس (2019) تمثيلًا هرميًا جديدًا لنموذج البيزي اللاسو التكيفي ككثافة موحدة، ممزوجة بالتوزيع الأسي القياسي بناءً على تحويل خليط الكثافة المنتظمة وصيغة توزيع جاما معينة مقدمة من ماليك ويي، (2014). ضع في اعتبارك نموذج الصياغة الهرمي الجديد المقترح والتوزيعات السابقة، بالإضافة إلى التوزيعات الخلفية الشرطية الكاملة الهيكلية في ظل عدم التكييف على σ2 مما يجعل عدم اليقين، من الخلف الكامل أحادي الصيغ، والتكييف على σ2 مهمًا، لأنه يضمن حديقة خلفية كاملة أحادية الوسائط وكاسيلا (2008). لذلك، يمكننا أن نستنتج أن عباس (2019) اقترح تمثيلًا هرميًا جديدًا باستخدام توزيعات مختلطة غير قياسية، والتي تحتاج إلى التعامل مع هذه المشكلة. لمعالجة هذه المشكلة، نأخذ في الاعتبار التمثيل الهرمي الجديد لنموذج Bayesian lasso التكيفي لنموذج Tobit بناءً على خليط مقياس الكثافة الموحدة ، والاختلاط مع التوزيع الأسي القياسي.

الكلمات المفتاحية: البيزي التكيفي،،ترك الرقابة نموذج الصياغة الهرمي، التوزيع اللاحق.

1. Introduction

Traditional least squares method used to estimate the mean of response variable in the regression function through the data provided by predictor variables. The least squares method provided best quality of the estimate especially when $n \geq p$ which implies that the smallest variance of the estimate and then the prediction accuracy will be adequate. However, when the number of predictor variables (p) approach the sample size (n), or when $p > n$, least squares cannot be used. Horel & Kennard [4] proposed a solution to address the problem of $p > n$ through introducing Ridge regression by adding the penalty function to the residual sum of squares (RSS),

$RSS(\beta)$ + penalty function $_{\lambda}(\beta)$,

where λ is the regularization parameter. Tibshirani [5] realized the Lasso method that set the value of irrelevant predictor variables to Zero (sparsity), as well as make shrinkage to the relevant predictor variables estimates. Lasso and Ridge have the same technique which minimize the penalized residual sum of squares, but with different penalty functions. Consequently, we can say that Lasso provide ”Variable Selection” techniques.
The widely popular regularization method least Absolute Shrinkage and Selection Operator (Lasso) [5] used in regression to shrinkage and variable selection. The difference between Lasso and Ridge method is the penalty function that refer to constraint of L_2 norm in ridge, whereas in Lasso the constrain is penalty function of L_1 norm. The Lasso estimate minimize the following penalized residual sum of square (RSS)

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{lasso}} = \text{argmin}_{\boldsymbol{\beta}}(\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta}) + \lambda \sum_{j=1}^{p} |\beta_j| \qquad (1)$$

where the regularization parameter is $\lambda \geq 0$. We can see from (1), the penalty function is consist of Lasso penalty L_1 norm. Since the penalty function is Lasso is non-differential at zero, then it has the ability to reduce the parameter estimates to be zeros. Here, X is the matrix of standardized predictors , y is the vector of centered value of the response variable .
Frank & Friedman [6] suggested the bridge regression with penalty function L_r−norm). Mienshausen [7] suggested new regularization method called relaxed lasso to control the bias in the Lasso estimate. Zou [8] proposed new penalized function which is called Adaptive Lasso (AL) to address the problem of bias in Lasso estimate by scaling each parameter with different weights, the AL coefficients are estimated by minimize,

$$\hat{\beta}^{n}{}_{\mathrm{AL}} = arg\,min_{\beta} ||y - X\underline{\beta}||_2^2 + \lambda_n \sum_{i=1}^{p} \hat{w}_j |\beta_j,| \qquad (2)$$

where $w_j > 0$ and $\hat{w} = 1/|\hat{\beta}$ ols|, with $\gamma > 0$.

Tibshirani et al. [9] proposed fused Lasso that handle predictor variables. It ordered in meaningful way. Bakin [10] proposed what is called the group Lasso. Zhoa et al. [11] introduced the composite absolute penalties as generalization of the group Lasso. Zou & Hastie [12] introduced the elastic net regularization as combination of Ridge and Lasso regression. We consider the above mentioned methods of regression parameters estimation as ”Frequentist” methods. Recently, the Bayesian methods have become widely used especially in regression techniques. Tibshirani [5] introduced the $|\beta_j|$ term in the penalty function as proportion to (−log) of Laplace density;

$$f(\beta_j) = \frac{\lambda}{2} \exp -\lambda|\beta_j|, \qquad \lambda \geq 0. \tag{3}$$

Tobin [13] defined the censored regression (Tobit model) as follows:

$$y_i = \begin{cases} y_i^* & ; \; y_i^* < 0 \\ 0 & ; \; y_i^* \geq 0 \end{cases}$$

where the latent variable $y_i^* = x'_i \, \beta + u_i$and $y_i = max\{0, y_i^*\}$ with $u_i \sim N(0,\sigma^2)$. Park and Casella [3] introduced the Lasso regression parameter estimates based on subjective Bayes [5] and show that the Lasso estimate as posterior mode estimate, where the prior density of β is Laplace density. Returning to [3] they introduced an important feature to the Bayesian analysis of the Lasso estimator under conditional Laplace prior β on σ^2,

$$\pi(\beta|\sigma^2) = \frac{1}{\sigma^2} \Pi_{i=1}^{p} \frac{\lambda}{2\sqrt{\sigma^2}} \exp\left(-\frac{\lambda|\beta_j|}{\sqrt{\sigma^2}}\right) \tag{4}$$

This feature (4) is provided the uni-modality of the full posterior of Lasso estimator. Leng [14] 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.

$$\hat{\beta} = arg\, min\|y - X\underline{\beta}\|^2 + \sum_{j=1}^{p} \lambda_j|\beta_j| \tag{5}$$

Griffin and Brown [15] 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.
Mallick and Yi [2] proposed new scale mixture of densities as new hierarchical representation of subjective Bayesian Lasso, i.e.,

$$\frac{\lambda}{2\sqrt{\sigma^2}} \exp\left(-\frac{\lambda\,|x|}{\sqrt{\sigma^2}}\right) = \int_{u > \frac{|x|}{\sqrt{\sigma^2}}} \frac{1}{2u\sqrt{\sigma^2}} \quad \frac{\lambda^2}{\Gamma(2)} u^{2-1} \exp(-\lambda u)\, du \tag{6}$$

which is the scale mixture of uniform density, mixing with Gamma(2,λ) the new scale mixture in (6) leads with the (4) to new hierarchical representation. In this paper, we proposed new representation of Laplace prior as scale mixture of uniform mixing with standard exponential distribution.

Prposition: A Laplace density can be written as a scale mixture of uniform density mixing with standard exponential distribution, i.e.,

$$\frac{\lambda}{2\sqrt{\sigma^2}} \exp\left(\frac{-\lambda|x|}{\sqrt{\sigma^2}}\right) = \int_{Z > \frac{\lambda|x|}{\sqrt{\sigma^2}}} \frac{\lambda}{2\sqrt{\sigma^2}} \exp(-Z)\, dZ \tag{7}$$

Proof: Mathematically, it is well-known that:

$$\exp\left[\frac{-\lambda|x|}{\sqrt{\sigma^2}}\right] = \int_{w > \frac{|x|}{\sqrt{\sigma^2}}} \lambda \exp(-\lambda w)\, dw \tag{8}$$

multiple both side of (8) by $\frac{\lambda}{2\sqrt{\sigma^2}}$, and letting $z = \lambda w$, then we can get:

ISSN (1681- 6870)

$$\frac{\lambda}{2\sqrt{\sigma^2}}\exp\left(\frac{-\lambda|x|}{\sqrt{\sigma^2}}\right) = \int_{z>\frac{\lambda|x|}{\sqrt{\sigma^2}}} \frac{\lambda}{2\sqrt{\sigma^2}}\lambda\exp(-z)\times\frac{1}{\lambda}dz$$
$$= \int_{z>\frac{\lambda|x|}{\sqrt{\sigma^2}}} \frac{\lambda}{2\sqrt{\sigma^2}}\exp(-z)dz$$

Hence, we get (7), we can say that the Laplace distribution is a scale mixture of uniform($-\frac{\sqrt{\sigma^2}}{\lambda}, \frac{\sqrt{\sigma^2}}{\lambda}$) mixing with standard exponential distribution. Abbas [1] introduced the following hierarchical model based on non-scale mixture of Uniform distribution mixing with standard exponential density, by reparametrization the scale mixture of Mallick and Yi [2],

$$\frac{\lambda}{2}\exp(-\lambda|\beta_j|) = \int_{z_j>\lambda|\beta_j|}\frac{\lambda}{2}\exp(-z_j)dz_j$$
$$\beta_j/\lambda \sim Uniform(-1/\lambda, 1/\lambda)$$
$$z_j \sim \text{Standard exponential density}$$

The proposed hierarchical model based on (9) consist of the probability density function of $(\beta|\lambda)$ which is uniformly distributed but does not involve the scale parameter σ^2. Following [2],[3]we can state that (9) is not conditional on σ^2, where [3] state that conditioning on σ^2 is important to guarantee the unimodal of the full posterior distribution and to make the point estimation more meaningful. In this paper we address this problem through adding σ^2 as conditional on β.

2. The Hierarchical Model and Prior Distribution

Following Park and [2,3,1] and by using (4) and (6), the hierarchical model formulation based on (7) is as follows:

$$y_i = \begin{cases} y_i^* & ; \; y_i^* < 0 \\ 0 & ; \; y_i^* \geq 0 \end{cases},$$
$$\boldsymbol{y}^* | \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2 \sim N_n(\boldsymbol{X\beta}, \sigma^2 I_n),$$
$$\boldsymbol{\beta}|\lambda, \sigma^2 \sim \prod_{j=1}^{p} \text{Uniform}\left(-\frac{\sqrt{\sigma^2}}{\lambda_j}, \frac{\sqrt{\sigma^2}}{\lambda_j}\right),$$
$$z \sim \prod_{j=1}^{p} \text{Standard exponential},$$
$$\sigma^2 \sim \pi(\sigma^2),$$

Where $\pi(\sigma^2)$ could be $1/\sigma^2$ or any inverse gamma prior for σ^2, where the inverse gamma for σ^2 is

$$\pi(\sigma^2) = \frac{b^a}{\Gamma a}(\sigma^2)^{-a-1}\exp(-\frac{b}{\sigma^2});\ \sigma^2, a, b > 0$$
$$\lambda_j \sim \text{Gamma}(h, d).$$

3. Full Conditional Posterior Distributions

Suppose we have the vector $z = (z_1, z_2, \cdots, z_p)^0$, then the full conditional posterior

$$\boldsymbol{\beta}|\boldsymbol{y}^*, \boldsymbol{X}, z, \boldsymbol{\lambda} \sim N_p\left(\widehat{\boldsymbol{\beta}}_{OLS}, (\boldsymbol{X}'\boldsymbol{X})^{-1}\sigma^2\right) \prod_{j=1}^{p} I\left[\frac{\sqrt{\sigma^2}\, z_j}{\lambda_j} < \beta_j < \frac{\sqrt{\sigma^2}\, z_j}{\lambda_j}\right]. \quad (9)$$

$$\boldsymbol{z} \sim \prod_{j=1}^{p} \text{Exp}(-z_j).I[z_j > \frac{\lambda_j |\beta_j|}{\sqrt{\sigma^2}}]. \quad (10)$$

$$\sigma^2|\boldsymbol{y}^*, \boldsymbol{X}, \boldsymbol{z}, \boldsymbol{\beta} \sim \text{InvGamma}\left(\frac{n+p-1}{2} + a, \frac{1}{2}(\boldsymbol{y}^* - \boldsymbol{X\beta})'(\boldsymbol{y}^* - \boldsymbol{X\beta}) + b\right). \quad (11)$$

$$\lambda_j|\beta_j, \sigma^2 \propto \text{Gamma}(h+1, d)\, I[\lambda_j < \frac{z_j\sqrt{\sigma^2}}{|\beta_j|}]. \quad (12)$$

See the Appendix (A) for the derivatives.

4. MCMC Sampling for the proposed Bayesian Lasso

Regularization path computations under Bayesian Lasso with Gibbs Sampler algorithm and based the scale mixture (7) and initial values for β, z, λ, and σ^2 are given with iterations the following steps:

1- Sampling the latent variable y : generating y from truncated normal $(X\beta, \sigma^2 I_n)$.
2- Sampling z : generating z_j as follows:

$$z_j = z_j^* + \frac{\lambda_j |\beta_j|}{\sqrt{\sigma^2}}$$

where z_j generates from left truncated exponential and z_j^* generates from standard exponential distribution with rate parameter equal to one in (11).
3- Sampling β: generating β can be done from truncated multivariate normal distribution proportional to the full conditional posterior distribution in (10).
4- Sampling σ^2 : generating σ^2 can be implemented from left truncated inverse gamma distribution proportional to the full conditional posterior distribution in equation (12). [2] generate (σ^{2*}) from right truncated gamma density,

$$Gamma\left(\frac{n+p-1}{2} + a, (y - X\beta)'(y - X\beta)\right) \text{I}[\sigma^{2*} < \frac{1}{\max_j \frac{\lambda_j^2 \beta_j^2}{z_j^2}}]$$

replacing, $\sigma^2 = 1/\sigma^{2*}$ then we can imitative sampling from truncated Inverse Gamma in (12).

5- Sampling the regularization parameter λ_j: we can update the regularization parameter λ_j from gamma distribution by generating the samples from (13).

5. Simulation Analysis

Simulation study is performed in two examples based the proposed scale mixture of unifroms and to identify many scenarios in which the New Bayesian adaptive Lasso (NBALasso) performs well. For simulated examples, we use the statistic "Mediam of Mean Absolute Deviations" (MMAD) to compare the performance of different regression models (Lasso, Bayesian Lasso (BLasso), Adaptive Lasso(Alasso), Bayesian Adaptive Lasso (BALasso), NBALasso). by using the following formula,

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

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

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

Where X is distributed from normal with mean zero and variance one, and the error term with following scenarios:

1. The distribution of the error is a normal distribution: N(μ, 1).

2. The distribution of the error is a mixture of two normal distributions:

   0.1N(μ, 1) + 0.9N(μ, 9).

3. The distribution of the error is a Laplace distribution: Laplace(μ, 1).

4. The distribution of the error is a mixture of two Laplace distributions:

   0.1Laplace(μ, 1) + 0.9Laplace(μ, 3).

The correlation between predictors Xi and Xj is $\rho^{|i-j|}$, and the matrix of predictor variable observations are $X \sim N(0, \Sigma)$, here $\Sigma_{ij} = 0.5^{|i-j|}$. Before carry out any regression model, we standardized the predictors values and centered the response variable values. The Bayesian lasso and the new Bayesian adaptive lasso estimates are the posterior means, we use the Gibbs sampler to implement the proposed conditional posterior distributions. In R package lars for lasso, we used the LARS algorithm to select the penalty parameter with (k = 10)-fold cross validation.

 ISSN (1681- 6870)

Example 1

In this example, we generate dataset of 50 each with 20 observations, this example used by [5] , here the true vector is $\beta = (0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85)$ and $\sigma = 3$. Table (1) Shows the proposed method (NBALasso) performed better with error distributed on (Normal, Laplace) than the other methods based on the values of MMAD criterion as well as based on the Standard Deviation (SD) values. But under (Normal Mix) the (BALasso) perform better and under (Laplace Mix) the (Alasso) performs well, so we can say the proposed new regularization method is comparable.

Table (1) MMAD values with SD

Methods	Error Distributions			
	Normal	Normal Mix	Laplace	Laplace Mix
Lasso	1.3491 (1.2154)	2.1028 (1.6769)	1.6834 (1.4071)	3.1969 (2.8276)
BLasso	1.2154 (1.1409)	2.3560 (1.7701)	1.7793 (1.6539)	2.6957 (2.3942)
ALasso	1.1883 (1.0739)	2.5364 (1.8482)	1.6026 (1.3479)	2.4231 (2.2947)
BALasso	1.1409 (1.0577)	1.9587 (1.3398)	1.5087 (1.1917)	2.5993 (2.3228)
NBALasso	1.0739 (1.0024)	1.9934 (1.3989)	1.3729 (1.0407)	2.4581 (2.2156)

The following trace plot figures (1) shows that the Gibbs sampler algorithm of the posterior distributions of the regression parameters convergence to the stationary distribution, there is no flat bits in trace plot figures of parameters pattern.

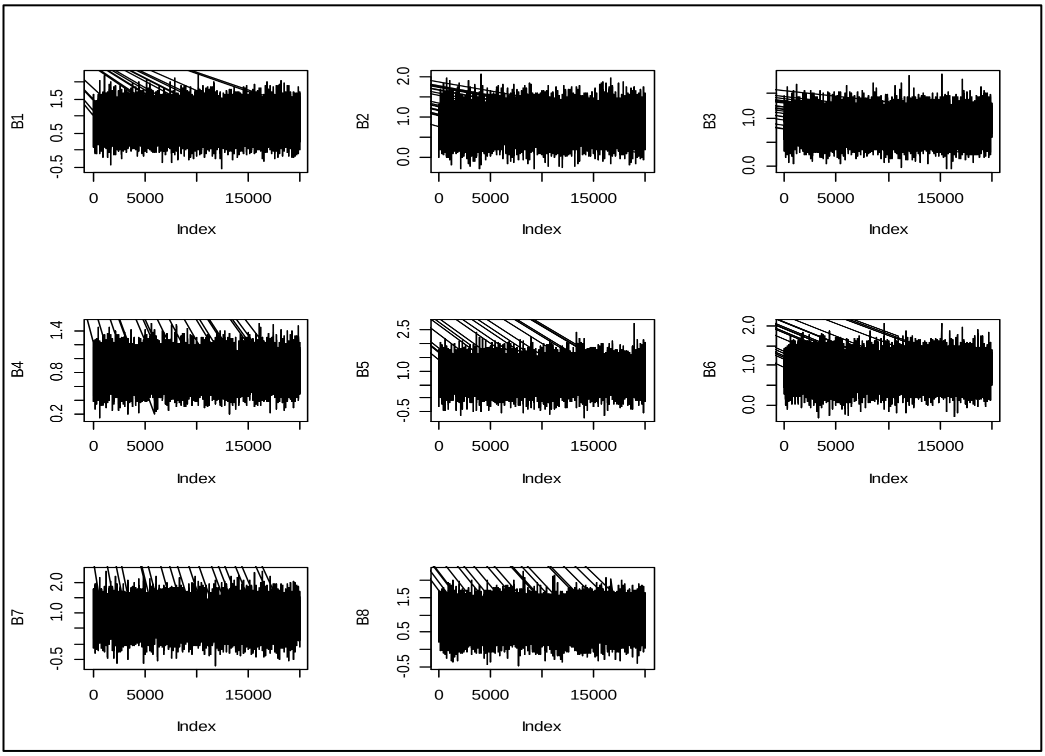
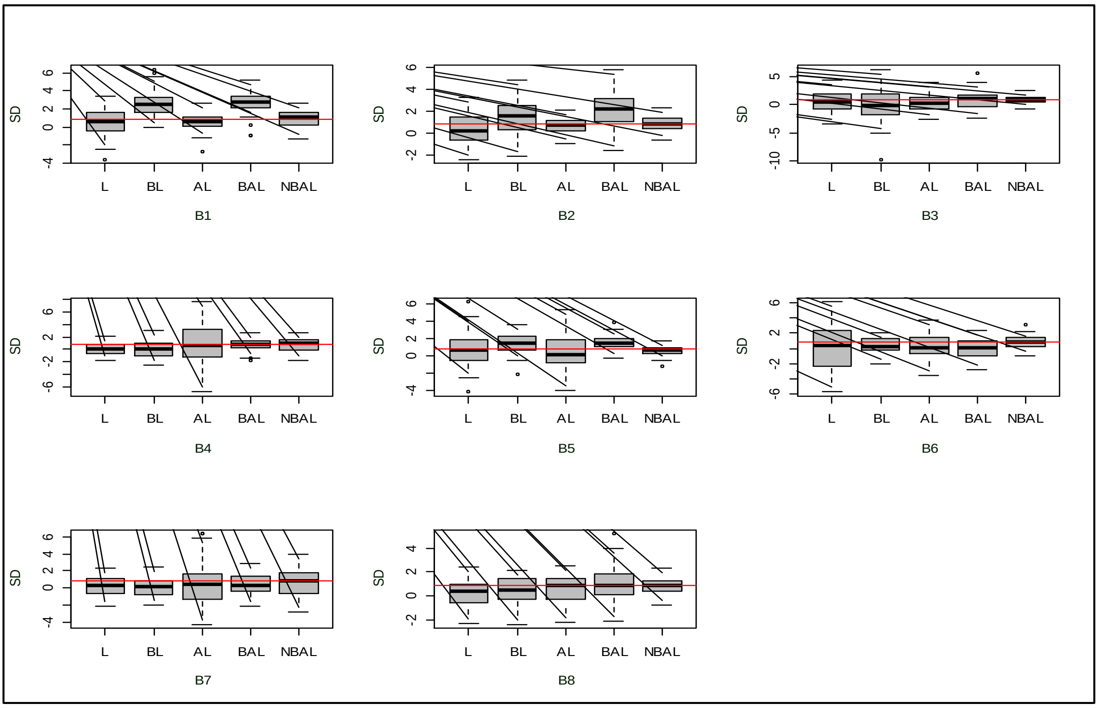


Figure (1) Trace Plots of Proposed NBALasso Tobit Parameters $\beta_1 - \beta_8$

Also, we provide our analysis with Boxplots figures, where figure (2) shows the boxplots of the Tobit regression parameters which exhibits that NBALasso does not suffer from the deviations pf the coefficients estimates compared with the other regression estimation methods, the value of the median of the estimated parameter values is very close to the true of the regression parameter (Red Line).



Example 2

In this example, we generate dataset of 50 each with 20 observations, this example used by [5] , here the true vector is $\beta = (5,0,0,0,0,0,0,0)$ and $\sigma = 2$. Table (2) Shows the proposed method (NBALasso) performed better with error distributed on (Normal, Laplace) than the other methods based on the values of MMAD criterion as well as based on the Standard Deviation (SD) values. But under (Normal, Laplace, Laplace Mix) the (NBALasso) perform better and under (Normal Mix) the (Alasso) performs well, so we can say the proposed new regularization method is comparable and performs better with the lower values of standard errors.

Table (2) MMAD values with SD

Methods	Error Distributions			
	Normal	Normal Mix	Laplace	Laplace Mix
Lasso	4.7317 (4.3920)	8.9440 (7.7034)	5.8908 (4.5547)	10.0509 (9.3654)
BLasso	5.0468 (4.6356)	7.2779 (6.6309)	4.7793 (4.3555)	9.6124 (8.8980)
ALasso	3.4735 (3.3835)	6.2129 (5.7648)	4.9195 (4.4090)	9.4318 (8.5260)
BALasso	3.3536 (3.2465)	7.1679 (6.6201)	4.1479 (3.7184)	9.0051 (8.3433)
NBALasso	3.2292 (3.0113)	6.4708 (5.8303)	4.1258 (3.6369)	8.7288 (8.0855)

The following trace plot figures (3) shows that the MCMC sampler algorithm of the posterior distributions of the regression coefficients convergence to the stationary distribution, we can see the well mixing of MCMC samples with target distribution.

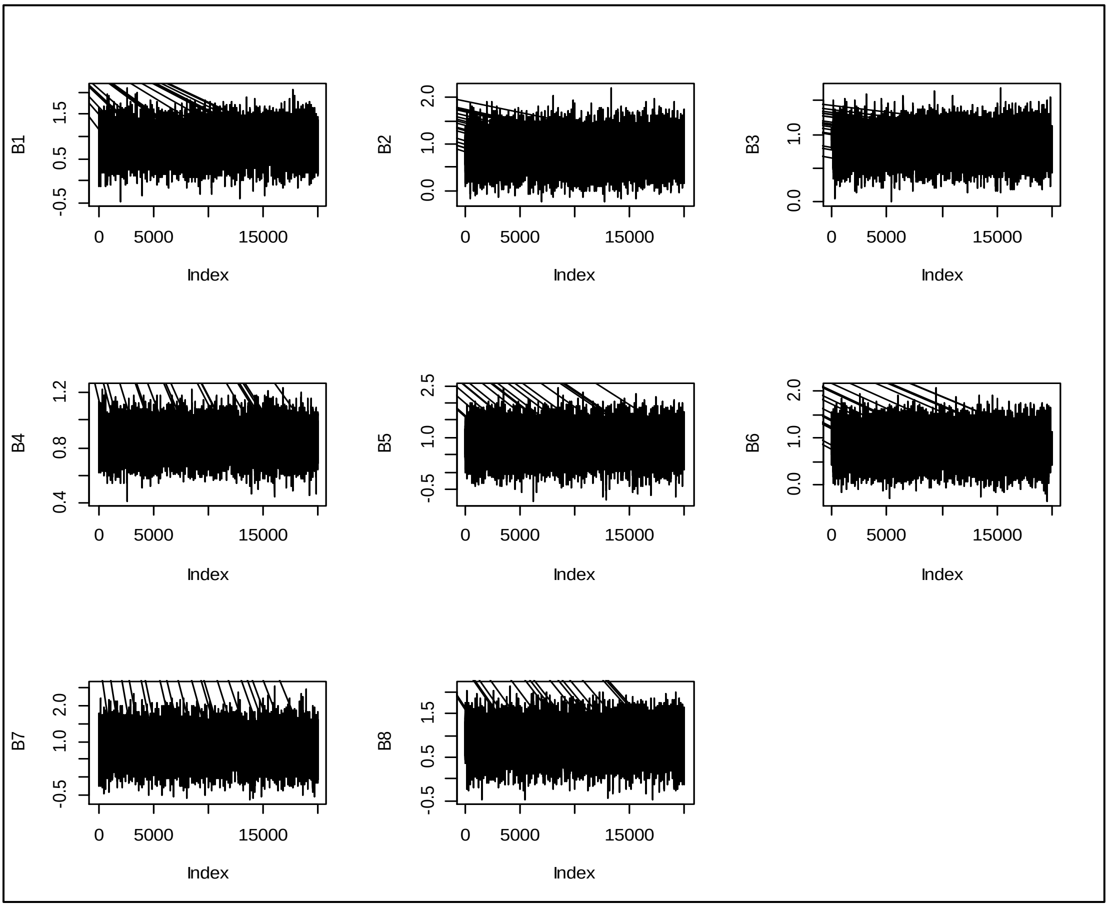


Figure (3) Trace Plots of Proposed NBALasso Tobit Parameters $\beta_1 - \beta_8$

Also, we provide our analysis with Boxplots figures, where figure (4) shows the boxplots of the Tobit regression parameters which exhibits that NBALasso does not suffer from the deviations of the coefficients estimates compared with the other regression estimation methods, the value of the median of the estimated parameter values is very close to the true of the regression parameter (Red Line).

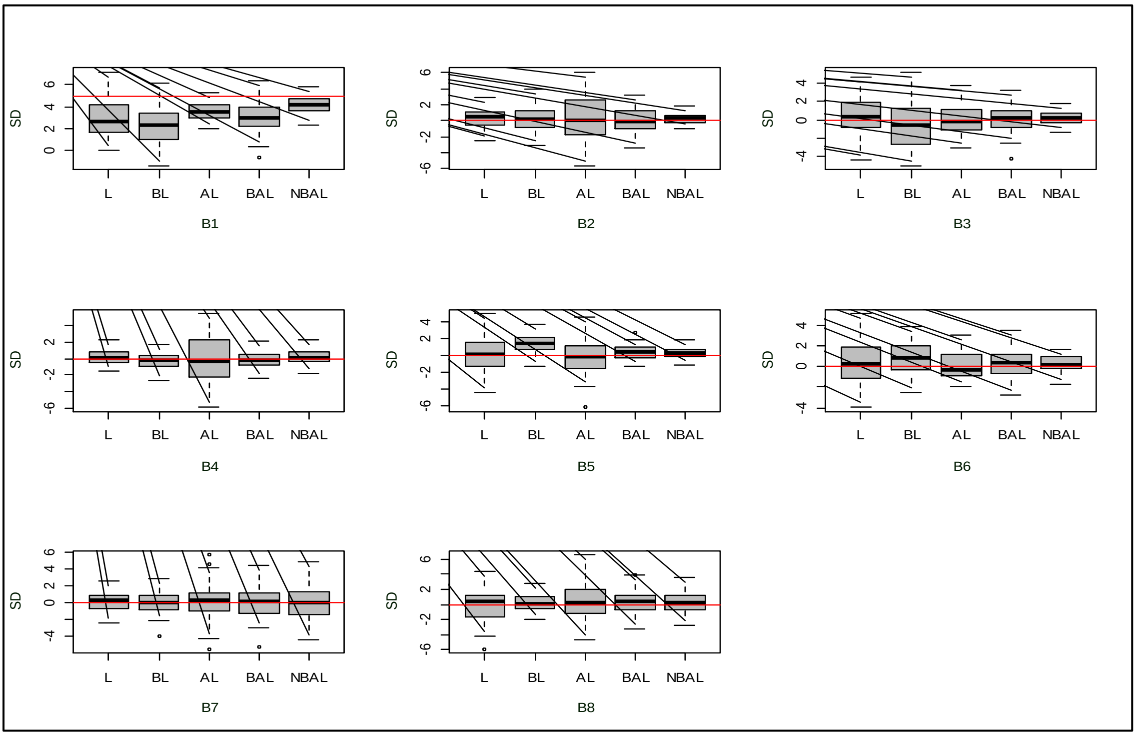


Figure (2) Compared of Performance with Different Methods along with $\beta_1 - \beta_8$

Conclusions

In this paper, new Bayesian lasso method for variable selection have proposed based on the Laplace prior distribution as scale mixture of Uniforms mixing with standard exponential distribution on their variances. New hierarchical model representation and new MCMC algorithm have developed. Two simulation examples conducted to explore the path solution of the proposed method. The results of simulation presented some evidence of competition of the proposed Bayesian Adaptive Lasso with the others methods.

References

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

[2] H. Mallick, & N. Yi, "A new Bayesian lasso," *Statistics and its interface*, 7(4), 571.2014.

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

[4] A. Hoerl. & R. Kennard," Ridge Regression: Biased Estimation for No orthogonal Problems.,"*Technometrics*, 12(1):55-67.1970.
[5] R. Tibshirani, " Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological),* 267-288.1996
[6] I. Frank & J. Friedman, "An Statistical View of Some Chemometrics Regression Tools," *Technometrics,* 35, 109-135.1993.
[7] N. Meinshausen, "Relaxed Lasso," *Computational Statistics & Data Analysis*, 52(1) :374-393.2007.
[8] H. Zou, "The adaptive lasso and its oracle properties," *Journal of the American statistical association,* 101(476), 1418-1429.2006.
[9] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, K. Knight, "Sparsity and smoothness via the fused lasso," *Royal Statistical Society: Series B (Statistical Methodology),* 67(1):91-108. 2005.
[10] S. Bakin, "*Adaptive regression and model selection in data mining problems,"* PhD Thesis. Australian National University. Canberra. Australia.1999.
[11] P. Zhao, G. Rocha,, and B. Yu, "The Composite Absolute Penalties Family for Grouped and Hierarchical Variable Selection," The Annals of Statistics, 37(6A), 3468–3497.2009.
[12] H. Zou, & T. Hastie, " Regularization and Variable Selection via the Elastic Net," *Journal of the Royal Statistical Society. Series B (Statistical Methodology),* 67(2):301-320.2005.
[13] J. Tobin," Estimation of relationships for limited dependent variables," *Econometrica: journal of the Econometric Society*, 24-36.1958.
[14] C, Leng, N. Tran and D. Nott ," Bayesian adaptive Lasso," arXiv:1009.2300v1.2010
[15] J. Griffin, and P.J. Brown ,"Bayesian adaptive lassos with nonconvex penalization," *Australian and New Zealand Journal of Statistics*, forthcoming. 2010

Appendix A

The derivations of the full conditional posterior distributions are as follows: We suppose that joint posterior distribution of all parameters is:

$$\pi(\beta,z,\sigma^2,\lambda|y,X) \propto \pi(y|X,\beta,\sigma^2)\pi(\beta|z,\sigma^2)\pi(z|\lambda)\pi(\lambda)\pi(\sigma^2)d\sigma^2.$$

Now we can write down the posterior distributions as:

1− The posterior distribution of β conditioning y,X,z,λ,σ^2 is:

$$\pi(\beta / y, X, z, \sigma^2) \propto \pi(y / X, \beta, \sigma^2)\pi(\beta / z, \lambda)$$

$$\propto \exp[-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)]\prod_{j=1}^{p} \mathrm{I}\ [\ |\beta_j| < \frac{z_j\sqrt{\sigma^2}}{\lambda_j}\ ]$$

$$\propto \exp[-\frac{1}{2\sigma^2}(y - \widehat{\beta}_{OLS})'X'X(y - \widehat{\beta}_{OLS})]\prod_{j=1}^{p} \mathrm{I}\ [\ |\beta_j| < \frac{z_j\sqrt{\sigma^2}}{\lambda_j}\ ]$$

Then,

$$\beta / y, X, z, \sigma^2 \sim N_p(\widehat{\beta}_{OLS}, \sigma^2(X'X)^{-1})\ \prod_{j=1}^{p} \mathrm{I}\ [\ |\beta_j| < \frac{z_j\sqrt{\sigma^2}}{\lambda_j}\ ]$$

Hence we have proved (10).

2− The posterior distribution of z is:

 ISSN (1681- 6870)

$$\pi(z / \beta,\lambda,\sigma^2) \propto \pi(\beta / z, \sigma^2, \lambda)\pi(z)$$

$$\propto \prod_{j=1}^{p} \exp(-z_j) \mathrm{I} \left[z_j > \frac{\lambda_j |\beta_j|}{\sqrt{\sigma^2}} \right]$$

Then

$$z / \sigma^2, \lambda, \beta \propto \prod_{j=1}^{p} \exp(-z_j) \mathrm{I} \left[z_j > \frac{\lambda_j |\beta_j|}{\sqrt{\sigma^2}} \right]$$

Hence we proved (11).

3− The posterior distribution of σ^2 is:

$$\pi(\sigma^2 / y, X, z, \beta, \lambda) \propto \pi(y / X, \beta, \sigma^2)\pi(\beta / \lambda, \sigma^2)\pi(\lambda)\pi(\sigma^2) d\sigma^2$$

$$\propto (\frac{1}{\sigma^2})^{\frac{n-1}{2}} \exp[-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)]$$

$$[\frac{\lambda}{2\sqrt{\sigma^2}}]^p \pi(\sigma^2) = \frac{b^a}{\Gamma a}(\sigma^2)^{-a-1} \exp(\frac{b}{\sigma^2}) \mathrm{I}[\sigma^2 > \max_j \frac{\lambda^2 \beta_j^2}{z_j^2}]$$

Then

$$\sigma^2 / y, X, z, \beta, \lambda \sim Inverse\,Gamma(\frac{n+p-1}{2} + a, (y - X\beta)'(y - X\beta) + b) \mathrm{I}[\sigma^2 > \max_j \frac{\lambda^2 \beta_j^2}{z_j^2}]$$

Hence we proved (12).

4− The posterior distribution of λ_j is:

$$\pi(\lambda_j / \beta_j) \propto \pi(\beta_j / \lambda_j)\pi(\lambda_j)$$

$$\propto \lambda_j \frac{d^h}{\Gamma h} \lambda_j^{h-1} \exp[-d\lambda_j]$$

$$\propto \lambda_j^{(h+1)-1} \exp[-d\lambda_j] \mathrm{I}[\lambda_j < \frac{z_j\sqrt{\sigma^2}}{|\beta_j|}]$$

Then

$$\lambda_j / \beta_j \sim Gamma(h+1, d) \mathrm{I}[\lambda_j < \frac{z_j\sqrt{\sigma^2}}{|\beta_j|}]$$

Where the prior density of λ_j has a Gamma(h,d).
Hence we proved (13).